universität wien

DISSERTATION

Titel der Dissertation

Computers and the Ability to See

Understanding the Negotiation and Implementation of

Image Processing Algorithms

Verfasser

Christoph Musik, Bakk.phil. MA

angestrebter akademischer Grad

Doktor der Philosophie (Dr. phil.)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 784 121
Dissertationsgebiet lt. Studienblatt:	Wissenschaftsforschung
Betreuerin:	Univ. - Prof. Dr. Ulrike Felt

Table of Contents

Table of Figures

Acknowledgements

In all probability it is impossible to thank all people who supported and accompanied the completion of this doctoral dissertation. Nevertheless I take this opportunity trying to do so in what follows.

First of all, I would like to thank my individual supervisor Ulrike Felt. She also made a substantial contribution towards realising the interdisciplinary research project "Identification practices and techniques in Austria, 18th to 21st century," which was funded within the scholarship programme 'DOC-team' by the *Austrian Academy of Sciences (ÖAW)* between October 2010 and October 2013, and which was the basis to realize my research. In this regard I would like to thank my project colleagues and friends Stephan Gruber and Daniel Meßner for a great time together and for all of their contributions to my work! The overall project was also supported and guided by Andrea Griesebner and Peter Becker, thanks a lot!

I thank Lucy Suchman who gave me the opportunity to discuss my project outline during a four month stay at the *Centre for Science Studies* at Lancaster University (UK) and I thank Susanne Krasmann and the whole team of the *Institute for Criminological Research* at the University of Hamburg for discussing my research during my two month stay in Hamburg in 2013!

I also thank Barbara Prainsack, Harry Collins, Iris Eisenberger, Nils Zurawski, Steven Hilgartner, and William J. Kinsella for comments and suggestions to my research. I thank Regula Valérie Burri who has kindly agreed to act as a second examiner of this thesis. Further, I would like to thank all current and former colleagues at the *Department of Science and Technology Studies* at the University of Vienna who commented and/or supported my research. Thank you Andrea Schikowitz, Anna Pichelstorfer, Anna-Lena Berscheid, Claudia Schwarz, Daniela Schuh, Dorothea Born, Florian Bayer, Gernot Rieder, Judith Igelsböck, Karin Neumann, Kay Felder, Lisa Sigl, Marlene Altenhofer, Maximilian Fochler, Michael Penkler, Michael Strassnig, , Philippe

B. Sormani, Sarah Schönbauer, Simone Schumann, Stefanie Schürz, Thomas Völker and Ulrike Ritter!

I also would like to thank Evelyn Zisch for proofreading and editing the English.

I give special thanks to my family and friends for any kind of support and assistance. Merci Mama, Papa, Lisa! Dankeschön Anni und Rudi!

This doctoral dissertation is dedicated to my daugther Flora and my wife Helene. You deserve the final acknowledgement as you both suffered most from its completion. I love you!!!

Chapter One

Introduction: Computers and the Ability to See

The German newspaper *Der Postillion* reported in an article in August 2012 that the human memory ("das menschliches Erinnerungsvermögen") offends against data protection regulations. In reference to a recent study presented by the *German Federal Commissioner for Data Protection and Freedom of Information*, it was argued that the human memory draws up movement profiles, recognises faces out of every thinkable angle and thus, it constantly collects and analyses every kind of information. The article brings in some examples of these abilities: The human memory, in contrast to video surveillance cameras is able to recognise behaviour patterns such as recognising that "this guy with the horn-rimmed glasses goes with the same S-Bahn every day". In addition, with the human memory it is possible to analyse the overall shopping behaviour of people in order to place personalised product recommendations. An example would be the situation in which a salesperson asks the customer "For you the same as yesterday?" Standard integrated face recognition in the human memory is also highly problematic. Indeed, does this tool facilitate personal salutation, for example when entering your own favourite pub. However, in all likelihood, behaviour patterns can be formed using the face recognition tool (e.g. "This guy gets drunk every night."). In order to cope with the human memory and its mania for collecting data, the Commissioner for Data Protection and Freedom of Information calls for a general rule that blindfolds and earflaps should be worn. Those items should be carried at all times, except in situations in which the other person has agreed to the storage and use of his or her personal data (so-called "opt-in"). For those that cannot wait for the regulation, the Commissioner advises carrying balaclavas or burka (so-called "opt-out").

As you may have recognised—at least if you are able to make use of human memory—the article, and the medium of its publication, *The Postillon*, are satiric. So obviously that means, there just is no such recent study presented by the *German Federal Commissioner for Data Protection and Freedom of Information* that reported about human memory offending against data protection regulations. As is the case with most satire, there is a lot of truth, or at least some elements of it, that refer to more serious issues. So this satiric article opens up some of my major research concerns and as such, concerns of this thesis. As you can expect from the title 'Computers and the Ability to See' this refers not so much to the earflaps as to the blindfolds. In short, my basic research interest is in computer vision and connected to it what could be called in reference to the satiric article 'computer memory' that in this context, is more than the physical device used in computers that stores programs and data; it is the ability of computers to remember or clearly recognise specific visual entities in the physical-optical world such as objects, people, cows, (individual) faces, (suspicious) behaviour patterns, facial expressions and so on.

What is it About?

This same interest in computer vision and computer memory is in human vision, and therefore in the human memory, as well. That means, computer or more generally, machine vision and human vision are not as clearly distinguishable from each other as one might expect. There is a close and ongoing relationship between computer vision and human vision, and as there is between computers and humans in general. Nevertheless, the boundaries between humans and computers and, ever since the late 1960s, also between human vision and computer vision are constant subjects for discussion and negotiation. To speak in Lucy Suchman's terms (2007: 226) we can ask the question how these boundaries between humans and non-humans, and adapted to fit my research interest between human and computer vision, are drawn and re-drawn? How are humans and non-humans, and how are human vision and computer vision (re-)configured in mundane practices by specific members of society? It is essential for the further reading of this thesis to note here that moments that address questions of computer vision are simultaneously always moments that address questions of human vision too. This is very briefly and broadly speaking because humans are inevitably involved in the design and development of computers or machines that are able to see. Humans are the ones teaching computers sight. Humans are also the ones to use and be affected by "seeing" computers. But what does "seeing" actually mean? What (different) understandings of seeing and closely connected to it, recognizing, do members of society have, especially when it comes to teaching computers to achieve sight? Is there one universal, global way of seeing that could be easily transferred to machines, or is seeing rather a diverse "situated" and cultural activity that hampers a simple and smooth transformation of this ability? In the context of computer vision, one has to ask how computers are taught to be able to see in which ways, by whom and how these processes might change our understanding of human vision and similarly of what is (perceived as) true and real in our world?

These introductory queries are the primary and fundamental questions that framed and guided my research and the writing of this thesis. However it would be presumptive to claim to answer these substantive questions anywhere near sufficiently deep enough within the frame of this thesis. As a consequence, it might provide a modest contribution towards a reflection on these fundamental questions, and might enrich the academic literature dealing with these questions by empirically exploring the social and political significance of Image Processing Algorithms in the aforementioned 'Human-Computer Vision (Re-) Configurations'. Before turning to the concrete research questions of this thesis and thus, the explication of how and where the thesis might provide a more major contribution, I shall provide in a nutshell, basic information on the background and embedding of the thesis that is further explained later on.

The thesis is based on an interdisciplinary, multiperspective approach that is framed by the academic fields of Science and Technology Studies (STS), Visual Culture Studies and Surveillance & Identification Studies. It especially is inspired by Lucy Suchman's work on 'Human-Machine Reconfigurations' (Suchman 2007) and the Visual STS approach of the 'Social Studies of Scientific Imaging and Visualization' (Burri & Dumit 2008). This links to what could be summarised as the theoretical frames of (feminist) post-humanism and material-semiotics, and connected to it, to the commitment "to empirical investigations of the concrete practices" of nonhuman entities and their specific agencies (Suchman 2007: 1).

The most relevant sociotechnical (transformation) processes that frame my involvement with computer vision and more specifically with Image Processing Algorithms are what could be condensed in the "grand narrative" (cf. Law 2008: 629) terms of surveillance society (especially what often is referred to as Smart CCTV or intelligent video surveillance) as well as the digitalisation, automatisation, and "smartisation" of social practices, artefacts and devices. On these grounds, the thesis explores 'Human-Computer Vision (Re-) Configurations' by analysing the negotiation and the development of Image Processing Algorithms in different sites from the computer vision laboratory to the news media. In doing so, my research followed a 'visiographic' strategy that will be explained in more detail in the last part of this introductory chapter.

To start with, in the hope of making the issue at stake more accessible to readers,[1] I shall try to lead to my concrete research questions and to explain their significance and my approach with a short (semi-fictional) story that will serve as an ethnographic *Gedankenexperiment* (cf. Gobo 2008: 151). It is "an attempt to solve a problem using the power of the human imagination and the head as laboratory" (ibid.). In doing so, I go back to the satirical example from the beginning.

The Story of the Good Old Times and the Smart New Future

Once upon a time, there was a person called Chris. Chris goes to the pub 'Good Old Times' every night. The bartender with the name Dani instantly recognises Chris when he enters the pub. A little glimpse is usually enough to see that the person coming in is Chris. An exception is at Christmas time when Chris dresses himself up as some kind of

[1] I am aware that writing a PhD thesis means writing primarily for a rather small circle of academic readers in the relevant fields of study. Nevertheless, I aspire to write at least some sections for a wider readership as well, which might have the additional advantage of questioning some of the assumptions usually taked for granted in the relevant academic fields.

elk and argues it is a reindeer costume. As this has happened for several years now, Dani and the other guests are instantly able to recognise it as Chris, as their experience has shown a recurring pattern that has been saved in their memories. Unless it is Christmas time, Chris sits down, orders and drinks two or three (sometimes four or five) special cowslip beers in the course of the evening and leaves the pub later on, staggering a little and weak-kneed. Dani as always is a bit worried about Chris finding the way back home without any serious incident, but as Chris usually comes again the next day in a good mood, everything seems to be all right in the 'Good Old Times'.

Then one day, the public health office commissioned and sent FUTURE VISION, a computer vision company, to the 'Good Old Times' pub to install ICVV (short for: "In Computer Vision Veritas"); a system that automatically recognises how many people leave the pub drunk, in order to evaluate alcohol and health behaviour of people in different areas and cities throughout the country and in the whole of the OECD. It was originally planned to connect the data with personal user profiles with the help of automatic face recognition technology (FRT), but by reason of the ORWELL protest - a local data protection activist group - this objective was banned because there were concerns that this very personal data could fall into the hands of intelligence agencies. Dani, the landlord of the pub, was quite glad about this, because if FRT was to be installed in the pub, there would have been need to install more light sources in the traditionally dimmed pub in order to make the face recognition work properly. Good illumination is the basis for this surveillance, as one of the FUTURE VISION guys told Dani on the quiet. Another reason for Dani to be glad about the fact that FRT was not installed was that Dani would have needed to ban all kinds of head and face coverings from the pub in case of its installation. How should Dani then have told Michele to remove her beloved hat for instance? Anyway, Dani had already prepared the back doors for good costumers such as Chris and Michele to use to exit the pub. In the end there was no way of hindering the installation of the ICVV system, but the back doors were there to do their job anyway.

Now, after the installation of the new ICVV system, there is a little black box on top of the official main exit door, that seems to have a camera inside, but the guys from

FUTURE VISION did not call it a camera, but a visual sensor. Next to the little black box is a sign that reads “For your personal health, a visual sensor is in operation”. Dani asked some of the guys from FUTURE VISION how the little black box up there works and they answered that it automatically sees when somebody leaves the pub drunk and then sends this data, encoded, to the health office. However Dani should not be worried, as it is just for statistics, in order to improve public health and the visual sensor does not save or send away any images or personal information. That is also the reason why it is called a visual sensor and not a camera. It does not save any images. Dani also asked why they do not use one of the more usual breathalysers and was told that the visual sensor system is more convenient, because the users—that is Dani and the pub guests—do not need to do anything. The ICVV system does everything fully automated and does not need any cooperation of users. They will soon see, the technology is so “silent” that they will forget that there is even something in operation. They just can enjoy their drinks as usual and will not be bothered by ICVV. Dani was left a bit puzzled by the miraculous visual sensor and for that reason he asked one of the guests, Steph - the one who always reads and preferably drinks Indian Pale Ale - about how the silent visual sensor in the little black box could work and how it recognises when somebody leaves the pub drunk. Steph, who had already read about this kind of system in the media, had the theory that it might work using path-tracking to recognise wiggly lines. That means, if somebody leaves the pub in a normal, straight line they are not drunk. When somebody leaves the pub in wiggly lines it means he or she is drunk. Well, Steph said, that it is just an everyday theory, but how else could a drunk person be recognized on sight. Everybody knows that drunk people sway uncertainly, and Steph added “look at Chris, staggering every night!”

In the following days the brand-new little black box was the centre of attention in the ‘Good Old Times’. One day Chris, Dani, Steph, Michele and the other guests started to make fun of the little black box. One after another, they positioned themselves in front of the visual sensor and staggered in clearly recognisable curves to the exit door. From that day on, every single one of the regular guests repeated the procedure of leaving the pub in clearly recognisable curves in a way they imagined looked like the perfect drunk. This procedure developed into a daily ritual and the pub even became well known for its guests leaving the pub in wiggly lines. After a while it attracted more and more guests and spectators and the 'Good Old Times' was crowded every night. Soon after, patrons of the other pubs in the area started to leave, staggering in wiggly lines and people had lots of great fun night after night.

One day, however, newspaper articles reported in reference to a so-called 'transdisciplinary study' of the public health office and the OECD about the area as the "World Capital of Drunks", the once quiet neighbourhood became known as a boisterous party and thus, drinking location. As a consequence to political pressure by the local authorities, the public health office initiated a costly anti-alcohol campaign in the area and in addition, the local police started a so-called 'Smart CCTV' scheme equipped with automatic FUTURE VISION criminal behaviour detection software. Its ICVV application having already proved the value of their systems as they provided the public health office with data about the grade of peoples drunken in public spaces throughout the neighbourhood.

The guys from FUTURE VISION, now well-dressed in business like suites, also dropped into the 'Good Old Times' again and offered Dani a new system they called 'The Smart Pub.' They told him about a prototype system they had already installed in a special university lab in the Smarter World Valley called 'The Pub of the Future' and reported about increased costumer and sales figures in this special pub. They told him how 'The Smart Pub' system automatically recognises how many people enter the pub, how old they are (in this regard they could also install the automatic pub gate in order to ban minors from the pub) whether they are male or female, in what mood they are and of course if they are already inebriated as well. A new special feature is the automatic recognition of body shapes in order to define people's general health situation. This data is subsequently automatically connected to current weather, biorhythm and sports events data and as a result, displays the perfect drinks and food recommendation on a mobile device called 'iPub 24' for every customer. The system already predicts what the customer wants, they said. In addition they are currently working on an add-on that automatically starts the desired beer pump once people enter the pub. Dani was a bit sceptical towards the Smart Pub system as it was not clear if the pub actually needed another increase in customers and sales. And anyways, folks like Steph would always drink IPAs and not change their drinking behaviour because of a “shrewd” pub. Three months later—things never turn out the way you expect—the ‘Good Old Times’ beamed itself into the future and was proud to finally call itself a SMART PUB©.

One night, in a sudden realisation in the wee small hours, Dani, Chris, Steph, Michele and the other guests recognised that their pub and the whole area has changed and they jokingly decided that the introduction of the mysterious little black box had been the starting point of this change. They were not sure if it had changed for the better as sales had increased significantly, especially soon after the newspapers had reported about the “World Capital of Drunks”. On the other hand, it had also changed for the worse because new and strange people, hipsters and bobos, had begun to turn up at the formerly quiet pub. The ‘Good Old Times’ had even made it into the ‘Lonely Planet’ guide. Steph even switched from drinking IPAs to carrot juice, following a quite annoying, recurring recommendation of the smart pub control panel. The smart pub—that was Dani’s guess—somehow must have recognised the bad condition of his eyes. Finally, all of them knew that something fundamental had changed. “That’s life” they said, and kept on drinking carrot juice and occasionally, as an act of resistance against the will of the ‘Smart Pub’, also IPAs.

– THE END –

From the Politics of Satire to the Politics of IPAs

In the Gedankenexperiment story there are several elements on different levels viewed together that are the subject of this thesis. A first glance at the story, especially when reading the end, reveals a strong, deterministic, technological view. That is, technological change becomes a process with its own internal dynamics; it happens without the involvement of people and is independent of human intent. Such a view also means that technological change determines social change (cf. Degele 2002: 24ff.). Dani and the pub guests have this kind of deterministic, technological view. Their point of view renders the situation comprehensible. They seem to have accepted the installation of the little black box and the smart pub system with a delay and realised that their behaviour and their lives have changed as a consequence of that. They unconsciously attributed agency to the little black box. From their perspective, it had influence and impact on them and changed their lives significantly. However, a closer look at the case shows that there were—to use the Social Constrution of Technology (SCOT) terminology—different relevant social groups (cf. Pinch & Bijker 1987) involved that actively shaped the technology and the concrete material manifestation of this little black box. Three of these groups were reported on in the story: The public health office that had an interest in evaluating alcohol and health behaviour of people, the computer vision company FUTURE VISION that promoted, installed, and also developed the little black box, namely the ICVV application, probably in cooperation with a university laboratory in Smarter World Valley. Finally, also ORWELL, the local data protection activist group that successfully vetoed the face recognition and personalisation component of the system. Connected to these groups are others that were not mentioned in the story. For example, think about the government or the parliament that set up a certain health agenda, or the researchers that developed the basic algorithms for the ICVV software. As one can see, there were many different social groups involved in shaping the composition and technology of the little black box. Simultaneously, the little black box shaped society as well by changing people´s lives, their (drinking) behaviour and the relations between different individuals and social groups. A simple techno-deterministic view—that STS always has rejected (Law 2008: 639)[2]—or socio-deterministic view—especially the aforementioned SCOT approach (Pinch & Bijker 1987) emphasising that human action shapes technology—on these relationships falls short. Instead, there is a need to consider both directions in relationship and to engage with the complex interactions between technology and society; between humans and technological artefacts. In order to understand these continuous and interwoven interactions between society and technology what was referred to as the 'co-production' of society and technology (Jasanoff 2004), and in further consequence, to make the best possible decisions, it is important to understand

[2] Reading Wyatt's (2008) contribution about varieties of technological determinism challenges Law's generalised view that STS always has rejected technological-determinism. Thus, I propose stating that STS has rejected radical technological-determinism in a very wide fashion.

how technologies, devices, humans, institutions, objectives, values, and so on, are interconnected with each other (cf. Johnson & Wetmore 2009). Such a focus on socio-technical systems assumes that material objects like the little black box, images or Image Processing Algorithms are inseparable from societal practices and relations, and that these societal practices generate the objects and attribute meaning to them. There is the basic STS assumption underlying this thesis that socio-technical systems consist of a complex interrelation of humans and things, whereas this complex interrelation is not neutral, but permeated with values. That means, values influence technology, and technology influences values (ibid.), an important notion to start with.

Introducing Non-Human Actors into the Social Sphere and the Social Scientific Analysis

Eventually, we are confronted with and are part of a complex actor network that consists of human and non-human entities and their connections to each other, as was especially demonstrated in the ongoing project of actor network theory (ANT) (cf. Callon 1986; Law & Hassard 1999; Latour 2005). Connected to ANT's call for "generalized symmetry" in analyses of human and non-human entities, contributions to social order (Suchman 2007: 261) are what Suchman refers to as 'discourses of agency' (ibid.: 225). This is, in my own words, the question of how much agency is allocated to humans and how much to non-humans such as computers, and if there are—contrary to ANT's call for "generalized symmetry"—differences between these types of actors and how these differences are (re-)drawn, or made real in everday practices? The struggle here is in Suchman's terms to both acknowledge "the sense of human-machine difference" and to consider "the distributed and enacted character of agency" (ibid.: 260). In reference to Latour's term of the "Middle Kingdom"[3] (cf. Latour 1993: 77f.), Suchman describes the need for

[3] Suchman refers to Latour's „Middle Kingdom" as a useful demarcation „with respect to human-nonhuman relations, within which he locates the space between simple translations from human to nonhuman, on the one hand, and a commitment to maintaining the distinctness and purity of those categories, on the other" (Suchman 2007: 260).

> "… developing a discourse that recognizes the deeply mutual constitution of humans and artifacts, and the enacted nature of the boundaries between them, without at the same time losing distinguishing particularities within specific assemblages." (Suchman 2007: 260).

The challenge here, is also to recognise and to consider human and non-human entities as distinct actors as well as taking account of the embedding of these distinct actors in what Suchman calls "sociomaterial assemblages" (Suchman 2008: 150ff.). The figure of the assemblage used here points to the "bringing together of things, both material and semiotic, into configurations that are more and less durable but always contingent on their ongoing enactment as a unity" (ibid.: 156).

It is here in the post-humanist[4] and material-semiotic[5] approaches of Science and Technology Studies that I principally position my engagement with "Computers and the Ability to See". While acknowledging central ANT conceptualisations such as its claim to integrate imperatively, non-human entities in social scientific analysis and also borrowing some of ANT's jargon, I do not share the view of "generalized symmetry" of humans and non-humans. Here I follow Suchman's claim for "a rearticulation (…) of dissymmetry" (Suchman 2007: 269) that recognises the "persistent presence" of human engineers, designers, users etc.[6] in technoscientific discourse and practice as articulation

---

[4] Following Orlikowski I understand post-humanism as an umbrella term for approaches in STS that „…seek to decenter the human subject—and more particularly, reconfigure notions of agency—in studies of everyday life" (Orlikowski 2007: 1437f.). For me „to decenter the human subject" does not mean to leave humans out and to centre exclusively on non-humans instead, but to allow both humans and non-humans to be focused on in social scientific analysis.

[5] Suchman refers to Haraway (1991: 194f.) in order to coin the phrase ´material-semiotic.´ In Suchman's words, by making use of ´material-semiotic' Haraway wanted "…to indicate the ways in which the natural and the cultural, or the material and the meaningful, are inextricably intertwined." (Suchman 2007: 261). Following Haraway (1997), Image Processing Algorithms (IPAs) could be understood as a "…construct or material-semiotic ´object of knowledge' forged by heterogeneous practices in the furnaces of technoscience." (ibid.: 129)

[6] However, as is argued by Sismondo (2010: 90), ANT also approaches the actions of scientists and engineers with particular interest when taking into consideration the subtitle of Latour's book (1987) *Science in Action*, that is: *How to Follow Scientists and Engineers through Society*

for “a durable dissymmetry among human and nonhuman actors” (ibid.: 270). I also share her view that the response to this observation of dissymmetry

> “…is not, however, to cry ‘Aha, it really is the humans after all who are running the show.’ Rather, we need a story that can tie humans and nonhumans together without erasing the culturally and historically constituted differences among them. Those differences include the fact that, in the case of technological assemblages, persons just are those actants who configure material-semiotic networks, however much we may be simultaneously incorporated into and through them.” (Suchman 2007: 270)

In this sense my point of view is to approach non-human entities or artefacts such as the little black box, images or Image Processing Algorithms as actors that all play an important role in the “show” called society. They might play an even more important role in the future once they are potentially configured, stabilised and “purified” as autonomous “smart” and “intelligent” entities, but nevertheless, as can be shown also on an empirical basis throughout this thesis, they differ significantly from human actors and their agency and thus, have to be treated differently in the social scientific analysis[7].

Understanding the “Silent” Politics of IPAs

Taking into account the Gedankenexperiment example of the little black box or visual sensor it becomes clear that this non-human actor or sociomaterial assemblage and its respective behaviour is elusive and difficult to understand for most humans using it, talking about it, or trying to imagine it. As one of the principal actors in this particular story of computers and their ability to see, the little black box or visual sensor is at the centre of attention. On the surface the little black box is actually both a camera and a visual sensor connected to computer hardware and software that integrates various Image Processing Algorithms that operate at the heart of this specific sociomaterial assemblage. The assemblage of the little black box can be considered—as was already implied in the story—as “silent” technology (cf. Introna & Wood 2004: 183). This

[7] This view also allows me to make use of the Sociology of Scientific Knowledge (SSK) standpoint on intelligent machines (cf. Chapter Three)

means, once in operation, it operates passively in the background and no user involvement is required (ibid.).

At this point it has to be made clear that it is a basic requirement of all attempts to configure computers with the ability to see, that these are in fact, attempts to produce, process and understand digital images by means of computer algorithms. This means, especially when starting from the point when we are confronted with a form of "silent technology" that it is important to understand exactly, the production, processing, and interpretation of digital images by algorithms where the semantic interpretation element is central in my involvement. Thus, these algorithms I am discussing are Image Processing Algorithms (IPAs); the main research focus of this thesis.

In contrast to what Introna and Wood call "salient" technology, "silent" technology is embedded and hidden, it has limited user involvement, which means it requires no participation or consent. It is open ended and flexible, which means that it can be used in several different contexts when adapted to these. It is mobile and maybe most importantly, it is obscure (ibid.). In the case of facial recognition algorithms, visual pattern recognition technology that is based on IPAs is employed. Introna and Wood ascribe its operational obscurity to two factors: First, algorithms usually are proprietary software and thus, "it is very difficult to get access to them for inspection and scrutiny" (ibid.). Second, most algorithms "are based on very sophisticated statistical methods that only a handful of experts can interpret and understand" (ibid.).

But why is it actually important to interpret and understand these sophisticated statistical methods and the operational obscurity of IPAs and the systems, networks or sociomaterial assemblages they are embedded in? The Gedankenexperiment story featured some possible fictional and non-fictional application areas such as behaviour pattern analysis in order to recognise drunkenness automatically, or for automated face recognition. Here it is important to note that the boundaries between fictional and non-fictional applications and conceivable application areas are continuously blurred. That is because narratives are of great significance in organising technologies (Bloomfield 2003: 195f.) such as IPAs. It can be said that the organisation of technology "in and through texts," e.g. in manuals but also in fictional texts, is a precondition for technology to be operational (ibid.: 197). It is important to make sense of the respective technology, especially if it is a matter of future technology that is planned to be integrated into society soon. In this regard, next to the showing and demonstration of technology and its worth (cf. Kirby 2011), the labelling or naming of a technology is a prerequisite to talking or writing about it. So, what am I talking and writing about? What are others talking and writing about? On a basic level, and that means put in a short sentence, that the matter here is that I am concerned with computers that are able, or at least in our imagination, able to see. But is it really seeing that is the right word in this context, the same seeing that we mean when we talk about humans that see? A seeing, for instance, that makes possible the reading of this text or the perceiving of the images that follow in the coming chapters?

Kelly Gates notes that computers 'see' only in a metaphorical sense, only in highly constrained ways, and only with a significant investment of human effort (Gates 2010: 11). She is certainly right as currently and also in the near future there is nothing known that comes close to what could be called a humanlike machine, one that is comparable to fictional characters such as the famous computer HAL 9000 from Kubrick's movie "2001: A Space Odyssey" that not without reason, will be of special interest in Chapter Three. As one will also see throughout this thesis, indeed 'seeing' in the context of computers is used only in a metaphorical sense. However, this metaphorical use of 'seeing' computers might not hide the fact that there exists—also literally speaking—computer vision and thus, computers that are able to see in our society: On the one hand this exists as an academic, scientific discipline as part of the computer sciences, on the other hand there are already some commercial applications in operation. Some of these applications are more prominent in public perception than others. For example, face recognition technologies ("Is this particular person the same person?") and connected to it face detection technologies ("Is there a face in the image?") are talked about usually in terms of face recognition or face detection[8]. This especially has been the

[8] In the use of German language in everyday life usually the term „Gesichtserkennung" is used for both face recognition and face detection. This vagueness in language does often lead to misunderstandings as face detection and face recognition are two different tasks on different levels. Correctly, face

case since the moment FRT and FDT were popularised in the form of consumer electronics (e.g., in digital cameras to detect faces in order to automatically focus on faces) or integrated into social networks (e.g., used with Facebook in order to find images of the same person automatically throughout the network). Many other similar pattern recognition technologies that are based on Image Processing Algorithms are not talked about as such. Which means, in the public discussion that they do not even have a specific name or label that distinguishes them from other entities. Instead, they are merely "hidden" parts of greater systems or entities such as 'Smart CCTV', 'Intelligent Video Surveillance' or 'Big Data', or of industrial and manufacturing applications. As a rule, one can say that anytime technology or a technological artefact is labelled 'smart' or 'intelligent', it is likely that some form of Image Processing Algorithm is part of this 'smart' or 'intelligent' article. In the case of 'Smart CCTV', what is usually meant is what Möllers and Hälterlein defined as "visual surveillance systems that analyze and interpret video footage by using pattern recognition technologies" (Möllers & Hälterlein 2012: 3). Which kind of pattern recognition technology is effectively part of 'Smart CCTV' systems is diverse and flexible. For example, Norris and Armstrong subsumed facial recognition, licence plate recognition, and intelligent scene monitoring under their umbrella term ‚algorithmic surveillance' (Norris & Armstrong 1999: 210) another maybe less normative term used for the automatic processing of video surveillance data that is restricted because it leaves other forms of IPAs out.

As I indicated earlier (cf. Musik 2011), the term algorithmic surveillance was also adopted by Introna and Wood for their analysis of the 'Politics of Facial Recognition Systems' (Introna & Wood 2004). They defined algorithmic surveillance in a literal sense as surveillance that makes use of automatic step-by-step instructions, especially of computer systems, to provide more than the raw data observed. Here, one certainly has to add the word visual or optical to raw data when talking about facial recognition technologies or any related pattern recognition technology or Image Processing Algorithms. Surette, using the metaphor of 'The thinking eye', introduced the term

---

detection is translated with „Gesichtserkennung", and face recognition with „Gesichtswiedererkennung" (wieder=re-).

‘second generation CCTV surveillance systems’ which are ‘smart’ and exploit digital technologies for artificial intelligence scene monitoring, e.g. the detection of people, unauthorized traffic, or unusual behaviour (Surette 2005). In contrast to this there are ‘first generation CCTV surveillance systems’ that are ‘stupid’, and based solely on human monitoring. To sum up, the terms algorithmic surveillance, ‘Smart CCTV’, and second generation CCTV surveillance systems have been widely used synonymously to describe the phenomenon of the automation of processing and interpreting video surveillance image data. But it makes sense to distinguish between different levels, because, for example, there is clearly a fundamental qualitative difference between automatically detecting a person in a specific scene, recognising who this person is, detecting in which direction this person is moving, or detecting that the person’s behaviour does not fit predefined norms. This qualitative difference becomes clear when taking a look at conceptualisations in computer science. For example, Turaga et al. (2008) note that the recognition of human actions and activities is processed in four consecutive steps that include three different levels:

1. Input video or sequence of images as a starting point
2. Extraction of concise low-level features (e.g. tracking and object detection)
3. Mid-level action descriptions from low-level features (action recognition modules)
4. High-level semantic interpretation of primitive actions

Following this scheme, low-level video analytics can detect that there is a person present and track this person in a specific area of surveillance. Mid-level surveillance systems can use the extracted low-level information and recognise that the person is walking or running. And finally, on a semantic high-level this walking or running could be interpreted—under certain circumstances—as suspicious behaviour, to give one example. But what really and definitely makes the observed behaviour suspicious is far from being clear and thus, is a matter of complex societal negotiation.

More particularly, it is especially the negotiation, development, and deployment of systems operating on the aforementioned semantic high-level that changes the relationship between humans and machines or objects, because interactivity between those is likely to change as well. Potentially, machines operating on a semantic high-level might be perceived as gradually becoming able to act more autonomously and by doing so also gain a higher grade of agency (Suchman 1987, 2007; Rammert 2007). It is clear that the ability to see, recognise and identify not only faces, but also facial expressions or human behaviour, is no longer exclusively reserved for humans, but it seems to be increasingly transferred to non-human actors such as computers as well.

As was argued earlier, once systems equipped with IPAs are in operation they are obscure (Introna & Wood 2004). That means, the system's specific mode of decision-making is black-boxed, but the decisions can have highly political consequences that can be extensive. It especially is these silent, mostly implicit Politics of Image Processing Algorithms[9] that potentially make them highly controversial when it comes to uncontestable and unbreakable sorting processes or decision-making affecting different people (or things) in different ways. This silent and implicit nature makes IPAs an important object of social scientific analysis. What is at stake is what modes of reality and truth are implicitly inscribed in IPAs as that can then obviously make a big difference when it comes to sorting things or humans out and making decisions out of what, for example, is currently called 'Big Data'. In a 2012 article (Feb 20, "Mining an Information Explosion") the *New York Times* even welcomed the "Age of Big Data." This welcome call referred to the constantly growing amount of data and stated, quoting an estimation by IDC, a technology research firm, that currently data is growing at 50 percent a year, or doubling every two years. According to the multinational corporation IBM, we produce 2,5 quintillion (10^{18}) bytes of data worldwide every day. The IDC study cited in the NYT article also reported that this data explosion is predominantely visual, with data deriving from various sources such as images, video clips, and surveillance streams. This means that coping with, and making sense of the ever increasing amount of data and information is not limited to numbers and texts, but does most of all, refer

[9] Following Winners ‚Do Artefacts Have Politics?' (1980), Introna and Wood argue that technology is political „by its very design", and it „includes certain interests and excludes others" (Introna & Wood 2004: 179). They also argue that the politics of software algorithms are different and thus, more problematic since particular algorithms are "silent technology" (ibid.: 182f.).

to visual data that is constantly produced in and about the physical world. Against this background, Image Processing Algorithms are essential in coping with, making sense of and understanding this ever growing amount of visual (big) data and materials.

Another current buzzword in the context of IPAs, computer vision and pattern recognition technologies is 'Ambient Intelligence'. Katja de Vries brings in the example of an intelligent coffee machine "that identifies you and anticipates how you would like to have your coffee" (de Vries 2010: 77), an example similar to the 'Smart Pub' scenario in the Gedankenexperiment story where the smart pub sensor anticipates what kind of food and drinks people would like to have. De Vries argues that such an imagined world of Ambient Intelligence "will largely depend on the development of technology that is capable of effectively identifying and distinguishing between people" (ibid.). In this regard it is not only the individual identity of a person, but also the categorical identity of a person as belonging to a certain category or group (e.g. age, gender, ethnicity, etc.) that matters. Therefore, this thesis proceeds on the assumption that there is a close historically constituted relationship between the individual and the categorical. Moreover, the history of identification and recognition of individual persons shows that it cannot be separated from the history of classification, categories and collectivities. The question, "Who is this person?" always involves the question, "what kind of person is this?" (Caplan & Torpey 2001: 3). In this regard, my research and thesis benefited greatly from being integrated into the interdisciplinary research project, "Identification practices and techniques in Austria, 18th to 21st century," which was funded within the scholarship programme 'DOC-team' by the *Austrian Academy of Sciences (ÖAW)*, because it was able to discuss and reflect on the historical dimensions of seeing and recognising.

Research Questions

Earlier, I presented some quite fundamental questions about the nature of seeing and how seeing might be (re-)configured in ongoing human-computer relations in order to introduce focus to the thesis. These questions framed and guided my research and the writing of this thesis and will also be further explored, especially in the theoretical-conceptual Chapters Two and Three. But, more concretely and as an outcome of my empirical research, this thesis explores the complex and multifaceted relation of computers and their ability to see, by focusing on the political and social significance of Image Processing Algorithms (IPAs). The analysis of IPAs is particularly suitable for exploring Human-machine Vision (Re-) Configurations, because IPAs act as powerful 'disclosing agents' (Suchman 2007: 226) that bring to light, not only the ways in which computers are (made to be) able to see, but also bring to light assumptions about human vision that are inscribed in IPAs and thus, in computers in particular ways. Therefore, the thesis proposes the following research questions:

How are Image Processing Algorithms (IPAs) developed, negotiated and implemented in different social situations, contexts, and sites? Who participates in these processes and on what (knowledge, information) basis are IPAs created?

How and where are IPAs (not) transformed into socially and politically relevant actors?

In what ways are specific forms of agency, authority, credibility, and "smarts" ascribed to IPAs or the sociomaterial assemblages they are part of and by whom?

How are particular human views and knowledge and as such, particular modes of reality and truth inscribed in IPAs and how are these human views and knowledge configured in these inscription processes?

On the way to viability how do IPAs implicitly or explicitly mesh with existing sociomaterial assemblages into which they are integrated?

And finally, how could societies and the sciences deal with IPAs in a responsible, reflective way of innovation?

What to Expect? The Outline of the Thesis

Below, I give an overview of the thesis in order to guide the reader through this text. Following the outline of the thesis as a final point in **Chapter One**, I describe and reflect on my methodical strategy of ´Visiography.'

The following **Chapter Two** entitled, “To See. To Recognise” deals with the question of what ‘Seeing’ and in close relation to it ‘Recognition’ means in the framework of current societies. I proceed to academic discussion in the area of visual cultural studies and note that images are the prime reference point of all discussion in the field. This is because “images represent social realities” and “shape the ways people think and interact” (Burri 2012: 46). However, it has to be noted here, that visual culture is not bound exclusively to images and visual representations, but is a permanent feature of the everyday practice of seeing and showing (Mitchell 2002: 170). What this all amounts to, is that human and thus, “social” vision turns out to be inevitably historically and culturally specific in all of its conceptions. It is, to use a term from cultural studies, ‘contingent.’ This observation provides an interesting challenge to assumptions about human vision in computer vision projects that stress the formal or pattern recognition model of seeing far more (Collins 2010:11). In the second part of the chapter I am concerned with the connections amid seeing, (re-)cognition and identification. Also, referring to the work of my ‘DOC-team’ project fellows, the historians Stephan Gruber and Daniel Meßner, I deal with the history of recognition by exploring the history of identification practices and techniques.

Following the historical analysis of seeing and recognising in Chapter Two, in **Chapter Three** I bring together conceptual frameworks that help understanding questions of ‘Human-Computer VISION (Re-) Configurations‘, which is also the title of Chapter Three. I am especially concerned with what Herbert Simon called the "Sciences of the Artificial“ in 1969, and what Lucy Suchman brought into the realm of ´Science and Technology Studies‘ (cf. Suchman 2008). In this context I also refer to the anthropological work of Diana E. Forsythe on the conceptualisation of knowledge in expert systems and discuss the standpoint of ´Sociology of Scientific Knowledge‘(SSK) towards intelligent machines and connect it to and confront it with computer vision and artificial intelligence literature. One of the most important characters in all of these discussions about intelligent machines and artificial intelligence is the "most famous computer that never was“ (*The Guardian* June 2, 1997); HAL 9000 from Kubrick’s film “2001: A Space Odyssey“. HAL 9000 is a cultural icon and “has come to serve as a leitmotif in the understanding of intelligent machines and the dangers associated with them"(Bloomfield 2003: 194). By reference to the fictional, but no less real character HAL 9000, vision, imagination, expectations and promises are discussed in the context of artificial intelligence and intelligent machines. In the second part of the chapter, I embed my thesis in a wider frame and address in brief the (US) history of the computer sciences and research in computer vision. Precisely because my empirical focus is on the relatively small geographical and cultural area of the nation-state Austria, local differences and particularities have to be taken into account in this history. In Austria, the specific techno-political identity of the nation is of significance as the country in the past has positioned itself regularly as a "Gallic Village" when it comes to the introduction and development of new technologies (e.g. nuclear power plants) (cf. Felt 2013: 15).

Chapter Four, the first of three empirical chapters relies on the previous theoretical-conceptual chapters. Because all attempts at giving computers the ability to see are, in fact, attempts at producing, processing and understanding (digital) images algorithmically, I understand these attempts as sociomaterial processes in which Image Processing Algorithms are developed, produced and implemented in devices or greater networks, promoted, used, critically evaluated and (re-)configured. In short, how Image Processing Algorithms (IPAs) are negotiated and implemented in different social situations, contexts, and sites. Consequently, computer vision laboratories and their output are not the only sites of interest. So, in this chapter I follow Nelly Oudshoorn (2003) and analyse the sociocultural (con-)testing of IPAs in newspaper articles and publicly available documents. I concentrate on one of Austria's first nationwide systems already in operation that makes use of image processing pattern recognition technology: the so-called ´Automatic Toll Sticker Checks' ("Automatische Vignettenkontrolle", in short: AVK) on Austrian motor- and expressways. The first mobile AVK device has been in operation since December 2007. The result of the automatic image processing ("Is there a valid toll sticker on the car's windscreen?") is essentially double-checked by a human operator in the assessment centre of ASFINAG, the operator of the system. This double-checking is a prerequisite, because dependent on the setting of the system parameters, there is always the possibility of false-positive cases. Meaning, cars with a valid toll sticker are then mistakenly recognised as invalid. Penalties would be the consequence. The aim of this case study especially, is to understand and reconstruct the progressive introduction and implementation of AVK and to analyse how this sociomaterial assemblage, in which IPAs play an important role, is described, framed and presented in the Austrian news. I am interested in which stories are told about the history, the relevant actors involved, the mode of operation, the capabilities and limitations, and the implications of the particular IPA technology AVK, in the news.

A re-occurring narrative in media articles states that it is the camera at the centre of attention. The camera, not the (silent) Image Processing Algorithm is positioned as central actor in the system and it is also the camera that recognises the presence and validity of toll stickers, whereas IPAs are widely neglected and blackboxed in favour of the 'automatic' and 'innovative' camera.

Chapter Five, titled "'Inscribing Social Order': Developing Image Processing Algorithms in Computer Vision Laboratories" deals with the question of how everyday patterns of seeing and recognition are related to and connected with Image Processing Algorithms. Of special consideration is the question how IPAs are developed and "constructed" in socio-technical laboratory constellations and how social order and social values are inscribed in IPAs. In this respect, the chapter is concerned with the socio-technical basic conditions of IPAs in the area of pattern recognition technologies: the most important basic requirement being the construction of what is called "Ground Truth" or "Ground Reality" by computer scientists. To give an example, the ground truth of suspicious behaviour corresponds with the question, "What does suspicious and thus, also non-suspicious, meaning normal behaviour, look like?" Which means that the construction of the ground truth in computer vision laboratories, is simultaneously the construction of social classifications, values and order. As such, the socio-technical construction of ground truth in Computer Vision projects can be regarded as a constituting social element that needs closer attention before it is "blackboxed" as was the case with the AVK system analysed in the previous chapter.

In Chapter Five, I elaborate on three different cases in which the sociotechnical construction of ground truth is analysed. In the first case, I refer to automatic facial expression recognition in order to show how the sociotechnical construction of Ground truth is between the two poles of expert and lay knowledge. In the second case, I give credit to the development of automated event detection in commercial banks and show that the sociotechnical construction of ground truth is also influenced by relevant social groups and their interests, as well as their concepts of knowledge. Finally, in the third case study, I emphasise the experimental character of these construction processes. To do so, I refer to a computer vision project with the aim of automatically detecting falls I followed in field work participant observation. In this specific case, future users and future sites of operation, imagined by computer scientists, played a crucial role.

In **Chapter Six**, the final empirical chapter, I engage with the question of what "functioning" means in the context of Image Processing Algorithms. How are IPAs made workable and viable? I concentrate on the negotiation of "functioning" IPAs in public computer vision demonstrations in which I participated and also acted as a test person. In the course of my field work in a computer vision laboratory in Austria it became clear that a constitutive practice in the lab was "to make things run or work." But what it actually means if something is "running," "working" or "functioning" is far from being self-evident. Rather, it is a continuous negotiation process. First, based on field observations and a group discussion with computer vision scientists, I describe their views of what "functioning" means and how the term, depending on the situation, is used in different ways. Within the academic field, the term is largely avoided. Rather, there is talk of probabilities and uncertainties. In contrast, when speaking to or in the wider public, the term "functioning" is used widely as a form of advertisement, amongst other things, in order to make sure of raising (future) research funding.

In the second part of Chapter Six, I deal with IT-demonstrations and presentations and describe the case of a public demonstration of automated detection of falls in which I participated. The early stage fall detection system was carefully designed and installed with an accentuation of certain functioning aspects that suppressed non-functioning ones, especially in regard to possible sites of application. The demonstrated fall detection system was embedded in meaningful narratives on future application areas. In this specific case especially, it was the vision of ambient assisted living (AAL) environments for the elderly.

Finally, **Chapter Seven** is not only the place for bringing together and summarising the previous chapters, but also suggests a conceptual and reflective framework for the further development and analysis of Image Processing Algorithms. Referring to the "Social Studies of Scientific Imaging and Visualisation" (Burri & Dumit 2008), I call this attempt "Social Studies of Image Processing Algorithms" (SIPA). The SIPA scheme also covers questions of governance, regulation, and ELSA (ethical, legal, and social aspects) of IPAs.

How was it Done? From a ‘Visiographic’ Research Strategy to the Analysis of Image Processing Algorithms

What follows, is a description of the methodical strategy, approach and background of this study. On the one hand, it is designed to enable the reader and other researchers to understand my course of action and analysis. Connected to this, on the other hand, it is a reflection on the use of methods in general and in this particular case.

It is well-known to any researcher that the search for the best method(s) is a basic requirement and normal practice in the course of research. Here, it has to be mentioned that the chosen research methods in most cases, have a long and winding history that is far from being obvious. One way to deal with this issue could be to hide this history and present a „smoothened“ and „glitzy“ story of the one perfect method that was planned and applied from start to end. But in my view, it is of importance to tell the story of personal research methods in a straighter and more honest way and not to hide the turnarounds, uncertainties and limits that come with every selection for, or against a specific method, strategy or approach.

Preceding Studies or the (Pre-) History of my Methodical Approach

Before this background, I start this section with a description of how I became involved and interested in the topic of computer vision and Image Processing Algorithms and how I started to research the subject. All of this can be seen as an integral part of this thesis, because the relevant preceding studies, on the one hand, influenced the further methods of analysis and on the other hand, are integrated into the empirical section themselves, especially in Chapter Five of this thesis.

It was in the summer of 2007 when I was first both fascinated and puzzled by a newspaper article in the German newspaper *Süddeutsche Zeitung (Nr. 195, 25/26 August 2007)* about automatic sex/gender recognition on the basis of faces. I asked myself, “How are they doing this? How is it possible to automatically recognise the sex or gender of a person with a computer?" At a first glimpse, the issue seemed not to be particularly crucial: feed the computer some training data, tell it who is male and who is female and it will estimate the sex/gender of a person later on automatically. I realised that it is not so easy and that we can find many socially and politically relevant transformations, standardisations, reductions, estimates and filtering processes in this usually blackboxed process. In the course of my Master studies in Sociology I started working on the related area of gender/sex recognition, namely 'Automatic Facial Expression Recognition' that was also the main issue of my Master thesis published in German at the University of Vienna in 2009 (cf. Musik 2009). For this purpose, I interviewed five computer scientists and behavioural scientists from Germany and Austria working in the field of Facial Expression Recognition and surveyed basic papers in this area of research. In addition, the history of facial expression research was traced back to its "beginnings," because the knowledge applied in Facial Expression Recognition systems is strongly grounded in its history. The case of Facial Expression Recognition researched in my Master thesis is one of three case studies referred to in Chapter Five of this PhD thesis.

I am referring to the second case study in Chapter Five when analysing the socio-technical construction of the ground truth which is also related to work done before I "officially" started work on this PhD thesis. The second case study in Chapter Five, 'Automated multi-camera event recognition for the prevention of bank robberies' presents empirical social scientific findings that were the output of an inter and transdisciplinary research project within the Austrian security research scheme KIRAS in which I was involved as a contract researcher at the Institute for Advanced Studies Vienna between 2009 and 2010. In the KIRAS scheme, projects developing security technology are obliged to integrate a partner from the Social Sciences and Humanities in order to ensure what is called "socio-political compatibility." In this case, the project consortium was managed by an Austrian Software Consulting company and was performed in cooperation with university computer scientists, a commercial bank, social scientists and the Austrian Federal Criminal Police Office. For the social scientists, the project was methodologically challenging as their role was far from being obvious at the start. This role could have been described as ranging from 'figleaf' or 'annex' to being a fully integrated and critically reflecting partner in this technological development.

A key question for the social scientists emerging over the course of the project was whether it is possible to identify and define suspicious behaviour in the context of a bank (more precisely the behaviour of robbers reconnoitring a bank) and if so, how this could be "translated" into an automatic computer vision system. This question was addressed by observing "normal" behaviour that is describing activities of bank customers in detail. Observations in two different branches of banks in Vienna, as well as video analysis of seven project-cameras installed in one additional branch, were performed. The method of non-participant observation was used, combined with video-analysis in Social Research (Knoblauch et al. 2006). Within four observation sessions a sample consisting of 236 people was obtained. To compare the observations of "normal behaviour" of bank clients with the behaviour of robbers reconnoitering a bank, records of interrogation footage of apprehended bank robbers were surveyed and security experts interviewed, as there was no video footage available to us.

"Going into the Wild:" From Lab Study Ethnography to Multi-Sited Visiography

I officially started this PhD project in 2009 when I got involved in the interdisciplinary research project 'Identification practices and techniques in Austria, 18th to 21st century' which was to be funded within the scholarship programme 'DOC-team' by the *Austrian Academy of Sciences (ÖAW)*, beginning in October 2010. When formulating the research proposal I noticed that the actual methodical approach is dependent on possibilities of field access and also on the path the relevant actors take. The general aim was to follow and observe the heterogenous actors and their respective paths and to describe their processes of networking. In this regard, I was aware of my own role as an integral part of these processes as I learned from Bruno Latour (Latour 2007: 50ff.). At this point in time, the overall research project and my own part in particular, oriented itself towards Actor-Network Theory (ANT) and I argued the application of ethnographic and technographic methods. That was because my central intention was to go where computer vision projects and objects were constructed, used and modified (Rammert & Schubert 2006: 12). My intention was to broaden my previous methodical approach and to gain a deeper understanding of computer vision and human-computer vision (re)configurations in this regard. To this purpose I reviewed methodical approaches that could be subsumed under the label of 'Lab Study Ethnography.'

In one of the most relevant studies for my project, Diana E. Forsythe analysed the complex relationship between computer scientists' beliefs and visions and the techno-scientific practice by ethnographic means, of fieldwork within the frame of an interpretative cultural anthropology approach (Forsythe 1993: 446). This approach did not only refer to the self-reporting of scientists (accessible especially through interviewing), but did also include observations of the specific practice (ibid.: 446). I decided to follow Forsythe, and thus chose laboratory-based participant observation as a research method to start with. This method did not include the lab as a secluded entity, but was also able to involve the activities and actions that take place outside the lab. This includes meetings with research promotion agencies, conference participation, public relations, events etc. and in doing so, follows Traweeks (1988) 'community centred approach,' an extension of the classic laboratory studies carried out by Latour & Woolgar (Latour & Woolgar 1986) and Knorr-Cetina (Knorr Cetina 1984). Moreover, Knorr-Cetina noted that it has been clear from the beginning of laboratory studies that external agents play an important role in the negotiation of scientific knowledge. This is what she called transscientific and transepistemic arenas (Knorr-Cetina 1984: 154ff.). Research promotion agencies, suppliers of equipment and materials, investors, politicians etc. (Knorr-Cetina 1995: 152) belong in these arenas. To sum up in a nutshell, "classic" laboratory studies showed firstly, in scientific research that actually everything (e.g. scientific facts, technological artefacts) is, in principle, negotiable (Bucchi 2004: 62). Secondly, the construction of facts and machines is a collective process (ibid.: 70) and thirdly, scientific practices in the lab are similar to practices in familiar ordinary life outside the lab (Sismondo 2010: 107; Beaulieu 2010: 454). This means there is no strict demarcation between science as carried out in labs and non-science outside the lab (Sismondo 2010: 107). So what is special about the technoscientific lab? Latour suggested moving outside it in order to find out how it had come to be seen that something special was going on inside (Latour 1983 cit. in Hine 2007: 659).

As I said before, the aim of my analysis was to follow and describe the paths of the heterogeneous actors connected by my guiding and framing research questions, with the computer vision laboratory as the starting point. In doing so, I wanted to note that the laboratory was not my central object of examination. I was not primarily interested in the laboratory itself and its specific culture, but my intention was to explore the place of the computer vision laboratory as one of the places where "seeing" is negotiated and where human and computer vision is (re-)configured. The question of how the computer vision laboratory is related to other entities was still significant. I asked myself if it is a 'Obligatory Point of Passage' (Callon 1986), a critical network channel often designed by the primary actor to ensure communication must pass through his or her domain (Bucchi 2004: 73). However, recent work in STS has challenged the centrality of laboratories and the notion of physical, face-to-face interaction in science (Beaulieu 2010). For example, when analysing humanities settings where should one look for the field (ibid.: 456)? Beaulieu called the concept of 'co-location' into question as a basic requirement for 'being in the field' and developed the idea of 'co-presence.' For the researcher this means not to principally ask 'Where do I go?' but to ask 'How can I establish co-presence?' Even if this approach was aimed at making highly mediated, distributed or non-lab-based fields such as the humanities accessible (ibid.: 453), it was still a useful notion for me, especially because it also challenged my relation to computer scientists working in computer vision. It offered me a way to enable research that is more cooperative than objectifying (ibid.: 462). This means: Not just going 'there' and studying 'them,' but to establish a form of togetherness[10].

How to Set the Research Agenda

As both technoscience and its analysis is a really complex and ever-changing endeavour, the discussion about methods in STS especially, seems to deal with how to set the

[10] On a related note Rammert argues for designing socio-technical systems together with technical sciences and called this method 'constructive sociology' (Rammert 2007: 36).

research agenda. Hine disapproves bounded territories as field sites and considers them somewhat artificial (Hine 2007: 655). That view fits the move in STS from going inside the lab to the exploration of sites outside the lab where science is practiced (as well). Thus, the spatiality of science becomes a topic of exploration in itself (cf. Law and Mol 2001). Hine advocates multi-sited ethnography and explains this by referring to the 'Zimbabwe Bush Pump' case (cf. de Laet & Mol 2000). The study of the Zimbabwe Bush Pump showed that technological artefacts are flexibly and variably defined and highly dependent on contextual judgement (Hine 2007: 663). Which is why it does not make sense to fix the site of research from the beginning, but to follow the path of the specific technical artefact if it is the artefact one is interested in. This again, is an argument for leaving the 'single case laboratory study' behind and doing research in a multi-sited manner. In my opinion, fixing sites for research is not done freely. Rather, it is based on experiences, ideas and an involvement with a topic in an iterative and cyclical way. It is part of the research process. If this is a call for being flexible, not to persist on one specific site from the beginning until the end, but to follow the actors as Latour (2005) suggests, then I share it. However if it is an argument for not fixing anything, it seems to be a bit exaggerated. Selections have to be made because it is not an option to follow everybody and everything.

In my view, this debate about how to set the research agenda quickly hides what issue is at stake. Nevertheless, I agree with John Law that we need to examine our methods for the directions in which they push us, and consider whether their biases and exclusions are desirable ones (cf. Law 2004 cit. in Hine 2007: 663). In this regard Law and Urry (2004: 393) referred to the performativity of research methods. This means that methods have effects and therefore can make differences. Opposed to the view that if a scientist follows the methodological rules she/he can understand reality properly (Law 2004: 5), Law argues that method is not "a more or less successful set of procedures for reporting on a given reality" (Law 2004: 143). Rather, method is performative and enacts realities, as well as generating non-realities or silences (ibid.: 113). Method simultaneously discovers and produces its research object (ibid.). This does not however imply arbitrariness in the use of method, but rather a recognition that different methods lead to different answers to different questions (Felt, Fochler & Strassnig

2010: 35). For example, the question which attitudes certain population groups have towards the phenomenon technology and what is the reason for these attitudes, can be answered entirely differently depending on the research methods used (e.g. survey or interviewing) (ibid.). In the case where research and scientific knowledge provide the basis for political actions and campaigns, the political dimension of methodological decisions is made clear (ibid.). The consequence is that it is no longer a question, as to what extent methods describe reality 'better' or 'worse.' Reflection on method is much more far-reaching (ibid.). Truth and politics go together one way or another (Law 2004: 149). But for Law, the question still arises whether one can recognise methods not only as purely technical procedures (ibid.: 143).

Hine argues that it may be dangerous to abandon methodological canon altogether (Hine 2007: 669). It makes sense also in my view, to deal with technical methodical issues. Methods are important for the social sciences and demarcate them from non-science and thus, offer advanced agency. More important is that methods show how the research has been realised and how specific insights or ideas originated. The crucial point is that one has to bear in mind that choosing a specific method does also mean choosing a specific political practice. Law also argues that standard methods are not wrong and still significant. In fact the problem with methods is the normativity that is attached to them in discourses about methods (Law 2004: 4). The consequence of this is that we should not abandon discourse on methodology. Instead of speaking about methods only in technical terms, we should learn to speak about methods in political terms as well.

As I have indicated before, the discussion about methods should not distract attention from the actual research interest or push it in undesired directions because the chosen method narrows down what is to be done. When I started, my research interest was basically in how computers are learning to see and how they are being taught to see, respectively. I was interested in the processes of configuration of vision between humans and computers, and proceeded in particular on the assumption that human vision is somehow translated and transformed into computer vision, which then influences human vision again. Thus, this might explain why my focus in the reseach process was first of all on vision as a more general concept. In my view, it was important to emphasise this focus and this was why I put forward the idea of "Visiography" as my primary methodical strategy. Visiography was thought of as a strategic tool or 'Gedankengebäude' to work with in the tradition of ethnographic inquiry. Therefore, using Visiography referred to established methodical traditions and made it possible to make use of well-established ethnographic methods, such as participant observation, document analysis or interviewing, while at the same time underlining my specific research interest in vision. Doing Visiography highlighted the core interest of my research; it clarified that this was not the laboratory itself (‚STS Lab Studies'), it was not about the folk or people (‚Ethnography'), it was not about social practices in general (‚Praxiography'), and it primarily was also not about one specific technology (e.g. face recognition) or the interactivities between humans and ready-made technology[11] (‚Technography'). Nevertheless, while stressing a conceptual boundary, the strategy of doing Visiography was ever entangled with the other sorts of "graphies" mentioned above because of its explorative character.

Approaching the Field, Data Collection and Analysis

After collecting the aforementioned material in the preceding studies, I started to collect further research interest-related, in-depth data following my visiographic strategy. To that end, I tried to establish contact to different relevant computer vision research institutions. In the beginning I got in touch with three major computer vision and pattern recognition institutions in Austria and decided to start with the laboratory that was most accessible to me. In a field preparation period, or what Hammersley & Atkinson call 'pilot research' (2007: 29), I spent one day of participant observation in a computer vision university laboratory and also interviewed the head of the lab on this occasion, in February 2011. I also obtained permission for participant observation there. Actually, I had also planned participant observation in one other computer vision

[11] Rammert & Schubert (2006), in their (German version of) ‚Technografie' understand technography as offering a methodological path that makes it possible to observe ready-made technologies (‚Techniken') closely in peoples´ everyday lives.

laboratory, but as the material turned out to be very comprehensive and my path had begun to lead in another direction during my period in the first computer vision laboratory, I changed my plans.

My observations in the chosen lab started on the 4th of October 2011 and ended on the 29th of November 2011. I did not only stay in the lab itself, but also went to other sites surrounding the lab. These included university classrooms, the office of a computer vision company connected to the lab, a funding call event at one of Austria's big funding agencies, the site of an exhibition of IT-innovations and adult education classes (Volkshochschule). Altogether in this period of time, I spent 177 hours in the field and produced 220 pages (A5) of handwritten fieldnotes, two half hour group discussions with about ten members of the computer vision laboratory and the aforementioned interview with the laboratory head (63min).

The two group discussions took place in the lab at the end of lunch break in one of the smaller rooms, which is the working place of four members of the team and also the lunch break room for most of the lab's members, especially as the coffee machine is situated there. The first group discussion took place on the 9th of November, so at the beginning of the second half of my fieldwork. Some days before the first group discussion, I asked the members of the lab during their weekly meeting, if they would be interested in a short group discussion about their work in combination with the input of my observations. The idea behind this, was not only to get information and views from the lab members, but also to inform them about and discuss with them my observations as a spectator in their daily working environment.

I decided on group discussions, because I had become very interested in some issues derived from my observations. My focus as a social scientist was to identify and propose problems that insiders were unaware of (Forsythe 2001: 137). This is why Forsythe stresses the analysis of the systematic comparison of the insider and outsider view, "This includes detecting tacit knowledge, something that by definition is invisible to insiders" (ibid.: 137). I wanted to examine and compare my outsider view to the insider views of the computer scientists and I also wanted to validate some of my emerging hypotheses. Through 'open coding,' I continuously examined my observational field notes[12] for categories and classifications following a similarity criterion (Gobo 2008: 228ff.). The aim of classification in ethnography is “to deconstruct the events and actions observed and segment them among a series of concepts” (ibid.: 234). I did so, by using terms employed by the lab members. For the first group discussion, I told the lab members at the beginning of the discussion (which I recorded on mp3 and transcribed afterwards) that I was interested in two basic terms commonly used in the lab: ‘automatic’ (“automatisch”) and ‘to work’ (“funktionieren”). When I mentioned ‘to work’ for the first time, the group discussion participants immediately started laughing out loud, as if everybody knew exactly why this term is significant. This collective laughing confirmed my aim of discussing this observational outcome within one of the two group discussions. I started by asking them to tell me about their experiences concering the context of when and how they use the terms ‘automatic’ and ‘to work’ (e.g. in public, at presentations, project proposals, in the lab) and what relevance these terms have for their work. Both the observations and group discussions provided me with comprehensive material that was the basis for the third case study in Chapter Five and for Chapter Six.

Doing ´Sampling’: Defining Units of Observation

“Traditional” ethnography books tell of the extreme importance of defining units of observation once you are in the field (cf. Gobo 2008: 98). For example, Gobo advocates clearly defined units such as persons, households, groups or associations (ibid.: 99). Others seem to be more flexible and note that the match between research problems and cases selected must be continually monitored (Hammersley & Atkinson 2007: 35). Fetterman (2010: 35) advises two approaches to the decision on how to sample members of the target population. First, choose who and what not to study. Or second, select who and what to study. That is, the sources that will mostly help in understanding life in a given community. Most common is however ‘judgmental sampling.’ This means a reliance on one’s own judgement to select the most appropriate members. The most

[12] Next to observational field notes I also set up methodical, emotional, and theoretical field notes (cf. Gobo 2008: 208ff.).

useful recommendations are given by Hammersley & Atkinson (2007: 28). They mention that the role of pragmatic consideration must not be underestimated (ibid.: 30) and suggest sampling within cases in three major dimensions: time, people and context (ibid.: 35ff.). Time is an often neglected dimension, but as they argue, attitudes and activities frequently vary over time. The aim is to establish adequate coverage, consider seasonal or annual cycles, pay some attention to special occasions and to include routine as well as observing the extraordinary. When talking about people they mention categories like gender, race, ethnicity, age, occupation or education which can be either 'member identified' or 'observer identified.' Context is the third dimension (ibid.: 39). They refer to Goffman's frontstage/backstage conception and note that for example, behaviour of teachers can differ in classroom and staffroom[13].

Sampling was an important process for me as my experience with computer vision laboratory work so far was that computer scientists in the lab work simultanously in many different areas and the areas change quickly also depending on the funding available. The question therefore arose whether I should concentrate on one specific project in the future, because it was really not easy to jump from one project to another within the same day.

Following Atkinson & Hammersley, a sampling dimension I was considering, was time: Different projects have different time-spans and as I was told in the lab, there are projects with more symbolic character and endless administrative work, and other projects with great flexibility and the freedom to experiment. Actually, I had planned to stay in the lab for about two to three months, but if no appropriate project were to have come up, I would have needed to stay longer. The question and uncertainty at the time, was if the computer scientists in the lab were going to be working on projects or issues that were especially interesting for me and my project. As it turned out fortunately, within two months of being in the lab, I was able to collect very comprehensive material.

[13] see also Chapter Six for more details on Goffman's frontstage/backstage conception.

Another important dimension of sampling was people. In my case, the most important criterion was occupation/education. Put simply, I experienced three different positions inside the lab: first, the lab head, second, researchers (especially PhD students and/or contract researchers), and third, one senior researcher who led and organised projects. When describing this position to Lucy Suchman in a conversation, she made use of the label 'storyteller' to describe him and similar persons she had got to know earlier. This term is sustained when looking at his list of presentations at the lab's webpage. Next to the conference presentations, there a number of presentations by invitation in different areas of society also is listed, especially in the areas of economy/industry, education and video surveillance/data protection.

As it turned out, I followed and was with four to five PhD researchers in particular that were working in lab on a very regular basis and on projects that attracted my interest. There was also the matter of access. I had the impression that these lab members did not mind my following them. Another pragmatic consideration was limited funding. I have to admit that it was impossible to follow the actors unrestrictedly, owing to lack of funds, for example when they left the city for project meetings or conferences. Nevertheless, by following these researchers I also got in touch with two other positions in the lab as they were frequently in touch with each other.

Context, the third sampling dimension mentioned by Hammersley & Atkinson (2007:39) was automatically relevant for my sampling, because by following the selected researchers I participated in different contexts, for example also outside the lab as mentioned above. In particular, computer vision presentations and demonstrations turned out to be special occasions that are centre stage in Chapter Six.

During the course of my research something else brought turbulence into my project and I would have eliminated it, but could not desist from including it in my thesis. It was during my field work that I came across a nationwide system already in operation that contains at its heart, Image Processing Algorithms that are designed to recognise patterns: the so-called 'Automatic Toll Sticker Checks' ("Automatische Vignettenkontrolle" - AVK) on Austrian motor and expressways. As it happened, I did not gain proper access to the system, because the operator ASFINAG rejected all my efforts to learn more about it. So, I was neither able to do any form of (participant) observation nor was I able to interview people about the system, so my analysis had to be narrowed down or shifted to the analysis of publicly available documents. Because this process of working out how to get access to the system was in itself a finding of my visiographic strategy in terms of a specific 'Culture of Secrecy,' I describe the precise methodical procedure for this particular case within Chapter Four. Notwithstanding that the case presented in Chapter Four is based on a different form of empirical material than is the case with my field work material in a computer vision laboratory and its surroundings, it nevertheless has to be seen as an integral part of my research. This is also in line with ethnographic tradition according to Hammersley & Atkinson, who state (2007: 3) that ethnographers do in fact gather "whatever data are available to throw light on the issues that are the emerging focus of inquiry." That meant that during the course of my research I realised that it was important to examine the issue of one of the first nationwide IPA systems in operation in Austria, the so-called 'Automatic Toll Sticker Checks.' As my only entry point was with the use of publicly available documents (no less interesting), the analysis and involvement using these available materials was the appropriate way to throw light on this issue from a different perspective.

Analysis

Analysis in ethnography "is not a distinct stage of the research" (Hammersley & Atkinson 2007: 158). It "preceeds and is concurrent with data collection" (Fetterman 2010: 2) and there usually are no "instructions on how to conduct data analysis" (Gobo 2008: 226). Hammersley & Atkinson suggest ignoring any accounts that try to offer a standard set of steps to be taken. Instead, they emphasise that it is most important to understand that "data are materials to think with" (Hammersley & Atkinson 2007: 158). They suggest "Grounded Theorizing" (not "Grounded Theory" which is the product of Grounded Theorizing), that is, "a way of working with data (…) in order to generate and develop ideas" (ibid.: 159). In consequence, a constant interplay and moving between data and ideas is at the heart of this procedure and enables the researcher to construct a theory or tell a story. In this regard, I oriented myself towards Latour's manner of proceeding. When he writes "if a description remains in need of an explanation, it means that it is a bad description" (Latour 2005: 137), he means that we have already given the explanation if we give a "good" description. So, giving a good and 'thick description' (cf. Geertz 1973) was at the heart of my data collection, analysis and writing process that finally turned me in the direction of my engagement with Image Processing Algorithms (IPAs) at the centre of my story. Retrospectively, I can note that what started with the methodical strategy of Visiography turned into the analysis of Image Processing Algorithms.

Chapter Two

To See. To Recognise.

To See.

Some years ago in 2010, I visited the *Biennale for International Light Art* that took place in the context of the *European Capital of Culture RUHR 2010*. The Biennale was titled 'open light in private spaces.' The exhibition project presented works of art by internationally renowned artists in 60 private spaces belonging to inhabitants of different cities and towns in the German Ruhr district. Next to impressive light art installations that caught my eye, the most rememberable piece of art was somewhat different. I happened to walk into one of the exhibition venues, an opthalmologist´s surgery, and took a seat in the waiting room. After a while, I was called into the consulting room. When the door of the room closed behind me I did not see anything at all; it was totally dark. I waited and expected some form of light to appear, at least something that my eyes could see, but suddenly a voice began to speak. It started to state clearly in German, words that could be translated as 'blood-orange,' 'basalt-grey,' 'curry-yellow,' 'tobacco-brown,' and so on. It was a fascinating experience for me, because with every different word, the description of a colour with a prefix from everyday life that specified the respective colour, I was able to imagine exactly what kind of colour was meant. In most cases, I saw amazingly clear pictures in front of me that were associated with the colour and the entity that described the colour. This artwork with the title 'dunkellampe' ('dark lamp') by the Reykjavík-based artist Haraldur Jónsson invited me to contemplate what it does mean to see and to recognise. What do we as a society actually mean if we speak about seeing or the ability to see and to recognise? The involvement with this question is the basis for understanding what it really means if the ability to see is to be transferred to computers and Image Processing Algorithms. In what follows, I first engage with some lessons learned from Blindness Studies and lead on to what can be termed Visual Culture Studies. Because images play such a crucial role in our contemporary visual culture, I also deal with a concept—the Social Studies of Scientific Imaging and Visualisation (SIV)—that analyses scientific images and their role in society from a sociological perspective. Because the ability to see, understood in sociological terms, is very much the ability to recognise something meaningful in a specific entity, the second part of this chapter engages with recognition.

Firstly, I deal with the history of recognition by referring to the history of identification practices and techniques and secondly, I connect this history with current technologies of facial and pattern recognition.

Seeing is more than just Seeing: Lessons Learned from Blindness Studies

An interesting and promising approach to this question is an involvement with blindness which would seem to be the opposite of seeing at first glance. But, considering my own experience in the dark room of the opthalmologist´s surgery, I started to question this contrast. Physiologically, I did not see anything with my eyes, but was still able to see clearly the colours and things mentioned. Did I draw on some kind of "saved" images in my mind that I automatically associated with the stated colours in the opthalmologist´s surgery? Also, blindness studies show that the physiological ability to see is not necessarily a prerequisite for participating in visual-aesthetic discourse (Mraczny 2012: 193). Mraczny interviewed blind people to learn about seeing and visual culture. He found out that practices of seeing that always include some form of visual meaning, are guided and produced through social institutions in the framework of a dominant visual culture. Visual culture acts as an ordering device that simultaneously defines practices of seeing and of blindness. It constitutes standards and instructions how and what has to be seen at what time (ibid.: 197). Simultaneously, in creating "normal" and "deviant" visual subjects it constructs blindness as a "disability," which also leads to discriminating effects (ibid.: 189). Länger (2002: 8f.) reported that blind speakers use the same visual language as non-blind people. This use of visual vocabulary shows the expertise of blind people in their involvement with visual situations and semantics (Mraczny 2012: 190), meaning that although blind people do not see physiologically, they see and think in visual categories and participate in the visual world. That is, amongst other things, due to the communication and interaction with non-blind members of society, who talk with blind people about optical and visual phenomena. Insofar, the visual becomes a highly relevant part of blind people's everyday lives. Especially in public situations when it comes to encounters with strangers, visual expertise is an essential condition for "correct" behaviour. For example, in German speaking countries where there are two different forms of second person singular pronouns (generally "du" for younger people or contemporaries and a more polite "Sie" for elders or people with a visibly higher status), the visual inspection of somebody in order to estimate his or her relevant status, age or sex, is the basis on which the decision on how to address the respective person correctly is made (ibid.: 195). For most non-blind individuals, seeing is a quasi-natural practice, taken for granted. It seems there is a continuously ongoing defence and reproduction of a visual order, because seeing does count as the "natural" and "realistic" way to perceive the environment (ibid.: 197). Also, academic writings continuously attest a predominance of sight over other senses in Western cultures (Burri, Schubert & Strübing 2011: 3; Sturken & Cartwright 2001: 300) and this is generally accepted:

> "Vision has played the role of the sovereign sense since God looked at his own creation and saw that it was good, or perhaps even earlier when he began the act of creation with the division of the light from the darkness." (Mitchell 2002: 174)

Next to Mitchell's definition of divine purpose there can be a discussion on whether the sense of sight is preferable to the other senses, being as it is, a reflection of broader visual culture. What would visual culture then be seen to be and how did such a dominant visual culture come about? Is it because sight and speech count as more public than the more personal senses of touch, taste and smell, or because pictures, graphs, and diagrams make it easier to show results and to present and create evidence (Burri, Schubert & Strübing 2011: 4) in areas such as the sciences.

Seeing with and in Images: The Visual Culture Studies Approach

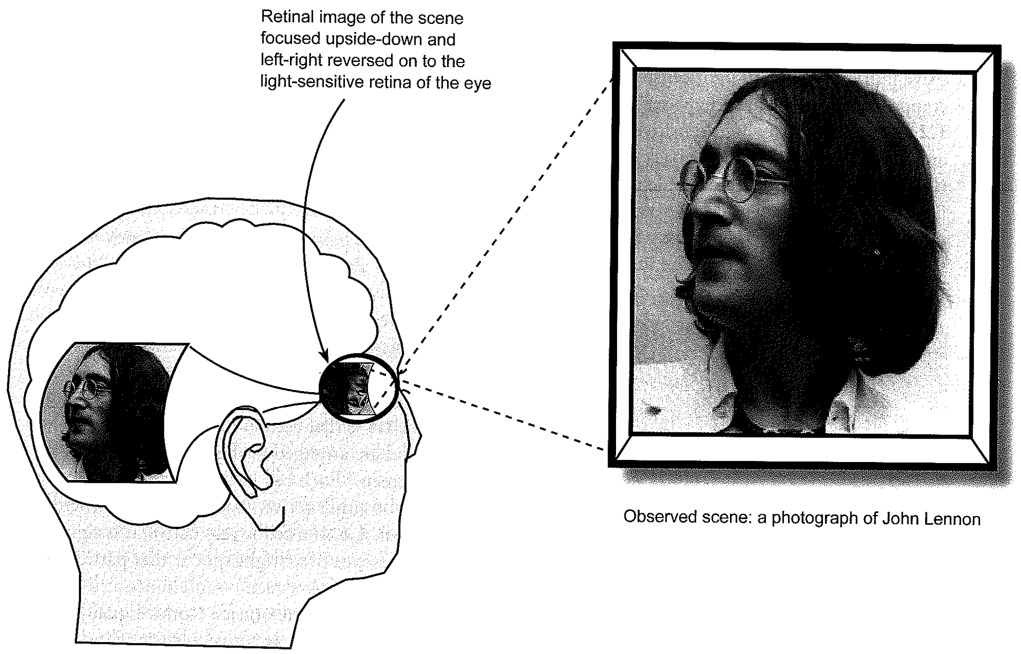
Consistent with these insights, academic discourse on visual culture is not so much about seeing in everyday life but about images and visualisations: seeing with, and in images. This kind of seeing seems to have become more significant since the advent of a widespread use of digital consumer cameras and the production of a vast number of digital images. For example, in December 2012 the internet image platform Flickr singularly hosted more than eight billion images and now as you read this text, this amount may have increased significantly. It would take about three to four westerners´ lifetimes to look at every single one of these images for just one second. When it comes to video image data these figures are even higher: for example, it would take more than 16 years of day and night watching to see only the new videos uploaded to the video-sharing website YouTube, in just one day in 2013. So we can ask if the fiction of a culture totally dominated by the visual is true and has become a real technical possibility of global extent today (Mitchell 1997: 10). Have we really witnessed a turn from words to images; what has been called an 'Iconic Turn' (Boehm 1994) or 'Pictorial Turn' (Mitchell 1994)? However, in this regard we have to mention here, that pictures and visualisations are not a new phenomenon of the digital world. This is why Boehm speaks of the return of images. Images have been of importance for many centuries. Remember, for example, the illustrations in the *biblia pauperum* that was a crucial means of communication in the Middle Ages, used to teach the Christian doctrine to illiterates (Schnettler & Pötzsch 2007: 2). Images also played an important role in institutional regulation and categorisation or archivation of people according to types, used for example in phrenology, craniology and physiognomy in the 19th century (Sturken & Cartwright 2001: 281). Starting from the insight that images and visualisations have been significant for centuries, the view that something is fundamentally new about the visual has to be questioned.

However the answer might be, today images represent social realities and shape the ways people think, interact (Burri 2012: 46) and see. Images do have different logics, one of them being that "images allow social actors to perceive visual information simultaneously" (ibid.: 49), independent from being together in one place. Another advantage is that images enable things to be seen at a glance, which is often set in contrast to other forms of perceiving social realities that can only be processed in linear sequence. Think about reading a text or listening to someone speaking. However, also images can be read in a nearly sequential way. The eye moves over the image on a structured path, in a not entirely coincidental way. The eye moves from one point to the next, though chronology can vary at different times and also differs in individuals (Breckner 2003: 37).

Nevertheless, images and visual information play a highly important role in the (re-) production of society. According to Burri, images can be defined in different ways. They can be seen either as 'inner images' or mental imagination or as physical expressions such as a bodily performance or, in a third perspective as 'the world as image' (Burri 2012: 46). Burri herself defines images as artefacts that are both visual and material objects and can be conceptualised as technical objects. In her work on the sociology of images, she suggests the concept of 'visual logic' to analyse images sociologically (ibid.: 45) from the perspective of social practices. Here, three different visual dimensions of images play a role: First, the visual value refers to the non-discursive characteristics of images and can be seen as the surplus value of images that makes images different from auditory, taste or tactile signs. "A picture is worth more than a thousand words" is a popular saying that describes the visual value of images quite well. The most important characteristic of the visual value of images for my understanding of seeing and images is "that images cannot entirely be transformed into textual or nummerical signs without losing some of their advantages" (ibid.: 50). In the context of 'computers and the ability to see' one has to ask, whether the view that human perception of a large amount of information by the means of images happens "at once" (ibid.: 50), is still maintainable, because what does "at once" actually mean? Does it, in fact, mean "extremely quick?" Does "extremely quick" mean at such a high speed that for most humans it appears to happen "at once?" But what actually goes on inside our heads when we see something like an image? How long does it take from seeing something to processing this something in our heads? A metaphor used many times for seeing is the photographic metaphor. It has its origins in the observation that our eyes are, in many respects, like cameras as both are equipped with lenses.

> "… and where the camera has a light-sensitive film or an array of light-sensitive electronic components, the eye has a light-sensitive retina, (…), a network of tiny light-sensitive receptors arranged in a layer toward the back of the eyeball (Latin rete—net). The job of the lens is to focus an image of the outside world—the retinal image—on to these receptors. This image stimulates them so that each receptor encodes the intensity of the small point of light in the image that lands on it. Messages about these point by point intensities are conveyed from the eye along fibers in the optic nerve to the brain.

The brain is composed of millions of tiny components, brain cells called neurons." (Frisby & Stone 2010: 2).



1.1 An "inner screen" theory of seeing
One theory of this kind proposes that there is a set of brain cells whose level of activity represents the brightness of points in the scene. This theory therefore suggests that seeing is akin to photography. Note that the image of Lennon is inverted in the eye, due to the optics of the eye, but it is shown upright in the brain to match our perceptions of the world—see page 8. Lennon photograph courtesy Associated Newspapers Archive.

Figure 1 An "inner screen" theory of seeing (Frisby & Stone 2010)

What Frisby and Stone refer to as the "inner screen" theory of seeing, does work in a similar way to photography. When optical information, for example

> "a photograph is being observed, (...) the pattern of activity on the "inner screen" resembles the photograph. (...) As soon as this pattern is set up on the screen of cells, the observer has the experience of seeing." (ibid.: 3).

The biological process of seeing is certainly the basis for understanding what seeing or the visual is. But from a sociological point of view we can start from the premise that seeing is subject to change, both culturally and historically (Tomomitsu 2011; Kammerer 2008; Burri & Dumit 2008, Rövekamp 2004). Seeing has been conceptualised in different ways in different centuries and connected to this, various technical phenomena appeared. For example, one point linear perspective in the Renaissance period, the invention of the microscope and telescope, the lighting of cities, photography, surveillance cameras, or computer-generated virtual realities (Rövekamp 2004: 15). "Seeing is not just the physical act of looking, it is also about doing, (...) it is about making inclusions and exclusions, presences and absences" (Tomomitsu 2011:14). Amoore (2007) notes that seeing is a specific form of visual culture and refers to Mitchell, who claimed not to limit visual culture to images and the media, but also to everday practices of seeing and showing (Mitchell 2002: 170).

Collins distinguishes a formal (or pattern recognition) model of seeing from an enculturational model (Collins 2010:11). The formal model of seeing "involves recognizing what an object really is by detecting its distinguishing characteristics." The enculturational model of seeing stresses that the same object may be seen as many different things. For example, as Goodwin notes (Goodwin 1994: 606), "an archaeologist and a farmer see quite different phenomena in the same patch of dirt." The farmer would be more interested in the quality of the soil, but the archaeologist in "artifacts that provide evidence for earlier human activity." For Goodwin, seeing emerges through the interplay between *a domain of scrutiny* (e.g. patch of dirt) and *a set of discursive practices* (e.g. highlighting) that are being deployed within *a specific activity* (e.g. planting crops). In the article *Professional Vision* Goodwin investigates seeing "as a socially historically constituted body of practices through which the objects of knowledge which animate the discourse of a profession are constructed and shaped (ib.: 606)." He attached importance especially to three practices of seeing: Firstly, *Coding Schemes* that are used to transform materials of the world into objects of knowledge, into categories, classifications and events that are relevant to the work of a specific profession (ibid.: 608). Secondly, *Highlighting,* that is, "making specific phenomena in a complex perceptual field salient by marking them in some fashion." (ibid.: 606). This can happen through coloured markers, post-it notes or handwritten annotations, for example. As a consequence this highlighting activity does not only shape one's own perception, but also that of others (ibid.: 610). Thirdly, *The Production and Articulation of Material Representations* such as diagrams, maps, graphs and photographs. Here, Goodwin refers to the central importance of 'inscriptions' in the organisation of scientific knowledge (cf. Latour & Woolgar 1979, 1986). According to Latour and Woolgar inscriptions can be summarised as all traces, spots, points, histograms, recorded numbers, spectra, peaks and so on, meaning everything more basic than writing (Latour & Woolgar 1986: 88). Inscription devices such as figures or diagrams transform pieces of matter into written documents, which are directly usable by one of the individulas working with these pieces of matter (ibid.: 51). They are especially useful in 'rhetorical situations,' because they are easy to transport and remain immutable and can be reproduced and altered in size (cf. Burri 2012: 48). Goodwin´s extraordinarily comprehensive text on *Professional vision* provides a great example of

> "… how the ability to see a meaningful event is not a transparent, psychological process, but is instead a socially situated activity accomplished through the deployment of a range of historically constituted discursive practices (ib.: 606)."

Goodwin exemplifies this with the so called 'King Trial,' in which four white policemen were charged with beating Rodney King, an African-American motorist, who has been stopped for speeding in the US in 1992 (ibid.: 606). The incident had been videotaped and for the prosecutor it was absolutely clear, objective evidence showing uncontrolled and brutal violence against Rodney King. However, the lawyers defending the policemen did not treat the tape as evidence that spoke for itself. Rather, they were able to transform the perception of the tape, in so far as it was evidence, into "a very disciplined and controlled effort to take Mr. King into custody" (ibid.: 617). With the help of a coding scheme delivered by experts that showed how police usually work, a 'perceptual transformation' had been accomplished (ibid.). Goodwin concludes that "the perspectival framework provided by a professional coding scheme constitutes the objects in the domain of scrutiny that are the focus of attention" (ibid.: 622).

Tomomitsu shows that it is not only discursive practices that shape seeing, but also a highlighting of the embodied material practices of seeing when analysing practices of scientific imaging (Tomomitsu 2011: 20), which I would identify as a specific form of professional vision. She describes three practices of how scientists enact 'seeing.' The first practice is *purification* that refers to "how scientists contain their objects through sample preparation in the laboratory." The second practice is *tinkering*, describing "the various ways that instrumentation is adjusted and fiddled with." Finally, the third practice is *touch* that refers to “how objects require physical handling or alteration to see and make visible” (ibid. 18).

During the production of images, aesthetic criteria are also applied to images and visualisations (Burri 2012: 50). This insight contradicts the view that images and especially scientific images are the quintessence of objectivity and that they have nothing to do with aesthetics. On the contrary, one has to ask, how far are aesthetics permissible in maintaining a status of objectivity? In the wake of the emergence of photography, possibilities of compensating for the weaknesses of the human eye occurred. For example, photography facilitated the permanent and objective securing and preservation of evidence in police work. Because photography replaced the practice of local inspection it acquired the label of being the most immediate and direct experience of truth (Heßler 2006: 26). Currently, we seem to live in a culture in which image evidence is crucial. The last bastions that struggle against image evidence are beginning to totter, exemplified by the decision to apply goal-line camera technology in world championship football matches as announced by the *International Federation of Association Football (FIFA)* and the *International Football Association Board (IFAB)* in February 2013.

Returning to the discussion of scientific images, Lynch (1990), in his ethnomethodological study of a laboratory demonstrated how visual representations are fabricated through processes of mathematisation and the selection of visual elements. Both of these modifying interventions aim at making the visual object more useful for the researcher by transforming, neglecting, or boosting visual signs (Burri 2012: 50). What is seen in an image also depends on how the image is interpreted. Interpretative practices are shaped by cultural traditions of seeing and by professional skills in reading images (ibid.: 51). How an image is interpreted is also dependent on the social status of the respective actors involved in the interpretation processes. At this point, we can propose the question of which social status, computers, machines or Image Processing Algorithms have achieved when they are involved in interpretation processes.

Next to the visual performance and the visual value of images, Burri refers to a third visual dimension of images that play a role in social practice; visual persuasiveness. It underlines the importance of visual information in communication as well as the rhetorical power of images (ibid.: 49). For example, medical images are regarded as scientific images and thus viewed as 'objective facts' with authoritative power (ibid.: 52). This notion of photographic truth and images as evidence, especially in connection with scientific images, hinges on the idea that the camera is an objective device for the capturing of reality, and that it provides this objectivity, despite the subjective vision of the person using the camera (Sturken & Cartwright 2001: 280). The idea of seeing farther, better, and beyond the human eye had tremendous currency. The camera was imagined as an all-seeing instrument, especially in 19th century (ibid.: 281), but still is today. In the 1890s X-Rays were seen to offer a new vision of the human body, today new imaging technology in medicine allows doctors to see their patients in a completely new manner - one that is way beyond human sight. Scientific images are thus understood as providing the capacity to see "truths" that are not accessible to the human eye (ibid.: 281). What is interesting about current developments in the automatisation of seeing, is that next to the purposes of exploring dimensions that are invisible to the human eye, there are more and more attempts at trying to be able to see in a manner almost self-evident to humans in their everyday practice of looking and seeing that does not need any professional skills; for example the detection or recognition of faces. I will come to this example more closely in the section about recognition, as it provides us with a very interesting discussion about what it means to be an "expert at seeing."

In scientific practice, images are used as arguments or evidence that Burri referred to as visual persuasiveness. Knorr Cetina describes these practices as 'Viskurse.' Viskurse are the interplay of visualisations and their ongoing embedding in communicative discourse (Knorr Cetina 2001: 307). With this point of view, visualisations can be seen as allies that hold the disposition of rhetorical evidence. That is because they act as inscriptions, signals produced by machines from nature that do not only describe nature (ibid.: 309). Scientific images as well as images in general, always refer to something. This referring is not restricted to "real" phenomena, but can also relate to other images or (measured) data. This indicates that images are involved in various transformations and are placed in chains of reference with the aim of making something visible. For that reason certain phenomena are constructed through scientific images (Heßler 2006: 28). For example, Hanke showed that around 1900 scientific visualisations within physical anthropology constructed the expected differences in "race" and "sex." Visualisations of human bodies had been adjusted and edited in such a way that differences became apparent. Through the use of mechanical recording devices, the resulting visualisations counted as objective, because human (mis)interpretaion was eliminated (Hanke 2006: 242f.).

A way to study Scientific Images: The Social Studies of Scientific Imaging and Visualisation (SIV)

Human vision is inevitably historically and culturally specific in all of the presented conceptions. These conceptions should provide a really interesting challenge to assumptions about human vision in computer vision projects that emphasise the formal or pattern recognition model of seeing much more. Burri puts her case for a sociology of images that must investigate the processes by which image interpretation is interactively negotiated in social practices. This means that the focus is not only on the images alone, but on social practice, and contexts of image production, interpretation and use have to be taken into account (ibid.: 53f.). This focus is similar to what Burri and Dumit developed in their concept of the *Social Studies of Scientific Imaging and Visualisation (SIV)* (Burri & Dumit 2008). The development of SIV has been a consequence of a general practice turn in social theory (Schatzki, Knorr-Cetina & von Savigny 2001) and they strongly refer to STS laboratory studies (Lynch, Knorr Cetina, Latour and Woolgar). However, their focus goes beyond the scientific laboratory and community. One major question is what happens when scientific images leave academic territories and extend or travel to other arenas and contexts? They talk about the "social life of images" that includes the total lifespan of images from production to use (Burri & Dumit 2008: 299f.). Mitchell also dedicates an individual "inner life" to images. He even conceptualises images as living organisms due to the possibility of lifetimes of images ending; images die when they are not used any longer (Mitchell 2005). Burri and Dumit discuss the social life of images on three thematic clusters that are artificially separated: production, engagment and deployment of images. It is important to note that because seeing is so often believing, one main SIV concern is to demonstrate how the making and using of images come together with seeing and believing in practices of scientific truthmaking and perceptual convention (Burri & Dumit 2008: 300).

Regarding the production of images Burri and Dumit ask the question "how and by whom images are constructed by analyzing the practices, methods, technology, actors, and networks involved in the making of an image" (ibid.: 300). The example of magnetic resonance imaging (MRI) shows that the production of images is dependent on a series of decisions concerning the machines, data, parameters, scale, resolution, and angles. These decisions and selections "do not depend on technical and professional standards alone but also on cultural and aesthetic conventions or individual preferences" (ibid.: 301). The production process of scientific images is far from being a neutral process, but is shaped by sociotechnical negotiation. This is also dependent on local variation in the question of who is able to read images and who is allowed to read them, because visual expertise is its own form of literacy and specialisation (ibid.: 302).

Once images have been produced, the question arises how images are used and how they are talked about or talked to. This is what Burri and Dumit are interested in when they conceive the notion of an engagement with images. The focus here is on the process of making data meaningful and thus, their becoming meaningful images. As a consequence, images should be treated as actors that are actively involved in scientific practices. In these practices the question arises of what impact images have on the objectivation of knowledge by reducing uncertainties that occur during observations (ibid.: 302).

Finally, the deployment of images invites a look at the trajectories of images as they leave the scientific and academic field. In areas outside academia, scientific images meet and interact with different forms of knowledge. In these various places away from science, the behaviour of images is characterised by contradictory attributes. On the one hand scientific images occur as expressions of nature that are associated with inducing efficacy. On the other hand, they are open to semiotics and interpretation. Representations are never completely self-explanatory. This can lead to a shift in meaning in different environments (ibid.: 304f.). As one could see in the King Trial mentioned above (cf. Goodwin 1994), courts are another key site where visual authority is regularly and formally challenged (Jasanoff 1998 cited in Burri & Dumit 2008: 306).

To Recognise.

One of the most famous court proceedings, where visual authority was challenged and visual ambiguity demonstrated, took place in the French Town of Rieux in 1560. The Trial at Rieux was about the imposture of Martin Guerre. In the 1540s Martin Guerre, a rich peasant living in the French village of Artigrat left his wife, child and property and only returned to his family after many years. During his absence and after three years of living in an agreeable marriage, the wife, Bertrande de Rols, said that she had been tricked by an impostor claiming to be her husband, Martin Guerre, and brought him to trial (Davis 1983). The trial at Rieux was a great example of how to establish the identity of a person without doubt, at a time when there were no photographs, tape recorders, fingerprinting, ID cards, birth certificates, and so on (ibid.: 63). How was it possible to identify without any doubt, and that means to clearly recognise, a person in the 16th century? During the trial many different methods were used, for example testing of the imposter´s memory. The insight which is not really surprising, was that most of the methods had to do with seeing and recognition and were thus of a visual nature. For example, his handwriting was examined, but it was not possible to find an earlier record of his writing. Witnesses were asked to identify him and special marks on his face and body were considered, thus making physical and bodily recognition was one of the most important factors in the trial (Higgs 2011: 23). The accused, Martin Guerre was also tested to see if he was able to recognise members of the family. In total one hundred and fifty people came to Rieux to testify on the identity of Martin Guerre or his possible impostor. The outcome of these testimonies was as following: “Forty-five people or more said that the prisoner was Arnaud du Tilh alias Pansette, or at least not Martin Guerre” (ibid.: 67). Conversely, “about thirty to forty people said that the defendant was surely Martin Guerre; (…) and these included Martin’s four sisters, and his two brothers-in-law“ as well as a member of the one of the most respected families in the area (ibid.: 67). This disagreement of witnesses on the identity of the accused person continued in the description and recognition of special bodily marks.

> "Some witnesses maintained that Martin had been taller, thinner, and darker than the accused, had a flatter nose and more projecting lower lip and a scar on his eyebrow that was nowhere to be seen on this impostor." (ibid.: 67)

While these witnesses challenged the identity of Martin Guerre, other witnesses confirmed his identity by insisting that he "had extra teeth in his jaw, a scar on his forehead, three warts on his right hand; and each of these was discovered on the prisoner" (ibid.: 68). Next to these witnesses who took up a clear position, most witnesses (around sixty or more) "refused to identify the prisoner one way or another" (ibid.: 68). To abridge the whole story here: In the trial at Rieux, the judge declared the accused person guilty of taking on the name and person of Martin Guerre, even though the evidence was hard to evaluate (ibid.: 71). During the second trial, "the condemned man appealed immediately to the Parlement of Toulouse, protesting his innocence" (ibid.: 72). During this trial the true Martin Guerre returned and the aforementioned and identified impostor, Arnaud du Tilh, was convicted and sentenced to death for fraud (ibid.: 73ff.).

What is especially interesting about the case of Martin Guerre is the demonstration of the difficulties in clearly recognising and thus definitely identifying a person and confirming his or her stated identity. Identifying a person can be regarded as one specific form of seeing and recognising where the focus lies on bringing together and matching a body and an individual identity. In the case of Martin Guerre, the main challenge was to identify whether the returned person that claimed to be Martin Guerre was one and the same as the one who had left. This involved a comparison between the Martin Guerre who used to live in Artigrat before he left and the Martin Guerre who returned to Artigrat several years later. The witnesses in the trial had to compare the appearance of the returned and present Martin Guerre to the picture of the Martin Guerre who had left that was somehow "saved" in their memories as there was no form of visual depiction available to the court. As described before, different strategies and practices were used to compare these two versions of Martin Guerre, especially the description and positive or negative recognition of special marks. Other stories of imposture also stress the performative factor of certain identities: Perkin Warbeck, who maintained that he was Richard Duke of York and claimed to be the King of England at the end of the 15th century, is one example of how different means and modes of performing were necessary to claim a particular, in this case, royal identity (Higgs 2011: 20ff.)[14]. Thus, the performative efforts made by the impostor of Martin Guerre, in acting like the peasant Martin Guerre may not be underestimated. From today's perspective, we are able to describe a long history of changing practices and techniques in recognising and identifying persons like Martin Guerre. An involvement with the history of identification practices and techniques enables an understanding of the specificity of today and future modes, practices, and techniques of recognition and thus, of seeing in general.

In what follows, I engage with the history of identification that in my interpretation can be understood as a history of recognition. In this history of recognition, wanted posters, the securing of names and surnames and their ongoing registration, photographs and photographic collections, anthropometry, dactyloscopy, and DNA analysis play central roles. An identification technique of the present and of imagined futures is automated facial recognition, to which I will turn my attention later and give special consideration to, as questions of seeing and recognition can be discussed with this example. But, as the history of identification as a history of recognition of individual persons will show, this history cannot be separated from the history of classification, categories and collectivities. The question "Who is this person?" does always involve the question "what kind of person is this?" (Caplan & Torpey 2001: 3). Therefore, I discuss the meaning of recognition also in the context of practices and techniques that focus on categorical recognition such as cyberspace identification, but also—and this leads to my central empirical research interest—different manifestations of pattern recognition such as behaviour pattern recognition that feature strong similarities and communalities with technologies for individual identification such as facial recognition. Furthermore, precursors of behaviour pattern recognition can be found in endeavours such as physiognomy or phrenology. Out of these insights I will develop an understanding of

[14] A more recent example for the performative factor of identity impostor is the story of Frank Abagnale that inspired Steven Spielberg's film „Catch Me If You Can" (2002).

what it means if we speak about recognition or the respective process of recognition, "to recognise".

The History of Identification as the History of Recognition

The anthology "Documenting Individual Identity" (Caplan & Torpey 2001) was a first attempt at analysing the field of person identification and the history of identification, systematically. In reference to this anthology 'Identification Studies' deal with practices, places, forms and material artefacts used in the description of people. One central aim of the description of people is to allow for recognition, for example in another place or at another point in time. This means that Identification Studies also deal with the storage and archiving of these descriptions, in documents, data bases, and registries.

One of the most important means of identification and recognition of people was—and still is today—the name. But how did it actually come about that the name became fixed as a legal and practical tool for recognition and identification? It was, for example, only from about the 11th century onwards that a second name or surname was added to the first name (Caplan 2001: 53f.). As we can recognise in many of today's surnames, see Caplan, they were generated from five different type categories, these being the filiation (patronymic), the place of origin (locative), the object (toponymic), the occupation and the sobriquet. For a long time, most of the European states at least, had no interest in establishing an obligation to hold a particular name. On the contrary; for example, in Germany around 1800, a personal name was seen as a self-evident right to be left to an individual´s own choice. The name was viewed as the property of each distinct person, to be taken and discarded at will. It was only when Napoleonic law was introduced into Germany that the state gained interest in personal names and it was 1816 before Prussia forbade the use of any name other than the registered. Following this, a permanent name was given, in order to hold individuals and their names accountable and recogniseable. This was important for several activities necessary in building a nation-state, such as the right to vote, for taxation, military conscription, legal processes, security and welfare policing. As a consequence 'Personenstandsgesetze' (laws on civil status) were introduced by Bismarck in 1875 and 'Standesämter' (registry districts and offices) were created (ibid.: 59ff.) that still exist today.

With the emergence and increase in the power of the state in Europe since the end of the 18th century, state authorities have struggled with the problem of how to secure, register, and locate their citizens. This was for very different purposes and targeted at constrasting groups such as criminals, foreigners, or minorities (Gruber 2012: 105). In this regard, for example, the police in Vienna established an approach that Gruber refers to as continuous evidence ('Ununterbrochene Evidenz'). The strategy of continuous evidence was designed for the purpose of unambiguously identifying, registering and locating persons and physical entities (e.g. houses) by establishing different means of resident registration, identity papers, or in the case of houses, house numbers (ibid.: 107). Every citizen was to be made visible to the state not by physical marks on the body, but by the indirect means of registration, passes and censuses (Caplan & Torpey 2001: 8). Gruber embeds the strategy of continous evidence in Michel Foucault's concept of "disciplinary space" (Foucault 1979), a logic of localisation that works by assigning a unique spot to every individual. Challenging questions arising from the concept are the relationship between the emancipatory and the repressive aspects of such registration and identity documentation. In this regard, Max Weber famously showed that the increasing bureaucratic handling of everyday governmental concerns was inevitable (cited in: Caplan & Torpey 2001: 5). Basic civil rights were created together with the fundamental advantages of individual registration and identification, however as history has shown, also nightmarish uses appeared, best exemplified in the Nazi use of population registers and identification documents to track Jewish and other "undesirables" (ibid.).

The importance and registration of names was again reflected by the creation of Article 24 of the United Nations International Covenant on Civil and Political Rights in 1966, that states that "every child shall be registered immediately after birth and shall have a name." Registration appears to be the vital essence of human existence as seen from this "modern" point of view. Without registration; no recognition (ibid.: 6). In this meaning of the word recognition - the acknowledgement of certain rights as a citizen of a certain state - a fundamental condition of recognition appears; the recognition or acknowledgement of rights is only possible when the individual person is included and secured, meaning registered, in a register or database, or more generally, in an entity of reference. This is made possible through the matching and comparison of the record in the entity of reference and the present condition of the record, or the correspondence between a person and a set of signs (Caplan 2001: 50) that went into the record in a specific form. Meaning that recognition does always refer to something that already exists, to which a current phenomenon can be confronted and compared with. This is valid for very different things such as names, fingerprints and faces, but also specific forms of behaviour. This insight leads to the fundamental question of what version of the phenomenon is the basis for recognition, what is saved in a certain form of registration, in what manner? What criteria are used and established, formalised and standardised? What is the best way of doing this?

According to Beatrice Fraenkel, the elementary signs of modern identity have come to be conventionalised as the name, the portrait, and the fingerprint. In this regard the fingerprint seems to be seen as the unofficial emblem of modern identity culture (Caplan 2001: 52f.). A look at the history of premodern practices (Groebner 2001: 19) regarding names brings in two other means of reference entities; insignia and passes. Insignia played a particular role in the form of courier or pilgrim badges and journeymen, servants, and travellers were required to carry papers with them stating who they were. This always came with the uncertainty of whether the name stated in the piece of paper was correct. Groebner refers to this phenomenon as a threatening gap between appearance and description and between person and paper (ibid.: 21). One means of closing this gap and of creating trust was the use of seals for the purpose of authentication. But also seals could be duplicated and forged (ibid.: 22) and did not solve, but only combat this threatening gap problem on the surface. Next to the widespread use of insignia and passes, the identification and recognition of persons in late Medieval and Renaissance Europe was left up to his or her own face and connected to it, portraits (ibid.: 22ff.). According to Groebner, from the 15^{th} century on, portraits played an important and complex role in court and city life. Even though the portrait gained more and more importance, images of sufficient quality to fulfill conditions of authentication and recognition were very difficult to reproduce in any quantity between about 1400 and 1700. However, as the example of "modern" facial recognition technologies shows, sufficient quality of images alone does not solve the problem of facial recognition, even if it is a crucial factor for successful and correct recognition. I will come back to this point shortly.

Portraits were important additional means for body description and depiction, for the purpose of recognition and thus, identification of persons. In the 19th century photographs were used more frequently, especially by the police as a memory aid (Jäger 2001: 28). Portraits, and images of the body in general, became promising tools in the course of the 19th century because they promised to achieve 'mechanical objectivity' (Daston 2001) in the identification of people. All forms of 'mechanical objectivity' were used in an effort to abandon language as a means of representation altogether. ´Communitarian objectivity' was a way of establishing a common language and a system of classification (Becker 2001: 141). Becker makes use of Daston´s conceptual framework to investigate analytical approaches to physiognomies and bodies used in police work in the 19th century (ibid.: 141ff.). As he notes, in the beginning of the 19th century, the "practical gaze" of police officers was based mainly on their experience. Into the second half of the 19th century this "practical gaze" became more and more structured due to the growth in physiological and anatomical knowledge; what Becker characterises as the decomposition of a particular appearance into meaningful elements through different discoursive practices. Along with the "scientific turn" taken by policing in the course of the 1870s, criminal police in France began to use photographs methodically to display effectiveness and less importantly as a real means of the detection of criminals. One reason for this precision was that police image collections had became cumbersome, and therefore, unmanageable (Jäger 2001: 27). For example, a photographic register was introduced in Paris in 1874 in which photographs of every person sentenced had to be integrated. After eight years, the image collection consisted of the unmanageable number of 75,000 portraits. The police department of Berlin learned of this and there were efforts at keeping the numbers of records as low as possible, but this only delayed the collapse of the system (ibid.: 39f.).

After a few years of the mainly useless, precisely because unmanageable, collection of photographs, new systems of classification and cross-referencing were needed (ibid.: 41). In 1879, Alphonse Bertillon, a clerk at the Préfécture de Police in Paris, presented a non-photographic solution to the organisation problem of thousands of photographic portraits and in addition, it was also an identification technique that promised a level of scientific precision. The so called 'Bertillonage,' or Bertillon system was the first "modern" system of criminal identification, emerging in European cities especially as a reaction to the recognition of the growing number of so called 'rècidivistes,' or repeat offenders (Cole 2001: 32ff.). The Bertillon system was based on anthropometry, the physical measurement of the size and proportions of the human body. Whilst most European anthropologists travelled the world "using anthropometry to quantify the physical differences between the 'savage' and 'civilized' races" (ibid.: 34), Bertillon applied anthropometry at home in European cities, not to delineate group identity, but for the purpose of individual identification. How did Bertillonage work?

> "A prisoner being Bertillonaged was first subjected to eleven different anthropometric measurements taken with specially designed calipers, gauges, and rulers by one of Bertillon's rigorously trained clerks, or 'Bertillon operators.' Each measurement was a meticulously choreographed set of gestures in which the exact positioning and movement of both bodies—prisoner and operator—were dictated by Bertillon's precise instructions." (Cole 2001: 34)

This "elaborate dance" also made use of a "precise" scientific language that was developed by him. For example, when describing eye colour, Bertillion did not only refer to the most commonly used colours blue, brown and green, but used a description system that made use of 50 types of eye colours (ibid.: 38). The "jewel" in Bertillon's morphological vocabulary was the description of the ear (ibid.: 40). He developed an enormous vocabulary that referred to every single part of the human ear; ear border, ear lobes, ear folds and so on. The translation of bodily features into such a universal language also had the advantage of being suitable for the transmission of physical descriptions by telegraph (ibid.: 46). This crucial element of the Bertillion system, the transformation of physical human bodies into language and further into transmittable and exchangeable codes, is to be found in many identification techniques that followed.

Fingerprinting is based on codes and also DNA analysis comes in the form of the specialised four-letter language of genetics (ibid.: 49). The scientific grounding of the Bertillon system, the attempt to describe the human body with "precise" morphological vocabulary and to measure it, following a strict order, was also the greatest problem of the system. "It was precisely in the diffusion abroad, however, that the Bertillon system suffered its greatest weakness." (ibid.: 52). For example, anthropometric practice was not nearly as precise in the United States as it was in France, the country of its origin (ibid.: 147). In comparison to fingerprinting, the Bertillon system was harder to implement, required extensive training for operators, was much more time consuming and thus, was by far the more expensive method. While anthropometry and the Bertillon system were seen as more scientific at that time, fingerprinting was seen as much more practical (ibid.: 151f.).

Simon Cole also showed that in the US, the implementation of fingerprinting in the police system was not only a matter of the authentication of identity, but also a means to establish a new order in existing databases. Because the police was not able to distinguish between "foreigners" without any reasonable doubt any more in their everyday practical work, (Asians or black people all seemed to look the same for the white policemen), fingerprinting was gradually introduced to facilitate recognition of these indiscernible populations (ibid.: 140ff.). Additionally, it offered a cheap means of rapid identification that was also sensitive enough—in comparison to anthropometric body measurements—to all requirements when female bodies needed to be physically identified (ibid.: 154). In Europe, dactyloscopy—the fingerprinting technique—was acclaimed enthusiastically as an objective and definite method of identifying individuals at the beginning of the 20th century. Problems and difficulties using fingerprints to identify suspects were mainly kept quiet or were not met with great interest (cf. Meßner forthcoming). For example, Theodor Harster, a senior official at the Munich records department argued in the 1910s, that no method of identifying persons existed which is comparable with the security, simplicity, cheapness and speed of dactyloscopy (ibid.). Many problems, difficulties, and failures[15] using fingerprints (and other biometrics) can

[15] For a detailed analysis of biometrics failures see Magnet (2011).

still be seen in today's discussions about forensic technologies and how they can be seen and used as evidence in legal cases.

Kruse (2010a), for example, dealt with fingerprints and DNA as forensic evidence in the Swedish legal system and Prainsack (2010) analysed the use of forensic DNA technologies in Austria. Both stress the ambiguity of forensic evidence and the need to transform crime scene traces in multiple ways into information before they count as forensic evidence. Fingerprints and DNA are not only used to connect particular bodies to particular identities, but they are both also used to connect particular bodies to particular crime scenes. Kruse emphasises that traces at crime scenes do not speak for themselves (Kruse 2010a: 366) and are useless on their own (ibid.: 369). In fact, forensic evidence:

> "is produced by an apparatus that involves traces, bodies (and not only criminal ones), forensic technologies and practices, law and legal practices, as well as less specialized cultural understandings" (ibid.: 366).

The view that every fingerprint is unique, has to be fundamentally questioned. A look at the police practice of fingerprinting shows that "no two prints made by the same finger are ever exactly identical" (ibid.: 368). As dactyloscopy is much less automated than constantly demonstrated in popular media series such as CSI, specialised and trained fingerprint examiners are still needed to compare examined fingerprints to fingerprints of suspects or those saved in police databases (ibid.: 368). Prainsack also mentions the "arduous, time-consuming and often monotonous nature" of forensic DNA work carried out in Austria, by law enforcers in her empirical example. "When you have a hit, this is when the real work starts" say law enforcers (Prainsack 2010: 161). She also recognised a power gap about the role of DNA evidence, between law enforcers and prisoners (both groups were interviewed). Whereas the law enforcement side showed "rather nuanced understandings of what forensic DNA profiling can and cannot do," the prisoners "tended to regard DNA profiling as infallible and true" (ibid.: 171). This power gap seems to be a constitutive element in the promotion, diffusion and public understanding of identification and recognition technologies. In the realm of criminal investigation "people in law enforcement authorities often state that the less ´the public' knows about police work at the crime scene, the better" (Prainsack & Kitzberger 2009: 72). One crucial resource for the public understanding of forensic evidence and identification technologies are TV series such as *CSI: Crime Scene Investigation*. In CSI, forensic science is the key to solving otherwise baffling crimes (Collins & Evans 2012: 905). In CSI, the dominant theme is that fingerprints and DNA traces reveal the absolute truth and deliver evidence that speaks for itself (Kruse 2010b: 80f.) That is in opposition to the need for DNA matches for interpretation and intervention, by humans. DNA matches are embedded in a complex chain of inference (Collins & Evans 2012: 906). In nonfictional forensic science, producing evidence is more complicated than the way it is presented in CSI (ibid.: 86). Finding DNA matches depends highly on the skilful and informed interpretation of images (Halfon 98: 805ff.). In nonfictional DNA practice, absolute certainty is unattainable and one must always refer to probabilities. In CSI, matches are synonymous with knowing for certain and thus, with absolute truth (Kruse 2010b: 86). A dominant message arising from CSI is "that it is easy, quick, routine and epistemologically very strong" (Ley, Jankowski & Brewer 2010: 13). This view leads to an asocial representation of science in the public that underpins the so called "CSI-effect" (Collins & Evans 2012: 906).

As these examples show, the threatening gap between appearance and description (Groebner 2001:21) is not (fully) eliminated with the introduction of forensic and biometric technologies. It is still a matter of complex interpretation and time-consuming human intervention in how far the two patterns of appearance and description, between body and registered identity coincide. It is a persisting process of negotiation taking place in sociocultural practices.

Facial Recognition (Technologies): From Individual Identification…

One of the most important means of identifying another person in everyday life is the visual recognition of their face. The face seems to be the "most common biometric in use by humans to identify other humans" (Introna & Wood 2004: 178). But how do people remember and recognise faces and how good are people at remembering faces? In this regard the paradox of face memory was observed. Generally spoken, while people are very good at remembering familiar faces (for example friends, family members, colleagues, but also well-known celebrities) they do, overall, less well in remembering (and thus recognising) unfamiliar faces (Bruce & Young 2011). Findings from several experiments on how our face memory works, show that the human ability to recognise faces from isolated features and configuration relations has to be questioned. Instead, human face recognition seems to work in a more abstract and holistic way and is also highly dependent on context information (ibid.).

It seems that at least in Western societies in everyday life, we carry around with ourselves the face as the most natural and direct way of showing our identity. There seems to be "an implicit common agreement to reveal our faces to others as a condition for ongoing social order" (Introna & Wood 2004: 178) that is challenged as soon as someone covers his or her face. As Introna and Wood note, once somebody hides their face "then there is almost an immediate assumption of guilt" (ibid.). There are however, also occasions in different cultural contexts where disguise is highly acceptable; think of the different forms of carnival around the world, whether in Rio de Janeiro, Venice or Cologne. In carnival practice, the identity of a particular person becomes secondary to some form of group identity that is achieved by wearing a mask with the respective trappings of a prince or princess, a pirate, cowboy, cat or mouse. While these carnival masquerade practices have the aim of temporarily leaving the common order of the bourgeoisie (Gruber, Meßner & Musik 2012: 224), there are also strategies and the need for disguise or hiding the face for other purposes, such as for religious reasons, for criminal acts, or for the resistance of facial recognition technologies (ibid.). As an example, the artist Raul Gschrey created and presented a composite mask that he titled 'The Typical German.' The mask is made out of 150 portrait images of males and females that come from German cities. It can be used and worn as a means of hiding from facial recognition technologies in public space (Gschrey 2012: 215). This is achieved, not only by hiding one's own face, but as Gschrey notes, by bringing average face characteristics to the fore. This is critical for facial recognition technologies, because these make use of deviations from the norm (ibid.: 216) in order to recognise a specific, individual face. Gschrey's artistic work carries on a long tradition of resistance to identification and surveillance practices and techniques by two means; on the one hand his artwork is an expression of active resistance by making recognition impossible, on the other hand it points to what might be called material resistance inherent to practices and techniques which means, to limitations in the use of a certain technology (Gruber, Meßner & Musik 2012: 219).

The recognition of faces is made impossible when wearing masks. The use of masks is also a prevalent theme in popular culture (for example in *The Phantom of the Opera*, *Batman*, *Zorro*, *V for Vendetta*) and does also impact real-life movements such as in the "hacktivist" group *Anonymous*. In the case of *Anonymous*, the so-called *V-mask* or *Guy-Fawkes mask* has become its trademark and is in fact "a brand or meme for all kinds of demonstrations and political events" (Manghani 2012). Differing from country to country, there are distince regulations regarding the wearing of masks. In Austria, for example, there is no general ban on wearing face coverings, but in the *Versammlungsgesetz*, regulating the right of assembly §9 (1)[16], the wearing of clothes or other items in order to disguise or mask the face for the purpose of averting recognition in the context of the assembly, during the assembly, is prohibited. If there is no threat to public safety and security, the enforcement of the ban can be suspended §9 (3)[17].

Apart from the use of masks, another method of disguise is the use of make-up and hairstyling. In this regard, a particularly interesting project of resistance towards facial detection and recognition technologies is Adam Harvey's project *CV Dazzle™*. CV Dazzle (short for 'Computer Vision Dazzle') "is a form of expressive interference that combines makeup and hair styling (or other modifications) with face-detection thwarting designs."[18] It has the aspiration of not being perceived as a mask or disguise but and rather a method of remaining inconspicuous. As Harvey explains:

---

[16] § 9. (1) An einer Versammlung dürfen keine Personen teilnehmen, 1. die ihre Gesichtszüge durch Kleidung oder andere Gegenstände verhüllen oder verbergen, um ihre Wiedererkennung im Zusammenhang mit der Versammlung zu verhindern oder 2. die Gegenstände mit sich führen, die ihrem Wesen nach dazu bestimmt sind, die Feststellung der Identität zu verhindern.

[17] §9 (3) Darüber hinaus kann von der Durchsetzung der Verbote nach Abs. 1 abgesehen werden, wenn eine Gefährdung der öffentlichen Ordnung, Ruhe und Sicherheit nicht zu besorgen ist.

[18] see http://cvdazzle.com/

> "The name is derived from a type of camouflage used during WWI, called Dazzle, which was used to break apart the gestalt-image of warships, making it hard to discern their directionality, size, and orientation. Likewise, the goal of *CV Dazzle* is to break apart the gestalt of a face, or object, and make it undetectable to computer vision algorithms, in particular face detection. Because face detection is the first step in automated facial recognition, *CV Dazzle* can be used in any environment where automated face recognition systems are in use, such as Google's Picasa, Flickr, or Facebook." (http://cvdazzle.com/)

Similar to the work of Raul Gschrey, Adam Harvey indicates with his project CV Dazzle how facial detection and recognition technologies work. According to Gross, Shi & Cohn (2001), Introna and Wood (2004: 185) note that there are two main categories of facial recognition algorithms; image template algorithms and geometry feature-based algorithms. The first, which can also be named the template-based method, determines the individual face identity by the difference or deviation from a general "standard" face. This category of facial recognition algorithm is the one artist Raul Gschrey referred to in his artwork "The Typical German." The second category, the geometry, feature-based algorithm points at key facial features such as eyes, nose or mouth. These points are connected to a network and distances and angles are measured in order to get a unique face print. As Introna and Wood note, both categories or methods have in common the issue of reduction. This means that in order to be efficient in processing and storing, certain visual information is reduced to a numerical representation and therefore some information is disregarded (ibid.: 186). This process of reduction, that could also be interpreted as a process of selection in which markers represent individuality, has consequences. As Introna and Wood note, in template-based algorithms "minorities tend to deviate the most from the standard template" and thus, minorities might become easier to recognise. Feature-based algorithms alternatively, have problems with larger databases and bad quality images.

Introna and Wood analysed the politics and implications of face recognition technologies, especially when implemented in CCTV systems (cf. Introna & Wood 2004). One of their central results was that facial recognition algorithms seem to have a systemic bias: men, Asian and Afro-American populations as well as older people are more likely to be recognised than women, white populations and younger people (ibid.: 190). A consequence of this bias could be, for example, that those with a higher possibility of being recognised are those with a higher probability of scrutiny or of setting off an alarm. This inherent tendency to bias has the capability for a "new type of digital divide" (ibid.: 192) that needs close attention in future research. Therefore Introna and Wood call for "bias studies," especially regarding the question of what can be done to limit biases (ibid.: 195). Additionally, they call for studies of actual implementations in order to elaborate on how to include facial recognition systems into larger security infrastructures (ibid.: 196) as technical artefacts like face recognition algorithms never act in isolation (ibid.: 195).

In 2009, Introna and Nissenbaum published a detailed socio-political analysis of facial recognition technologies (FRT) that bridged the technical and social-scientific literatures (Introna & Nissenbaum 2009: 3). They reported on FRT performance, FRT evaluations, FRT in operation, connected policy concerns as well as moral and political considerations. Regarding the performance, meaning which "types of tasks can current FRT successfully perform, under what conditions?" (ibid.), they state that "image quality is more significant than any other single factor" (ibid.). FRT works well with small populations in controlled environments when the aim of FRT is the verification of identity claims. In contrast to this, it showed rather poor results in more complex attempts, for example in uncontrolled environments. That means, the performance of FRTs is especially dependent on environmental issues like background, lighting, camera distance, size and orientation of heads. In addition, image age, gallery use and a consistent camera use, play a role in performance (ibid.). In close relation to the performance is the evaluation of FRTs. Introna & Nissenbaum divide evaluation into three different types: technological, scenario, and operational. In technological evaluations the performance of algorithms is tested. Such an evaluation is "normally performed under laboratory conditions using a standardized data set that was compiled in controlled conditions" (ibid.: 21). The main purpose and advantage of technological evaluation is the high degree of repeatability, but they are not designed to be evaluated under different conditions and settings. That means that in technological evaluations such as the widely reported *Face Recognition Vendor Tests* (FRVT) in 2002 and 2006[19] “many of the potential factors that may influence recognition rates were kept constant or controlled for” (ibid.: 28). The best algorithms in the FRVT 2002 produced 80% accurate results, in the FRVT 2006, the best algorithms produced 99% accurate results. These results indicate “a massive improvement in the technology” (ibid.), but a closer look at the evaluation showed significant test differences, especially regarding the quality of images used (ibid.). In the words of Introna and Nissenbaum, “the information available (at the pixel level) to the algorithms in 2006 was potentially twenty-five times greater than that of 2002” (ibid.). Introna and Nissenbaum also indicate the need for careful interpretation of such high accuracy numbers such as 99% in the FRVT 2006:

> “Finally, it is also worth mentioning that technology evaluations are just one element of an overall evaluation. The really significant results, with regard to the feasibility of the technology, are the performance of these algorithms as part of specific scenarios in operational conditions.” (ibid.: 29)

The first step out of a laboratory environment is towards scenario evaluation. They are designed to model and simulate real-world environments and populations in certain scenarios, but cannot be compared to full operational conditions. Nevertheless, in these scenario evaluations[20] “the performance of the systems were in fact significantly lower than in the technology evaluations” (ibid.: 32). The best of the four tested systems correctly identified volunteers in only 36% of the time in the verification task (ibid.).

---

[19] The *Face Recognition Vendor Tests (FVRT)* were independent assessments „performed by NIST and sponsored by organizations such as the Department of Homeland Security, the Director of National Intelligence, the Federal Bureau of Investigation, the Technical Support Working Group, and the National Institute of Justice.“ (cf. Introna & Nissenbaum 2009: 27ff.)

[20] Introna & Nissenbaum refer to the BioFace evaluations that were „joint projects of the Federal Office for Information Security (FOIS) in Bonn, Germany, and the *Federal Office of Criminal Investigation (BKA)* in Wiesbaden, Germany, with additional assistance provided by the *Fraunhofer Institute for Computer Graphics Research (IGD)*“ (cf. Introna & Nissenbaum 2009: 30ff.).

Finally, operational evaluations test systems in situ, in their actual operational conditions (ibid.: 22). Introna and Nissenbaum note that ideally, FRT systems start with technology evaluation, followed by a scenario evaluation, and finally by operational evaluation (ibid.: 21). This means, for full understanding, if—in this case facial recognition technology is viable - it has to be tested not only in the lab, but also in situ. However, currently there is a lack of publicly available data on operational evaluation of facial recognition systems. One of the few operational evaluations was conducted by the *German Federal Criminal Police Office (BKA)* in Mainz rail terminal between October 2006 and January 2007. Two hundred test subjects acted as "suspects". The daily average of people passing through specific bottlenecks was 22,673. Under these conditions by daylight, recognition rates were 60%, by night, recognition rates dropped to 10-20%. That meant that daylight was the most significant factor for high recognition rates. In addition, recognition rates were 5-15% lower at the bottlenecks on stairs, than on the more canalising escalators (ibid.: 37). Introna and Nissenbaum note that overall "the report concludes that FRT is not yet suitable as a system for general surveillance in order to identify suspects on a watch list" (ibid.). In addition to this, Introna and Nissenbaum conclude that vendors of FRT use technology evaluation without providing the wider context of these evaluations: the consequence being that misleading conclusions about the efficacy of FRT occur. They also recommend an appropriate use of evaluation data by the media (ibid.: 38). In this regard Kelly Gates noticed the common perception that FRT "are either already deployed in a variety of settings, or that their deployment is happening at a rapid pace" (Gates 2011: 5). She also notes that these assumptions are not only made by industry and the media, but also by the surveillance studies literature. Overall, Gates claims that FRT so far "do not work very well outside constrained settings" (ibid.). But the question of how well technologies in general, and FRT in particular, work is very much dependent on how they are "viewed as obvious, necessary, and inevitable next steps in the technological trajectory" (ibid.: 24). In the case of FRT, the most important questions are "how much accuracy would be necessary?" and "what exactly constitutes an accurate facial likeness" (ibid.: 48). When studying the history of FRT, Gates comes to the conlusion that the "social construction of the accuracy of FRT became central to its development" (ibid.). While Gate's and also

Introna and Nissenbaum's insights suggest that FRT are not accurate, at least once deployed in scenario or operational settings, especially in the "watch list" scenario, the accuracy of FRT was attested by its proponents and mediated by the media by themeans of "rhetorical closure" (Pinch & Bijker: 1987: 44), that means by simply claiming that FRT works and is accurate. Still, the ongoing funding of FRT research projects is a good indicator of defective accuracy and therefore for the necessity—if the goal is to improve accuracy—of further research and development. For example, in 2010 the *German Federal Ministry of Education and Research (BMBF)* in their scheme *KMU-innovativ*[21] funded the research project "Parallele Gesichtserkennung in Videoströmen" (concurrent face recognition in video streams) with 1.2 million Euro. As the participating *Institute for Anthropomatics* at the *Karlsruhe Institute for Technology (KIT)* reported on their website[22], the project had the goal

> "... to improve face detection, tracking and recognition by means of parallelization and make it real-time capable for realistic scenarios. Face tracking and recognition in real-work scenarios are very challenging due to many different factors. In the project, we supply the core computer vision components and algorithms. We are working on improving the state-of-the-art for meeting the challenges presented in realistic settings."

The project is particularly interesting, because a planned field test in the *Wildparkstadion*, a football stadium in the German city of Karlsruhe, was stopped in July 2011 due to the protest of supporters of the football club *Karlsruher SC (KSC)* and the data protection commissioner of the federal state Baden-Würtemberg. Obviously, as was argued only on the web page of the football club KSC, misunderstandings in communication led to this situation. In fact, the football club KSC did not give assent to the field test. Thus, the central goal of the project of making FRT real-time capable of realistic scenarios, such as in a football stadium on match day, could not be achieved. The answer of the *Home Office* in Baden-Würtemberg to a minor interpellation

[21] KMU-innovativ is a research funding scheme for small and medium-sized enterprises (SMEs). See http://www.bmbf.de/en/10785.php?hilite=kmu [Jan 30,2013] for further information.

[22] http://cvhci.anthropomatik.kit.edu/project/pagevi [Jan 30,2013]

(Drucksache 15/470/ 01. 09. 2011) in the *Landtag* of Baden-Würtemberg in October 2011, reported that the project partners were currently considering alternatives. Potentially, it was stated, that an entire abandonment of field tests was being considered, but without threatening the success of the project (p.3). In my view, this statement is completely paradoxical, because the goal of the project, to make FRT real-time capable of realistic scenarios, could never be achieved without testing FRT exactly in a realistic scenario such as the field test in the surroundings of a football stadium. More to the point, the case showed that firstly, possible future FRT applications in the public or semi-public space do not seem to meet high enough data protection and privacy requirements and thus, basic human rights. Secondly, the football supporters' active resistance to a FRT field test challenges the social acceptance of FRT fundamentally. Nevertheless, the cancellation of the field test did not give researchers the chance to test their algorithms with real-time, operational data. Instead, as I was told by the participating *Institute for Anthropomatics*[23] test data was recorded on the private grounds of one of the project partners with about 50 study volunteers. Meaning that the project algorithms could only be tested in a scenario evaluation instead of a full operational evaluation.

At an earlier time and in another cultural context, a FRT operational field test was made possible and actually performed, but the outcome was rather unsatisfying in comparison to the promises made when installing the system. As Kelly Gates reports, in Florida in June 2001, the *Tampa Police Department (TPD)* integrated FRT[24] into their existing CCTV system in a neighborhood called Ybor City. It was the first urban area in the United States to be equipped (Gates 2010: 68). But in August 2003, "after a two-year trial period, the TPD abandoned the effort to integrate facial recognition with the CCTV system" (ibid.). What was the reason for the abandonment of a promising system that was said to "do the watching for the CCTV operators" (ibid.: 79) in a "non-discriminatory fashion" and "free of human prejudices" (ibid.: 80)? The system did not

[23] Phone call on February 5, 2013 with Prof. Rainer Stiefelhagen, director of the *Computer Vision for Human-Computer Interaction Lab* at the *Institute for Anthropomatics* at *Karlsruhe Institute for Technology (KIT)*

[24] The System used called „FaceIt" was delivered by the corporation Visionics (Gates 2010: 78)

identify a single wanted individual. The TPD Police Captain was cited in the news saying that the FRS "was of no benefit to us, and it served no real purpose" (Stacy 2003 cit. in Gates 2010: 85). That means, that FRT was on the one hand, not a practical solution for the police, on the other hand, the system did not deliver one single success story that could display FRT as a symbol for technological sophistication used by a modern police force (Gates 2010: 86). This outcome was also in contrast to promises made about FRT articulated in the post-9/11 period. The recorded video images of the major 9/11 terrorists such as Mohammad Atta and Abdulaziz Alomari, showing them passing through airport security, suggested that FRT "may have helped avert the September 11 terrorist attacks" (Gates 2011: 1). FRT "emerged as an already existing, reliable, and hight-tech solution to the newest, most pressing problem facing the nation" (ibid.: 2). The 9/11 terrorist attacks in the US happened at a time when the business of security began to overlap with the business of IT (ibid.: 99). FRT began to become "a homeland security technology" that was "automatically identifying the faces of terrorists" (ibid.: 100). The central FRT and biometrics industry and the rhetoric of proponents at this time, positioned FRT "as the solution to the new terrorist threat", amongst other things, by claiming "that such systems could have prevented at least one if not all of the hijackings" (ibid.). In this regard, 9/11 offered the opportunity to catapult FRT from a

> "set of technological experiments into something more closely resembling what Bruno Latour calls a 'black box' – a functioning technology positioned as virtually indispensible to a secure, technological future." (ibid.: 101).

It is exactly here that technical and socio-cultural endeavours meet. Having a look back at the scenario and operational evaluations of FRT helps to understand what the problem is: FRT was black-boxed too soon, meaning that in the years following 9/11, a wide array of different actors, in major technological evaluations delivered at least promising results. Here, it has to be noted again that technological evaluations are "normally performed under laboratory conditions with FRT using a standardized data set that was compiled in controlled conditions" (Introna & Nissenbaum 2009: 21). That means that on some sites and in some situations, especially in the lab, FRT might work very well. However on many other sites and in many other situations, especially in real-world environments, FRT might work very badly. That has a lot to do with the organisation of the recognition process; remember the case of Martin Guerre at the beginning of this section. The witnesses at the Guerre trial had to compare the appearance of the returned and present Martin Guerre to the picture of the "old" Martin Guerre that was somehow saved in their memories. There was not any form of visual depiction available, such as an image, to the court and the witnesses. This might also be the case in court proceedings today when it comes to eye witness accounts. In theory, one might think an image of a person would help to identify the actual person. But this might, amongst other things, be dependent on the quality of the image. A standardised, carefully placed, high-quality, high resolution, up-to-date image of a person would seem to be the most promising and even a prerequisite when it comes to automated facial recognition. This is why, for instance, international passport photo standards as well as standards for the exchange of biometric data were created in recent years. The *International Standard Organization (ISO)*, for example, has listed the standard *ISO/IEC 19794-5:2011* for "face image data" in the area of biometric data interchange formats. Today, all around the world passport photo standards can be found, carefully outlining what kind of passport images are suitable. For example, the guidelines of the *UK Home Office* consist of a list of 22 different requirements such as image size and background, illumination, image age, head positioning and so on[25]. The requirements also list twelve incorrect examples showing people that have been photographed from too far away, from too close or looking away from the camera. The standardisation of face image data is a crucial factor for the proper functioning of FRT in its verification duty. In this regard, the *International Civil Aviation Organization (ICAO)*, the UN agency responsible for security standards on passports, was one of the central actors (LSE 2005: 48). Already in the late 1990s, ICAO researched "the potential uses of biometrics and other forms of digitisation of passport information but, in the years that followed, little progress was made." (ibid.) It especially was an outcome of 9/11, when

[25] See http://www.direct.gov.uk/prod_consum_dg/groups/dg_digitalassets/@dg/@en/@travel/documents/digitalasset/dg_174925.pdf [April 29, 2014] for the full list of requirements!

> “in May 2003, in line with US initiatives, the ICAO published new standards for MRTD (machine readable travel documents), which introduced biometric technologies in order to facilitate global interoperability in border-control identification. Under these standards, the face has been selected as the primary biometric, in the form of a high-resolution digitized image, which will be stored on a contactless chip.” (Higgs 2011: 197).

That means, for the task of verification, a high resolution digitised facial image saved on a passport is compared to an image taken at the site and time of control. An example are the automated passport control gates (‘e-passport gates’) installed by the *UK Border Agency* at several British Airports[26]. Here, travellers do not need to touch anything with their fingers or other parts of the body but there still has to be physical contact and cooperation by the passport holder. First, by putting the biometric passport onto a scanner, followed by a highly standardised way of getting through the gate using direct eye contact with a screen while standing on footprints on the floor and waiting perfectly still for a certain period of time. That means for the verification task, FRT does not only need an internationally standardised image saved on the passport, but also a highly standardised setting for the taking of the reference image.[27] Clearly, standards and standardisation are central to the viability of FRT. Moreover, in general, standards are amongst the building blocks of our “modern” society (Busch 2011). As Busch notes, “standards are the recipes by which we create realities” (ibid.: 2), “that we use to hold the world of people and things together“ (ibid.: 5) and “order ourselves, other people, things, processes, numbers, and even language itself” (ibid.: 3). Standards “help regulate and calibrate social life by rendering the modern world equivalent across cultures, time, and geography” (Timmermans & Epstein 2010: 70). Standards are “a way of classifying the world” (Bowker & Star 2000: 13). But as much as standards regulate, calibrate,

[26] See http://www.ukba.homeoffice.gov.uk/customs-travel/Enteringtheuk/e-passport-gates/ [Feb 5, 2013]

[27] I tried the ‘e-passport gates’ two times at Manchester airport in 2011. The first time I needed human operator assistance and had to change border gates. The second time it did not work with my biometric passport at all, and I had to see an officer to check my passport. In both cases I would have been faster getting through border control than the people in the “normal” queue as most people used conventional, manual control.

classify and order the world, these orderings are “always partial and impermanent (...) and never complete ” - the world is just to messy (Busch 2011: 6). On the other hand, as Busch’s study of standards shows, standards are very quickly taken for granted. Thus, societal values and power that had influence on the creation of a standard, are also fixed and taken for granted (ibid.: 268). Meaning that standards are not politically neutral, but pose “sharp questions for democracy” (Timmermans & Epstein 2010: 70). In the case of FRT, we have to ask who is affected adversely by face image standards and how they are affected, as these standards assume that every person has a machine-readable face as shown on the passport photo standards and requirements of the *UK Home Office*, for example. This is the case also for other biometrics, as Irma van der Ploeg notes. This is paradoxical, because biometrics on the one hand assume “that every individual is physically unique” but on the other hand, there is the expectation of similarity, assuming that every human has “a clearly audible voice, a set of ten fingerprints, two irises, and a recognizable face, and so on.” (van der Ploeg 2011: 30) Thus, a face image standard as a prerequisite for the verification of identity at passport control or other checkpoints is likely to discriminate a wide array of people that do not fit into these standards requirements, by prescribing special treatment for these people.

This does also count for FRT watch-list or CCTV video surveillance scenarios. Moreover, in these cases, cooperation of the persons of interest cannot be expected, and thus images that comply with the requirements and standards of ISO or ICAO are hard to achieve. As Introna & Nissenbaum note, it seems obvious that such affordable standardised images “will not be easy to capture without the active participation of the subject.” (Introna & Nissenbaum 2009: 19) However standardised images of a high quality are a requirement for high recognition rates. In this regard, pose variation (rotation of head) and illumination (the existence of shadows) are two crucial factors (ibid.), because variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity (Moses, Adini, & Ullman 1997 cit. in Introna & Nissenbaum 2009: 19). “Pose variation and illumination problems make it extremely difficult to accurately locate facial landmarks.” (ibid.). Meaning that the correct identification of a person by means of automated face recognition is not solely dependent on whether the face on the two images is the same face, but also or even more important, on the images used and how the respective images were produced. In this respect, images can range in the best case from a standardised, high quality, high resolution, up-to-date image of a person that is in compliance with international standards such as the ISO*/IEC 19794-5:2011* for "face image data", taken in a lab, to a low quality, low resolution, out-of-date image of a person taken in real life, in the worst case. While the possibility of procuring a best case image is higher for a verification task, it is very low in a watch list scenario. Especially in the watch list scenario, the respective image database is of great importance as well. In this context, the question arises of why a certain person and specific information about this person is in the database. How did this person come to be integrated in a database or to be put on a watch list? These questions point to the close relation between the individual and the categorical mentioned above. Therefore, in the last part of this section I discuss the meaning of recognition in the context of practices and techniques that focus on categorical recognition, such as behaviour pattern recognition. This is a specific manifestation of Pattern Recognition Technologies that feature strong similarities and share communalities with technologies for individual identification such as FRT.

...to Categorical Identification: Pattern Recognition (Technologies)

Automated Facial Recognition is a Pattern Recognition Technology. The pattern to be found in the image of interest, connecting facial landmarks to a network and measuring distances and angles (Gross, Shi & Cohn 2001 cit. in Introna & Wood 2004: 185) is compared to the pattern derived from the reference image. However, it is not the image itself or the visual representation of a network of facial landmarks that has to be matched in order to evaluate the grade of sameness. Moreover, the two templates are transformed into binary codes. Thus, binary codes are the basis for the matching of two facial images and therefore for the decision of whether a certain face representing an individual person can be recognised in another image.

Behaviour pattern analysis works in quite a similar way. Its central goal is to find out if a certain behaviour observed—not the facial image of interest—corresponds with a specific expected behaviour—not the reference facial image. That means, a reference behaviour pattern is needed to evaluate the behaviour of interest. When I was an undergraduate student in sociology, a professor of one of the introductory courses sent us to a popular public transport hub close to the university. He sent us there just to observe and our task was to describe what we saw. What happened to all of the student groups was as one might expect: sensory overload. There just was so much to see, but also so much to listen to and to smell that it was almost impossible to see something meaningful. On the other hand, the obvious and therefore not so noteworthy observation was of people waiting for and entering trams, or trams arriving and people leaving the tram. Retrospectively, I can clearly see different patterns that occured at this public transport hub. For example, it is a pattern that at night time between 1 a.m. and 5 a.m. there are no trams and therefore also no people waiting for or leaving or entering trams. However, it might also be the case that some person would still wait for a tram to arrive, because he or she does not know that there are no trams between these times. This might be interpreted as an exception or deviation to a certain behaviour pattern and could therefore attract attention. On the other hand, it might be of no interest at all. It moreover depends on whether this behaviour pattern deviation is a relevant "domain of scrutiny" (Goodwin 1994: 606). For example, the scene could be of interest because of public safety and order, and therefore security guards or police forces would observe the relevant domain of scrutiny with a specific set of discoursive practices (ibid.) in order to maintain this. The watching might also be done from afar with the help of video surveillance or CCTV cameras which has been referred to as "governance from the distance" (Krasmann 2005: 308ff.). Norris and Armstrong, in their book *The Maximum Surveillance Society* (1999) analysed the behaviour of CCTV control room operators in the UK and how suspicion is socially constructed (ibid.: 117ff.) in these control rooms. One of their main findings was that the differentiation of whether somebody or something is suspicious "is not based on objective behavioural and individualised criteria, but merely on being categorised as part of a particular social group" (ibid.: 150). In fact, they compiled seven types of suspicion, with three of these types being commonest: categorical suspicion meaning suspicion based on the style of clothing, race or subculture was the commonest one (34% of the observed cases). Nearly as common as categorical suspicion was transmitted suspicion, for example, initiated by the police or store detectives (31%). In 24% of the cases the type of suspicion was behavioural. In such cases, for instance, CCTV control room operators were acting following fighting incidents. Locational suspicion (4%), personalised suspicion (3%), protectional suspicion (2%), and voyeuristic suspicion (1%) were subordinate (ibid.: 112). That means that the social construction of suspicion is not one dimensional (ibid.: 121), but in almost all of the cases, refers to the “normal ecology” of an area (Sacks 1978: 195 cit. in Norris & Armstrong 1999: 118). Following ethnomethodologist Harvey Sacks, Norris and Armstrong argue that for CCTV control room operators “the normal ecology of an area is also a ‘normative ecology’ and thus people who don’t belong are treated as ‘other’ and subject to treatment as such” (ibid.: 119). But what counts as “normal” and ”normative” is far from being obvious, moreover it is dependent on the social and cultural interpretation of a situation or, in Goodwin’s words, dependent on “a specific set of discoursive practices” (Goodwin 1994: 606). Norris and Armstrong, for example, observed that running and loitering—two types of non-criminal behaviour—were “seen as potentially indicative of criminality” (Norris & Armstrong 1999: 131). As this kind of behaviour was often observed in undesirable social groups such as the homeless, vagrants or alcoholics, it “had less to do with their criminogenic potential but more to do with the capacity to convey a negative image of the city” (ibid.: 141).

As Norris and Armstrong show in their analysis, in the case of CCTV control room operators, the reference behaviour pattern needed to select the behaviour of interest is based on the operators view of the “normal” and therefore of the “normative” ecology of the area of observation. Interestingly, Norris and Armstrong observed that in most cases the selected behaviour of interest did not lead to any deployment. This ocurred particularly because the “CCTV operators could not themselves intervene nor could they demand intervention by the police” (ibid.: 198). They argue that “suspicion rarely had a concrete, objective basis which made it difficult to justify to a third party” (ibid.). One consequence of this approach to the problem was the development of informal working rules that classify “people and behaviours as worthy of attention” (ibid.: 200). The formalisation of working rules, or in other words making the tacit rules explicit, will be of central consideration in the next chapter when I discuss the computerisation of seeing and recognition activities in the context of human/machine vision (re)configurations. In the context of video surveillance and CCTV, Norris and Armstrong made 'automated algorithmic surveillance' a subject of discussion (ibid.: 210ff.). They claim that

> "the true panoptic potential is only fully realised when the mass of the citizenry is not only subject to being watched by the cameras, but when the information derived from surveillance can be documented and dossiered." (ibid.: 210).

In their view, Pattern Recognition Technologies like facial recognition systems, licence plate recognition, or intelligent scene monitoring are an answer to the limitation and cost problem of human monitoring, information processing and handling (ibid.). Intelligent scene monitoring is an example for automated classification based on behavioural clues. A consequence of this automated classification is the inclusion and exclusion of certain people or social groups (ibid.: 220): what David Lyon referred to as 'Social Sorting' (Lyon 2007a: 163).

Conversely, these classifications might also be "blind to a person's social characteristics" (ibid.: 225) and therefore would be seen as "neutral" technology. In contrast to this neutrality view of technology and algorithmic decision-making, Introna & Wood's insights on facial recognition algorithms showed that FR algorithms are systematically biased (cf. Introna & Wood 2004). That means, algorithms that distinguish between one group and another are far from being neutral. Moreover, they seem to be similar to CCTV control room operators: social actors influenced by certain assumptions about the "normal ecology" of an area (Sacks 1978: 195 cit. in Norris & Armstrong 1999: 118) of interest. In this context, Katja de Vries made use of Derrida's use of the term ‚Shibboleth'[28] (Derrida 1992 cit. in de Vries 2010: 76) in her paper on profiling algorithms and ambient intelligence. In her interpretation a shibboleth is then

---

[28] The term ‚Shibboleth' can be found in the Christian Bible in ‚The Book of Judges', 12, 5-6, Old Testament: „(5) The Gileadites captured the fords of the Jordan leading to Ephraim, and whenever a survivor of Ephraim said, "Let me cross over," the men of Gilead asked him, "Are you an Ephraimite?" If he replied, "No," (6) they said, "All right, say 'Shibboleth.'" If he said, "Sibboleth," because he could

> "a device used to decide who is in and who is out; who is us and who is them; who is likely to be a good customer and who is not; who is allowed to pass the border and who is not." (de Vries 2010: 76).

In this conception of an algorithmically constructed shibboleth (ibid.: 83) it is not the human operator or guard who has the ability to judge if and under which circumstances a person belongs to a certain group or category. It is the algorithm that owns this ability and therefore is also responsible for futher consequences or treatment. In my interpretation, an algorithmically constructed shibboleth (ibid.) is a device that determines and decides about characteristics and peculiarities in the behaviour of people or things that indicate or even make obvious the belonging of the particular behaviour—connected to a person or thing—to a certain social group or category. In this context, as de Vries notes, "inductive reasoning" and thus, the "assumption regarding the sameness of different people" and the "sameness in time" play a crucial role and can also "lead to irrational stereotyping and discrimination" (ibid.: 80). That means, to come back to my transport hub example: if the behaviour of all people waiting for a tram between 1 a.m. and 5 a.m. is considered to be a problem for public safety and order by an algorithm—for example by sending an alarm to the police—all people waiting at this specific site at this time, no matter what their intent, are suspicious and will be under scrutiny. De Vries also raises normative questions in the context of algorithmic categorisations and indicates the possibility of "a serious increase in unwarranted discrimination in general" following the "increased use of machine profiling" (ibid.: 83). She claims that "further academic and societal debates (...) needed to decide which forms of discrimination are warranted and which are not" (ibid.). That is in my view an important claim as we know that not all classifications are discriminatory. When Bowker and Star note (2000) that "to classifiy is human", they point to the omnipresence of classification work in everyday life and the far-reaching impact of classifications. Classification and categorisation create social and moral order that influences human lifes significantly. They can either "give advantage or they give suffering" (ibid.: 6). In Bowker and Star's definition, classification "is a spatial, temporal,

---

not pronounce the word correctly, they seized him and killed him at the fords of the Jordan. Forty-two thousand Ephraimites were killed at that time."

or spatio-temporal segmentation of the world“ (ibid.: 10). The first consistent, unique classification principle is the “genetic principle of ordering”, creating temporal and/or functional order. Classification is also always dependent on the creation of residual categories that are mutually exclusive. At least in an ideal world, these categories are clearly demarcated from each other. The third criterion for classification is the completeness of a system, totally covering the world it describes (ibid.: 10f.). In practice, these ideal classification characteristics are hard to achieve, but in some cases there is plainly a need to achieve them. Bowker and Star demonstrate this amongst other things using the involvement with racial categories under apartheid in South Africa and the associated quiet politics of classification (ibid.: 195ff.). They argue, “For apartheid to function at this level of detail, people had to be unambigously categorizable by race” but “this task was not to prove so easy” (ibid.: 201). This especially was due to several deviations from pure types, as a “lack of a scientific definition of race” appeared repeatedly (ibid.: 202). The technologies of classification were rather to be found in the everyday understanding of people. For example, “combs were sometimes used to test how curly a person’s hair was”, or a pencil was stuck in someone’s hair in order to test if the person’s hair was very thick indicating a black person (ibid.: 210). The consequences of these rather arbitrary technologies of classification were enormous resulting in “disparities in power and privilege” and it was “not surprising that so many coloured people wanted to pass as white” (ibid.: 216).

Chapter Three

Human-Computer Vision (Re) Configurations

What seems to be retrospectively, rather an arbitrary technology of classification in the case of race classification in apartheid South Africa, but nevertheless had enormous discriminating consequences, could also be transferred to what seems like an innovative and sophisticated system of race classification. In theory even if there were the possibility of estimating the curliness or thickness of hair automatically, by fictional, sophisticated "smart," advanced technology, one still has to question whether this is the right method of evaluating race or ethnicity of a person. In this case, it becomes clear that this fictional, sophisticated, advanced technology used for recognising the curliness or thickness of hair is not "better"for evaluating the race category of a person than is the out-dated pencil test described in the final part of chapter two. As real and far-reaching the arbitrary pencil test was in apartheid South Africa, so too would any fictional, smart, advanced technology be today. It depends on how much authority, power and truth is allocated to the method or theory and by whom. In this regard the widespread view of technical authority and neutrality (cf. Gates 2010: 10) should not be underestimated. Technology and machines commonly count as "more accurate and objective" and "less subject to prejudices" (ibid.) than humans. Technologies like Facial Recognition or other Pattern Recognition methods that purport to possess human vision and recognition abilities are in line with what Gates calls the "digitizing impulse". It is just "another way of standardizing images and eliminating individual human judgement in their interpretation" (ibid.). Thus, the digitizing impulse is a re-articulation of the mechanizing impulse in the 19th century (Daston & Gallison 1992 cit. in Gates 2010: 10). When Gates identifies digital technologies as "being used to make newly empowered claims to the truthfulness of visual images" (Gates 2010: 203) she argues against Mitchell who sees digital technologies as subversive to traditional notions of photographic truth (Mitchell 2001: 223 cit. in Gates 2011: 203). Also, Borck argues for digital computer technology to be in line with the tradition of 19^{th} century image production and engagement where the interference of human involvement was thought to have been eliminated (Borck 2001). He notices a "paradox of technical transparency" (ibid.: 388): On the one hand in the production of images, more and more digital technology is used, whereas exactly this increased use of digital technology leads to the view that images are extraordinarily real and represent reality exactly as it is. But—as Donna Haraway famously noticed—knowledge is always situated and perspectives are always partial (Haraway 1988). That means that we cannot assume that there is "a universal, disembodied, objective form of vision, outside of any particular vantage point or subject position" (Gates 2010: 10), but that humans as well as machines "must necessarily embody particular ways of seeing" (ibid.). Gates questions the view that "objective, all-seeing machines that function much better, more efficiently, and more powerfully than human perception alone" can exist. Instead, she promotes the view that "computers 'see' only in a metaphorical sense, only in highly constrained ways, and only with a significant investment of human effort." (ibid.: 11). So for example, for a machine as is also the case for a human being, it is not always clear how to differentiate between men and women. Both machines and humans have to master the recognition of this difference and have to learn that there is a difference and that this difference is of importance in most societies. Both machines and humans also have to learn what the characteristics and features that indicate or determine the gender or sex of a person are. Questions of resemblance and difference—in this case which characteristics exemplify resemblances within a group of men or within a group of women? Which characteristics differentiate these groups: are key in constituting what is real? (Suchman 2008: 140) However, these resemblances and differences are not set in stone, they have to be continuously enacted (ibid.: 141). What seems to be clear and obvious for most adults in everyday life when located in familiar surroundings, is a real challenge for machines; a challenge that cannot be solved without significant investment of human and societal effort.

From today's perspective it is especially in the fields of computer science and computer vision and their connections to research in Artificial Intelligence (AI) that is engaged in this challenge and in research activities. Thus, in this chapter I first deal with what has been famously called "The Sciences of the Artificial" by Herbert Simon in 1969. The term was taken up by Lucy Suchman (2008) and thus, brought into the realm of (feminist) Science and Technology Studies (STS). In this regard I will bring together computer science and AI literature with STS exploration into these areas. One of the most popular and recognised actors and the main reference point in the discussions about AI is the fictional character *HAL 9000* from Kubrick's movie *2001: A Space Odyssey*. Therefore, HAL will be of special interest when I discuss visions, imagination, expectations and promises connected to AI and computer vision, and connected to the larger societal transformation processes of digitalisation, automatisation and smartisation.

Subsequent to this analysis of the interconnections between science fiction and science fact, I engage with the essentials of computer science and computer vision history in brief and argue that these endeavours cannot be regarded as universal and global. Moreover, local differences and particularities must be considered as well. This is of great importance as the special focus of my study is on the relatively small geographical and cultural area of the nation-state Austria that often positions itself as a "Gallic village" when it comes to the deployment of certain new technologies (Felt 2013: 15). Therefore, I will also elaborate in brief on the history of computer vision in Austria in, by referring to the formation of the most important umbrella organisation, the *Austrian Association for Pattern Recognition (AAPR)*. The national techno-political identity of Austria (Felt 2013) and in addition to this, the techno-political identity of Europe as a whole, is connected to the current state of computer vision and pattern recognition in Austria which will also be addressed in this section.

The Sciences of the (SM)Artificial: From the Human-like Machine to Smart Environments

In Science and Technology Studies (STS) the automation of human abilities, tasks and procedures such as human vision has been discussed and analysed under the term 'Sciences of the Artificial' (Suchman 2008) until now. The term was coined by Nobel Prize Laureate, economist, computer scientist, psychologist, and management theorist Herbert Simon in 1969. Suchman, from her feminist perspective, contrasts "Simon's conception of relations of nature and artifice" with his effort to overcome this boundary by exploring the historically relevant practices that created this boundary between nature and culture (Suchman 2008: 141). By doing so, she questions "antecedents and contemporary figurings of human-technology relations" (ibid.: 139). Suchman especially was concerned with questions of "what understandings of the human, and more particularly of human action, are realized in initiatives in the fields of AI and robotics" (ibid.: 144). Projects within the Sciences of the Artificial that aim at humanlike machines bring up the question of what it actually means to be human or humanlike. What is characteristic of being human and acting like a human? Adapting this to my research interest, I need to ask the question: what is then characteristic of human vision and recognition?

Questions on how boundaries between humans and non-humans are drawn, re-drawn and modified in these projects ensue after asking initially what is characteristic for human vision and recognition. Research in AI and robotics always contain repetition or mimicry. The machine is becoming a powerful 'disclosing agent.' Assumptions about the human and what it means to be human come to light. Thus, a way to break down these assumptions is to explain how computer scientists and engineers, the very people that are extensively working on constructing robots and human-like machines, imagine being human (Suchman 2007: 226) and imagine human vision and its relationship to computer vision. Referring to Donna Haraways notion of 'figuration' (Haraway 1997: 11), Suchman observed that the prevalent figuration of the human-like machine in Euro-American 'imaginaries' is one of an autonomous, rational agency. AI projects have simply reiterated these culturally specific assumptions (Suchman 2007: 228). In these

Euro-American imaginaries the figuration of the child is of importance as well. In this conception the learning child is confronted with a specific trajectory of development that brings with it unavoidable periods and stages (ibid.: 237). With her book *Plans and Situated Actions* (1987) Suchman has given the most significant social scientific contribution to the field of computer science so far (Collins 1995: 292). Suchman showed that plans of action prescribed by computer programmes and machines can be applied to human action only in retrospect. Persons react moreover, to set actions in everyday life in an unrestricted manner because everyday situations are just too unpredictable for plans. This "emphasis on sociality" to be found in Suchmans work stands, as one might expect, "in strong contrast" to the "fixation on the individual cogniser as the origin point for rational action" (Suchman 2008. 144). Suchman and what she calls the 'feminist frame' argue that

> "the universal human cognizer is progressively displaced by the attention to the specificities of knowing subjects, multiply and differentially positioned, and variously engaged in reiterative and transformative activities of collective world-making." (ibid.).

This critical thinking about how humans and machines are continuously being (re)-configured calls for thought on how human-machine (re)configurations could be conceived differently (ibid.: 153). This claim does also connect to the broader aim of Science and Technology Studies, namely the

> "understanding of science as culture, as a way of shifting the frame of analysis from the discovery of universal laws to the ongoing elaboration and potential transformation of culturally and historically specific practices to which we are all implicated, rather than innocently modest, witnesses." (ibid.: 153f.).

If we connect this aim of leaving the discovery path of universal laws to one of the central insights in chapter two: that is human vision is inevitably culturally and historically specific, we can ask what this means for the creation of computers and machines that are able to see? It is this question that will be tackled in what follows.

The Sociology of Scientific Knowledge Standpoint on "Intelligent Machines"

Knowledge is one of the central resources of the 'sciences of the artificial' in its project to build human-like machines with the capability to see, as it is also in STS. From the *Sociology of Scientific Knowledge (SSK)* standpoint, knowledge is always social, therefore a computer cannot show the whole range of human abilities, but only that kind of knowledge that can be computerised (Collins 1995: 298). From this perspective, the key to understanding the possibilities and limits of intelligent machines is 'tacit knowledge' (Collins 2010). In his book *Tacit and Explicit Knowledge* Collins (2010) extends the term 'tacit knowledge'—that was introduced by Michael Polanyi[29]—and shows how it consists of three elements: 'relational' (contingencies of social life), 'somatic' (nature of human body/brain) and 'collective' (nature of human society). In my view especially interesting and the "irreducible heartland" of the concept of tacit knowledge (ibid.: 119) is collective tacit knowledge (CTK). In Collins' argumentation the individual can acquire this specific kind of knowledge only by being embedded in society (ibid.: 11) and by having what Collins calls 'social sensibility' (ibid.: 123)[30].

For Collins, CTK is "strong knowledge", because there is no way known "to describe it or to make machines that can possess or even mimic it" (Collins 2010: 11). This notion not only draws a clear distinction between humans and machines, but also emphasises the uniqueness of humans (ibid. 123). For example, when Collins shows the impossibility of socialising pets such as cats and dogs (ibid.: 125). Collins questions the AI dream that humans and machines "will come to be able to communicate in a way that is indistinguishable from the way humans who are part of the same cultural group

[29] By tacit knowledge Polanyi (1966) means that 'we can know more than we can tell'. In his conceptualisation, tacit knowledge is not captured by language or mathematics, but has to be performed in everyday interactions.

[30] Almost 100 years earlier, Charles Horton Cooley, a precursor of symbolic interactionism (cf. Helle 1977) distinguished between spatial/material and personal/social knowledge (Cooley 1926: 60). The former, based on sense perceptions, gives rise to exact or quantifiable natural science. The latter—in close connection to Collins' term ‚social sensibility'—only emerges in the negotiation and communication of other people's way of thinking.

communicate" (ibid.: 52). On the other hand, Collins does not deny that certain human activities can be transferred to and mimicked by machines (ibid.: 55)[31]. This is the case when actions "mimic the world of mechanical cause and effect" that "are called mimeomorphic actions" (ibid.). An example for a mimeomorphic action is the 'salute.' A 'salute,' for example in military contexts, could be executed better by a fictional saluting machine than by a human. In contrast to the salute being an example for mimeomorphic actions the 'greeting' in everyday life is as an example for polimorphic[32] action. These are actions "where the associated behaviours are responsive to context and meaning" (ibid.). Actions like 'greeting' "exist because they are collectively constituted" (Collins & Kusch 1998: 23). In principle, for specific activities such as transportation, mimeomorphic procedures could substitute polimorphic actions (Collins 2010: 170). Collins describes such a scenario:

> "If every car, human, dog, and cat was fitted up with a responder that could be read by satellite, and nobody minded changing the car from an instrument of gratification to a pure instrument of transport, it would be conceptually easy to automate the entire journey to work while avoiding running over humans, cats, and dogs." (ibid.: 169f.).

What Collins describes as a rearrangement of the world by satellite-steered responder technology in order to automate transport, comes close to contemporary visions of the so-called 'Intelligent Road' or 'Intelligent Transportation Systems' (ITS). What could be read as an attempt to boost traffic efficiency and security could be also read in terms of surveillance. In this context, Collins refers to Orwell's famous novel *1984* in order to envision such a scenario in which "it is a matter of the extinction of troublesome cultural diversity and context sensitivity." (ibid.: 170).

Intelligent Transportation Systems as one potential application area of Image Processing Algorithms are part of new infrastructures and new mobilities, and thus,

---

[31] Collins here refers to the theory of the shape of actions, or action morphicity that has been developed by himself and earlier by Martin Kusch (cf. Collins & Kusch 1998).

[32] The term polimorphic does not refer to the prefix "poly" that indicates "manyness", but refers to the prefix "poli" in order "to connote that the appropriate behavioral shape of such an action has to be determined by reference to the society (polis)" (cf. Collins & Kusch 1998: 33)

part of new sociomaterial assemblages. In this regard Sheller and Urry (2006) connect surveillance and infrastructures of security with what they call the 'new mobilities paradigm'. They assume 'multiple mobilities'—these are all actual and possible movements—that increasingly shape, organise and structure all areas of social life. This involves multiple ways of moving such as walking, or running, movements enhanced by various forms of technology such as bicycles, busses, cars, trains, planes and ships, but also movement of pictures and information like letters, postcards, fax, email, telephones etc. as well as 'many-to-many' communication in the form of networked and integrated computer systems (ibid.: 212). Connected to these 'multiple mobilities' are 'nodes', for example stations, hotels, motorways, resorts, airports and so on. Airports are especially interesting places in which new technologies such as CCTV, GPS, or biometrics are first trialed "before moving out as mundane characteristics of cities" (ibid.: 220). Typical of these new mobilities and infrastructures are also computers that "make decisions in nanosecond time, producing instantaneous and simultaneous effects" (ibid.: 221). These effects are reflected in what Adey referred to as 'secured and sorted mobilities' (Adey 2004). Airports, as symbols of mobility and emblematic of our post-modern world (ibid.: 500) and their concentration on the border are one of those places, in which "undesirable mobilities may be distinguished from the desirable" (ibid.: 502). Here we are exactly at the point that Collins calls "the extinction of troublesome cultural diversity" (Collins 2010: 170). As these sorting processes are continuously transferred to machines and computers, or to be more precise, as these sorting processes are newly arranged in specific 'sociomaterial assemblages' (Suchman 2008: 150ff.) in which "smart machines" are going to play vital parts, it is important to understand how these automated processes work, on what basis they were created, how they are embedded in existing 'sociomaterial assemblages' and how much authority and credibility, and level of "smartness" are attributed to them by whom. Exactly these questions will be approached in the following empirical chapters.

"What everybody knows": The Conceptualisation of Knowledge in Expert Systems

Diana E. Forsythe explored the creation and development of knowledge in the context of expert systems in computer scientist laboratories (Forsythe 1993). Thus, her writings are highly relevant for approaching the question on knowledge categories and how this knowledge is used in the creation of Image Processing Algorithms. Forsythe has been interested in the ways knowledge is conceptualised by specific groups of scientists and how these concepts are realised in the practice of knowledge production. Her research showed that computer scientists considered knowledge acquisition as problematic, because "inefficient" humans had to be involved in this process (ibid.: 454). This problematisation of the human was due to the specific conception of knowledge inherent in the computer scientists she followed. In this conceptualisation, knowledge was understood as formal and codified. By contrast "what everybody knows knowledge" was not defined as knowledge per se (ibid.: 458). Following this insight one might ask how computer scientists deal with rather more informal, fluent and changing forms of knowledge that could be called 'tacit' or 'non-explicit' knowledge. To give an example from my fieldwork observations, the ability to recognise whether something is machine written or hand written might be clear for most literate people that are used to both types of writing. There might also be tacit agreement about this recognition task, meaning that any form of expert knowledge for this specific recognition task would not seem to be needed. The ability to recognise and differentiate between machine written and hand written texts does not appear as something specific, but as something self-evident ("what everybody knows"). In this regard, coming back to Forsythe, she observed introspection as a method of research in the process of engineering (ibid.: 458). That means engineers of expert systems relied on their own unproblematically perceived views instead of seeking out empirical data. She identified an engineering ethos with a clear technical orientation in problem solving practices. In this regard trial and error as "practically rather than theoretically problem solving" was preferred (ibid.: 456). It became apparent that it was "better to build a rapid prototype and see how it turns out than to map it out exhaustively beforehand" (ibid.). However, these conceptualisations and engineering assumptions and the common knowledge connected to it, might be biased in a particular direction. For example, they might be predominantly gendered – meaning from a specific male viewpoint, as was also observed by Forsythe—whereas a female view is neglected. This would not be as problematic, as such gendering would be open to scrutiny and perceived as one particular perspective or one "particular way of seeing" (Gates 2010: 10). On the contrary, it might be the case that the specific male view is perceived as universal und neutral. Furthermore, from a social scientific perspective, knowledge is never self-evident, but must be interpreted (Forsythe 1993: 453) in different social constellations and situations. Instead of the conceptualisation of knowledge in expert systems being static, formal and codified, Forsythe's anthropologically informed social scientific view is that individuals' beliefs "are modified through negotiation with other individuals" in everyday life (ibid.: 466). A view I share with her.

Two other characteristics of knowledge in expert systems observed by Forsythe worth mentioning here were first, the brittleness of background knowledge that has been taken for granted, and second, its narrowness following the involvement of only one expert view (ibid.: 467). In conclusion, Forsythe notes that any knowledge-based system necessarily involves selection and interpretation. In reference to Bourdieu (1977) she argues that knowledge engineers exercise power, because "the ability to decide what will count as knowledge in a particular case is a form of power" (Forsythe 1993: 468). Consequently, "the exercise of this power is to some extent invisible" (ibid.: 469). This means that the engineers' specific (e.g. male, western etc.) 'situated' view with all its tacit values and assumptions is being black-boxed and stabilised over time. Nevertheless, it is perceived by the user of such a system as being 'correct' and 'true' in every sense.

> "Furthermore, while system-builders know that every knowledge base has its limitations, they do not appear to be aware that members of the public may not know this." (ibid.)

This means, similar to the power gap mentioned in Chapter Two about the role of DNA evidence between law enforcers with their nuanced understanding and prisoners with their absolute black and white view as recognised by Prainsack (2010: 171), there might also be a power gap between the capabilities and limitations of data processing ‘expert systems’ of system builders and members of the public. This power gap and connected to it, the public understanding of science and “smart” technology will be addressed in more detail in Chapter Four.

Finally, Forsythe brings in the political aspect of the power exercised by knowledge-based system engineers by asking “whose interests are embedded in that ‘knowledge’?” (ibid.). This does not mean that all knowledge-based systems are problematic from the beginning, but following this conceptualisation they are—as are humans —political and moral. Nevertheless, it is important to analyse (empirically) how certain assumptions are inscribed in technological artefacts such as Image Processing Algorithms and how these inscribed assumptions might affect society. In what follows, I shall give an example of what this means by looking at computer vision and image processing literature, before my empirical analysis in this regard is presented in Chapter Five.

What does a Cow Look Like? Challenges and Possibilities of Computer Vision

Sonka, Hlavac & Boyle (2008) for example, in their introductory book on image processing and analysis explain the challenges and possibilities of computer vision using the example of a cow (ibid.: 2). They describe that following a ‘training phase’ in which the system is taught what a cow might look like in various poses, a model of a cow in motion could be derived. In consequence:

> “these models could then be fitted to new (‘unseen’) video sequences. Crudely, at this stage anomalous behavior such as lameness could be detected by the model failing to fit properly, or well.” (ibid.2).

One central assumption in this statement is the assumption of similarity (cf. van der Ploeg 2011: 30). They describe that when the system is taught what a cow might look like, it is assumed that there is only one universal cow. One look at the global databank of animal genetic resources shows that there are 897 reported regional cattle breeds, 93 regional transboundary cattle breeds, and 112 international transboundary cattle breeds (FAO 2007: 34ff.). That means, there is certainly not one universal kind of cow, but there is a reported number of 1102 different cattle breeds worldwide. This example makes clear what Forsythe's insight into the brittleness and narrowness of background knowledge that is taken for granted (Forsythe 1993: 467) means for computer vision. In order to teach the computer what something, e.g. a cow, looks like, the human computer scientist has to give example data about the object of interest in a training phase. For instance, if the computer scientist is based in Austria and predominantly uses images of cows showing the most widespread Austrian 'Fleckvieh' cattle in order to teach the computer how cattle in general look, the possibility of recognising the 1101 other breeds such as, for example, the Ugandan Ankole cattle might be lower and thus, the algorithm discriminates all but Austrian Fleckvieh. In such a case Austrian Fleckvieh cattle would be the standard and norm of what a cow looks like.

Sonka, Hlavac & Boyle (2008) refer to this problem in their introductory book in more technical terms as a "local window vs. need for global view" problem (ibid.: 5). They note that "it is often very difficult to interpret an image if it is seen only locally or if only a few local keyholes are available." (ibid.). This problem also has in their view, a long tradition in Artificial Intelligence research. Especially the formalisation of context was (and I may add, still is) a "crucial step toward the solution of the problem of generality" (ibid.). Sonka, Hlavac & Boyle in their attempt to answer what they call the "philosophical question" of 'why is computer vision difficult?' offer next to the "local window vs. need for global view" problem five other ways of answering this question. In my view, two of these other answers are especially interesting regarding assumptions about human vision.[33] First, the "loss of information in 3D to 2D" as the geometrical properties of "typical image capture devices such as a camera or an eye (...) have been approximated by a pinhole model for centuries (a box with a small hole in it, called in Latin a 'camera obscura'".(ibid.: 3)

---

[33]The three other answers to the question "why is computer vision difficult?" that I do not dwell on here in detail are "noise," "too much data" and "brightness measured." (Sonka, Hlavac & Boyle 2008: 4)

> “The main trouble with the pinhole model and a single available view is that the projective transformation sees a small object close to the camera in the same way as a big object remote from the camera.” (ibid.)

This means the need for using (pinhole) cameras that produce 2D images in order to realise any form of computer vision is perceived as a basic disadvantage in comparison to 3D human vision. Second, the “interpretation of image(s) constitutes the principal tool of computer vision to approach problems which humans solve unwittingly” (ibid.: 4). It is argued that “when a human tries to understand an image then previous knowledge and experience is brought to the current observation.” In consequence, the argumentation goes on, the “human ability to reason allows representation of long-gathered knowledge, and its use to solve new problems.” In contrast, it is noted, even several decades of Artificial Intelligence research on the issue, “the practical ability of a machine to understand observations remains very limited.” Sonka, Hlavac & Boyle in this regard, note that the “interpretation of image(s) in computer vision can be understood as an instance of semantics”, that is the study of meanings that analyses relations between expressions and their meanings (ibid.). Hence, the interpretation of images in computer vision is concerned with the establishment of relations between expressions and their meanings, an ability that is assumed in this textbook that humans solve unwittingly. Therefore, computer vision is “an attempt to find a relation between input image(s) and previously established models of the observed world” (ibid. 5).

Depending on the means of how these relations are analysed and integrated in Image Processing Algorithms in order to interpret and understand images, these activities can be interpreted as actively (re)configuring and (re)arranging the world and thus demonstrate how the world is perceived by its inhabitants. This means, to come back to the cow example, if computer scientists teach the machine how cows look and for this endeavour predominantly use images of Austrian ‘Fleckvieh’ cattle, the world of cows and how cows are perceived is (re)configured in such a way that Austrian ‘Fleckvieh’ cattle might become the standard and possibly even the norm of how a cow looks and thus, of what is characteristic for cows in general. It might be the case that many of the 1101 other breeds of cattle are of a very similar shape as Austrian ‘Fleckvieh’ and thus, the possibility of recognising these ‘correctly’ as cows is high as well. However, there might also be at least some of the 1101 other breeds of cattle that have a different shape and thus, the possibility of recognising these ‘correctly’ as cows is rather low. Of course, this decision is also up to societal negotiations. In this case the setting of a specific threshold, for instance, stating “a cow is a cow if the algorithm recognises a match of 67% to the standard template of a cow.”

Again, this example makes clear that computer vision projects take part in (re)defining standards, classifying and differentiating among groups. On the one hand, computer vision projects (re)configure and (re)arrange the world; how it is ordered and thus; how it is pereceived and experienced by its inhabitants. On the other hand, computer vision projects (re)configure and (re)arrange humans and machines and their relationship and (re)draw boundaries between them.

The crux of the matter is that the work of computer scientists and coding in general, especially relating to classification such as social sorting (cf. Lyon 2003) never occurs in an objective or neutral way, but is embedded in specific, socially situated practices and actions. Bowker and Star (2000) see computer software in many ways as “frozen organizational and policy discourse”, in which policy is coded into software. In this view, software like technology, is “society made durable” (Latour 1991). This means that specific social practices, normative notions of what cows look like, about good and bad behaviour, political and moral assumptions, and cultural values are either consciously or tacitly inscribed in the software (Graham & Wood 2003). Moreover, “algorithmic systems thus have a strong potential to fix identities as deviant and criminal”—what Norris calls the “technological mediation of suspicion” (Norris 2002). However it is not only the individual person that is singled out for attention. In some circumstances coding and classification processes may have profound effects on the shaping and ordering of human life in general, creating new social classes (Lyon 2003). But, as was already noted in the introduction to this thesis, such a techno-deterministic view that highlights the determining effects of technologies on society falls short. Rather, I started this thesis from the premise that technology and technological artefacts such as machines or computers and human beings are closely related and intertwined. They are "co-produced" as Sheila Jasanoff has famously called this process. In her view, technology "both embeds and is embedded (…) in all the building blocks of what we term the social" (Jasanoff 2004: 2). This means that technology is inseparably intertwined with the "ways in which we choose to live in" (ibid.). The development and use of technology both shapes and is being shaped by the specific societal and political context in which it is embedded. We can witness "multiple moments of co-transformations of both the technological and the political, and their relation is under continuous redefinition" (Felt 2013: 17). What follows from this is that talking and thinking about technology is simultaneously always talking and thinking about society and "how to live" in this society. Connected to and concerned with this question of "how to live" are the 'ethics of the good life' (Verbeek 2011: 155ff.). As human existence always takes shape in relation to technology, "the question of the good life concerns the quality of the ways in which we live with technology" (ibid.: 156). Following the classical Aristotelian principle of finding the right middle ground between two extremes, Verbeek develops an understanding of the ethics of technology, and thus of the ethics of the good life, in which two extremes should be avoided:

> "It (…) aims at developing ways to take responsibility for our technologically mediated existence. It wants neither to let technology determine humanity nor to protect humanity against technology; its goal is to develop a free relation to technology by learning to understand its mediating roles and to take these into account when designing, implementing, and using technology" (ibid.: 158).

Verbeek's approach to the ethics of technology as intertwined with the ethics of the good life is in contrast to what he calls a 'dialectic approach' "which sees the relationship between humans and technology in terms of oppression and liberation" (ibid.: 155). In the context of Artificial Intelligence and the sciences of the artificial such a dialectic approach is often applied. In this regard, narrative devices shaping the understanding of the relationship between humans and "intelligent" machines are central (Bloomfield 2003: 193). As Bloomfield notes, "the focus of numerous utopian and dystopian visions" is a trajectory from mechanisation to automation to machine intelligence that "represents the increasing displacement of human skills and agency" (ibid.). In this regard one of the most popular and recognised characters and the "primary reference point of all discussions about AI" (ibid.: 194) and "intelligent" machines is the fictional character HAL 9000 from Kubrick's film and Clarke's novel *2001: A Space Odyssey*.

HAL 9000 as a Cultural Icon and the Prototype of Intelligent Machines

HAL 9000 is a cultural icon "both in making sense of the relationship between technology and human beings and as a milestone in the appraisal of technoscientific development" (Bloomfield: 193f.). HAL "has come to serve as a leitmotif in the understanding of intelligent machines and the dangers associated with them" (ibid.: 194). Or, as Bloomfield puts it simply in reference to the British newspaper *The Guardian* (June 2, 1997): "HAL 9000 is the most famous computer that never was".

Bloomfield in his involvement with HAL 9000 points to the 'significance of narrative' in organising technologies (ibid.: 195f.). His argument is that a precondition for placing technology is its organisation "in and through texts" (ibid.: 197) such as manuals but also fictional texts. In this regard, according to Bloomfield, a simple narrative structure is to be observed "when it comes to making sense of technological development" (ibid.):

> "the old (anachronistic), current (soon to be displaced), and latest (but almost in an instant already fading) technologies are interrelated, recognised and understood as such." (ibid.)

In the case of robotics and artificial intelligence, development is embedded in a path connected to previous machines or technologies, as Bloomfield notes for example, 18th century automata, Jacquard's loom and the digital computer. Finally, this path "leads to the ultimate goal of 'true' artificial intelligence" (ibid. 198). What this "true" artificial intelligence could look like with regard to computer vision was perfectly displayed in the character of HAL 9000. In *2001: A Space Odyssey* HAL

> "displayed image understanding capabilities vastly beyond today's computer systems. HAL could not only instantly recognize who he was interacting with, but also he could lip read, judge aesthetics of visual sketches, recognize emotions subtly expressed by

> scientists on board the ship, and respond to these emotions in an adaptive personalized way." (Picard 2001)

HAL 9000 showed capabilities that strongly refer to Facial Recognition Technologies (FRT), Automatic Lip-reading, Motion Analysis, Behaviour Pattern Analysis (BPA), or Facial Expression Recognition. This means that HAL was constituted from a wide array of combined and sophisticated computer vision technologies that go beyond most human vision capabilities. For example, in one of the key scenes of the film, when the two astronauts Bowman and Poole talk about disconnecting HAL, they are not aware that HAL is able to understand everything they say by lip-reading as they have unplugged the microphones in a certain protected area. This means HAL displayed an ability to read and understand complex words and sentences only by the visual technology of lip-reading that most humans could not. As the movie demonstrates this ability perfectly well, the question of plausibility and thus, future realisation arises. Seeing HAL and his computer vision abilities as part of his "true" artificial intelligence can be interpreted as a 'diegetic prototype' (Kirby 2011: 193ff.) as it demonstrates to a large public audience the utility and viability of this technology (ibid.: 195). In contrast, to strategically,[34] further demonstrate the harmlessness of a technology by means of a diegetic prototype in order to promote a specific technology (ibid.), HAL's depiction is more in the tradition of the "narrative of technology going wrong or out of control" (Bloomfield 2003: 194). No matter whether films are a means of promoting or of warning of certain technological developments, "in the popular imagination today's science fiction is expected to become tomorrow's science fact" (ibid.: 199). Bloomfield indicates the close interrelatedness of science fiction and science fact, or technology fiction and technology realisation.

> "What is key here is that such narratives shape our expectations of the future and in addition form part of the ideas or cultural material out of which the future is realized or constructed." (ibid.)

[34] For example, Kirby reports that there is an increasing number of scientists who consult fictional media projects. This cross-consulting has become standard practice in the entertainment industry in recent years (Kirby 2003: 258) also in order to promote certain new and emerging technologies.

Bloomfield demonstrates this close relation by referring to the publication of a scientific book, *HAL's Legacy: 2001's Computer as Dream and Reality* (Stork 1997) in celebration of HAL's official "birthday" on January 12th, 1997. According to Bloomfield, HAL and its part in 2001 provided a challenge to the scientific community and invited scientists to make it become fact (Bloomfield 2003: 202). The book is reviewed by Bloomfield as an attempt to assess and evaluate "HAL's capabilities in the light of current knowledge in computer science and other fields associated with artificial intelligence" (ibid.). A central insight of this benchmarking was that some of today's (n.b. in 1997) computers such as the famous IBM chess computer Deep Blue, are very good at certain tasks such as playing chess. In comparison to HAL these real computers lack human general intelligence and characteristics. The "mundane tasks of daily human existence"—which I interpret together with Collins (2010) as 'collective tacit knowledge'—appeared to be a much greater challenge than was the challenge of playing chess (ibid.: 203). But HAL's significance was derived from his abilities to be and act in a human-like way. For example, HAL's voice was friendly, warm and emotional rather than a typical, mechanical, computer voice as it was spoken by a human. This depiction led to the view that computers are going to talk like humans in the future (ibid.: 205).

The omnipresence of HAL in discussions about artificial intelligence and intelligent machines shows that such a narration of the future is "no mere passive reflection on what is to come but rather part of the process through which the future is realised" (ibid.: 211). As there are principally many ways of how the future can be realised, Bloomfield notes that "each narration of the future seeks to bring it to the present—a present future—and in so doing thereby assists in making it less open." (ibid.).

HAL's significance in the discussion about artificial intelligence and intelligent machines can be explained to a large extent by its capacity as a "cultural icon of the computer age" (ibid.: 2010). This was mirrored in wider culture as HAL offered "a ready-made storyline or narrative device through which developments in computing and instances of computers going wrong could be related." (ibid.). For example, as Bloomfield explains, HAL as the best known archetype of computers going wrong, was a perfect narrative device for the understanding and making sense of the so-called 'millenium bug' or 'the year 2000 computer bug', which resulted from the abbreviation of the representation of years in four digits to two digits. The 'millenium bug' reminded people that "microprocessors had become near ubiquitous components in the technological infrastructure of the modern world", and that "computers had infiltrated so many everyday machines and systems on which the developed world depends" (ibid.). In this regard HAL's depiction in the movie—and here it should be noted that it was 1968 when it was released—represented a networked and ubiquitous technological infrastructure in strong contrast to other depictions of intelligent machines as anthropomorphic robots such as those seen later on in the 2004 movie 'I, Robot' for instance. HAL's depiction as an omnipresent sensor, embodied and "visually represented as a red television camera eye located on equipment panels throughout the ship"[35] in the technological infrastructure of the spaceship 'Discovery', can be seen in close relation to what could be called a 'smart environment', or 'pervasive computing' or 'ubiquitous computing'.

HAL's Transformation to the Smart Home: The Simpsons Ultrahouse 3000 Vision

Nevertheless, HAL has, in contrast to 'distributed systems' that consist of "the coupling of a number of simple, decentralized agents", (Weyer 2006: 132) a "centralized brain" (ibid.), depicted in the film as a large-scale processor. However, HAL's interconnected "senses" are distributed everywhere throughout the spaceship. This arrangement of HAL inspired the authors of the famous animated sitcom *The Simpsons* to a parody of the movie *2001* and HAL in the first episode of the thirteenth season *Treehouse of Horror XII* that was first broadcast in the United States in 2001. In the second instalment of the episode *House of Whacks*, the Simpsons purchase the so-called *Ultrahouse 3000* that is according to the Simpsons Wiki, "a fully-automated house

[35] cf. Wikipedia 'HAL 9000' http://en.wikipedia.org/wiki/HAL_9000 [March 14, 2013]

produced by the ‘Ultrahouse’ company, which is a division of ‘Mega House’.”[36] The fictional *Ultrahouse 3000* is described in the Simpsons Wiki the following way:

> “A house with the Ultrahouse installed has a high-tech, futuristic look: all of the doors are automated and red camera lenses resembling eyes can be found everywhere. In general, the whole interior of the house is a robot. The house can speak (although it's unclear where the voice comes from) and move things around by itself, with the aid of robot arms that are equipped with all kinds of tools and can appear from anywhere.”[37]

In addition to this general description, the Simpsons Wiki lists some of the most important tasks *Ultrahouse 3000* can perform: “General household chores, including cleaning”, “opening and closing doors, pouring drinks, giving backrubs, preparing baths”, “checking blood alcohol content, analyzing bathroom "leavings" (to determine favorite foods and possibly general health)”, “administering medications by dispensing pills or giving injections.”, “cooking, serving food, and doing the after-meal washing up” and as a consequence of its artificial intelligence it is able to “socially interact”, meaning “it can be a pretty good conversationalist depending on which voice option is used.” The voice options available in the Simpsons case are particularly interesting, because they take up the theme of HAL’s human-like voice in the film *2001*. As "Marge doesn't like the Ultrahouse's standard mechanical voice” she and her kids “explore the other voice options and settle on having it use Pierce Brosnan, one of the many actors to play the famous character of the fictional British Secret Service agent James Bond.”[38]

Regarding the visual capabilities of *Ultrahouse 3000* the Simpsons Wiki makes mention of one particular scene in which

> “…the Ultrahouse seems to be taken with Marge, watching her get into the bathtub (to her initial embarrassment), lighting candles in the bathroom for her, and turning on the tub's bubbles, which she enjoys.”

---

36 http://simpsons.wikia.com/wiki/Ultrahouse_3000 [March 18, 2013]

37 http://simpsons.wikia.com/wiki/Ultrahouse_3000 [March 18, 2013]

38 http://simpsons.wikia.com/wiki/Treehouse_of_Horror_XII [March 18, 2013]

As Marge gets into the bathtub, “Pierce”—as the *Ultrahouse 3000* is charmingly called by Marge—notes that she does not need to cover up and characterises himself as “merely a pile of circuits and microchips.” Just following this comment when Marge undresses in order to get ready for the bathtub, the *Ultrahouse 3000,* HAL-like red camera eye focuses on Marge by zooming in for a close-up view. Connected to this, “Pierce” comments on what he sees with a highly impressed, eroticised voice articulating his attraction to Marge with a spontaneous “wow, yes!” Meaning, in contrast to his own description of being “merely a pile of circuits and microchips” the *Ultrahouse 3000* is captivated by Marge while she is undressing. In this sense, the fully-automated *Ultrahouse 3000* computer is acting in a way that could be associated with the charismatic, human character of James Bond as personated by Pierce Brosnan. The *Ultrahouse 3000* “Pierce” feels sexual desire for Marge following watching her undressing and being naked. Thus, he displays typical Western-liberal human emotions and sexual desire. As displayed in this particular Simpsons episode, the *Ultrahouse 3000* “Pierce” is able to associate and relate the visual perception of a specific naked human body (bodily expression) to a verbal comment stating attraction to this specific naked human body (social meaning of the bodily expression).

In detail, this association is achieved in at least four steps. Firstly, “Pierce” detects a human being entering the bathroom and recognises Marge as being Marge. He recognises that Marge is undressing when she is in front of the bathtub. Secondly, he recognises that Marge is naked. Thirdly, he sees at once that this act of undressing and being naked is of special interest and thus, it is reasonable to inspect the scene more closely by focusing on the undressed naked body and zooming in. And fourthly, as the scene is finally inspected more closely, “Pierce” simultaneously recognises that Marge’s naked body is attractive to him, which he audibly articulates. Meaning that the smart machine “Pierce” is able at least to understand the particular social meaning of undressing and nakedness in a liberal Western society. Furthermore, he seems to have an understanding of what kind of nakedness (Marge’s specific female body) is attractive to him.

The intelligent machine – HAL 9000 from *2001* - that shows "true" artificial intelligence is brought, in a Simpsons parody, from a spaceship travelling in outer space in the far future, down to earth and the middle-class, everyday environment of the fictional Simpsons family in the episode 'House of Whacks'. It transfers the 'smart environment' of pervasive and ubiquitous computing from the future spaceship to the present 'smart home'. In respect to this, the Simpson's *Ultrahouse 3000* and its characteristics remind very strongly of conceptions of 'Smart Homes' that have appeared increasingly in recent years. For example, computer scientists such as Augusto and Nugent (2006) describe smart homes as

> «the enrichment of a living environment with technology in order to offer improved habitual support to its inhabitants and therefore an improved quality of life for them.» (Augusto & Nugent 2006: 1)

> «The term Smart Homes creates expectations of an environment which is capable to react 'intelligently' by anticipating, predicting and taking decisions with signs of autonomy. From a computational perspective there is a natural association between this expectation and the use of techniques from Artificial Intelligence (AI).» (ibid.)

> «With such a radical change in living environment and lifestyle it is possible to witness an improved quality of life, an improved level of independence and finally an extended amount of time where the person can remain within their own home without the need of institutionalisation.» (ibid.: 2)

It is interesting to see how expectations and promises of "an improved quality of life" or independent living are articulated in these descriptions of smart homes in this piece of computer science literature. What is interpreted as an improvement here, especially for the elderly of remaining in their own homes for a longer period of time, instead of having to move into a care institution, could be on the other hand be also interpreted as Deleuze does, as new "mechanisms of control that are equal to the harshest of confinements" (Deleuze 1992: 4). This confrontation of these two views of smart homes shows what David Lyon has named a basic characteristic of surveillance: It is both care and control, and an unavoidable aspect of living in contemporary societies (Lyon 2007b: 177).

Sociotechnical Imaginaries and Located Accountability

As was demonstrated in the previous section, narrations and imaginations of the future play a crucial role in the development and the negotiation of technology both in fictional and non-fictional texts. Visions of the future drive technical and scientific activity (Borup et al. 2006). The economy of these imaginations and future scenarios is a powerful way to shape certain techno-scientific present scenarios (such as automated 'smart homes') and to disregard others (Felt 2007: 302) such as leaving home for institutionalised care. Having the notion of co-production in mind, this economy of imaginations and future scenarios is not only about the techno-scientific present and future, but also clearly involves cultural und societal spheres as we can see in the example of smart homes very well. How is such a collective imagination of the future constructed? In this regard, Jasanoff and Kim make use of what they call 'sociotechnical imaginaries' (Jasanoff & Kim 2009). These are "collectively imagined forms of social life and social order reflected in the design and fulfillment of nation-specific scientific and/or technological projects (Jasanoff & Kim 2009: 120)."

For Jasanoff and Kim, scientific and/or technological projects "are almost always imbued with implicit understandings of what is good or desirable in the social world" (ibid.: 122) and exemplify this with the question of "how science and technology can meet public needs and who even are the relevant publics" (ibid.: 122f.). That is why 'technoscientific imaginaries' are simultaneously also 'social imaginaries'. An important note is that 'imaginaries' are not to be compared to policy agendas, as they are less explicit, less issue-specific, less goal-directed, less politically accountable and less instrumental. For Jasanoff and Kim, imaginaries operate "in the understudied regions between imagination and action" (ibid.: 123).

What follows from the concept of 'sociotechnical imaginaries' is to take note of the conditions of place (Livingstone 2003) and nation (Hecht 2001). So, for example, what are Austrian specific 'sociotechnical imaginaries' in the context of computer vision? This means it is not only a matter for the laboratory as a special place of knowledge production, but it is also about the embedding of these techno-scientific activities in a nation or culture.

Livingstone reminded us that the modern invention of the scientific laboratory "can be interpreted as a conscious effort to create a 'placeless place' to do science, a universal site where the influence of locality is eliminated" (Livingstone 2003: 3). However, he reminded us as well that the familiarity with certain local customs in the scientific laboratory and elsewhere, is "fundamental to sorting out the coded messages within which communication is embedded" (ibid.: 6). Suchman's concept of 'located accountability' attempts to grasp this aspect of the local as well. With 'located accountability' she means the

> "ongoing engagement with the question of just how imagined futures are shaped by their particular circumstances, and how they work to reproduce and/or transform more extensive historical, cultural, political and economic arrangements. The concept emphasizes particularities of innovation by drawing attention to the situatedness of knowledge production" (Suchman et al. 2008: 4).

To give an example, Austria's data protection law is often regarded as 'strict' in comparison to other countries[39], which affects the work and projects of computer vision researchers in a significant way. When the amendment of the Austrian data protection law became effective in 2010, one subparagraph especially created uncertainties within the computer vision community:

> §50a (7) Data collected of data subjects concerned by video surveillance may not be analyzed by comparison with other picture data and not be searched using sensitive data as selection criteria. Special duty of documentation and deletion.

Because this subparagraph is somehow unclear[40], there is still wide scope for interpretation in particular cases (in Austrian law: 'Anlassfälle' – reference cases). In

---

[39] As I do not compare Austria or its data protection law systematically to one specific country here, it begs the question to which extent it is possible to grasp specific sociotechnical imaginaries through cross-national comparison. I would argue that this is still possible through the relevant actors' awareness and positioning of Austria in comparison to other nations and their data protection laws.

[40] Privacy groups informed me that this probably refers to Face Recognition Technologies (FRT), but Computer Vision researchers feared that it also affects 'Multi Camera Tracking' (comparing picture

dealing with data protection and privacy issues, the negative side of a data-intensive technological future such as computer vision is also demonstrated and discussed. Thus, it is not only a matter of looking at optimistic scenarios, but also at more pessimistic images of the future (Tutton 2011:1) and how they are related to the optimistic ones (ibid.: 2). Therefore, it is important to be aware of different places and activities, where the relevant actors are discussing imagination, promises and expectations of research and technology. Borup et al. (2006: 292) note that there are quite contradictory expectations amongst people closely involved in scientific work: they state that "when wearing a public entrepreneurial hat they might make strident claims about the promise of their research", but "when among research peers, they will be much more cautious and equivocal, though publicly still committed to the promises associated with their field". This difference in what is talked about where and when, will be a central theme in the following empirical chapters, especially in Chapter Six. First, I turn to the significance of place in computer vision by exploring its history in a global and local Austrian context.

---

data from one camera to another in order to track objects or persons) or to 'Behaviour Pattern Analysis'. In this case, the subparagraph would really challenge computer vision work in general.

Putting Computer Vision in its Place: Explorations into the Global History of Computer Science and its Local Particularities

Following Jasanoff & Kim's notion of nation-specific 'sociotechnical imaginaries' and Suchman's concept of 'located accountability', I shall engage in brief with the history of computer science and computer vision in particular, and argue that these endeavours cannot simply be regarded as universal and global. Moreover, local differences and particularities as well as the historical circumstances must also be considered. This is of significance as the focus of my study is especially on Austria, a relatively small country which often positions itself as a "Gallic village" when it comes to the deployment of certain new technologies (Felt 2013: 15). Therefore, as I already stated in the introduction to this chapter, the history of computer vision in Austria and connected to it, the national technopolitical identity (Felt 2013) will also be a topic in this section.

The Beginnings: Explorations into the (US) History of Computer Science and Computer Vision

As was described in the previous chapter, human vision is inevitably culturally and historically specific. So also is the case with computer and machine vision because the history of human vision and computer/machine vision cannot be clearly separated from one another. Thus, it is hard to determine where exactly the history of computer vision starts. In addition to this, there is currently no systematic analysis of the history of computer vision available. Only fragments about the beginnings and the historical development of computer vision in the form of presentation slides and notes can be found. In one of these few resources, Kropatsch (2008) in his personal perspective on the history of computer vision, gives credit especially to two publications: First, Azriel Rosenfeld *Picture Processing by Computer* in 1969 and second, Linda G. Shapiro *Computer Vision Systems: Past, Present, and Future* in 1983. In his 1969 book, considered to be the first computer vision textbook ever, Rosenfeld noted that over "the past 15 years, much effort has been devoted to developing methods of processing pictorial information by computer" (1969: 147). That means according to Rosenfeld that the first attempts at image processing and computer vision started in the mid to late 1950s. Shapiro, in her review of the most important early computer vision systems, remarked that the Ph.D. thesis by Lawrence G. Roberts, published at MIT in 1965 "is generally credited with designing and implementing the first computer vision system" (Shapiro 1983: 200). In 1968, the first journal in this field was launched with the title *Pattern Recognition*. Two years later, in 1970, the first *International Conference on Pattern Recognition (ICPR)* was held and in 1977, the first *Computer Vision and Pattern Recognition (CVPR)* conference took place (Kropatsch 2008: 7). In 1978, the *International Association of Pattern Recognition (IAPR)* came into official existence[41]. According to their own description today, on their official website, IAPR

> "is an international association of non-profit, scientific or professional organizations (being national, multi-national, or international in scope) concerned with pattern recognition, computer vision, and image processing in a broad sense."[42]

These major dates in the early history of pattern recognition, computer vision, and image processing coincide with the history and establishment of the computer sciences. As Ensmenger shows (2010: 131), in the late 1960s computer science became "normal science" (Kuhn 1962 cit. in Ensmenger 2010: 131). That means, in this period of time in the late 1960s with the establishment of the first journal, textbook, and conference, the disciplines of computer vision, pattern recognition and image processing also became "normal science", at least in the United States. It is interesting to note here that in the case of computer vision many terms meaning literally the same were coined in the beginnings and are still used today. For example, the official history on the website of

---

[41] A document of the detailed history of the IAPR from its founding in 1973 to 2008 can be found on http://www.iapr.org/aboutus/history.php [March 26, 2013]. On page 15 of the document they state that "following its formal organization in Kyoto in 1978, the *International Association for Pattern Recognition* (IAPR) was recognized as a full-fledged international organization, created to serve the field of pattern recognition in a manner similar to the way the International Federation for Information Processing (IFIP) was serving the computer field."

[42] http://www.iapr.org/aboutus/ [March 26, 2013]

the *Austrian Association for Pattern Recognition (AAPR or OAGM)* lists eight different concepts “that all somehow define very similar fields”[43]:

1. Image Processing
2. Image Understanding
3. Image Analysis
4. Image Pattern Recognition
5. Computer Vision
6. Machine Vision
7. Robot Vision
8. Industrial Vision

Before I return to the specific Austrian history of Pattern Recognition, I shall first have a closer look at the establishment of computer science and its role in what could be called the “great computer revolution” (Ensmenger 2010). Ensmenger locates this “great computer revolution”, or “the larger process of computerization of modern society” in the mid-to late 20th century not so much in the computer hardware itself, but more in computer software and thus, in its history (ibid.: 5f.). He notes, for the computer to work “what matters is that it is programmable” and that it is software “that makes a computer useful” (ibid.: 5). Software for Ensmenger is

> “an ideal illustration of what the historians and sociologists of technology call a sociotechnical system: that is to say, a system in which machines, people, and processes are inextricably interconnected and interdependent.” (ibid.: 8).

Thus, the involvement with software makes the “social dimensions of technology (...) particularly apparent” (ibid.). Also, in the history of the computer revolution, software became more and more significant. Ensmenger remarks that

> “by the end of the 1960s software had become the computer: software, rather than the computer, had become the focus of all discussions, debate, and dissension within the computing community” (ibid.: 9).

[43] http://aapr.icg.tugraz.at/history.php [March 20, 2013]

That meant following the early years of the computer revolution with a focus on the computer as a machine and hardware, computer software had become the main object of interest. In the US, on the way to this change of attitude in the 1950s, computer programming was still an "inherently undisciplined and unscientific activity" (ibid.: 128). There was an ongoing conflict between theory and practice and between academic computer scientists and professional business programmers (ibid.). Within the 1960s in the US, computer science had to undergo a turn towards information processing. In Europe, what was to be called computer science had been organised earlier "around the study of information" as was reflected in the German word "Informatik" that was widely used throughout European countries (ibid.: 130). But in the end, as Ensmenger notes in his US perspective, it was not information but the algorithm that came to be the "foundational concept of modern computer science" (ibid.). In this regard, it was the computer scientist Donald Knuth of Stanford University who claimed that the study of the algorithm defined the modern discipline of computer science (ibid.: 131). Knuth in his work on *Fundamental Algorithms* (1968) that was the first volume of a series called *The Art of Computer Programming* saw the origins of the discipline in the work of 9th century Persian mathematician Muḥammad ibn Mūsā al-Khwārizmī. The modern word 'algorithm' was created from his name (Ensmenger 2010: 131). That means at the beginning of the 1970s, the algorithm became the fundamental unit of analysis in computer science (ibid.). It "provided clear and well-defined problems (along with some exemplary solutions) for students of the discipline to study and pursue" (ibid.: 132). In the mid 1970s computer science "established itself as a mathematically oriented discipline with real scientific credibility" (ibid.: 135). Contradictory to this development there was still a certain mismatch between the output of the computer scientists and the needs of business in the US. The business sector—corporations such as IBM or Bell—criticised the newly established discipline of computer science for being too theoretical and not useful for the real world (ibid.: 134). This criticism was directed at computer programmers and their personal styles in their work in and for corporations. The view that this new species known as programmers that emerged in the course of the 1950s and 1960s was problematic was widespread in the industry (ibid.: 149). Problematic in this case meant that a certain lack of sociability and a disinterest in fellow humans was ascribed to computer personnel in general (ibid.). On the other hand, the work done by computer programmers required a high degree of creativity and ingenuity as Ensmenger notes:

> "Translating even the simplest and most well-defined algorithm into the limited set of instructions understood by a computer turned out to require a great deal of human ingenuity" (ibid.: 151).

This expression of computer programmer creativity in the application of algorithms was not the most challenging part. Moreover, "the process of constructing the algorithm in the first place turned out to be even more challenging" (ibid.), explaining the importance of the mathematical and scientific basis of algorithms. As Ensmenger explains:

> "(...) Even the most basic human cognitive processes are surprisingly difficult to reduce to a series of discrete and unambigious activities. The skills required to do so were not just technical but also social and organizational." (ibid.: 152).

The view of some influential scholars such as Herbert Simon, that improvements in Artificial Intelligence and the development of 'thinking machines' "would lead to the elimination of the computer specialist altogether" (ibid.: 155) was not shared outside the community. To date, this has never been realised as such. Instead, computer programmers "assumed a position of power" (ibid.: 29) as "programming had been identified as a key component of any successful computer installation" (ibid.: 29) as soon as the beginning of the 1950s. And it seems, to make a long (hi)story short[44], that this "success story" has continued. Ensmenger in his US perspective summarises that in "recent years 'computer people' have become some of our wealthiest citizens, most important business leaders, philanthropists, and most recognized celebrities" (ibid.: 1). But whereas a few US computer stars such as Bill Gates, Steve Jobs, and Larry Ellison have been "analysed" in a number of mostly biographical books,

> "little has yet been written about the silent majority of computer specialists, the vast armies of largely anonymous engineers, analysts, and programmers who designed and

[44] For a detailed in-depth description see Ensmenger (2010)

> constructed the complex systems that make possible our increasingly computerized society." (ibid.: 3).

It is undeniable that such a study of the silent majority of computer specialists is important to understand our computerised society today in more detail; and Ensmenger elaborates the issue very well from a United Sates perspective. However this thesis is at its heart, neither concerned with computer specialists nor with computer celebrities. Instead—as was outlined in the introduction—it is particularly concerned with the political and social significance of Image Processing Algorithms (IPAs). Of course, computer specialists are nevertheless crucial actors in the development and the implementation of IPAs, but it is rather their views and everyday actions shaping the design of IPAs than their sociodemographic constitution as a specific social group that will be of interest in the following empirical Chapters Five and Six.

Before advancing into the empirical chapters, I will concern myself in brief with a situated and fragmentary perspective on the history of computer vision in Austria. My engagement with this history encompasses the specifically Austrian, technopolitical identity (Felt 2013) that is important for the understanding of particularities of innovation (Suchman et al. 2008: 4) such as in the field of computer vision in a specific region such as Austria.

Keeping Technologies Out and Bringing them in? On Austria's and Europe's Technopolitical Identity

The history of the formation of the technopolitical identity of Austria and as such, partly of Europe that is important for the understanding of technology development, implementation and use in Austria and Europe is deeply connected to what Felt calls,

> "a very specific imaginary of technological choice, namely collectively keeping a set of technologies out of the national territory and becoming distinct by not embracing them." (Felt 2013: 3).

What Felt characterises as a specific form of sociotechnical choice instead of following the standard interpretation of

> "technological resistance as a form of technophobia which might threaten the innovation-friendly climate constructed as crucial to the development of contemporary Europe" (Felt 2013: 3)

can be demonstrated using the opposition of the Austrian population to a nuclear power plant in Austria since the 1970s and the rejection of genetically modified food/crops since the 1990s (ibid.: 4).

Felt outlines the case of nuclear power—to digress into Austria's technopolitical history—where, in the early 1970s, construction work for the first nuclear power plant in Austria was started in Zwentendorf some 40 kilometres northwest of the capital Vienna. The nuclear power plant in Zwentendorf was initially based on the consent of all political parties, but in the mid 1970s the foundation of an umbrella organisation named the "Initiative of Austrian Nuclear Opponents" challenged the construction and organised Anti-Zwentendorf demonstrations. Finally, in a referendum "on November 5, 1978, with an extremely thin majority of 50.5%, voters said no to Zwentendorf, bringing the Austrian nuclear power program to a halt." (ibid.: 10). The anti-nuclear power position gradually became part of technopolitical culture in Austria, and especially the events of the 1986 Chernobyl nuclear power plant accident cemented this Austrian position (ibid.). In 1999 there was even a "Constitutional Law for a Nuclear-Free Austria" passed (ibid.: 11). Following this representation of "Austria being free from nuclear power plants", the same happened again by keeping out genetically modified food/crops in the 1990s (ibid.: 12ff.). Following these developments, Austria was regarded by the media as some kind of "Gallic village" when it came to the deployment of certain new technologies (ibid.: 15). In reference to the popular French comic series *Asterix and Obelix*, Austria was described as a "clear white spot in Central Europe" similar to the famous, fictitious village of indomitable Gauls within the Roman Empire (*Profil* 31/03/2005 cit. in Felt 2013: 15). That means,

> "keeping these specific technologies out created the imagination of a well-delimited Austria, in its sociotechnical practices different from 'the others'. Thus, a national technopolitical identity had been created, a new self-understanding of Austria as a small nation which can manage to choose a different sociotechnical trajectory than its more powerful neighbors." (Felt 2013: 16).

The ability to choose a different sociotechnical trajectory than others certainly gives an impression of freedom and self-determination; important tools for the building and sustaining of a nation state. Sovereignty in Austrian politics is a recurring theme when it comes to the positioning of the country within the European Union. As we have seen in this regard, the specifically Austrian, 'sociotechnical imaginary' of "collectively keeping a set of technologies outside national territory" plays an important role in this positioning within the EU. In line with this 'sociotechnical imaginary' is what I wrote about the stringent, Austrian data protection law in the previous section of this chapter. Referring to §50a (7) of the data protection law[45] it is likely, but still up for negotiation that certain computer vision and pattern recognition technologies such as Facial Recognition (FRT) are excluded from Austrian national territory. This excluding of FRT stated in the Austrian data protection law refers to video surveillance applications, however FRT used in internet services such as *Picasa* or *Facebook* is not covered by the paragraph. It may be a coincidence or again part of this Austrian 'sociotechnical imaginary' of keeping out certain technologies that it was an Austrian law student that started the initiative "europe-v-facebook.org" (*Europe versus Facebook*). Amongst other things, in August 2011, the Austrian law student Max Schrems formally complained against "Facebook Ireland Ltd." about a new Face Recognition feature available on *Facebook* to the *Irish Data Protection Commisioner* (DPC). *Facebook* alongside its registration in the US and Canada, is also registered in Ireland, making the company subject to Irish and European Union data protection laws. In the complaint he stated that the feature breaches the Irish Data Protection Acts. It is "an inproportionate

[45] §50a (7) Data collected of data subjects concerned by video surveillance may not be analyzed by comparison with other picture data and not be searched using sensitive data as selection criteria. Special duty of documentation and deletion.

violation of the users right to privacy" and "proper information and an unambiguous consent of the users is missing."[46]

Without going into details, the complaint of the Austrian law student made an impact. In reference to several media reports in September 2012 *Facebook* announced that

> "… it would delete all facial recognition data it stores about its European users, going beyond recommendations made by the Irish DPC to adjust its privacy policies. Facebook said at the time it had already turned off the facial recognition feature for new users in the E.U., and said it would delete templates for existing users by Oct. 15."[47]

This means, the initiative of the Austrian law student Max Schrems to keep face recognition technologies (FRT) out of Facebook and out of his home country, Austria was adopted and pursued by the Irish DPC and the European Union and finally led to the exclusion of the Facebook FRT service in Europe. Thus, at least in this specific case, the Austrian 'sociotechnical imaginary' of keeping out certain technologies and being free of privacy infringing face recognition technology in the online world of *Facebook*, became a European 'sociotechnical imaginary' of keeping out FRT from *Facebook* generally in the EU. So, in reference to Felt (2013: 16) one can state that in the case of Facebook's privacy issues, a European technopolitical identity had been created, a new self-understanding of Europe as a union which can manage to choose a different sociotechnical trajectory than 'the others,' especially the United States[48].

Still, the ban of the Facebook face recognition service from European computers does not mean that FRT has been entirely kept out of the European Union. To use another example of FRT services: within the Google image organiser Picasa, I am currently (March 26th, 2013) able to make use of—but could also deactivate—the Picasa 3.9. face

---

[46] http://europe-v-facebook.org/EN/Complaints/complaints.html [March 26, 2013]

[47] http://www.cfoworld.com/technology/57103/facebook-deleted-all-eu-facial-recognition-data-regulators-confirm [March 26, 2013]

[48] This demarcation of the EU versus the United States by developing a separate (continental) European technopolitical identity was certainly boosted significantly in the course of the global surveillance disclosures (or United States National Security Agency (NSA) leaks) beginning in the summer of 2013 and initiated by Edward Snowden.

recognition feature in order to sort pictures by faces, automatically. This means, both the excluding of face recognition in video surveillance applications regulated in the Austrian data protection law, as well as the ban on the Facebook FRT feature from the European Union are specific manifestations of culturally and geographically situated technological choices. These choices are not about resisting FRT in general[49] but only in exactly those specific 'sociomaterial assemblages' (Suchman 2008: 150ff.) mentioned above. Meaning that, the kind of 'sociomaterial assemblages' we choose to live with, we support and finance, we develop and implement, is a matter of continual societal negotiation that differs from place to place and is subject to change over time.

What is characteristic for these negotiation processes in the computer sciences and in computer vision is what Ensmenger described as "an ongoing conflict between theory and practice and between academic computer scientists and professional business programmers" (Ensmenger 2010: 128). This conflict will be a recurring theme throughout the following empirical chapters. However, academia and the business sector are as much connected to each other as is conflict and reconciliation, or cooperation. This can also be demonstrated in the history of the *Austrian Association for Pattern Recognition* (AAPR or OAGM) that somehow represents the official history of pattern recognition, computer vision and image processing in Austria. This is because, on the one hand there is a lack of a more systematic analysis of this history, and on the other hand it is the most important institutionalised organisation in this field in Austria.

Computer Vision in Austria: The Austrian Association for Pattern Recognition

The *Austrian Association for Pattern Recognition* (AAPR), or in German, *Österreichische Arbeitsgemeinschaft für Mustererkennung* (OeAGM), is the Austrian division of the

---

[49] In general, it must be stated that Austria positions itself as very sympathetic to Information and Communication Technologies (ICT) as it demonstrated, for example in the prime aim of the *Kompetenzzentrum Internetgesellschaft* (KIG) of pushing Austria to the top of the ICT nations (see http://www.kig.gv.at/Portal.Node/kig/public/content/zielekig/52087.htm [April 22, 2014]

International Association for Pattern Recognition (IAPR). Researchers of all major pattern recognition and computer vision groups in Austria are organised within this association. Currently, the website of AAPR/OeAGM lists 14 research groups throughout the country[50]. One of these research groups is the *Institute for Digital Image Processing (DIB)* of the non-academic research institution *Joanneum Research* based in Graz. According to Kropatsch (2008) on the history of computer vision, the 'Digitale Bildauswertung Graz' (DIBAG) at *Joanneum Research*—the precursor of DIB—was the first institution in Austria of its kind, founded in 1980. The founder and first director of DIBAG was Prof. Franz Leberl who was also one of the founders and the first chairman of the *Austrian Association for Pattern Recognition (AAPR)* that was founded one year later in 1981. It was Franz Leberl who presented a retrospective of pattern recognition activities in Austria at the occasion of the 25th AAPR Meeting in 2001 The history of AAPR/OeAGM is presented in this retrospective, on its website[51]. As Leberl notes in this perspective on the history of the AAPR/OeAGM,

> "... it was the then new Institute for Digital Image Processing and Graphics (...) at Joanneum Research which provided the initial organizational backbone to OeAGM. This may have caused an orientation towards the applications."

What started with an orientation towards practical applications and thus, towards business, went on to become more academic as the story continues:

> "With the creation of the only Austrian Professorship for Pattern Recognition at Vienna Technical University, that Institute became the driving force of OeAGM. As a result, the OeAGM became more academic, and we see the effect represented in the audience of today's 25th meeting."

Characterising the current constituency in 2001, Leberl referred to "probably two or three academic institutions that 'carry' the OeAGM". Amongst these, are the *Pattern*

---

50 http://aapr.icg.tugraz.at/research.php [March 27, 2013]

51 see http://aapr.icg.tugraz.at/history.php for a more detailed description of the history of AAPR/OeAGM [March 27, 2013]

Recognition and Image Processing group at *Vienna University of Technology* and the *Institute for Computer Graphics and Vision* at *Graz University of Technology*. In addition to these,

> "the growing number of non-University research centers and research companies in nearly all Austrian provinces should produce a noticeable diversification of this support."

That means, that the OeAGM as Leberl noted "was meant to be both academic and industrial and to draw together from both arenas and backgrounds". As such, the situation in Austria was similar to the United States characterised by the continuous interaction of theory and practice and of business and academia. When following Leberl´s opinion, this interaction seemingly was one of cooperation rather than of conflict, but more historical research is needed to verify this hypothesis.

Whereas the empirical Chapters Five and Six deal predominantly with the academic sector and its interconnections to industry and business, the following empirical Chapter Four deals with business and its interconnection to the media.

Chapter Four

It's the Camera! The Blackboxing of Image Processing Algorithms and their Uncertainties

A Reconstruction of the Implementation of 'Automatic Toll Sticker Checks' (AVK) on Austrian Motorways in the News

Giving computers the ability to see is a complex sociotechnical process. As elaborated upon earlier, all attempts at this, are in fact attempts at producing, processing and understanding (digital) images algorithmically. Therefore, it makes sense to understand the process of giving computers the ability to see as the sociomaterial process in which Image Processing Algorithms are developed, produced and implemented in devices or in larger systems; advertised, used, talked about, criticised, configured, in short; materially and semiotically negotiated and formed in varying sites and in different situations. This makes clear that computer science laboritories in university or industrial settings are not the sole sites of importance when analysing the construction of Image Processing Algorithms. Making IPAs can be understood as "…a practice of configuring new alignments between the social and the material that are both localized and able to travel …" (Suchman, Trigg & Blomberg 2002: 164). They "…take their shape and meaning not in any single location but through their incorporation across diverse milieu" (ibid.). Thus, the focus on technology (invention) in the lab and on technoscientific experts (cf. Pinch 1993) was continuosuly broadened by STS scholars in the last years. For example, Nelly Oudshoorn, in her influential book *The Male Pill* (2003), analysed the testing of technology also in the media. In her understanding, journalistic and scientific texts are equally important for analysing the technology testing (ibid.: 192). In her area of interest, male contraceptive technology, "… journalists have played an active role in articulating and demarcating the identities of the potential users of this technology-in-the-making." Thus, journalists as well as scientists played an important role in the assessment of this new emerging technology. When tracing the path of a scientific report to a press release to media reports, Oudshoorn shows how the media accounts differed from the original scientific ones. While the scientific report stressed the prototype character of the technology as being far from the finished product, the press bulletin reported on a major breakthrough (ibid.: 205). The British and Dutch media accounts analysed by Oudshoorn, presented the technology in a significantly different way, namely as painful and problematic for its users. In doing so, the newspapers shaped the scientific claims, contesting them "by providing an alternative testing of the new technology" (ibid.: 206). The media articles did not question the technical, just the cultural feasibility of the technology (ibid.: 207). This was exceptional, as Oudshoorn notes, because more often it is the case that journalists shape scientific claims by uncritically replicating what scientists tell them. This often leads to a "simplified and overly optimistic picture of what has been claimed" (Fox and Swazey 1992 cit. in Oudshoorn 2003: 207). What follows from this insight is the recognition of a gap between different groups of people in their ability to know what specific technologies consist of and are able to do. This connects to what was elaborated upon in Chapter Two when referring to the power gap that arises between law enforcers and prisoners about the role of DNA evidence (cf. Prainsack 2010): whereas the law enforcement side (those with specialist insider knowledge) showed "rather nuanced understandings of what forensic DNA profiling can and cannot do", the prisoners (those without specialist insider knowledge) "tended to regard DNA profiling as infallible and true" (ibid.: 171). That means, those people that only perceived a simplified and overly optimistic picture of DNA profiling via the media were not able to develop a more nuanced understanding—and this includes a critical assessment—of what the technology of forensic DNA profiling was able to accomplish. Thus, they were put in a position of less power because of less knowledge in comparison to the law enforcement side. I have identified this power gap between different groups (e.g. those surveilling and those being surveilled) as a constitutive element in the promotion, diffusion and public understanding of identification technologies.

Public Understanding of Science and Technology and Public Understanding of Uncertainty

The 'classic' deficit model in the public understanding of science and technology demonstrates that public discomfort with science and technology is caused by a lack of specialist knowledge (Collins & Evans 2008: 283). Meaning that, employing the deficit model, it is assumed that people need to be educated in and about a specific scientific development or a specific technology in order to be able to accept it. Following Collins and Evans, in the context of IPAs, this classic understanding of the deficit model can be reframed in two different ways, the first way being that public discomfort with science and technology would be caused by the transfer of specialist knowledge. Correspondingly, public comfort with science and technology would be caused by a lack of specialist knowledge. Therefore, it could be an appropriate strategy in science and technology promotion that people should not to be enlightened on a specific scientific development or specific technology in order to be able to accept it. This would not mean that there is no education or knowledge transfer at all, but that there is a different way of imparting this specialist knowledge. It is well-known that in the case of identification technologies, pattern recognition and forensic science, this kind of knowledge among the general population is principally derived from TV series such as CSI. In CSI, the dominant theme is that fingerprints and DNA traces reveal the absolute truth and deliver evidence that speaks for itself (Kruse 2010b: 80f.): what Michael called "technoscientific police procedural" (Michael 2006:90). This is in opposition to the need for DNA matches—as is the case for a wide array of other pattern recognition applications such as IPAs operating on such a level—for interpretation and intervention by humans. It is embedded in a complex chain of inference (Collins & Evans 2012: 906). In nonfictional forensic science, producing evidence is more complicated than presented in the CSI version (ibid.: 86). Making DNA matches depends highly on the skillful and informed interpretation of images (Halfon 1998: 805ff.). In nonfictional DNA practice, absolute certainty is unattainable and probabilities must always be referred to. In CSI, matches are synonymous with knowing for certain, and thus, for absolute truth (Kruse 2010b: 86). A dominant message arising from CSI is "that it is easy, quick, routine and epistemologically very strong" (Ley, Jankowski & Brewer 2010: 13). This view leads to an antisocial representation of science in the public that underpins the so called "CSI-effect" (Collins & Evans 2012: 906).

This brings me to the second way in which the 'classic' deficit model can be applied to this case. In this second approach, the deficit model shows that a lack of specialist knowledge in the general public, representing science and technology as antisocial, causes on the one hand, overestimation and unrealistic expectations of what science and technology are able to do, as the example of CSI shows, and on the other hand leads to public indifference, as science and technology are experienced as antisocial and are thus, processes that are out of reach to normal citizens. In the context of *Hawk-Eye* and similar visual decision aid technologies that are increasingly part of televised sports coverage such as tennis, Collins and Evans claim they are "making their capacities and technological limits more clearly visible" (Collins & Evans 2008: 284), in order to "promote a more nuanced and widespread understanding of the statistics of uncertainty" (ibid.) to the public. In this regard, Science and Technology Studies could become a public engagement activity that is science and technology education with public participation. In any case, citizens would "benefit from the opportunity to experience science as a social practice rather than a set of facts to be learnt" (Collins & Evans 2012: 905). As a consequence, science and technology might become more open and democratic instead of being perceived as a highly sophisticated, isolated activity performed by a handful of people in powerful positions.

The news media clearly play a role in informing the way people understand science and technology. A study of the relationship between the media coverage of science and public understanding of it, showed that "most people are aware of the main themes or frameworks of media coverage of science related stories". (Hargreaves, Lewis & Spears 2003: 52). People are not only aware of, but can also be regarded as active participants in the process of interpreting science-related stories from their own points of view (Irwin & Wynne 1996: 139). What people do with media accounts of science and how they make sense of theses accounts is not solely up to media articulation, but rather to their own experiences with science and the media. As a consequence, negotiation processess of science (and technology) do not only take place in the media, but also in the engagement of readers (Felder 2010: 32) that manifest in quite different ways. Nevertheless, media reporting is in some cases the only entry point for citizens to gain information about and engage in science and technology. Clearly, in these cases the media significantly shape the meaning of science and technology and how they are imagined. As such, they not only transport how people perceive and understand science and technology, but are also a way to domesticate (cf. Silverstone & Hirsch 1992) and normalise specific scientific and technological projects. That is why a focus on media articles and other public documents about science and technology is important. These are places where scientific and technological developments are critically tested and contested or conversely, where they are noncritically not tested at all. Thus, media articles and public documents pave the way for the acceptance or not of technologies by a wider public.

Culture of Secrecy

The reports in the media on science and technology might also be the only entry point for social scientists for studying their area of interest in technology, because it is not possible to get access to technology by other means. Torin Monahan in his book *Surveillance in the Time of Insecurity* (2010) reported on his efforts to study *Intelligent Transportation Systems* (ITS) in the United States and noted: "It is more difficult than one might expect to obtain access to ITS control centers", and "the obstacles to learning about ITS were inordinately high." (ibid.: 103). He reported on declined and unanswered phone calls and e-mail. In some cases an initial contact proved impossible to follow up. Monahan's attempts to explain this behaviour was partly seen in government employees having insufficient time, or was due to the general "firewall culture" and their trying to avoid unnecessary scrutiny. As Monahan argues, the main reason for this "Culture of Secrecy" is down to the fact "that ITS operators knew that their centers had the look of surveillance and that they wanted to distance themselves from that characterisation of their work" (ibid.). In the context of video surveillance, Kammerer in a similar way considered this lack of knowledge, misinformation and superficial knowledge about the realistic potential of it, as strategically functional, producing consent and public comfort (Kammerer 2008: 83). In the course of these processes, as Kammerer notes, information politics of security managers are contradictory. On the one hand, enormous media campaigns are supposed to gain the consent of the population. On the other hand, they are designed to deter. Of course, the principle of 'security by obscurity' should not be underestimated as details could fall into the wrong hands (ibid.). This, however, is no reason to seal oneself off completely from any form of communication and information about the respective technology in use. Nevertheless, this constellation makes the scientific analysis of technological systems highly difficult, if not to say impossible, in some cases. From my point of view this issue of what Monahan named 'the culture of secrecy' and its far-reaching implications should be on the future research agenda as a topic in itself, as it happens too often that social scientists are confronted with this attitude and hindered in their work. In what follows I describe my experiences with the culture of secrecy in Austria.

The Case Study: The implementation of 'Automatic toll sticker checks' (AVK) on Austrian motorways

During my field work I realised that in contrast to my expectations, in Austria most of the relevant image processing projects in the area of pattern recognition, especially those of Behaviour Pattern Analysis are at the best still at the prototype or field test stage. I asked myself, if there actually are projects that are already in operation as it would make sense to analyse and learn from these projects in operational conditions. As it is a characteristic of ethnographic field work it was by default rather than by design, but also as an outcome of my deeper understanding of computer vision and image processing over time that I came across a nationwide system in operation that contains at its heart, Image Processing Algorithms that are designed to recognise patterns: the so-called 'Automatic Toll Sticker Checks' ("Automatische Vignettenkontrolle", in short: AVK) on Austrian motorways and expressways. My first experience of AVK was in a newspaper article and I started to find out more about it, searching and reading different articles in the press and publicly available documents. As my questions about AVK grew the more press accounts I read, I quickly realised that it would be necessary to talk to ASFINAG[52], the operator of AVK. In March 2012, I wrote an email with a request for a scientific interview about AVK to a press spokesman who I had seen mentioned regularly in newspaper articles on AVK. I explained that following the information about AVK in the press, I would like to gain a comprehensive, objective picture from the operator ASFINAG and outlined that I was interested in both the 'history' of the implementation of AVK and in the technical and practical mode of operation. The email

[52] ASFINAG plans, finances, maintains and levies tolls on the entire Austrian motorway and expressway network covering 2,175 kilometres. ASFINAG was established in 1982 and is wholly owned by the Austrian Federal Government. A contract signed in 1997 between the Federal Government and ASFINAG gave the company additional powers and responsibilities: By virtue of this contract, ASFINAG holds usufruct rights related to land and facilities belonging to the primary federal road network and owned by the Federal Government and has the right to collect tolls and/or charges from those who use such land and facilities. As a user-funded company, ASFINAG has committed itself to utmost efficiency in managing its financial resources. ASFINAG does not receive any money from the federal budget (Source: http://www.asfinag.at/about-us).

remained unanswered until two weeks later I wrote the same email to a press spokeswoman, also mentioned in many of the articles. This time I got a quick reply with the information that my request had been forwarded to the executive office responsible. As I did not get a reply for two and a half weeks I sent a short reminder and again got a quick, friendly reply stating that on behalf of the executive office I would receive the most important data and facts about AVK in the following days, but that an interview would not be possible. One month later, in May 2012, I received a polite email including the following information about AVK[53]:

> The first AVK device went into operation in December 2007 and **at the moment there are five devices in operation**
>
> **AVK – for traffic safety and as a supplementary measure to manual control:** AVK is a digital camera system and can be viewed as an additional or supplementary monitoring procedure to the manual toll sticker checks by the toll monitoring and enforcement unit. In places where pulling-over is not possible due to traffic safety, AVK is in operation (e.g. on multi-laned motorways, urban areas or motorways without a hard shoulder). Thus, AVK serves to ensure the safety of both our customers and employees. AVK is not a substitute for manual checks by the toll monitoring and enforcement unit. AVK checks take place randomly with frequent (weekly) changes in location.
>
> **AVK – Boosting toll sticker morale:** Toll sticker morale is actually quite high at about **98%**. AVK is designed to further boost toll sticker morale, especially in the interest of our accountability to the law: **Fair treatment of all motorway users, so that the case does not arise that some users of the network without a toll sticker, do so at the cost of others.**
>
> **Data protection**: The technology in operation is in constantly changing locations on the whole nationwide motorway and expressway network. Only those vehicles that verifiably offended against the obligation to pay a toll in Austria are registered and fined. A general image including the number plate and also a detail image of the

[53] Translation by author. The original text in German can be found in the appendix. Text marked bold as in the original email text.

> windscreen are recorded. The angle of the camera is set so that the faces of driver and co-driver are not recognisable. The data won is also checked manually for monitoring purposes. Our rule is to give the customer the benefit of the doubt! AVK is based on the regulations of the data protection law and was duly declared and registered in the 'Datenverarbeitungsregister' (the central registration office for data processing applications in private companies).

After a short time of disappointment about the information on AVK I had received being reminiscent of a press bulletin, I replied to the email and thanked the press spokeswoman for sending me the information. But I also mentioned that I had known most of the facts from articles in the media. I asked again about the possibility of an interview and also attached a PDF to the email with my questions about AVK in order to give a better impression of what exactly I was interested in. I requested a written answer to my questions should an interview still not be possible. To date, the email has remained unanswered.

Four months later, in October 2012, I was able to visit the *19th ITS World Congress* in Vienna. There, ASFINAG was present with an exhibition stand where I tried to talk to representatives, but they did not tell me anything new and referred me to the relevant people, for example press officers. Nearby, there was also an exhibition stand belonging to the company EFKON AG, the producer of the newest AVK devices that were also displayed prominently on the stand. I was able to talk to a representative who was able to bring me in contact with a product manager. After an introductory talk I asked for an interview about their AVK device. To my surprise, I was immediately encouraged to write an email to request an interview. One month later I was able to interview the same product manager about AVK. One week before this interview with the EFKON product manager took place, I tried to ask ASFINAG again about an interview with them. Meanwhile, a colleague of mine from another research institution in Vienna sent me the details of a contact close to the *ASFINAG Maut Service GmbH* executive office. In the email to this contact that was similar to my first request, I also mentioned the interview with EFKON as I thought this would increase the possibility of a positive answer. This was not the case. To date the email has remained unanswered but it did not go unnoticed. Some days after the EFKON interview, the product manager I had interviewed got in touch with me as the interview had been noticed by someone outside EFKON. Due to this, I was politely asked not to cite any customer-related information that had not been published elsewhere until then. From my point of view and in my interpretation this means: ASFINAG knew about my EFKON interview; highly probable after my email request that remained unanswered, and wanted to make sure that no ASFINAG-related information about AVK would get into my possession through EFKON. Obviously, ASFINAG was taking a great interest in absolute secrecy regarding AVK. As I was never informed of the reasons an interview or the answering of questions about AVK was not possible, it has remained unclear to me what it is exactly that has to be kept secret in this way from a social scientist and the wider public.

These developments made the case even more interesting for me and I decided not to delegate it to the realm of unpublished academic work. ASFINAG's information and secrecy policies should be seen as a result in themselves, confirming Kammerer's insights experienced in Germany (cf. Kammerer 2008) and even going beyond what Monahan (2010) reported in the US. Under these conditions I decided to focus on the detailed analysis of all publicly available documents about AVK, concentrating especially on newspaper reports. Therefore, the aim of the case study went in the direction of understanding and reconstructing the incremental introduction and implementation of AVK on the basis of publicly available documents and to analyse how it is described, framed, imagined and tested or non-tested in the Austrian news. I am particularly interested in what stories are told in the news about the history, the relevant actors involved, the mode of operation, the capabilities and limitations, and the implications of AVK.

The case study is also designed to broaden my perspective on this research topic, because it analyses Image Processing Algorithms that are already in operation using publicly available newspaper articles and documents. That means they already make a difference and have an impact on the daily lives of people driving on Austrian motorways. AVK can be seen as a pioneering system in this regard. Additionally a lot can be learned about the potential of algorithmic identification technology, which "is already being integrated in to Automated Social Technical Systems for enforcement purposes" (Lyon 2003: 274). In principle there is not much difference to the Pattern Recognition systems aimed at humans (e.g. fall detection, event detection, facial expression recognition, face recognition and so on), because in all of these cases, visual patterns are analysed and compared to templates. It is no wonder that the first systems in operation are aimed at cars, and in this case at toll stickers, because it is easier to automatically detect and recognise standardised patterns attached to cars in comparison to detecting ambiguous human behaviour patterns. Amongst other things this is because cars and their individual elements are easier to distinguish from their environment and are usually driven in highly standardised settings (e.g. on clearly marked lanes on motorways).

Certainly, the heading of this chapter anticipates some of the results presented in what follows. It is the camera that is the focus of attention in the newspaper articles. The camera is positioned as the central actor in AVK. It is the camera that recognises the presence and validity of toll stickers, whereas Image Processing Algorithms are widely neglected and blackboxed in comparison to the 'automatic' and 'innovative' camera. Another central theme in many of the newspaper articles is the presentation of AVK as a ready-made and autonomous camera system. This is in contrast to the necessity of double-checking the AVK results manually in an enforcement centre by human operators. Error rates, probabilities, uncertainties, false positive or false negative cases are not made a subject of the discussion. Instead, AVK is mainly evaluated and tested by its economic success. In this context a recurring theme is the presentation of exclusively provided detection numbers and sales figures.

The media articles to be analysed were researched with the help of the online databank *WISO* search[54], in which 118 German language newspapers and magazines and a total of 115 million articles are listed (Dec 2012). Amongst these are the most important Austrian newspapers such as *Kronen Zeitung*, *Kurier*, *Der Standard*, *Die Presse*, *Salzburger Nachrichten*, *Oberösterreichische Nachrichten*, *Tiroler Tageszeitung*, *Wiener Zeitung* or *Kleine Zeitung*. The free, daily newspapers *Heute* and *Österreich* are not included in the database. Additionally, there are also some of the most important Austrian magazines

[54] http://www.wiso-net.de

in this database such as *Profil*, *News*, *Format* or *Falter*. In addition to WISO, I also made use of *Google* and *APA-OTS*[55], the original text service of the *Austrian Press Agency*. This made it possible to also include in the analysis, online articles of the Austrian Broadcasting Corporation (*ORF*), press releases and other publicly available documents (especially parliamentary questions and answers).

I searched the WISO databank, Google and APA-OTS with the search terms "AVK", "Automatische Vignettenkontrolle" and "Automatische Vignettenkontrollen" (only in German, literal translation: automatic toll sticker checks) The time span within which the articles and documents were found, ranged from May 2007 until October 2012.

The core sample consists of:

- 13 lead stories
- Four shorter lead stories
- 26 brief notes, or side notes within other stories
- Two press releases
- Two Parliamentary Questions (Interpellation) and Answers

Background material:

- ASFINAG website http://www.asfinag.at
- ASFINAG Geschäftsberichte (ASFINAG Business Reports)
- Mautordnung (Toll Regulations) beginning with Vol.1 (1st Sept. 2003)

Following the collection of data I started to scrutinise it using open coding (Gobo 2008: 227) in order to examine the media articles for themed and temporal phase clusters. The outcome was ten different themed, temporal phases, starting with a pre-implementation period and ending with AVK performance reports. In a next step, I developed a framework for analysing the media articles following the open coding and my questions on the material obtained. This framework consisted of seven levels: labelling of AVK, relevant social groups, temporality, mode of operation, problem definition, evidence, and the wider frame. In a next step I started to describe the ten identified themed and temporal phases within the analysis framework. In what follows,

[55] http://www.ots.at

an expansive and detailed description of the incremental introduction and implementation of AVK is presented to facilitate a comprehensive understanding of how AVK has been portrayed, framed, imagined and tested or not, in the Austrian news and in publicly available documents over time. The description of the phases is followed by an analytical conclusion discussing two central insights.

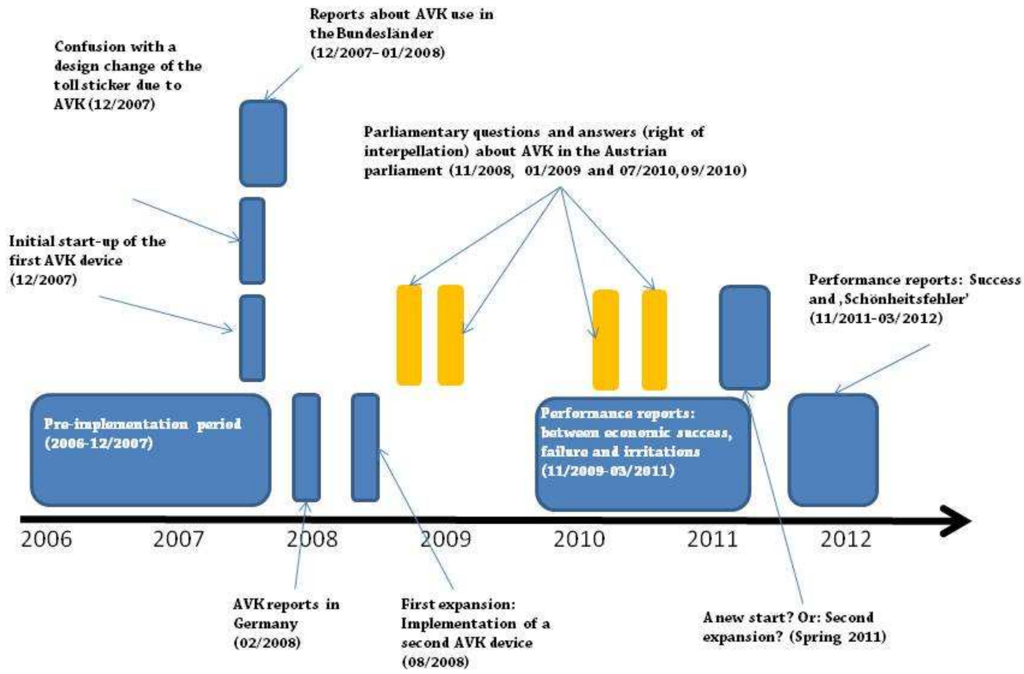


Figure 2: Overview of the Implementation Period of AVK in Austria

Pre-implementation period (2006 - 12/2007)

The first time that the Austrian public was confronted with AVK was about seven months before the first AVK device was introduced on Dec. 12th, 2007. In reference to a radio feature in the *Ö1 Morgenjournal* programme, an online article (a01) was published on May 2nd, 2007 (oesterreich.ORF.at) about the AVK system test mode. The article referred to AVK as an 'automatic check' ("Automatische Kontrolle") and 'camera system' ("Kameraanlage"). The time-frame for the introduction of AVK was presented in the first paragraph as impending ("künftig"). Later, the text stated that the test mode would start mid year; that is mid 2007. It further explained that there was only a test mode of the system planned. This first test mode with a different version or system had not shown satisfying results so far, it was argued.

Regarding the mode of operation, the article described it the following way: cameras capture the windscreen of a vehicle including an enlargement of the toll sticker (if the toll sticker exists). The only data saved derives from cars that raise suspicion of not having a toll sticker on the windscreen. It is described as a mobile system, which can be installed in different places. It is however, planned for implementation especially on heavily trafficked motorways in the greater Vienna area. The plan is to operate the system on 80 days a year. In the article, the problem of very few regular checks carried out on the heavily trafficked motorways of the greater Vienna area was described. Here, in the future, automatic checks are planned to replace the ASFINAG toll monitoring and enforcement unit.

Next to ASFINAG, other relevant social groups mentioned in the article were 'toll sticker offenders' ("Vignettensünder"), 'car drivers' ("Autofahrer") and 'toll sticker offenders from abroad' ("Vignettensünder aus dem Ausland"). Toll sticker offenders not showing a valid toll sticker on the windscreen had risen by 20 percent in the previous year (2006). 80 percent of the offenders is from abroad. It has to be noted here that it is not clear from the article if the 80 percent relate to the total number of toll sticker offenders, or to the 20 percent increase that took place in 2006.

At that point in time the *Ö1 Morgenjournal* radio feature and the online article about the programme were the only media reports. It took almost four months before AVK was mentioned as a side note in an article (a02) about intelligent roads and how cameras contribute to these (*Der Standard*, Aug. 29th 2007, p.9). Nevertheless, this side note included AVK within the greater topic of a vision of intelligent roads. The article presented a range of technologies on this subject. The tenor of the article was that, with the help of cameras and video images, traffic density, traffic jams, speed and distances can be measured automatically. In contrast to the first media report on AVK, the side note in this article stated that the automatic toll sticker monitoring system had been tested on the A23 (that is an urban motorway in the south-east of the city of Vienna often referred to as "Südosttangente") for one year at that time. It was not obvious from the article if the same system referred to in (a01) is meant. In this article, the new system was presented as a method of identifying 'toll evaders' ("Mautpreller"). Alongside this relevant social group, (just another name for 'toll sticker offenders') (cf. a01), and ASFINAG, another is introduced in the side note: ARBÖ[56]. ARBÖ is cited as having no objections to the new AVK system as long as all data protection requirements are watertight ("hieb- und stichfest").

A very brief note (a03) in the newspaper *Kleine Zeitung* (Nov. 13th 2007) added to the information, reporting that AVK consists of two camera systems that capture an image of the overall car and an image of the windscreen and that also wrongly pasted toll stickers can be recognised. Just before the implementation of the first AVK device on Dec. 12th 2007, the newspaper *Tiroler Tageszeitung* reported on Dec. 11th 2007 (a04) that 'this week', ASFINAG is going to present an automatic toll sticker check ("eine automatische Vignettenkontrolle").

Here we can see that there was no controversy about AVK in the Austrian press in this pre-implementation period, even though AVK was in operation in the context of an expansion of (toll sticker) checks. This means that in the articles there was no relevant social group identified with objections against it. The only conceivable problem in the context of AVK was maybe seen to be data protection, but as ARBÖ's statement in the press release shows, everything is all right as long as data protection is guaranteed.

Initial start-up of the first AVK device (12/2007)

The first AVK device was implemented in Austria on Dec. 12th 2007. On this and the following day there was wide media coverage of AVK in the Austrian news. Altogether eight different media organisations reported on the implementation of AVK. I was able to identify four main articles (two in the same newspaper), two shorter main articles,

[56]ARBÖ is an Austrian association of car drivers

and three short notes. As the shorter main articles and short notes only repeated aspects reported in the main articles, the analysis is particularly, but not only focused on these four main articles.

AVK was mainly referred to as an 'automatic toll sticker checking system' ("Automatische Vignettenkontrolle") or AVK, but there were also other terms or names used: 'intelligent system' ("das intelligente System," cf. a05, a07), 'electronic eye' ("elektronisches Auge," cf. a07, a09), 'all-automatic toll sticker checking facility' ("vollautomatische Vignettenkontrollanlage.", cf. a06), 'modern cameras' ("moderne Kameras,", cf. a06), and 'electronic assistant' ("elektronischer Helfer," cf. a10).

The commencement date for AVK was announced as Dec. 12th 2007 in all articles. Most of them state that primarily there was one device installed and in the following two years, ten more devices were to follow (a05, a07, a09, a10, a11, a12). Two articles noted that the device was to be in operation for three days (a06, a07). Almost all articles named the A23 ("Südosttangente") as the location of the device, and many articles stated that basically there were plans to operate the device at 15-20 different sites in the whole of Austria on 80 days a year. These sites were to be in predominantly urban areas. In some articles the exact site was indicated as being outside the *Asperntunnel* - a tunnel on the A23 (a05, a10, a11, a12).

The introduction of AVK on urban motorways was, as argued in most of the articles, a consequence of the emerging problem that hardly any checks had been carried out in these areas, particularly in the greater Vienna area (e.g on the A23) as there were no opportunites there due to the nature of the highway, for stopping and checking cars. AVK was presented primarily as a solution to this problem. Apart from this, AVK was seen as useful for boosting toll sticker morale ("Vignettenmoral") in general. The articles argue in reference to ASFINAG that toll sticker morale was already high at around 98 percent, but the aim was to reach 100 percent. Other reasons referred to in the articles of why AVK was introduced was on the one hand, support for the ASFINAG[57] toll

[57] ASFINAG toll monitoring and enforcement personnel are referred to as 'toll sheriffs' ("Mautsheriffs") in one article.

monitoring and enforcement unit and on the other hand, due to its cost-efficiency ("kostengünstig", cf. a05). This brings me to the wider frame of which AVK is a part, as reported in the articles following the initial start-up. In this same article (a05) AVK is presented as a measure taken to promote the economic efficiency of ASFINAG. Meaning that AVK is supposed to generate additional revenue towards the financing and maintenance of the motorway network in Austria. In figures, the expenses of the AVK device were specified as 230 000 Euros, while the expected revenue in fines was specified as 10 000-20 000 Euros per day. This means that the AVK device would pay for itself within a year, the article reported. In another article (a06) the expected revenue in fines was declared as 10 000-20 000 Euros within a year.

Another frame was established in the headline of one article (a05). According to it, toll sticker offenders did not stand a chance any more with these 'first automatic toll sticker checks.' ("Die erste automatische Vignetten-Kontrolle: Pickerlsünder sind nun komplett chancenlos"). AVK was presented in this article as the ultimate means of combating toll sticker offenders.

Particularly in the shorter articles and notes, a third frame presented was automatisation (e.g. a08). Here it was argued that toll sticker offenders would no longer only be detected by the human eye, but also by the electronic eye of AVK (a10, a12, a13). This metaphor was used where there was only limited space allocated and was obviously a means of communicating what AVK is, in shorter articles and notes.

In two articles (a06, a07) the mode of operation of AVK was said to be similar to that of speed cameras ("Radarkasten"), the main difference being that the AVK device was installed on overhead gantries ("Überkopfbügel") above the motorway. In (a06), the head executive of ASFINAG, Klaus Schierhackl, explained the mode of operation as follows: "The modern cameras detect when a car is without a toll sticker or if the validity of the affixed toll sticker has expired. These cases are passed on, double-checked and the culprits fined."[58] It was further stated that the cameras take pictures of all cars. These pictures capture the windscreen and number plate. According to Schierhackl the test run was satisfactory: "We now have very good images, on which the toll stickers and car number plates are clearly recognisable."[59]. From this statement and how it was presented in the article, it was not clear for the reader, if the term "clearly recognisable" referred to the human eye or to the electronic eye. Nevertheless, a favourable impression of the system was given in this account.

With regard to the relevant social group of foreign car drivers, the article stated that these were checked randomly ("punktuell"). In such cases the toll sticker inspector read the respective data in realtime and informed their colleagues, waiting 5 to 10 km further along the motorway of the toll sticker offender. They could then stop them and enforce a penalty. A particularly relevant social group within the foreign car drivers were German —or as named in other articles, 'German tourists' or 'German toll offenders'. In this article, as well as in two others (a05, a10) they reported that due to an international agreement a fine could also be sent to German toll sticker offenders.

This same article also explained that saved images were protected by a forgery-proof code. Additionally, it stated, that metadata (car registration number, date/time and affixing of the toll sticker) was to be saved for three years. In (a05) they further stated that this was due to the legal time limit for appeal.

Article (a05) went slightly more into detail. In this description of the mode of operation, the project leader and the executive director of ASFINAG Maut Service GmbH are cited as saying that the device was equipped with two cameras and thus, two pictures were taken: an overview image that included the number plate as well as a detailed image of the windscreen. The second step in the process was for the system to search for the toll sticker and ascertain its validity, if affixed. In consequence, only the images of

[58] "Die modernen Kameras erkennen, wenn ein Auto keine Vignette hat bzw. die geklebte Vignette nicht mehr gültig ist. Diese Fälle werden dann weitergeleitet, gegengeprüft und zu einer entsprechenden Anzeige gebracht."

[59] "Inzwischen haben wir sehr gute Bilder, auf denen sowohl die Vignette als auch die Autokennzeichen gut erkennbar sind."

suspicious cars were saved; all other cars with a valid toll sticker were immediately deleted by the intelligent system. The suspicious data was transmitted to a PC, on which the data was checked manually with special software. In doubtful cases (“im Zweifelsfall”), meaning in cases in which the image was not completely in focus, the executive director is cited as saying that ASFINAG made a decision in favour of the client; meaning in favour of the driver. In this wording the relevant social group of car drivers was described as being clients of ASFINAG. The post-processing was described in more detail in (a09). Additionally to the points made in (a05), it was stated that before ASFINAG could request the payment of a substitute for toll evaded, the images were double-checked by two employees. In this process, ambiguous cases were rejected; for example in the case where a windscreen wiper covered a part of the toll sticker. Again, it was argued that in such doubtful cases the decision was in favour of the client.

To sum up, the newspaper articles on the initial start-up of AVK showed different understandings of what AVK is and how it works: On the one hand there were reports describing it as an ‘all-automatic’, autonomous system. On the other hand, there were detailed descriptions of a division of responsibility between the two cameras integrated in an AVK device and of human operators double-checking the data (in particular digital images) produced and analysed by the camera. However, also in these more detailed descriptions no precise observations about the mode of operation could be found, which kind of technology (e.g. image processing, pattern recognition etc.) was used, or where the technology came from (e.g. distributor), not to mention capacities and limitations, e.g. possible error rates or restrictions. All in all, AVK was represented as a ready-made and unproblematic system that was able to increase toll sticker checks, boost toll sticker morale and generate additional revenue.

Confusion with a design change of the toll sticker due to AVK (12/2007)

Only one week after the initial start-up of the first AVK device, on Dec. 20th 2007, and on the following day, four Austrian newspapers printed short reports (a15 to a18) about confusion with the design of the new annual toll sticker for 2008. The information was provided by an *ARBÖ* press release (a14) on Dec. 19th 2007. The annual toll sticker entitles motor vehicles to use toll roads during the calendar year as indicated on the toll sticker. It is valid from December 1st of the preceding year until January 31st of the following year[60]. This meant in this case that the annual toll sticker 2008 was valid from December 1st 2007 up until, and including January 31st 2009. The confusion reported was due to two different versions of the toll sticker as can be seen on the images below:

Figure 3: **Toll sticker design change[61]**

As reported in all four short notices, the version on the left showed two white stripes and in between, one red stripe behind 'B 08' designating the year of issue. In the second version on the right, all three stripes behind the 'B 08' were white. As reported in the short notices, both versions were valid, according to *ARBÖ*.

The *ARBÖ* press release reported that the two design versions proved necessary because of the newly introduced automatic toll sticker checks. It was further argued that due to the additional "lane" (the white stripe in the middle on the second, new version of the toll sticker) the imprint '08' could be recognised and read more easily by the automatic

[60] See http://www.asfinag.at/toll-stickers-and-rates

[61] Source: http://www.auto-motor.at/Auto-Service/Mautstrecken/Autobahn-Vignette-2008-Aussehen.html

camera[62]. While (a15) only reported on the necessity of the two versions because of AVK, (a18) additionally informed that the three white stripes (note: no mention of the imprint ‘08’) could be recognised by the camera better. (a16) and (a17) both reported that the design of the toll sticker was changed during the production process in order to ensure better legibility for the newly introduced automatic toll sticker checks. (a16) added that due to this the number ‘08’ was more recognisable for the automatic camera.

What is interesting about the press release and the short notes is that all referred to the (automatic) camera that recognised or read the toll sticker. Therefore, the camera was positioned as the acting agent of the AVK system (and not a network including the camera, computer hardware and Image Processing Algorithms) that is able to recognise and to read what is on the toll sticker. In this regard, in contrast to (a14, a16 and a17), where the three white stripes and the number ‘08’ were said to be more recognisable or legible, (a18) reported that (only)the three white stripes were better recognised by the camera.

Finally, there is one very small, but crucial difference between what was reported in the *ARBÖ* press release and in three of the four short notes. In the press release (a14) it is stated that the imprint ‘08’ can be recognised and read “even better“, but in (a16, a17, and a18), the word “even” in addition to better, is missing. That means, while the press release does confirm that AVK was already able to recognise and read the toll sticker well, before the design change, the short newspaper notes only stated that following the design change the AVK camera recognised and read it better than before. In my interpretation, the term “even better” used in the *ARBÖ* press release signals its implicit approval of the AVK system as the design change is presented as improving an already efficient system. These reported insights also proposed problems arising from the “necessity” statements made in the press release and in all short notes. In the end it is not clear to the attentive reader, if the design change was a necessity, or if it was only a way of improving the system. It has to be noted here that the comparative of the word

[62](a14): “Durch die zusätzliche Fahrbahn kann der Aufdruck "08" noch besser von der automatischen Kamera erkannt und gelesen werden.“

good - that is, “better” - used in the press release and in three of the short notes, pointed to the probabilistic nature of the system. Contrary to the media reports of the initial start-up period that presented AVK as a properly functioning black box, the use of the word “better” is an indicator that there were also cases in which AVK did not function as well as in other cases. Here, for example, it seemed to be the case that toll stickers with the new design (three white stripes behind ‘08’) are more recognisable or readable than the ones with the old design (white, red, and white stripe behind ‘08’).

Reports about AVK use in the Bundesländer (12/2007 – 01/2008)

In the initial start-up period the A23 (“Südosttangente”) in Vienna was named as the site of operation. However, many articles also stated that basically there were plans to operate the device at 15-20 different sites in the whole of Austria. These sites of operation were supposed to be predominantly in and around urban centres. On Dec 25th 2007 the Styrian and Carinthian regional newspaper, *Kleine Zeitung,* published an interview (a19) with Alois Schedl and Klaus Schierhackl of the then new ASFINAG management board. In the interview, amongst other things, they were also asked about the new automatic toll sticker checking system and if there were plans for “total surveillance” (“totale Überwachung”). Schierhackl explained that within the following two years there were to be a maximum of ten AVK devices deployed. He continued, saying that full coverage made no sense (“flächendeckend”). At the end of the interview he answered the question of whether AVK was also to be implemented in the Austrian Federal States of Carinthia and Styria with a clear yes, and added that ASFINAG would report the implementation there when the time came[63].

In January 2008, two different regional papers reported on the implementation of AVK in their specific regions. *Tiroler Tageszeitung* (a20) also reported on the two design versions of the 2008 toll sticker which was due to the introduction of new surveillance technology (cf. a20), saying that AVK surveillance cameras were more able to pinpoint and identify this new design. In this context, the article also reported on future checks

[63](a19) SCHIERHACKL: „Ja. Wenn es so weit ist, sagen wir es.“

of toll stickers that were to be executed in a fully automated manner. Regarding the mode of operation the article explained that the cameras took two images (windscreen and number plate) of all passing vehicles. If the toll sticker was not recognised correctly by the system, the respective vehicle had to be checked manually. As a consequence, they reported penalty notifications could also be sent to toll sticker offenders without stopping them on the spot. Interestingly, the *Tiroler Tageszeitung* article is the first article that critically discussed issues of privacy and data protection. It brought up for discussion the statement that a comprehensive and temporally continuous checking of vehicles, coupled with the retention of this data created problems with data protection regulations. They added that according to the data protection law everybody has a right to secrecy regarding his or her personal data. Even so, the article did not go into detail (e.g. what this really meant etc.) and also did not elaborate on the issue. It is interesting to note here that in the course of events, not a single newspaper article followed up this faint call for critical discussion of data protection issues concerning AVK. All in all, my sample showed that following the introduction of the first AVK device on Dec 12th 2007, no public dialogue or controversy about AVK in the Austrian press took place. The side note in the *Tiroler Tageszeitung* remained the only case in which the possibility of data protection problems was brought into the discussion.

In the week following the publication of (a20), the regional paper *Vorarlberger Nachrichten* reported on the installation of a device for toll sticker checks in Tyrol and the westernmost Federal State of Austria, Vorarlberg (a21). The article titled "Toll offenders don´t stand a chance"[64] stated that there was a test device installed on the motorway at Hall in Tirol and that it was going to be installed as a trial on the Walgau and Rheintalautobahn (N.B.: A14) in Vorarlberg. This article, published on January 18th 2008, was the last article on this issue in my sample to be found in Austrian newspapers until November 2009[65]. That means that in this period of one year and ten months there was not any newspaper article about AVK published in Austria.

[64](a21): „Keine Chance für Mautsünder"

[65]There was one appearance of AVK within this period, in May 2009, but it was only mentioned as a very brief side note in a short *Tiroler Tageszeitung* article (a25).

AVK reports in Germany (02/2008)

In February 2008, three German newspapers reported on the implementation of AVK in their neighbouring country of Austria. While it was mentioned only as a short note in the *Leipziger Volkszeitung* and in the Cologne newspaper, *Express* (a22, a23), the Munich-based newspaper *Abendzeitung* reported on the Austrian AVK system in a longer main article (a24). This article with the title 'High-tech hunt for German toll evaders' („Hightech-Jagd auf die deutschen Maut-Muffel") introduced readers to the subject with citations from users of the Austrian Broadcasting Corporation internet forum on AVK. There, words like 'robber barons' ("Raubritter"), 'modern highway robbery' ("moderne Wegelagerei") or 'horror scenario' ("ein Grauen") were used when AVK was being written about. In contrast to these user comments, the ASFINAG press spokesman, Marc Zimmermann was quoted as saying that the automatic checking system is not a 'rip-off' ("keine Abzocke"), but a way of boosting toll sticker morale, especially that of German car drivers which was said to be particuarly poor ("besonders mies"). In numbers; out of annually 85 000 toll offenders, one third came from Germany, for example tourists or those living close to the Austrian border.

Regarding the mode of operation, the system was described as consisting of a high-tech camera, later named a 'toll sticker blitz' („Vignetten-Blitzer") installed above the motorway on bridges or steel gantries taking one picture each of windscreen and number plate. As to the characteristics of the system, a speaker for the Austrian association of car drivers *ÖAMTC*, was cited as saying, "the thing is pretty clever"[66] as for example, it also did well in recognising the mini-perforations in the short-term toll stickers which signalise their validity. Here it should be noted that this is the toll sticker of choice for German car drivers who only travel on Austrian motorways occasionally. The article explained that once the camera had detected an offender there were two possibilities: either an immediate check and penalisation, or the authorities sent fine notification to Germany, where it is also enforceable by law. However – and here I follow the general drift of the article - German authorities often ask for a clearly identifiable

[66](a24): „Das Ding kann ganz schön viel."

image of the driver. If such an image was not included, it might be that German toll offenders could escape with a good lawyer.

First expansion: Implementation of a second AVK device (08/2008)

As mentioned earlier, the *Vorarlberger Nachrichten* article published on January 18th 2008 (a21) was the last article about AVK in my sample of Austrian newspapers until November 2009. This is particularly interesting, because in August 2008, a second AVK device was installed on Austrian motorways. Retrospectively, one can find information about the use of two devices in the publicly available answer of the *Austrian Federal Ministry for Transport, Innovation and Technology* to a parliamentary question (atopq1; January 26th 2009) on which I will focus in the next paragraph. The first reference to the second AVK device in Austrian newspapers was in an article in the *Oberösterreichische Nachrichten*, sold in all of Austria, on November 25th 2009 (cf. a26). It was only on July 8th 2010 that an article (a28) appeared in the *Wiener Zeitung*, also sold in all of Austria, citing the ASFINAG press spokeswoman as saying that there had been a second AVK camera in operation since August 2008. That means, the first mention of the second AVK device in Austrian newspapers was about 15 months after its implementation and it was about 23 months after this implementation that another newspaper reported on it additionally mentioning the implementation date of August 2008.

Parliamentary questions and answers (right of interpellation) about AVK in the Austrian Parliament (11/2008, 01/2009 and 07/2010, 09/2010)

It is remarkable that in the period in which my sample of Austrian newspaper articles did not show any reports on AVK: that is over one year and ten months between January 18th 2008 and November 25th 2009[67], there was a parliamentary question time

[67]Again, there was one appearance of AVK within this period in May 2009, but AVK was only mentioned as a very brief side note in a short *Tiroler Tageszeitung* (regional newspaper) article (a25).

(interpellation) on AVK (pq1; November 27th 2008) with answers to it (atopq1; January 26th 2009) in the Austrian National Parliament. To come straight to the crucial point: there has not been a single account about this interpellation process or any information provided about AVK in answer to the parliamentary questions, in Austrian newspapers.

According to the website of the Austrian Parliament, parliamentary questions are a means for the parliament to exercise political control over the work of the Federal Government and its members. Parliamentary questions are described there the following way:

> Under Art. 52 B-VG (n.b. B-VG is the Federal Constitutional Law) the National and Federal Councils may examine the activities of the Federal Government and interrogate its members on all matters of execution (right of interpellation) and demand all requisite information. This also applies to all enterprises in which the Federal Government holds a majority interest. In principle, the persons interrogated are under the obligation to answer truthfully. If the desired information cannot be given the reasons must be stated.[68]

In this specific case, the Member of Parliament (MP) Harald Vilimsky and colleagues from the *Austrian Freedom Party FPÖ* asked 13 written questions on the issue of 'fully automated toll sticker checks' ("vollautomatische Vignettenkontrolle") to the Federal Minister for Transport, Innovation and Technology, Werner Faymann (*Social Democratic Party of Austria SPÖ*) on November 27th 2008 (cf. pq1)[69]. On January 26th 2009, the new minister Doris Bures (*SPÖ*) answered the questions (cf. atopq1).

The first questions asked were on basic information about AVK, or as named in the interpellation 'fully automated toll sticker checks' ("vollautomatische Vignettenkontrolle"). To sum up the answers to the first questions (1-6): The first AVK

---

[68] http://www.parlament.gv.at/ENGL/PERK/KONTR/POL/1INTERPELLATIONSRECHT/index.shtml [Nov 29th 2012]

[69] Four days earlier, on November 23rd 2008, Werner Faymann announced the continueation of a 'Grand Coalition' with the Austrian People's Party ÖVP following the general election on September 28th 2008. This also meant that Faymann was going to be the new Federal Chancellor of Austria and therefore he left his position as the Federal Minister for Transport, Innovation and Technology.

device was implemented on December 12th 2007 and at the present time (January 26, 2009), there were two devices in operation. These two devices were in operation all over federal territory subject to a sampling plan. The sites of operation changed weekly as a rule. Once the AVK device was in operation it was able to record images on one traffic lane in good visibility conditions; that is in daylight. During summertime it could be in operation about 16 hours and during wintertime about eight hours a day. All vehicles on the respective lane were captured. The original cost of one AVK device was EUR 223 351.75 (net) and annual maintenance costs, EUR 4 220 (net). The latter, payable only once the guarantee had expired. Until then there had been no information about the operational life span of an AVK device.

Question seven in the interpellation was about the mode of operation. The answer was the following:

> "The surveillance system produces two images. One image is for number plate recognition, the other one of the windscreen is for the assessment of whether a toll sticker has been duly attached. The angle of the camera is set so that persons in the car are usually not recognisable on the windscreen images. Data is electronically encoded, transferred to an external storage device and transmitted to the enforcement centre of the ASFINAG Maut Service GmbH. The AVK system complies with the strict requirements of the data protection law as does the entire toll system. It was reviewed and authorised independently by the data protection commission."[70]

Questions eight to 13 were about AVK system statistics. The answers to these questions can be summed up as follows: There were no figures available on how many cars were checked in total, as images that do not show an infringement of the toll law (n.b.

---

[70] (Translation by author): „Das Kontrollsystem generiert zwei Bilder, ein Bild zur Kennzeichenerkennung sowie ein Bild der Windschutzscheibe zur Beurteilung, ob eine Vignette ordnungsgemäß angebracht wurde. Der Winkel der Kamera ist so eingestellt, dass bei der Windschutzscheibenaufnahme die Personen im Auto in der Regel nicht erkennbar sind. Die Daten werden elektronisch verschlüsselt, auf ein externes Speichermedium übertragen und dann an die Auswertezentrale (Enforcement Center) der ASFINAG Maut Service GmbH übermittelt. Das automatische Vignettenkontrollsystem entspricht – wie das gesamte Mautsystem – den strengen Bestimmungen des Datenschutzgesetzes und wurde von der Datenschutzkommission gesondert begutachtet und genehmigt."

meaning images with valid toll stickers) were automatically deleted by the system, or in case of doubt, by the manual follow-up check. Since the date of implementation (Dec 12th 2007), a total of 12 869 vehicles with an invalid toll sticker were detected (answer to question eight). Out of these 12 869 vehicles, 5 299 were domestic and 7 570 were foreign (answer to question ten). Usually, images of vehicles with invalid toll stickers were recorded and analysed in the ASFINAG enforcement centre. Should the result show there was no valid toll sticker attached, a written request to pay 'compensatory toll' ("Ersatzmaut") was sent to the client. It was however also deamed feasible to conduct a selective pulling over of clients ("gezielte Ausleitung") following an immediate online-evaluation (answer to question nine).

In answering question eleven; if and how many cars displayed more than the two permitted annual toll stickers on a windscreen, the Minister noted that there was no investigation in this regard, as to have more than two toll stickers was only an explicit recommendation and not a breach of regulations ("Verwaltungsübertretung"). Question twelve and its answer deserve more scrutiny, because the answer in my opinion does not satisfy the legal requirement of answering truthfully or if the desired information cannot be given, to state the reasons for that. The question asked was about the error rate of the system. Firstly: in how many cases was no toll sticker detected even though there was one attached, and secondly: were there any cases in which a toll sticker was detected even though this was not the case. To sum up in my words, numbers of false positive and false negative cases were requested. The answer to the question was: "Altogether there were 159 cases, in which the absence of a valid toll sticker was not clear. In such cases the decision was in favour of the client."[71] In my interpretation that means that there were 159 false positive cases that could also not be clearly recognised by the human operators. There were no numbers provided about false negative cases and additionally there was no reason given why this information was not shared. Obviously the information provided in answer eight was relied on—that there were no figures available on how many cars were checked in total, as images with valid toll

[71]"Es gab insgesamt 159 Fälle, in denen das Fehlen einer gültigen Vignette nicht ganz eindeutig nachgewiesen werden konnte. In diesen Fällen wurde zugunsten des Kunden entschieden."

stickers are automatically deleted by the system—as being evidence enough for not answering the false negative question. If there was no valid toll sticker image, there can be no figures available about false negative cases, meaning those cases in which a valid toll sticker was detected even though there was none. Usually in technology evaluations (cf. Introna & Nissenbaum 2009: 12) of IPAs, false positive and false negative rates are necessarily specified during system tests and evaluation in order to prove viability. This means in my interpretation, that information about false negative cases was available at the time of interpellation but it is highly probable that it was not provided.

The answer to question 13, asking how many objections against detected offences were filed, stated that in a total of 1 359 cases, customers complained against the demand for a 'compensatory toll' ("Ersatzmaut"). Regarding the question of burden of proof in such a case, it was noted that the decision to initiate administrative prosecution ("Verwaltungsstrafverfahren") was based on the two available images which could be provided to the district administration ("Bezirksverwaltungsbehörden") on request.

As already indicated, it is astonishing that there were no reports in Austrian newspapers of the interpellation process on the AVK system. Also, a second parliamentary question and answer time following the 'fully automated toll sticker checks' („vollautomatische Vignettenkontrolle") question and answer time, remained unnoticed in the Austrian press. On July 9th 2010, the MPs, Mayerhofer, Vilimsky and colleagues from the *Austrian Freedom Party FPÖ* raised 16 new questions in writing, on the issue of 'toll stickers – quality and penalisation despite a valid toll sticker' ("Vignette – Qualität und Bestrafung trotz vorhandener Vignette") with the Federal Minister for Transport, Innovation and Technology, Doris Bures (*Social Democratic Party of Austria SPÖ*) (cf. pq2). On September 9th 2010, Minister Bures answered these questions (cf. atopq2). They fell in line with question 13 and its answer (pq1), taking up again the issue of client complaints, as clients objecting to their valid toll sticker having been recognised as invalid, maintained that this was due to a manufacturing error. Due to this, parts of the toll sticker were missing once affixed to the windscreen.

> "In 2009 for instance, practical problems arose with peeling away and affixing the toll sticker. The stickers did not separate completely from the backing film and parts of

letters and numbers went missing on toll stickers attached to windscreens. These flawed stickers were recognised and registered by the AVK as invalid. Affected car drivers using (needing to use) motorways equipped with AVK on a regular basis, were repeatedly registered and fined for displaying an invalid toll sticker."(pq2[72])

The questions following the description of this problem of a production error leading to unjustified penalisation by means of AVK were mainly about toll sticker production conditions (1-8), critically investigating quality control of the manufacturing process. Questions nine to eleven were about the AVK system: How many devices are in operation at the moment? (9) Where are these devices in operation? (10) Are there significant differences regarding the number of penalised car drivers following a check by enforcement authorities ("Straßenaufsichtsorgane") or by AVK? Interestingly, the Minister did not answer these questions, in contrast to similar questions in (pq1), arguing that they were not the responsibility of her Federal Ministry (cf. atopq2: 2 for the detailed answer and arguments of why this was the case).

Questions twelve to 16 (pq2) were raised about possible multiple penalisation ("Mehrfachahndungen"). This was the case when one and the same vehicle was detected with an invalid toll sticker multiple times a year. According to ASFINAG, the answer to this set of questions was that there were 864 multiple detections in 2009 using AVK. The answer also referred to toll regulations that included the request to affix the toll sticker undamaged and directly onto the inner windscreen so that it was clearly visible and could be checked from the outside. If a toll sticker was affixed with the intent to mislead or deceive, and this intent was clearly proven, this could lead to a demand for compensatory toll and penalisation.

---

[72] (Translation by author) „Konkrete „Probleme“ mit dem Ablösen und Anbringen der Vignette gab es beispielsweise im Jahr 2009. Vignetten haben sich nicht zur Gänze von der Folie gelöst, Teile von Buchstaben und Ziffern haben bei der auf die Windschutzscheibe aufgeklebten Vignette gefehlt. Dies wurde von der automatischen Vignettenkontrolle erkannt und registriert. Betroffene Autofahrer, die Autobahnabschnitte mit automatischer Vignettenkontrolle regelmäßig benutzen (müssen), wurden mehrfach als mit „ungültiger“ Vignette zur Anzeige gebracht und mehrfach bestraft.“

Again, although there were critical questions about AVK and its implications such as error rates, multiple detection, or toll sticker production errors in the two parliamentary questions raised, and although there was further information about AVK provided in the answers, Austrian newspapers did not report on this discussion. That means the discussion about AVK in the Austrian Parliament and its implications went entirely unnoticed by the wider Austrian public.

Performance Reports: Economic Success, Failure and Irritations (11/2009-03/2011)

Almost two years after the implementation of the first AVK device on Austrian motorways (12/2007) and 15 months after the implementation of the second AVK device (08/2008), a first account of the performance of AVK was published in the nationwide newspaper *Oberösterreichische Nachrichten* (a26) on November 25th 2009. The central theme of the article was that 12 200 more toll sticker offenders were caught in the first two quarters of 2009 than in the same period in 2008 as a consequence of AVK implementation. A *Kurier* article (a27, January 5th 2010) agreed that AVK had been a success, stating that toll sticker offenders were on the retreat. Ingrid Partl, of the ASFINAG Maut Service GmbH argued in the article that the reason for this trend was seen as an increase in car drivers´ awareness, due to permanent checking. It was further argued that AVK also contributed to this trend. About six months later a *Wiener Zeitung* article (a28) continued the individual accounts of AVK success when commenting 'A new Record in the Hunt for Toll Sticker Offenders' ("Neuer Rekord bei Jagd auf Vignetten-Sünder") as the headline. The teaser informed the reader that new cameras were catching almost 2 000 toll sticker offenders a month and that the system was going to be extended. In the main body of the text they reported, in reference to the new ASFINAG statistics that 1 800 toll sticker offenders (this was the more concrete number instead of the almost 2 000 toll sticker delinquents mentioned in the teaser) had been detected by AVK. As the compensatory toll was 120 Euros, the annual revenue generated by AVK was calculated as being roughly 2.6 million Euros. Meaning that one device (230 000 Euros) had already paid for itself. AVK was therefore presented as a sort of magic weapon ("Wunderwaffe") in the fight against toll sticker offenders. As a consequence, the article went on, ASFINAG was thinking about the expansion of the AVK system as they were 'very confident' about it. However the article also explained that the development of the system had taken a long time and had even been in danger of failing due to technical problems. Contrary to this, they also reported that the system had been continuously refined ("technisch immer weiter entwickelt"), which lead to a higher success rate ("höhere Quote“).

Three months after (a28), on October 11th 2010, an article (a29) in the daily newspaper *Kurier* announced in a short note within a longer article about ASFINAG that they were currently testing a new toll sticker surveillance device, because the existing device in operation on the *Südosttangente* (A23 in Vienna) 'never really had functioned perfectly' ("nie wirklich perfekt funktioniert") for years. As such, the note in the article contradicted the view of AVK success as reported in the previous articles (a26-a28) and presented the existing two devices as failures. This short notice remained the only one questioning the proper functioning of AVK.

In the context of reports about the expiration of the 2010 toll sticker on January 31st 2011, a *Salzburger Nachrichten* article (a31) did not mention failure, but reported the expansion of the system from two existing AVK devices to five altogether, in 2011. A positive picture of AVK was also voiced in a *Tiroler Tageszeitung* article (a34) when ASFINAG executives declared themselves satisfied with AVK performance.

In this period, two articles reported on irritations in the context of AVK. On February 9th 2011 an article in the daily newspaper *Kronenzeitung* (a32) told the story of a car theft that ended with a compensatory toll demand sent to the actual owner of the car, because the stolen car had been detected by AVK without a toll sticker. In the article this was seen as an injustice, because the car owner eventually had to pay the compensatory toll. Another irritation with AVK was reported in *Kleine Zeitung* (a33) in that same month. It told of a reader who forgot to attach the toll sticker to her car for one week. As she was not stopped by any authority during this period, she thought she had been lucky, but after a while she received several demands for compensatory toll in the post, as she had driven right into the trap ("in die Falle getappt"). The article addressed the delay between being detected by AVK without a toll sticker and the moment at which the compensatory toll demand is received, as this delay is new to the Austrian people who are only used to on the spot checks and fines. It further explained that once the compensatory toll is definite, it is valid for the day of detection and the day after. Both accounts (a32 & a33) pointed out the rigidity of the toll system that made no exceptions, strictly following toll regulations. While the injustice seen in these accounts was principally sympathised with, the articles also implicitly display the AVK device itself as being successful and well functioning technology, able to detect toll sticker offenders perfectly in both cases.

A new start? Or: Second expansion? (Spring 2011)

A *Salzburger Nachrichten* article (a31) reported on the planned addition to the existing two AVK devices to altogether five devices in 2011 on February 2nd. About four and a half months later, on June 21st 2011, the relatively small regional newspaper *NEUE Vorarlberger Tageszeitung* reported of the implementation of three new AVK devices on Austrian motorways (a36). Thus, ASFINAG would operate five AVK devices altogether. According to ASFINAG in the short article, camera technology had improved in comparison to prior devices. The article also referred to a press release (a35) by the producer of the new AVK devices, the company *EFKON*. In the press release, the article explained, *EFKON* pointed out that the compact design of the new devices made installation easier and therefore enabled more frequent site changes. Aside from that, the system was able to detect toll sticker offenders automatically without interrupting the flow of traffic. Regarding the mode of operation, the article cited the press release when describing the process in which the system had an overview of one lane from an overhead position in order to capture images of the front of all passing vehicles. Subsequently, the images captured were analysed with regard to the existence of a valid toll sticker. The *NEUE Vorarlberger Tageszeitung* article (a36) remained the only Austrian newspaper to directly refer to the *EFKON* press release. One week later, on June 27th 2011, in a short note, the newspaper *Neue Kärtner Tageszeitung* (a37) reported that latterly ("neuerdings“) there had been a focus on mobile toll sticker checks and that was going to be expanded over the next few years. One day later, the headline of a *Kronenzeitung* short note (a38) announced the fight against toll offenders using mobile toll sticker checks ("Kampf gegen Mautpreller mit mobiler Vignetten-Kontrolle"). It stated that until then 15 000 toll offenders had been caught with the help of the new mobile, automatic toll sticker checking system operated by ASFINAG. So far, the short note reported, five such 'surveillance cameras' had been in operation. They announced further that the 'special camera ("Spezialkamera") recognised instantly if a valid toll sticker was affixed to the vehicle. About two weeks after the *EFKON* press release was published, on July 5th 2011, an online article on the German online tech-news platform *heise.de* (a39) reported on the *EFKON* press release in a side note, providing a link to the English version of the press release on the *EFKON* website. It stated that a few days preceding this, the Austrian company *EFKON* had announced that their AVK device was now ready for mass production („die Serienreife erlangt hat"). Until then, the article claimed, automatic systems for inspecting toll stickers had not gone beyond test installation due to high error recognition rates.

The *EFKON* press release (a35), also available in German on the website of its parent company *STRABAG* (a huge construction company operating mainly in Austria and Germany) neither reported on the readiness for mass production nor on other automatic toll sticker inspection systems that had not gone beyond test installation due to high error recognition rates. Thus, it remained unclear where *heise.de* article (a39) had obtained its reference from. Nevertheless, it is interesting to have a closer look at the *EFKON* press release seen from this angle, as it represented a new start rather than an expansion in the history of AVK. The press release did not directly mention the existence of the two original AVK devices. It only stated that "in addition to toll enforcement and monitoring officers, ASFINAG also uses automatic enforcement and monitoring systems (Automatische Vignettenkontrolle, AVK)". The overall impression of the press release was that the new *EFKON* devices were neither a replacement nor an expansion of the two existing ones, but were something completely new. In the press release, AVK was mainly referred to as "automatic toll sticker checking", but additionally it was also named a "mobile system" and "innovative system". The latter strengthens the impression of something new.

In the *EFKON* press release, AVK was described as solving three different problems: it enabled free flow of traffic and thus, it also contributed to increased traffic safety as it was "used in those places where safety reasons do not permit manual monitoring". Thirdly, as part of "an efficient, automatic toll sticker enforcement and monitoring system" it led to more fairness and is more just "in the interest of those road users who comply with payment of the time-limited toll". Seen in the wider frame, the press release also validated the Austrian toll sticker system when presenting it as "easy to use and economical." In comparison to the newspaper articles reporting on the initial start-up of the first AVK device in 2007, the problems addressed in the *EFKON* press release differed significantly. The newspapers had reported that AVK was a solution to the problem of missing opportunities for stopping and checking vehicles on urban motorways, especially in the wider Vienna area due to safety reasons, whereas now the new system was presented as a contribution to increasing road safety. That meant that AVK was no longer being presented as a means of checking vehicles in places where it had not been possible before due to safety reasons, but as a safety measure in itself. A second shift took place when the *EFKON* press release raised the issue that the new AVK system led to more fairness and justice. Such argumentation in the initial start-up reports in 2007 had been lacking. Instead, in these earlier reports it had been seen as a measure for boosting toll sticker morale.

Regarding the mode of operation, the description in the *EFKON* press release did not differ significantly from the newspapers´ descriptions of the "old" AVK devices. The press release explained that "the system recognises the toll sticker and automatically checks its validity." It further argued:

> "From an overhead position, the system overlooks one lane of the roadway and photographs the front view of all passing vehicles. The images are then checked for the existence of a valid toll sticker. This innovative system from EFKON is capable of independently determining and monitoring the existence and the validity of the Austrian toll sticker on the vehicles without interrupting the flow of traffic. A special high-resolution light-sensitive camera is for the system to determine whether the sticker is a valid Austrian one." (a35)

Interestingly, in the *EFKON* press release describing the mode of operation, the human operator was completely missing. This was underlined when 'the capability of the system to independently determine and monitor' the existence and validity of the toll sticker was described. Similar to the press accounts in the initial start-up period in 2007, it was the camera—in this case a special high-resolution, light-sensitive one— that was positioned as the central actor in the AVK system that was able to recognise and to read whether the toll sticker existed and was valid. Only indirectly, when the Senior Vice President of *EFKON* thanked his development team at the end of the press release for having "done an outstanding job" and having "successfully incorporated (...) many years of experience with various toll systems, camera technologies and image processing, does the reader gets to know a little bit more about the technology behind or in the camera: that is, image processing. It is the only moment in which 'image processing' as a term came into play in all of the publicly available accounts of AVK in my sample.

Performance reports: Success and 'Schönheitsfehler' (11/2011-03/2012)

Following the new start or second expansion of AVK in Spring 2011, in the context of the sales start of the 2012 toll sticker on December 1st 2011, an article in the *Oberösterreichische Nachrichten* (a40) reported about the performance of AVK. Again, as was the case with the first expansion in August 2008, when the same newspaper (and same journalist) reported successful performance (cf. a26), the *Oberösterreichische Nachrichten* article (a40) was the first account of the performance of the newly implemented AVK devices in Austrian newspapers. Again, the central argument in the article (a40)—as was the case in the earlier one (a26)—was that 16 000[73] more toll sticker offenders had been caught by September 2010 in comparison to the same period of time in 2009, as a consequence of AVK implementation. This was explained in the article, in reference to ASFINAG press spokeswoman that prior to this („zuvor") there had been only two AVK devices in operation. After the end of March (2011) there were

[73]This number refers to the total number of toll sticker offenders, that means, deriving from both manual and automatic checks.

then five devices in operation. As (a40), published in November 2011, seemed to be a copy of (a26), published in November 2009, just with different numbers, it was interesting to note that the same article was missing in November 2010. The reason might be that there had been no increase in the number of detected toll sticker offenders, and therefore—at least this reason would seem logical—no reportable success story was available in 2010. A closer look at other newspapers confirmed this possibility: In 2008, there were 97 000 toll sticker offenders reported (cf. a28). In 2009, (a28) reported 106 000 offenders. In 2010, there were 103 146, and in 2011, 130 903 toll sticker offenders (cf. a43). That means that in 2010, when a decrease in the number was forseeable, there was no report about AVK in the newspaper *Oberösterreichische Nachrichten*.

However, as the presented numbers indicated, the minor success story of AVK in Austrian newspapers continued: On March 16th 2012, a main article in the nationwide newspaper *Wiener Zeitung* (a43) reported a new record in toll sticker offenders ("Neuer Rekord an Vignetten-Sündern"). In reference to ASFINAG annual statistics—not publicly available, but passed on exclusively to the *Wiener Zeitung* journalist—a jump ("sprunghafter Anstieg") in toll sticker offenders was revealed. The increase in AVK devices was named as having substantially contributed ("wesentlicher Grund") to the increase in detected toll sticker offenders. It also reported that the AVK devices detected 4 500 toll sticker offenders a month, which is about 54 000 a year. The article noted that following a long development period and several initial difficulties, AVK was slowly starting to fulfill all expectations as a magic weapon ("Wunderwaffe") in the fight against toll dodgers. The same expression had already been used by the same journalist in an earlier *Wiener Zeitung* article (a28) in the first performance report period. As a consequence of the success of AVK the paper reported there were plans to extend the system with three more devices, (a43). As the first part of (a43) presented AVK as a success story, the second part contextualised the success with a mention of so-called "Schönheitsfehler" – minor flaws. In spite of the jump in detected toll sticker offenders (n.b. +26.9%), the overall ASFINAG revenue from fines remained at nearly the same level as in the year before (n.b. +1.19%). For ASFINAG, the article states, the reason for this discrepancy was difficult to explain, because the overall revenue from fines consisted of fines from both cars and trucks. The journalist considered the reason for this discrepancy to be the absence of agreements with other nations. In reference to an *ÖAMTC* lawyer, the article argues, toll sticker fines are under private law and not road traffic offences. Meaning that fines resulting from AVK detection are not enforceable outside Austria. This implied, as a subtitle of the article confirmed, that the absence of bilateral agreements did complicate the enforcement of fines, especially those of non-Austrian drivers detected by AVK. Here the question arose of whether the presented detection numbers also included foreign car drivers that cannot be fined due to missing international agreements. If this was the case it might explain why there had been a higher increase in detection numbers in comparison to a far smaller increase in sales. It is also interesting to note here that in these performance reports, only the increase in detection numbers was highlighted, while the previously mentioned aim of AVK of boosting toll sticker morale was not an issue any more.

Conclusions

The aim of the case study focused on in this chapter was to reconstruct the staggered introduction and implementation of the so-called ‚Automatic Toll Sticker Checks' („Automatische Vignettenkontrolle", in short: AVK) in Austrian motorways and expressways, one of the first nationwide systems based on Image Processing Algorithms in operation in Austria. I presented how AVK had been described, outlined, introduced and tested or not in the Austrian news and in publicly available documents.

My exclusive focus on media reports and publicly available documents was an outcome of the ASFINAG—the operator of AVK—strict information and secrecy policy that I presented as a result in itself at the beginning of the chapter, also confirming what others in academia (Kammerer 2008; Monahan 2010) have described as a 'culture of secrecy' when it comes to technologies that have a touch of surveillance about them. That is also the reason why I did not refer to an interview I conducted with a product manager of EFKON, the producer of the "new" AVK devices introduced in 2011. After the interview was held I was politely asked by the product manager, probably on the initiative of ASFINAG, not to cite any customer-related information (in this case: ASFINAG related information) that had not already been published elsewhere. Due to the strict ASFINAG information policy there were also no possibilities to observe and analyse situations in which human operators in the ASFINAG enforcement centre double-checked AVK results by reviewing the transmitted images of potential toll sticker offenders. Such a focus on the human side of image and IPA result interpretation would have been necessary in order to be able to analyse the relation of human vision to computer vision in more detail once Image Processing Algorithms have already been implemented.

In what follows, I shall present two analytical conclusions based on the previous description and analysis of the staggered introduction and implementation of AVK. Firstly, I shall concern myself with the blackboxing of Image Processing Algorithms and their uncertainties as was performed in the analysed media reports. And secondly, my theme shall be the non-testing and non-contesting of AVK in the media accounts and in connection to this, its presentation as a moral agent.

The Blackboxing of Image Processing Algorithms and their Uncertainties in the AVK case

The first central insight of the analysis of media articles about AVK is that the camera is positioned as the central and most powerful actor in the whole story. It is the camera that detects and recognises the presence and validity of toll stickers, whereas Image Processing Algorithms and other relevant actors such as human operators are widely neglected and blackboxed in favour of the 'automatic' and 'innovative' camera. First of all, this way of describing new AVK technology can be interpreted as being over simplified. That is, the focus on the camera as the central actor refers to a well-known technological artefact; the camera, which makes it easier for people to understand what this is all about. What is transported to people not familiar with the system is that actually the innovation is not something radically new and as such it might not be something problematic, but follows other similar technologies such as speeding cameras. It presents AVK as a speeding camera not for checking speed, but for checking the right to be allowed to drive on the motorway by ownership of a valid toll sticker. This way of framing AVK does conceptualise something new and unknown such as AVK, as something very familiar, but at the same time this well known something comes with a look of innovation about it, by labelling it "intelligent". Thus, it was remarkable that the media did not report what is actually inside the blackbox of this 'intelligent' and 'innovative' camera, deeming it self-evident. Meaning that it was no matter of concern what actually made the camera an intelligent camera. It was therefore already a closed, black-boxed actor in the system. It was presented as a matter of no concern that Image Processing Algorithms in particular, make the camera an intelligent camera. By not mentioning them at all, Image Processing Algorithms were made part of the background, they were made completely invisible. One might say, for normal citizens—and most of these will never be confronted with AVK at all—it is of no importance to know what kind of technology something such as AVK is based upon. However - and here I dispute this line of reasoning - knowing what kind of technology is used is the basis for critical questioning and discussion. For active participation in (highly technological) democratic societies, the capacities and limits, as well as the uncertainties that come with this technology have to be addressed. As much as IPAs were kept part of the background, so also were their uncertainties and failings left completely invisible in the media. Some of these uncertainties were mentioned during parliamentary questions and answers: For example, it was stated that the system (in 2009) only worked in good visibility conditions that is, in daylight. Also error rates, false positive and false negative cases were asked about. The answer was that there had been 159 false positive cases in a specific period of time, but no numbers about false negative cases were presented. Instead of making capacities, limits and uncertainties of IPAs a matter of discussion, most of the newspaper articles presented AVK as a stable, fully developed, ready-made and almost autonomous camera system by referring to it as an “intelligent system”, “electronic eye”, or “modern camera” that is in actual fact, able to detect and recognise the presence and validity of toll stickers automatically. That AVK is able to do so was shown by presenting its economic success. It was a recurring theme in the media articles that detection numbers and sales figures that had been exclusively provided, were presented to the public. This focus on economic success, amongst other things, might be the case, because many articles were published in the business section of the newspapers[74]. It is interesting to note that there was not one single article published in the science & technology section of a newspaper, transporting the message that AVK is not a matter of science and technology, but very notably an issue of business.

What does this mean for the public understanding of science and technology in general and for Image Processing Algorithms or computer vision in particular?

[74] Apart from the business section, AVK articles were also found in the transport & automobile section, in the local or domestic section, and in the news section of newspapers.

The Non-(Con-)Testing of AVK as a Moral Agent

Connecting to my first central insight, my second central insight in this chapter is that AVK was neither tested nor contested in the media articles at all. It was framed as unproblematic, ready-made and familiar camera technology that makes sense especially in economical terms, by facilitating the collection of more tolls and toll fines, and moreover, acting as a moral agent in accomplishing more justice on Austrian motorways. What happened was that people were told and taught of the positive economic and fairness effects ("boosting the toll sticker morale of car drivers") of the ready-made technology of AVK in order to accept it as useful and plausible technology. This was done especially by ASFINAG, the operator of AVK. Because ASFINAG was in a position to decide which information and communication about AVK was provided or not provided, e.g. to the general public, to the media or to a social scientist, they certainly can be regarded as an 'Obligatory Point of Passage' (Callon 1986) through which information and communication must always pass. ASFINAG provided the media with exclusive and filtered information about the economic success and the moral aspirations of AVK, while not providing information about the uncertainties, potential errors and the organisation that come with it. This information was partially provided in the publicly documented, two parliamentary question and answer times, but astonishingly no single media article reported on the interpellation process, nor on the information provided in the answers.

Instead of this, the media - with the exception of one article (a29) questioning the proper functioning of the first AVK devices but not taking this any further - just uncritically replicated what ASFINAG had claimed and did not show any interest in other aspects that were not brought up or provided by the operator ASFINAG. Thus, the media articles led to the impression that AVK was infallible and the ultimate means on" combating toll sticker offenders, maybe best expressed by one headline that associated AVK with toll sticker offenders now having zero chances ("Die erste automatische Vignetten-Kontrolle: Pickerlsünder sind nun komplett chancenlos"). This implied that AVK was a perfect moral agent ensuring that nobody would ever drive without a toll sticker on Austrian motorways again. It put AVK in a position in which its viability is neither tested nor contested. It was just taken for granted. The impenetrable and powerful black box, AVK and how it was framed in the media was completely in line with Latour's 'Berlin Key' (Latour 2000) or the 'Sleeping Policeman' (Latour 1999: 186). Similar to speed bumps forcing people to reduce their driving speed, as otherwise they would damage their cars when driving over them too fast, so also does AVK force people to put a valid toll sticker on their windscreens while driving on Austrian motorways, because all-seeing AVK could be anywhere, anytime - a message delivered to car drivers particularly via newspaper articles. On the one hand, the way in which AVK was presented by the media, with the camera as its central performer, made it appear like "magic technology" because it was never made clear to the audience how the intelligent camera really works. On the other hand, the mere presence of one of the most powerful symbols of objectivity - the camera cf. Daston & Gallison 2007) - might be sufficient explanation for the viability and 'raison d'être' of this "magic technology" that is AVK. A closer look into that magic black box and the sociomaterial assemblage of which it is part, could definitely endanger the magic and power of the moral agent, at least seen from the point of view of the operator. From my point of view, a closer look into that magic black box and its place within a larger sociomaterial assemblage would put those affected into a more impartial position from where they could participate in discussing, assessing, and contesting not only the specific technology of AVK, but also (future) "intelligent" or "smart" technology, and sociotechnical development in general.

Chapter Five

'Inscribing Social Order': Developing Image Processing Algorithms in Computer Vision Laboratories[75]

In daily life, most technology works in the way we expect it to and therefore we are usually not interested in how exactly specific technology functions. If a technological artefact or process does not perform adequately, more often than not we are unable to fix it ourselves, because we have neither sufficient knowledge nor the skills for doing so. The technological artefact similar to the AVK device focused on in the previous chapter, appears as a quasi-natural 'black box' and thus, as 'ready made technology', which seems to operate in a fixed and predictable manner (Schulz-Schaeffer 2000).

In this chapter, the approach to technology and Image Processing Algorithms will be, to use the words of Bruno Latour (1987), through the back door of technology in the making and not through the more grandiose entrance of ready-made technology. The main purpose of the chapter is to understand how Image Processing Algorithms, operating on a semantic level in the area of pattern recognition are being designed in computer science laboratories. It approaches one of the main questions of this thesis, that is: how are distinct human views and knowledge and as such, distinct modes of reality and truth inscribed into IPAs and how are these human views and knowledge configured in these inscription processes? As such, it also reflects on the question of how everyday patterns of seeing and recognising are interwoven, configured and

[75] This chapter is a reworking und continuation of my paper: MUSIK, Christoph (2011): The thinking eye is only half the story: High-level semantic video surveillance. In: *Information Polity* 16/4: 339-353.

reconfigured using Image Processing Algorithms and their larger sociomaterial assemblages. The chapter is concerned with the sociotechnical requirements of IPAs. One basic and fundamental requirement of all IPAs in the area of pattern recognition is the sociotechnical construction of what computer scientists call 'ground truth' in their everyday work. For example, the ground truth of suspicious behaviour detection corresponds with the question of what suspicious behaviour and hence normal behaviour looks like. The sociotechnical construction of ground truth is simultaneously the construction, production and reproduction of social classification and social order. Therefore, the ground truth of Image Processing Algorithms is a crucial and constituting element of society that needs closer attention before it becomes black-boxed and moves into the realm of things that are taken for granted. Understanding the sociotechnical construction of the ground truth is key to understanding and learning about society and social order assuming that what is inscribed in ground truth has real implications once an IPA has been deployed.

In further consequence the chapter shows how the complexities of human vision and recognition can be transformed and reduced in these processes of constructing ground truth and how this might impact our conception of how we see and recognise. By doing so it draws on discussions from the fields of Surveillance Studies as well as Science and Technology Studies, refered to in previous chapters. The first section of this chapter is a reminder that brings together these insights. It summarises in short, the co-production processes of technology, knowledge, code, algorithms, and society. One conclusion of the chapter will be that complexity transformation and reduction in the development of semantic IPAs have an upskilling effect rather than the reverse, where the implementation of IPAs in larger sociomaterial assemblages is concerned.

Empirically, the chapter is concerned with three different "domains of scrutiny" (cf. Goodwin 1994). All these domains of technology in the making have digital images in common, whether they derive from static images or from dynamic video images as well as being produced and processed by algorithms in order to analyse behaviour patterns automatically. This also means that in all of the three cases the above-mentioned ground truth has to be constructed for the purpose of comparing and matching the observed scene to the ground truth template once the IPAs are in operation. With all of the three cases another aspect of importance for the sociotechnical construction of the ground truth is emphasised. Firstly, after introducing what is meant by the term ground truth I shall lay emphasis on the construction of ground truth in the area of tension that lies between expert and lay knowledge by referring to Facial Expression Recognition Technologies that I researched in the dissertation I wrote for my Master's degree (cf. Musik 2009). Here I showed that it is a matter of selection which approach in forming a ground truth is given preference and subsequently applied. Secondly, I shall refer to what is called 'Automated Multi-camera Event Recognition' for the prevention of bank robberies and other incidents in the context of commercial banks. I refer to an applied research project in which I was involved as a social researcher. With this example of automated event recognition it can be shown that on the one hand, an attempt at the sociotechnical construction of a ground truth is influenced by relevant social groups and their interests, and on the other hand connected to it, how this construction is a matter of setting boundaries and preferences between the areas of formalised, explicit and everyday, tacit knowledge. I came across the third example I shall refer to during my fieldwork in a university computer vision laboratory: Automatic Fall Detection. This case particularly, stresses the experimental character of sociotechnical ground truth construction as it occurs within the framework of computational scripts and the idea computer scientists have of possible users and their respective behaviour, not to forget the imagined places of application.

The Co-Production of Technology, Knowledge, Code, Algorithms, and Society

Analysing semantic Image Processing Algorithms brings up the question of what kind of knowledge and computer codes are similarly being applied, transformed, and co-produced. Both the use of knowledge and computer codes can be regarded as social activities and thus, they are not "natural"; meaning that they did not just drop from the sky without the involvement of human and societal practices. Both knowledge and computer codes are constructed and produced; they are in a way "manufactured". Here, ‘laboratory studies’ (Knorr-Cetina 1981, Latour & Woolgar 1979) in which the manufacture of techno-scientific facts and knowledge was analysed in situ in scientific laboratories can be drawn from. As Knorr-Cetina observed: “Facts are not something we can take for granted or think of as the solid rock upon which knowledge is built” (Knorr-Cetina 1981: 1). Knorr Cetina gave meaning to the myriad decisions and selectivity of fact-fabrication. Thus as she notes, it is important “to study the process by which the respective selections are made” (ibid.: 7).

In the context of semantic IPAs, the specificity of knowledge has to be brought to mind. The concern is basically the imperative to translate implicit into explicit knowledge. This imperative, which can be found in more and more areas of social life today, is generated by the increasing application of computer-based information technology (IT). What is behind this trend is that so far, most decisions and activities have been based on implicit or tacit knowledge of the experts involved.

These ever increasing activities and decisions based on implicit or tacit knowledge are delegated to IT systems such as those that integrate Image Processing Algorithms. In this process, the implicit or tacit knowledge has to be made explicit. Thus, rules applied to activities and decisions have to be identified and specified in a way suited to computer programmes. In further consequence, they have to be formalised and codified (Rammert 2007). But as was outlined in Chapter Three in the Sociology of Scientific Knowledge standpoint on “intelligent machines”, it was assumed that only a specific kind of knowledge can be computerised (Collins 1995: 298); one that mimics “the world of mechanical cause and effect“ (Collins 2010: 55).

This process of making tacit knowledge explicit in the case of IPAs was also described as a process of reduction. As I explained in Chapter Two, this issue especially refers to Facial Recognition Technologies (Introna & Wood 2004, Kammerer 2008). It is important to note that this process of reducing complexity has consequences: In the case of FRT for example, minorities are easier to recognise. The problem with all algorithmic systems is that the issue of reducing information is a basic requirement, because systems only operate with the binary digits of 1 and 0 (Graham & Wood 2003). It was noted earlier that getting inside the production of these computer codes (that distinguish between one group and another) is becoming more and more important (Lyon 2007b). Moreover, these processes are now the only components "completely open to human discretion and shaping" (Graham & Wood 2003). This is especially important when code is written far removed from the point of application (Lessig 1999) and is therefore in danger of ignoring the specialities and peculiarities at the point of application. The crux of the matter is that coding, especially that relating to classification such as social sorting (Lyon 2007b) never occurs in an objective or neutral way, but is embedded in specific social practices. Bowker and Star (2000) see software in many ways as "frozen organizational and policy discourse", in which policy is coded into software. In this view software, like technology, is 'society made durable' (Latour 1991). This means that these specific social practices, normative notions of good behaviour, political assumptions, and cultural values are either consciously or tacitly inscribed into the software (Graham & Wood 2003). Moreover, "algorithmic systems thus have a strong potential to fix identities as deviant and criminal"—what Norris calls the technological mediation of suspicion (Norris 2002). However it is not only the individual person that could be singled out for attention; in some circumstances coding and classification processes may have profound effects on the shaping and ordering of human life in general, creating new social classes (Lyon 2007b).

A Matter of Selection: The Sociotechnical Construction of Facial Expression Recognition 'Ground Truth'

A relatively new field of research in computer vision is Facial Expression Recognition or, as Gates referred to it, Automated Facial Expression Analysis (Gates 2011: 151ff.). These technologies aim at determining the mood and emotions of a person automatically and in real time. A specific facial expression is related to a specific basic emotion like happiness or anger. First attempts at facial expression recognition emerged in the 1990s. Today research in this area can be found in at least 70 research institutions around the world. The application of facial expression recognition exists especially in two areas: Human-Machine Interaction (HMI) and Video Surveillance (VS). In the case of HMI, machines (robots, computers etc.) are required to be able to detect and interpret human facial expressions automatically. The aim is to improve interaction between humans and machines in general (Ioannou et al. 2007), because it is argued that humans expect machines to behave like humans (Wimmer et al. 2008). Facial Expression Recognition technologies could, for example, be integrated in ticket machines or personal computers so that they recognise when the user becomes frustrated and can then provide help as a result of this recognition.

The second area of application is Video Surveillance. Facial Expression Recognition is intended for workplace monitoring systems, research on the impact of advertisements on consumers in public as well as in private spaces, consumer research (one example is the commercial software FaceReader™[76]) and also in the detection of terrorists, e.g. under the US security program SPOT (Screening Passengers by Observational Techniques) which was introduced at 14 US airports in 2006[77].

Historical Embedding of Facial Expression Recognition

Science has been trying for a long time to make human beings, and especially their bodies, readable. The human face was, and still is, of special interest in this regard. It has been measured not only for identification, but also in the hope of gaining access to the inner workings of humans. This has a long history starting with ideas of the ancient world from Aristotle´s *Historia Animalium* to pre-Confucian China with its face readers and "the study of the systematic correspondence of psychological characteristics to facial features or body structure"[78] known as physiognomy. In the past, physiognomy

[76] According to the producer „…the world's first tool that is capable of automatically analyzing facial expressions“ Noldus Information Technology, http://www.noldus.com/human-behavior-research/products/facereader

[77] The SPOT programm is „a behavior observation and analysis program designed to provide the Transportation Security Administration (TSA) Behavior Detection Officers (BDOs) with a means of identifying persons who pose or may pose potential transportation security risks by focusing on behaviors indicative of high levels of stress, fear, or deception.“ (Homeland Security (2008): Privacy Impact Assessment, p.2); available at http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_tsa_spot.pdf [July 11, 2013]

[78] Encyclopaedia Britannica, http://www.britannica.com/EBchecked/topic/458823/physiognomy.

was always situated between the poles of the sciences and the arts and is today, said to be clearly non-scientific. On the other hand it is firmly grounded in daily life, because we are not able to go through life without being confronted with physiognomy (Schmölders 1997).

In the late 18th century the Swiss founder of what then counted as scientific physiognomy, Johann Caspar Lavater, claimed to be able to recognise the human character in the outlines of the human face. Lavater worked with graphics and illustrations that were produced by different artists. These artists also had the task of standardising and homogenising the heterogeneous visual material for further usage which can be best described with the German term 'Umzeichnen' or redrawing, in the words of Swoboda (2002). The artistic image had to be transformed into a scientific image for further analysis. Lavater also produced graphics of his own; most importantly images of the silhouette, which he produced with the help of a special silhouette backlighting machine ("Schattenrißmaschine"). So in this way objectivity and thus authority was reached by mechanical picture-making (Daston & Gallison 2007). The next step was to produce lines and angles out of the mechanised images that allowed mathematical calculation, classification, and a specific order (Swoboda 2002).

The era following Lavater can be characterised as a pathway leading from physiognomy to facial demeanour and expression, seen especially in Charles Darwin's studies. His book *The Expression of the Emotions in Man and Animals* published in 1872 has to be read in the physiognomic tradition, even though there is a radical move (Person 2005) away from the steady parts of the body and physiognomy (frame and skeleton), to the flexible parts of the body and the face, pathognomy and expression (Brednow 1969). Classical physiognomy was also pursued on a more direct route, particularly in the phrenology of Franz Joseph Gall (cf. Brednow 1969, Schmölders 1997).

Darwin's book *The Expression of the Emotions in Man and Animals*—which was published only four months after his more prominent book *The Descent of Man,* and which had actually only been planned as a chapter of the latter—was revisited by American psychologist Paul Ekman under a 100 years later in the mid 1960s when he started his research on facial expressions and emotions. Ekman and his colleagues created the

Facial Action Coding System (FACS) on which virtually all efforts to recognise facial expressions are based. At the beginning of Ekman's research, the fundamental question was whether facial expressions are universal or specific to each culture. The result arrived at, was that specific facial expressions are linked to a corresponding emotion in every examined culture (Ekman 1988). According to Ekman there are six basic emotions that are expressed in the same way in every culture worldwide: Anger, Disgust, Fear, Happiness, Sadness, and Surprise. However, emotions are not only biologically determined, they are also culturally influenced. For example, there are different display rules in every culture, that are informal norms about when, where, how, and to whom one should express emotions (ibid.). Subsequently, Ekman focused on physiology and especially on facial muscles. In 1978 he, together with Wally Friesen developed a tool for measuring the face—the Facial Action Coding System (FACS)—which was revised in 2002 by Ekman, Friesen and Hager. The FACS is a mode of coding the over 10 000 possible, human, facial expressions and is based on the human anatomy of facial musculature. According to Ekman, FACS today is used "by hundreds of scientists around the world to measure facial movements"(Ekman 2007). In addition to this, "computer scientists are working hard on how to make this measurement automatic and speedy" (ibid.). The aim of FACS was to create a comprehensive system of categories that can be used for defining all muscle movements of the face that are distinguishable with the human eye (Ekman 1988). The movements of the face were summarised into 44 Action Units. With the help of FACS, experts can describe facial movements; these have to be measured and classified, and only then can a specific emotion be interpreted by the human FACS expert[79].

Between expert and lay knowledge

The basis for teaching a machine or a computer to recognise facial expressions automatically is the engineering of a so-called 'ground truth' or 'ground reality' of how a specific facial expression, for example anger, can look. But what does ground truth

[79] cf. Gates 2011: 170 for a description of how people become FACS experts

mean? In the following interview passage a computer scientist working with facial expression recognition explained to me:

> I3a[80]: "…And maybe back to the ground truth question. That is, I said all our algorithms are based on machine learning and for machine learning you supervise machine learning that means that you give the machine example data to train from. So for instance if you want a machine to recognise a specific person then you show the machine images of this person and you tell the machine that this image shows that person. You give the correct answer already in the training phase. If you want to recognise laughing or fear or whatever, you show the machine images of laughing or afraid persons and you tell the machine these images show laughing or afraid persons. And so the machine can recognise it later. But in the training phase this information has to be given and this is called ground truth."

The ground truth, as one might expect, does not exist from the beginning, but has to be constructed. The machine has to learn from the computer scientist first. The computer scientist teaches the machine what ground truth looks like, for example what the facial expression of fear looks like by referring to images that are supposed to show fearful people. This means it is the computer scientist who defines how fearful people look by referring to images displaying fearful people. In another passage, the comparison of machine learning and human learning is quoted as an example:

> I3a: "… but it's pretty much to human learning. If you learn vocabulary then you have been given vocabulary. You have to match the German word to the English word. If you don't know the vocabulary and you hear the English word, you know the German word, you don't have to see it anymore. But during learning, of course, you have to match it. That's what the machine does."

As is argued in the interview passage, two things that are supposed to mean the same have to be matched. The assumption is, just as a German word has an equivalent in

[80] Quotations from the Facial Expression Recognition interviews mentioned in Chapter One are coded with I (for Interview) and the number of the Interview (1-5) at the beginning of the interview passage in bold letters. Interview 3 (I3) was held with two computer scientists, which are marked I3a and I3b.

English and vice versa, an emotion, e.g. fear, has an equivalent in a bodily expression shown in the human face, and displayed on a digital image. But what does this equivalent look like? Who tells the machine which specific facial expression corresponds to which emotion? In my interview data two different approaches to this question were found. One approach is what I call the 'FACS expert approach' and the other what I call the 'FACS non-expert' approach.

a) Facial Action Coding System (FACS) expert approach:

> I3a: "Cohn-Kanade. That's a really standard database. Many, many people are working with that."
>
> I3b: "These databases are generated on the basis of ground realities. Cohn-Kanade facial expressions database is connected with the Facial Action Coding System."
>
> INTERVIEWER: "What does ground reality mean?"
>
> I3b: "For facial expressions there is a full coding of a face. That if you move your one eye up and if you are smiling your lips go up. Databases are generated by persons sitting in front of a camera."
>
> INTERVIEWER: "But people are told to make facial expressions?"
>
> I3a: "In Cohn-Kanade they are advised to give an extreme facial expression. So if they have to smile, in the first streams they are neutral and at the end they are really smiling. So they generate a smile, it is not natural like when I am talking with you, but they are really forced to laugh."
>
> INTERVIEWER: "Is any expert checking these expressions? Like in an experiment if somebody tells me to smile and is there anybody who says, yes, that's a correct smile or too fake?"
>
> I3a: "It depends on the database. In the Cohn-Kanade database it is like that. There is an annotation which tells you in which ways the person is smiling and it has been annotated by an expert giving evidence."

INTERVIEWER: "And do you know more about these experts, from which profession do they come? Are they psychologists?"

I3a: "Usually they are FACS experts, usually they annotate such data."

In the FACS expert approach above, the ground truth of facial expression is constructed using experimental, artificial image data that has to be annotated by experts. In order to produce training images for the computer, people who are not computer experts are asked to give an extreme facial expression in a laboratory setting. Even if this is not actually a "natural" expression but one that was asked for, FACS experts annotate these facial expressions that are said to be caused naturally and biologically (Ekman 2007) in exactly this way. Thus, in this approach a ground truth is constructed by a FACS expert together with any untrained person using their dramatic abilities in showing "natural" facial expressions. The resulting training images are then used to illustrate the given facial expression, and FACS experts interpret them in a laboratory. What counts as an emotion, e.g. fear, is thus a coproduct of "artificial" facial expression and expertise based on biological and natural facial expression.

b) FACS non-expert approach

In this approach, the training images used do not come from a FACS database, but rather from several other image collections:[81]

I4[82]: "It is a mixed collection of images from very different sources. It starts with any database, progresses with pictures we took ourselves, collected from the internet and ends with pictures from the TV..."

---

[81] The following interview passage was originally in German (see below), English translation by author

[82] (Original Quotation/Translation by author) "Das ist eine Mischsammlung aus allen möglichen Quellen, die man so auftut. Und das fängt an bei irgendwelchen Datenbanken, geht weiter bei Bildern, die wir selbst fotografiert haben, das sind Bilder, die wir aus dem Internet sammeln und endet bei Bildern, die wir aus dem Fernsehen gewinnen."

> I4[83]: "…Our procedure for recognition of the training library is a fully untrained approach. This means that these people are not trained, they are just like you and me, annotating on the basis of their practical experience."

The second possibility I came across in my interview data for constructing a ground truth for facial expression is the approach of someone who has nothing to do with FACS and operates with practical and tacit knowledge. The individuals annotating the data; for example stating when they recognise fear in a specific training image, are computer scientists and students. They have no special training in facial expression recognition. This ground truth is based on a library of images from many different sources. The aim of the image library is to have a variety of different origin and not to have images that were produced in laboratories under specific conditions, annotated by FACS experts. The computer scientist interviewed explains the rough estimation of facial expression data with the absence of FACS experts in the annotation process. On the one hand in this case, real-time ability is more important than more exact results and therefore the system has to work with simpler features. The essentiality of real-time processing in this system can, on the other hand be interpreted as a demonstration of the "naturalness" of the system. The argument was that in the need for recognition to be as fast as human recognition, exact (FACS) expert knowledge could actually slow down this fast real-time processing within the system.

What one can see in these two different approaches ('FACS expert approach' and 'FACS non-expert approach') is that the sociotechnical construction of the Facial Expression Recognition ground truth is a matter of selection and not a natural given. There is no "truth" from which to start, but a truth that has to be negotiated and created by humans, their selections and interpretations, especially their selection of images and with what kind of coding schemes (cf. Goodwin 1994: 608) these are interpreted in the manual annotation process. The two different approaches towards constructing a ground truth show that it is necessary to test and choose what kind of knowledge is

[83] (Original Quotation/Translation by author) "Bei uns und bei dem Verfahren, wie wir diese Bibliothek trainiert haben ist ein völlig laienhafter Zugang, d.h. die Leute sind nicht trainiert, das sind einfach nur Menschen wie du und ich, die einfach aus ihrer Erfahrung heraus die Annotation durchführen."

"workable" in the transformation process and in what way it should be processed and codified. In contrast to this ambiguity, the term ground truth itself refers to the longing for some kind of universal truth found in the human face, in order to recognise facial expressions, human emotions, and in further consequence also in the biological body in general. But, one might ask, who is right or what approach is ultimately better? Are there differences between the two approaches once deployed? What approach is more accurate or viable? Is it the knowledge processed in the (FACS) expert approach that presumes to know exactly where to look in order to be able to recognise a real emotion? Or is it the practical everyday knowledge and common sense that is used to recognise emotions in facial expressions in everyday life? Without being able to answer these questions properly here due to the lack of empirical data, a preliminary answer might be: it depends. It depends on what the concrete aim of an automated facial recognition process is (e.g., rough estimation or more exact results) and it also depends on where and in what way it is going to be used and applied. Of course, it also depends on how humans evaluate the results. Here, Kelly Gates points to the problem of "intercoder agreement" (Gates 2011: 170):

> "Maximizing the level of agreement among human coders concerning the appearance, timing, and intensity of facial actions becomes a means of establishing the accuracy or 'ground truth' of facial coding for AFEA algorithm development." (Gates 2011: 171)

This means that once an evaluation of the different approaches outlined and the resulting algorithms for facial expression recognition can be implemented in order to analyse how well they work and if they work in different ways, the crucial point is that it depends on "consistent results no matter who is doing the coding" that give the method "…credibility as an accurate measure of reality." (Gates 2011: 171). Thus, if human coders (the ones who previously annotated the facial expressions) doing the evaluation come to an agreement that the tested algorithm was accurate in any random case, then the algorithm is accurate no matter what the person being analysed showing a specific facial expression says. Therefore it is a "closed loop" (ibid.) system that works "in terms of the system's internal agreement—that is, whether the classification system consistently agrees with itself" (ibid.).

A Matter of Boundary Work: The Sociotechnical Construction of the 'Ground Truth' of Suspicious Behaviour

The second case study I am referring to, dealing with 'Automated multi-camera event recognition for the prevention of bank robberies' could not draw on existing expert knowledge like the FACS framework for Facial Expression Recognition, but had to generate new empirical data itself, using interviews, observation, and document analysis. These methods were already mentioned in Chapter One in more detail. In the following section, the sociotechnical construction of a ground truth for the suspicious behaviour of prospective bank robbers reconnoitering Austrian bank branches and the impossibility of such a task is described.

Between tacit and explicit knowledge

As was already referred to in Chapter Two, Norris and Armstrong (1999) demonstrated how CCTV control room operators construct categories of suspicion in social interactions. One insight of their influential study was that "CCTV operators bring with them taken-for-granted assumptions about the distribution of criminality within the population" (ibid.: 118). This entails that the selection and differentiation of people observed "is not based on objective behavioural and individualised criteria, but merely on being categorised as part of a particular social group" (ibid.: 150). It is clear that such practices of selecting people are of a discriminatory nature. With semantic Image Processing Algorithms that have been, amongst other things, developed to automatise observation of video surveillance monitors, the process of coding and programming, and therefore computer scientists themselves seem to act in place of CCTV control room operators in constructing categories of suspicion, and potentially, also in discriminating between people. In this specific case, it is not only the computer scientist constructing what might count as suspicion—and in the same way normality—but also others that are involved in the development and implementation process. Other relevant actors involved were security experts of a commercial bank and the police, who were seen simultaneously as possible end-users of efficient and cost-effective solutions for issues of security and public order. A commercial vendor and software consulting company for computer vision systems was also involved in the project that had the economic interest of developing and marketing a relevant technical system in the form of a ready-made product. Another contributor was a group of social scientists to that I belonged to. The contribution of the social scientists to the project tended to be twofold: on the one hand the collection of data for generating a ground truth was what the computer scientists asked for, on the other hand the process of generating the respective ground truth was critically studied. For the social scientists the project was methodologically challenging as their role was far from being obvious at the beginning of the project. This role could be described as ranging from 'figleaf' or supplementary, to being a fully integrated and critically reflecting partner of the technological development.

The afore-mentioned contributors were all part of a project consortium within the framework of the Austrian Security Research Program KIRAS[84]. In this programme projects developing security technology are obliged to integrate a partner from the Social Sciences and Humanities in order to ensure "socio-political compatibility". This aim was also reached by integrating other relevant contributors to the development process: both civil rights campaigners and the National Data Protection Agency, as well as employees of a commercial bank working in the public areas of bank branches. Finally, non-human actors such as technical equipment (e.g. cameras, computer hardware) and the media were also considered part of the project.

What does 'Suspicious Behaviour' Look Like?

At the beginning of the project both the security experts from the bank and the police security experts strongly put forward their point of view that it is possible to uncover potential bank robbers in their process of reconnoitring bank branches when deciding which branch they will rob. Furthermore they suggested that potential robbers will desist from robbing one particular branch if they are addressed in person, in or in front of the respective branch. In their view, automatic detection of such exploratory

[84] KIRAS (acronym from the Greek kirkos for circle and asphaleia for security) supports national research projects which aim to increase the security of Austria and its people. The protection of critical infrastructure was selected as the first thematic focus. The programme started in 2005 and is scheduled for a duration of 9 years. KIRAS is an initiative of the Federal Ministry of Transport, Innovation and Technology (BMVIT) managed by the Austrian Research Promotion Agency FFG. For more information see http://www.ffg.at/en/kiras-security-research

behaviour could assist in averting anticipated criminal action. Therefore, further procedure for the social scientists in the project, was to interview these security experts to make use of their described "gut feeling" (or implicit knowledge) in order to help design an applicable system that was planned to be able to detect suspicious behaviour in bank branches, automatically.

The outcome of the interviews was the definition of a set of suspicious behaviour terms assembled in collaboration with all project partners. This set included the action of lingering in the bank foyer without using a machine (e.g. the ATM), or of interacting with a member of staff over an extended period of time, or also of staying at a machine for an unusually long period of time. However, it was not possible to gather more accurate knowledge about the actual behaviour of bank robbers reconnoitring locations. In addition, there were also no sequences of images available, showing bank robbers reconnoitring locations that could help to gain more accurate knowledge. It quickly became clear to the social scientist that there seemed to be a limit to the specification of suspicious behaviour in the context of a bank branch, because the assumptions of the security experts about suspicious behaviour were rather abstract, at least in terms of being able to inscribe these assumptions into IPAs. Thus, the determined criteria remained questionable regarding their relevance for implementing an effective and applicable automated system.

Nevertheless, until this point in the project several selections and decisions had been made that were the basis for proceeding. This had been done on the one hand, on the basis of the security experts' interest in preventing bank robberies and their assumptions that loitering and interaction/interactivity are potential indicators of suspicious behaviour; on the other hand this was because computer scientists had indicated that these criteria seemed to be technologically feasible. That meant, in theory that it is technologically possible to automatically track individual persons throughout a bank while also measuring the time they are inside the branch. It also seemed possible that this tracking and measuring of the loitering time is done for specific areas only, for example indicating the use of a specific machine or waiting in a queue. From the point of view of the participating computer scientists the predefined set was very welcome, because retrospectively I learned that they were especially interested in working on the basic computer vision problem of "multi camera tracking" of persons and objects.

In a further step the behaviour of people visiting a bank branch was observed and analysed by the social scientists to learn about the "usual" or "normal" behaviour of bank customers. These observations showed that the average time people stay in the bank foyer (where ATMs, bank statement printers and bank counters can be found) was 03min 08sec (median 01min 58sec). 38 of the total sum of 236 people (16%) observed stayed longer than 5 minutes and 10 out of 236 (4%) stayed longer than 10 minutes. This means, following the observations of the social scientists that the outlier percentage concerning attendance time of bank clients was rather high. What followed from these insights was the finding that a simple detection of attendance time would not make sense in the everyday practice of the bank branch. There was still also no evidence at all indicating that longer attendance times are unusual or even constitute suspicious behaviour. In fact, all longer attendance times of clients observed could be explained by the social scientists. These longer attendance times appeared to be usual behaviour, at least non suspicious behaviour: for example, many bank clients had to wait in line in front of the ATM machine or bank counter, others had problems with a machine, requiring the assistance of a bank employee and some just took their time filling in a transfer form. That means, there was a considerable number of people that stayed for "longer" in the bank branch than the average majority. Thus, even if there had been evidence that longer attendance times indicated unusual or suspicious behaviour (what was not the case), the detection of these longer attendance times would also have brought to the fore, the clients that stayed longer in the bank branch than the average, although they had just acted in an absolutely normal way.

As a consequence of these insights, people just staying in the bank foyer over an extended period of time without using a machine or without interacting with a member of staff were also interesting to observe and examine. 17 out of 236 (7%) people observed, did not exhibit any typical bank branch activity; almost 50% of these accompanied somebody performing usual activity. This means that also the behaviour of about half of the inactive people could easily be explained by the human observer. The other half of the people observed that did not exhibit any usual activity and were inside the bank foyer only for a very short period of time. Most of these clients just glimpsed inside, observed the long queue and left. As such, also this “inactive” behaviour can be regarded as normal practice in the context of a bank branch instead of being suspicious reconnoitring.

One main conclusion of the observation was that usual behaviour in bank foyers is very diverse, although the specific context of a bank determines expected human behaviour to a considerable extent. There was a great range of different behaviour patterns to observe, making the detection of unusual or suspicious behaviour difficult. As a consequence, we discussed detecting those clients showing a specific deviation from the average attendance time. This was however, questionable, as there was no evidence that they were actually exhibiting suspicious behaviour. Additionally, there even was information that many bank robbers reconnoitring a bank branch behave like perfectly ordinary customers or are in fact themselves, bank customers. Then, in the case of detecting only those with specific deviation from the average, the usual might become the norm. In the case where a system is in operation this may have serious consequences for those being watched. The pressure to conform to ‘normal’ behaviour may increase for those entering a bank foyer, and if they do not behave faultlessly they might provoke adverse consequences. Meaning that those simply diverging from the norm would attract attention instead of possible real bank robbers that are more likely to behave in an ordinary way.

Ultimately one must also consider that (automated) detection of the suspicious behaviour of potential bank robbers is like finding a needle in a haystack. There were estimates made for this project that in the 512 banch branches in Vienna there were 70 million people entering and leaving in the course of one year. In contrast, statistics show that there were 63 bank robberies in Vienna in 2008. The probability of picking out just one of these 63 cases out of 70 million is very low, especially when considering that a distinction between normal and suspicious behaviour in the context of bank branches is extremely difficult for human observers and even more difficult for observing computers. To sum up; within the project it was not possible to find a satisfactory way of constructing a ground truth for suspicious behaviour that could be effectively deployed within a surveillance system in the context of a bank branch.

A Matter of Experimental Imagination: The Sociotechnical Construction of a Fall Detection 'Ground Truth'

The third "domain of scrutiny" (cf. Goodwin 1994) I am dealing with, is the sociotechnical construction of a fall detection ground truth. This case stresses the experimental character of sociotechnical ground truth construction processes as it happens within the framework of computational scripts and the image computer scientists have of possible places, users and their respective (fall) behaviour. As I mentioned before, the ground truth has to be constructed for the purpose of comparing and matching the observed scene in the field of application to the ground truth template in order to recognise the relevant event in the respective sequence of images available. Recognition does always refer to something that already exists to which a current phenomenon can be confronted and compared. So, in the case of fall detection the question is what version of the phenomenon of 'falling' is the basis for the detection of these falls? What criteria are used and established, formalised and standardised? Thus, in the end, the formulation of the fall detection ground truth template does also define what it means to fall. Also in reference to the social construction of Facial Expression Recognition ground truth and the insight that it is a matter of selection, it has to be added here that different versions of constructing a ground truth of fall detection may exist. What follows refers to my two months field work, participating and observing in a computer vision laboratory (see methods section in Chapter One for a detailed description) in which I more or less accidentally (there was no strategy planned) stumbled upon the case of fall detection during my stay there. This was, on the one hand a matter of the accessibility of this case, as the responsible researcher encouraged me and on the other hand, fall detection attracted my attention due to its relevance to behaviour pattern recognition.

A 'significant fall' in this specific context of fall detection meant that the falls to be detected occurred within a specifically situated application field, namely in the framework of «Ambient Assisted Living». The concept of «Ambient Assisted Living» is understood by the «Ambient Assisted Living Joint Programm», a very large European funding activity, as a way[85]:

- to extend the time people can live in their preferred environment by increasing their autonomy, self-confidence and mobility;
- to support maintaining health and functional capability of the elderly individuals,
- to promote a better and healthier lifestyle for individuals at risk;
- to enhance the security, to prevent social isolation and to support maintaining the multifunctional network around the individual;
- to support carers, families and care organisations;
- to increase the efficiency and productivity of used resources in the ageing societies.

Fall detection in the context of AAL is especially directed at elderly people and their homes. The hope is that automatic fall detection can reduce fears and support elderly in living in their preferred living environments. Fall detection developed in the context of AAL also has consequences for the universality of an imagined future. As it is designed with just this context in mind, it is questionable whether the system could also be easily implemented in other application areas such as in public spaces or directed at different age or activity groups, for example. The setting also had consequences for the hardware used and researched in this specific, during my observations. In what I experienced, the chosen computer hardware scripted the research process to a considerable extent and thus, was a constituting element in inscribing social order to the respective Image Processing Algorithm in the making. In what follows, I shall describe the experimental character of this sociotechnical ground truth construction of fall detection as it happens within the framework of computational scripts and imagined users, places and behaviour.

[85] See http://www.aal-europe.eu/about/objectives/ [April 24, 2014]

Between the Computational Script and Imagined Users, Places and Behaviour

During the first few days of my field work in a university computer vision laboratory in Austria, I encountered what was referred to as 'fall detection' (or as it is called in German, 'Sturzerkennung'). There was one young researcher at PhD level in particular who worked under the supervision of a senior researcher on the issue of fall detection. By keeping company with Jonas, which is how I shall name the young researcher, I learned continuously about fall detection and in my view, became significantly involved in the research process over the course of about two months.

The Computational Script of the Xbox Kinect Sensor

One of the first occasions at which I became aquainted with fall detection in the laboratory was during an experiment. For this experiment that took place in one of the three major rooms of the lab, Jonas had brought an old foam rubber mattress and placed it on the floor just in front of a Microsoft Xbox Kinect sensor. The Kinect device was already standing on a permanent shelf affixed to the wall. It was then connected to his laptop placed on a desk close to both mattress and Kinect. He turned the screen of the laptop in the direction of the mattress. He then simulated a fall directly onto the mattress and stayed lying there for a while. On the laptop screen I could see the shape of a body that from my point of view showed Jonas´ body and its movements. The body shape on the laptop screen was filled-in white, but was later changed to filled-in blue. In addition to this shape there was also another visual representation of Jonas´ body that I noted down in my field notes as a "simple pattern of lines following the major body parts". Later, I learned that this visual representation of the body was referred to by Jonas and the other researchers as a "skeleton" because of its resemblance to a human skeleton. As I saw Jonas fall onto the mattress, almost simultaneously the visual representation of his body on the laptop screen also fell. This impression of Jonas´ body on the laptop screen was more a movement (or sequence of single images) of the representation of his body from a more vertical to a more horizontal position. However, as he had obviously fallen down, I was able to also recognise the change in the orientation of his visual body shape depiction as something that looked very much like falling down. At that particular moment it was clear to me that what I had observed on the laptop screen was a digital pictorial representation of the act of falling down. It would be interesting to know if I would have recognised this same act of falling down on the laptop screen if I had not seen the original, live falling down of Jonas at that same moment or slightly previously.

A little later Jonas got up again, took his laptop, disconnected it from the Kinect sensor, and returned to his desk in the same room. He had recorded the scene in order to ascertain whether Kinect had recognised the fall. Or more concretely, if Kinect had recognised the positional change of the skeleton on the recorded sequence of images that then could be used to determine the fall. Soon after, while watching the recording of the scene, Jonas realised that the skeleton did not fall down, meaning that it had not worked the way he had intended.

So to my mind, Jonas prime interest in the experiment was to explore how the hardware used, the Microsoft Xbox Kinect sensor works, in order to see if it was suitable for his central concern of automated fall detection. It was crucial to understand what was possible with the hardware at hand and how it would react to the specific "domain of scrutiny" of falling. In general from my experience, a great part of the work in the computer vision laboratory was understanding what the computer or the hardware in use is capable of. Why had it acted in a particular way and how could this be made available to the researcher in his intention of detecting falls. In this first instance, the young researcher Jonas had wanted to find out just how well the Kinect sensor and the visual representation of what it records and depicts on the laptop screen represents what he had done shortly before; namely falling down or at least what he understood falling down to be. As soon as he had realised that it simply had not reacted in the way he had wanted it to, that is the skeleton did not change its orientation from a more vertical to a more horizontal position, he started looking for alternative ways. He had tried to explore how he could have made use of the specific script of Kinect, in order to detect falls with it.

On that same day, quite late in the afternoon Jonas worked again on the issue of fall detection. I sat next to him at his desk and observed how he browsed chronologically through single frames of the video that he had recorded in the experiment before. I was able to see on the computer screen the coloured shape of a person that with all probability, was Jonas himself standing in an upright position. What was new to the depiction in comparison to the one I had observed before, was an almost vertical straight line connecting particular points on the body depicted. As I got to know later, Jonas had been using the skeleton data to find another way of representing the body and did so by calculating a straight line out of the available skeleton data. That means, what I had observed were several bodily transformations: first of all from Jonas himself, to the visual filled-in shape depiction prescribed by the Kinect sensor and then from this shape to the skeleton representation; and finally, from the skeleton representation to the depiction of connected points in a straight line that had been programmed by Jonas. This transformation process proved promising as at the end of the day, it had succeeded. At the end of the single frame sequence both the shape of the body and the straight line changed from a more vertical to a more horizontal position. What had happened was a concurrence of several transformation levels, all present or potentially present on the laptop screen. That meant in the words of the researcher that the specific selected points of the skeleton that were connected to the straight line followed the visual representation of the body[86]. Jonas had recognised the problem of nonconformity of the different levels and had been able to resolve it by converting different image solutions. However, as I then discovered this success did not mean that falls could now be recognised with the help of the Kinect sensor. It had just been an important step in the right direction.

The next problem needing to be solved was of the floor of the room and its correct IPA recognition, especially as the background to the scene had been irregular. That meant that it was not one uniform area, for example a completely grey area but instead, consisted of many different parts; in this case all kinds of laboratory furniture and equipment. Again, it was part of the work of the researcher to find out why the computer at hand performed in the way it did, and how this behaviour could be changed to go in a desired direction. This depended to a large degree on the ingenuity and creativity of the researcher and how he was able to “domesticate” or encode the Kinect

[86] in German original: “Die Skelettpunkte gehen jetzt mit dem Körper mit.“

sensor and its behaviour. In doing so, he depended very much on the computational script that he followed throughout this process. At this point it would be legitimate to ask why the researcher did not actually use custom-made hardware and software and why he chose the Xbox Kinect sensor instead. He could have designed hard and software specially for his fall detection case. This task is usually however, not the work of a computer vision researcher and additionally, economic and thus also application criteria played a decisive role in the basic research I was able to follow and observe. Generally speaking, it was often said in the laboratory that they needed to work with keenly priced, affordable hardware in order to keep down future prices. In many other projects too, especially those in the field of video surveillance, they tried to use off-the-shelf consumer or network cameras and explored all possibilities for using these cheaper hardware products that already existed. Highly specialised, expensive cameras were mostly not considered applicable as these would add to future costs and thus, would not be used. That meant that the work in the university computer vision laboratory I had been permitted to observe, was highly orientated towards future application, in particular to the consideration that expensive systems would only stand a slight chance of being sold and used. In this regard, the well-known mass product Microsoft Xbox Kinect was regarded as highly affordable[87] and thus, highly suited to research into fall detection[88]. Next to the relatively low price the Kinect sensor also had the advantage of delivering 3D information using a standard 2D camera and thus, so-called 'depth data',

[87] For example, Amazon.de Price for Xbox 360 Kinect sensor: €89, or, Kinect sensor for Windows Price: €198 (July 2013)

[88] During my fieldwork I also discovered that Microsoft has an interest in researchers and companies exploring and using the possibilities of the Kinect sensor. On the Kinect Website, Microsoft explains this: "Kinect for Windows gives computers eyes, ears, and the capacity to use them. With Kinect for Windows, thousands of businesses and developers are creating applications that put people first—allowing their customers to interact naturally with computers by simply gesturing and speaking." http://www.microsoft.com/en-us/kinectforwindows/ [July 4, 2013]. For this purpose, Microsoft offers for free a so-called SDK (Software Development Kit) to download on their website. The SDK enables developers to use different programming languages "to create applications that support gesture and voice recognition by using the Kinect for Windows sensor and a computer or embedded device." http://www.microsoft.com/en-us/kinectforwindows/Develop/developer-downloads.aspx [July 4, 2013]

meaning that depth of field made it possible to estimate distances. It was also privacy enhancing, as with depth data images individual persons cannot be easily recognised as only a body shape is depicted. To summarise, the Kinect sensor promised several advantages for the researcher but in order to benefit from these, it was necessary to first understand and then “de-script” (Akrich 1992) the specific script of the Kinect sensor. So all that followed in the research process took place within the framework of the Kinect sensor and its SDK (Software Development Kit) script. As such, the existing product Microsoft Xbox Kinect and its configuration was “inscribed” into the ground truth of fall detection from the very beginning.

Imagined Users, Places and Behaviour

Following the process of getting to know how the hardware-software package available (Kinect for Windows) worked, the PhD researcher Jonas continued his experimentation. The most important part of this research process was to differentiate between a fall and other movement that was not a fall. That meant, to decide and subsequently teach the computer if and when there is a proper fall. As I wrote before, the most relevant question in this context for me was what version of the phenomenon of “falling” was the basis for detecting these falls. Which criteria were used and established, formalised and standardised in this specific setting?

Once Jonas had recorded his training data (the sequence of images that showed his fall) most of the following work was carried out at his computer using as a basis, the different forms of manipulated and transformed images. Conceptually it all went in the direction of a mathematical equation: Jonas strategy that had emerged from his experimentation process was to find a way of detecting parallels between the plane of the floor and the straight line depicting the human body which again, was a transformative product based on the skeleton representation available in Kinect SDK. Later I discovered that Jonas had programmed the straight line by selecting four or five prominent points on the body viz. the head, the nape of the neck, the middle of the hip, and the knees. For a while he worked on establishing a match between the four or five points on the body and the straight line he had already programmed.

His aim was to to formulate an equation describing the relationship between the plane of the floor and the straight line of the body. Jonas assumed that the more similar the orientation of these two elements was, the more likely it was that a fall had occurred. That meant as I understood it, that the Image Processing Algorithm to be programmed was mainly thought of as an equation that recognised when two geometric elements that represented the floor of a room and the body of a human in a three dimensional space were parallel. In his first attempt at fall detection, Jonas measured the distance between the head point, meaning the very top of the body, and the floor and his assumption was that if this distance was virtually zero there had been a fall. However first tests produced results that were not robust enough, so Jonas rethought his approach to the straight line formed out of the body points mentioned before. It also seemed unworkable to make use of the shape depiction or of the skeleton depiction for the equation. Both were not stable enough, meaning that both were unsuitable for using in an equation and additionally, the visual depictions did not follow the also visually represented, real falling body precisely. For example, as I was able to observe several times on the laptop screen while Jonas was falling down, the form of the skeleton depicted imploded at that very moment. That is, it lost its form as a skeleton in an uncontrolled, inexplicable way instead of following the path of the real Jonas falling down from a more vertical to a more horizontal position.

Once Jonas had formulated the equation describing the angle between the two geometric elements that represented the floor of the room and the body of a human and whether they are parallel or not, he started with further experiments. He did so in the same way he had before with the experiment consisting of the mattress on the lab floor, the Kinect sensor standing on a permanent shelf on the wall, and a laptop standing on the desk close by. Jonas started the recording and then walked around the mattress in a circle. After a while, he fell down directly onto the mattress. He remained lying for some seconds and then stood up again. When Jonas fell down onto the mattress, a female colleague of his I shall name Johanna, laughed aloud. I could not find out her reasons for this, but at least it was some kind of feedback about the fall. Once Jonas was up again he asked Johanna if she could also fall down for him, because she was a little smaller than him and it was an opportunity to test the system under slightly different conditions. So Johanna also fell down. Before falling, she raised her arms above her head in front of the Kinect sensor in a "hands up" way in what Jonas called the "Y-Position". Jonas had also used this movement as a method of calibration before, because the Kinect sensor needs to detect the body of the individual subject in order to perform. Once her body had been detected and was being tracked correctly by the Kinect, she also walked around the mattress and fell down.

Following this, Jonas brought to the site of the experiment three large, black, hard plastic boxes, usually used for transporting equipment and stringed the boxes together. My estimation was that the boxes were about 40cm high and in length and tied together, an adult like Jonas was able to lie on them. He then continued to make something like a bed, placing the mattress on top of the three boxes. After that, he walked around the bed, raised his arms in the Y-position in order to calibrate the Kinect and lay down on the bed. Meanwhile, Jonas´ experiment using the bed had raised the attention of some of his colleagues from another lab room. As they watched the scene, they got in his way and he jokingly told his colleagues that he could not work under these circumstances. Two of them imitated the Y-position and were asked not to, as the Y-position only works with the subject being tested. As they watched, Jonas prepared to fall again, this time putting the mattress just behind the boxes (seen from the Kinect perspective) onto the floor. He fell again. His colleague Oskar commented the fall with, "You didn't fall very realistically!" said in a snippy way. Jonas reacted to this comment by inviting Oskar to show him what a realistic fall looked like, adding the term "simply spontaneous". After the calibration (Y-position of the arms) Oskar fell down and asked his colleagues if his fall had been better. They all answered promptly with a loud and clear, "No!"

I found the observation of this scene of high revelance as it was clear evidence that the researchers needed to discuss and negotiate what it really meant to fall down realistically. Even if it had all been conducted in a jokey way, it had shown the difficulty of first, defining what realistic falls are, seen from the human perspective and how they look and second, how to authentically imitate such realistic falls in the lab in order to teach the computer what a realistic fall looks like and how it can be recognised. At this moment of time in the research process, Jonas and his colleagues were not referring to the falling of elderly people in their homes, they were just playing around with their own ideas of what falling down means and looks like. In my view, there were two reasons for this. First, Jonas was just at the beginning of his research and was looking for a basic way to detect falls, that in his case was to detect the difference between walking or standing in an upright position and lying on the floor. This was also his reason for including the situation of someone lying on a bed, temporarily constructed by stringing together the three black boxes with an estimated height of 40cm. So the boxes and lying on them was a test for the imagined case in which the test person did not fall down, but was simply lying on something like a bed or a couch. The assumption here was that there is critical falling, represented by falling down onto the floor, as well as uncritical "falling" represented by falling down or better, lying down on a bed or a couch. Here the pertinent question was whether difference in height between lying on a bed and lying on the floor could be detected with the help of the available hardware-software package. A meaningful and as such detectable fall was therefore defined exclusively, as a fall onto the floor. For the mathematical equation these insights meant that a meaningul fall occurred not only if the floor and the straight body line were parallel, but also if the difference between floor and straight line were as close as possible to zero.

What was missing at this point in the research process was an appropriate threshold value for the grade of parallelism between floor area and straight body line and what height difference between these two elements would constitute a meaningful, critical fall. Following this, Jonas tried to find out how his equation correlated with the experimental data he had recorded before. Once he was able to understand the results presented in numbers, he would also be able to define thresholds for the specific point of time at which a fall had occurred. So at this point in his research, the setting of threshold values was still open and none had yet been fixed. During the course of my field observation this changed temporarily when the researchers demonstrated their fall detection system to the wider public in a presentation. Exactly this demonstration will be of central concern in the following chapter.

Later in the research process, Jonas evaluated his experimental data. For this purpose he first had a look at two different images on his computer screen. In one, a specific area predominantly in the lower half, was coloured blue, while the same area on the other image was red. However these coloured areas were not clearly demarcated from the other areas and there were also some uncoloured parts in the lower half and some parts in the top half that were coloured blue or red. I discovered that the coloured areas represented the floor, but because of occlusions some areas in the lower half were not recognised as such. Also, some areas in the non floor area in the top half were depicted as being part of the floor as they probably looked similar to this to the Kinect sensor. I also learned that an empty room with just floor and walls would be a best case scenario for computer vision, while a room full of objects would be the worst case.

At this point in the programming process, the real living environments of elderly people were not being considered at all, but later within the project, I was told they were planning some field testing in real living situations. Concerning this issue I will also go into more detail in the next chapter, because these questions are related to questions about the functioning of the system. In the laboratory, the experiments had not taken place in an empty room, but the Kinect sensor had been focused exactly on the mattress that had been placed in an exposed area of the lab. Location had not been an issue at that time. What had been an issue were the falls themselves in the sense that the central concern was to detect the difference between what was seen as a real or realistic fall by the researchers and everything else that was defined as not being a fall. In the end I got the impression, the researchers' ideas and the practical and experimental realisation of these ideas of what counts as a realistic, meaningful fall was clearly inscribed into the ground truth of fall detection. As the researchers did not have images or video sequences of elderly people actually falling and as it was also not possible to allow elderly people to fall under experimental conditions in the lab, it had become clear to me that this thought up, contrived manner of experimenting had been the obvious mode of procedure from the perspective of the researcher.

I too became a popular experimental "faller" in the course of the days I spent at the lab. Together with Jonas I tried out all variations of falling. Some of these, Jonas had found in a paper that had been used for another project and he compiled a checklist of which falls had either been detected correctly or had been false positive. For example, sneezing where the upper body suddenly moved forward, or just sitting down on a chair were correctly detected by the algorithm as not being falls. At that time, lying down on the bed was detected as being a fall. Jonas explained this case with the incorrect setting of the threshold value describing the angle between the depiction of the subject and the plane of the floor, and said he would see how he could alter this. But, and this was one of many uncertainties at this time, he was not sure what consequences this change in threshold value would have on the detection of other types of fall. Something else that caused troubles at this stage were falls that happened in the direction of the Kinect sensor. In such cases, Jonas explained the missing detection with the behaviour of the straight line representing the human body as it did not move in a way the equation could recognise and so the triggering threshold value was not reached. So, for the human observer it was obvious that falling towards the Kinect sensor was also a fall, but the hardware-software package did not see it this way.

In conclusion, there were several cases in which the algorithm did not detect falls or did detect falls where there were none. It was then Jonas´ task to find out why the algorithm he had programmed within the framework of the Kinect sensor script had not behaved as predicted. When I left the lab there were open questions and a general feeling of uncertainty about whether fall detection would really work in authentic environments with actual elderly people falling. This question of viability will be central to the following, Chapter Six.

The Sociotechnical Construction of the Ground Truth

This chapter has explored IPAs operating on a semantic level in the area of pattern recognition in the making, using three case studies. This empirical analysis of IPAs allows reflection on how everyday patterns of seeing and recognising are interwoven, configured and reconfigured in the 'making of' IPAs. Of central concern was what kind of social order is inscribed into these and in what ways. The central site where these inscription processes are to be found is the sociotechnical construction of the ground truth. This IPA ground truth is a crucial and constituting societal element I considered very closely in this chapter.

The first case study 'Facial Expression Recognition' showed that there are different ways of constructing and building a ground truth of specific facial expressions. At the same time, depending on what approach was chosen, what counts as a specific facial expression and simultaneously as a specific emotion, for example happiness, is defined. Thus, the ground truth of facial expressions and emotions is a matter of selection and it is not clear what the "right" or "better" approach is. As I found in my interviews with computer scientists both approaches had different goals: the first strove for precision in the way it determined facial expression while the second one pursued speed and real-time ability for the purpose of a better user experience. One reason for this setting of priorities was that the first approach operated with "external" expert knowledge that predefined how the specific facial expressions were supposed to look on a biological/natural basis while the second was based on the "internal" everyday, common sense knowledge of the computer scientists that defined how facial expressions look in the course of annotating the training data in the computer vision laboratory.

The second case study of 'Automated multi-camera event recognition for the prevention of bank robberies' expressly underlined the importance of context information to the understanding of specific situations. In the interdisciplinary research project I participated in, it was not possible to define clear categories that represented the "suspicious" or "normal" behaviour of reconnoitring bank robbers. This may have stemmed on the one hand from missing concrete, expert knowledge about their behaviour and on the other hand from the fact that the "normal" behaviour of bank clients was observed to be just too diverse. Thus, it became impossible to differentiate between "normal" and "suspicious." Though it was technically feasible to measure the attendance time of bank clients automatically, this information did not provide substantial information about how "suspicious" or "normal" behaviour looked, as a deviation from the average attendance time can have many different reasons. This meant there were clear limits to the application of IPAs in cases where hardly any differences between the behaviour pattern of interest and other forms of behaviour could be found.

The same can be said for the third case study that was based on field observation in a computer vision laboratory. In this case of fall detection, a large part of the research process was engaged in finding out whether there are clear differences between what was thought to be a critical fall and other types of falling or lying down, for example lying down on a bed or couch for the purpose of resting. In this specific case, the ground truth was constructed within the framework of the computational script of the Microsoft Xbox Kinect sensor and its SDK as well as the ingenuity and creativity of the researcher in de-scripting, rewriting and making use of the hardware-software package. The other fundamental element was an experimental investigation into the dynamics of falling down within the context of AAL environments as there was no substantial or applicable knowledge available on how “real” or “realistic” falls of elderly people look. The ground truth was defined within this framework as a mathematical equation that was based on the relationship between two central visualised geometric elements; first, the plane of a floor area and second, a straight line visually representing the human body.

These three empirical cases demonstrated that the sociotechnical construction of a ground truth is significantly shaped by the search for resemblance and difference (Suchman 2008: 140) within the specific “domain of scrutiny” (Goodwin 1994). Therefore it is an essential procedure on the way towards teaching computers to see, to explore and define characteristics that stand for resemblance. For instance what characterises a “critical” fall in order to distinguish it from an “uncritical” fall? The exploration and definition of characteristics that differentiate between “critical” and “uncritical” falls has to be researched. As such, these practices that are a basic requirement for the sociotechnical construction of the ground truth, are key in constituting what is real (Suchman 2008: 140) when it comes to computer vision. So for example, they were essential in defining what happiness is, what suspicious behavior is, and what a (critical) fall is. What seems to be clear for humans and for human vision in the many situations of daily life, and this might be because it is a matter of continuous negotiation and discussion, seems in many cases impossible to teach a computer. Nevertheless, this might be possible in a particular 'situated way'. Although these elements and how they are constituted are not immutable, their manifestation in the form of a ground truth used for the processing of images and video sequences, could be perceived not as something specific to IPAs once they have been deployed, but as even more authoritative, neutral or objective than for example, an individual human being. What was demonstrated with these three different cases was that in contrast to the general view of technical authority and neutrality, specifically situated and subjective views that were negotiated in different sociotechnical constellations in and around computer vision laboratories, were inscribed (or were going to be inscribed) in the respective ground truth, and thus inscribed into the ability of the computer to see and recognise. As a consequence, similar to human vision the processing of images by algorithms is a situated, interpretative practice that is shaped by cultural traditions of seeing in the field of computer vision and professional skills in reading images (cf. Burri 2012: 51). My field observation in a computer vision laboratory showed especially that the visual material worked with was interactively negotiated within the framework of the computational script of the software-hardware package (Kinect sensor for Windows in this case), the researchers ingenuity and creativity in de-scripting and rewriting the computational script, and the researchers ideas on the specific phenomenon being investigated (in this case, falling down). In this regard the computer vision reseacher had the (implicit) power to decide and select which kind of knowledge to draw on, altough this decision was influenced by the imagined future application area of AAL and limited by the computational script to a considerable extent.

Transformation and Reduction of Complexity

The ground truth and its sociotechnical construction process contain various transformations on the way from transferring human ways of seeing to computer ways of seeing. Assuming "that images cannot entirely be transformed into textual or nummerical signs without losing some of their advantages" (Burri 2012: 50) the question is, to what extent the visual value of images (and also the visual value of live, everyday situations) gets lost in the process of sociotechnically constructing an IPA ground truth and how this impacts the ways humans see and recognise. All of the cases presented in this chapter displayed different steps in the transformation process as well as reductions in complexity that can be categorised on at least three analytical levels:

First, on the conceptual level the problem is the isolation and fragmentation of the (human) body and the missing integration of holistic, contextual information available in the original scene. For example, in the case of Facial Expression Recognition, similar to Face Detection and Face Recognition, only the face is incorporated into the analysis. All other body parts, the situation and the site of the Facial Expression Recognition are neglected. In the other two cases presented, the human body is not fragmented but processed as a coherent unit. As such, it is isolated from the rest of the scene in order to be detectable and able to be processed for the IPA. In the case of fall detection, the unit of the human body needed to be put into relation to another element available in the scene, in this case, the floor. All other elements were perceived as a disturbance and thus, actively neglected or seen as irrelevant in this context. That means, the elements of interest were not only isolated from the rest of the scene, they were highlighted in the sense of how Goodwin described it (cf. Goodwin 1994: 606). That is, the specific phenomenon of falling down was highlighted by making the human body, the floor and their relation to each other the salient features. Depending on the level of transformation, the body was highlighted as a body shape, filled-in with a specific colour, as a skeleton or as a straight line. As such, the perceptual attention of both the researcher and the computer algorithm was targeted at the reduced (straight line) visual representation of the body and its movements in relation to the reduced (plane) visual representation of the floor area.

The second transformation and reduction level is the engineering of visual data. Here, the displacement of frames of reference can be witnessed[89]. This means that in comparison to the original scene, the frame of reference moves away from the complex sociocultural situation in reality to the (no less real) digital image and thence to the

[89] For a more detailed involvement with reference in science see Latour (1999: 24ff.). He understands reference as „…our way of keeping something constant through a series of transformations“ (ibid.: 58)

(again, no less real) representation of the central visual elements as material representations, to the (once more, no less real) representation in numbers that can be used by the computer for further processing and calculating. This process was best exemplified in the case of fall detection: several body transformations took place from the original, physical body to its flat, coloured shape depiction, to the skeleton representation, to the depiction of connected points, to a straight line that then could be processed by the computer in a mathematical equation.

The third transformation and reduction level is the immediate processing of visual input data. During this process the observed data in the form of input images have to be analysed in comparison to the respective ground truth template. They have to be manipulated, filtered and smoothened in order to both be "workable" for the IPA and to obtain individual results. In turn that means that specific thresholds have to be predefined by the computer scientists; for example when precisely a facial expression really represents the emotion anger or what is a threshold value for a "critical" fall. In the fall detection case that could mean, for example, that the threshold value is set at a fictional value of 9, assuming that 90 is absolutely nonparallel and 0 is absolutely parallel. If this is the case and the result of the IPA analysis is 9 or below, then the assumption is that a critical fall took place. But, if the result is 10 or above the assumption is that no critical fall occurred. This might be especially challenging when the calculated value is particularly close to the threshold value when a sequence of images is analysed. In such a case it is necessary to filter and smoothen the results over time, in order to decide whether there has been a critical fall or not.

Context is Crucial

In summary, it can be stated that in all of the presented cases, the technical component of vision and regulated recognition that is closely related to what Collins called the formal or pattern recognition model of seeing (Collins 2010: 11) dominates, and therefore the social and interpretative components (what Collins calls enculturational model of seeing) have been ignored. The whole complexity of human vision and recognition is simulated in its structure and framework. However the involvement of complex information about the context of a situation has been widely denied. This is of importance, because information as well as human action and activity are not self-explanatory but rather, are negotiated out of social context (Richter 2002) in situated actions (Suchman 1987, 2008). This also concerns face-to-face control, which is negotiated; not absolute. Furthermore it is "based on a complex moral assessment of character which assesses demeanor, identity, appearance and behavior through the lens of context-specific relevancies." (Norris 2003: 276).

Image Processing Algorithms focus on specific visually observable objects and body movements when it comes to the processing of visual data. This hightlighting hides the processes of negotiating the meaning of these visually observable objects and body movements in face-to-face interactions, because it operates with the pre-defined ground truth template representing a particular view. But as was shown in Chapter Two, human vision is not just the sum of isolated, observed components. Instead it was assumed that vision is subject to change, both culturally and historically (Tomomitsu 2011; Kammerer 2008; Burri & Dumit 2008, Rövekamp 2004). As was noted in Chapter Three, Charles Horton Cooley (1926: 60) distinguished between spatial/material and personal/social knowledge. The former, based on sense perception, gives rise to better, or more exactly, to quantifiable natural science. The latter only emerges through negotiation and communication with other people and their ways of thinking. 'Social knowledge' is in close relation to what Collin's calls 'Collective Tacit Knowledge' (CTK): his argumentation being that individuals can acquire this specific kind of knowledge only by being embedded in society (Collins 2010: 11) and by having what he calls 'social sensibility' (ibid.: 123). In contrast to human vision, IPAs do not integrate this kind of social knowledge or social sensibility, both of which are of outstanding importance for understanding the complexity and ambiguity of human behaviour or facial expressions in a specific situation.

Under these circumstances, IPAs that transform and reduce the complexity of reality in a significant way give cause for reflection on what integration of these into larger systems or networks in our contemporary society means, and what kind of new order will appear once these systems have been integrated into social life. For example, automatically measuring attendance times, facial expressions or the relationship between the body and the floor can provide indications for suspicious behaviour, for the emotional constitution of a person, or for a critical fall. They are partial aspects. The crucial point is that they are not equivalent. Suspicious behaviour, the emotional constitution of a person, or a critical fall are complex entities comprised not merely of attendance time, facial expression, or a body-to-floor relationship, but rather made up of many different elements that are subject to continuous negotiation. Furthermore, attendance time, facial expression or body-to-floor relationships are ambiguous; they are context and situation dependent. To get a realistic impression of what IPAs are able or not able to do, it is of great importantance to make that difference clear.

Another important conclusion is that transformations in complexity and reductions in the development of semantic IPAs have an upskilling rather than the opposite effect, concerning the implementation of IPAs in greater sociomaterial assemblages and sociotechnical networks. It can be assumed that along with the implementation of such algorithms in real-world scenarios and systems, human operators and users have to be trained in order to manage, work and interact with these algorithms and systems rather than fully abandoning the human factor. It is however, not only operators that are concerned with such systems. They depend on the actual application, the site of application and who is involved and in what manner. Whoever is involved in, or affected by these algorithmic sorting and decision-making processes has to be understood in best possible detail in order to handle these systems and minimise possible risks such as false positive findings.

Chapter Six

How 'Functioning' is Made: The Negotiation of 'Functioning' Image Processing Algorithms

In one of the group discussions during my participant observation in a computer vision laboratory,[90] one of the computer scientists said that if something does not function that is generally seen in a negative way by us; and in saying “us” he meant workers in computer vision. In response to this statement, his colleague even strengthened the argument by saying that not functioning “does not exist at all.” A third one added: “... I think it is not a matter of something functioning or not functioning, because in principle, everything functions, it only is a question of where and how well it functions.”[91]. And finally the second computer scientist completed the sentence with: “When and how...”[92]

But what do they actually mean when computer scientists working in the field of computer vision and image processing, talk about something functioning? Moreover, what does it mean if we in general speak about something that functions or something that works? Generally spoken, one could say if something functions, certain

[90] See Chapter One, section ‚How was it Done?‘ for the methodological background of the group discussion

[91] (Original Quotation/Translation by author) “... ich glaub es gibt eigentlich nicht etwas funktioniert oder etwas funktioniert nicht, weil prinzipiell funktioniert alles, es ist nur die Frage wo und wie gut.”

[92] „Wann und wie...“

expectations are fulfilled. For example, at the moment I expect my keyboard to display characters and subsequently words on the screen of my laptop and right now this is going absolutely smoothly. It works. Generally, I do not think about what it actually means to say 'something functions'. I would normally not ask myself whether typewriting on my laptop functions; I just take it for granted, until the time may come when the displaying of certain characters does not function any more. In everyday life, that is my assumption, we usually talk more about something 'not functioning' than about something 'functioning'. One of the computer scientists was certainly right when he said that it is not a matter of something functioning or not functioning because in principle, everything functions, it is only a question of where and how well it functions. Functioning can be regarded as something highly dependent on situation, place, and time. However in the group discussion quoted, there is also a second version of 'functioning' mentioned. This was when one computer scientist said "not functioning does not exist at all". An understanding of these two versions of functioning seems to be quite easy at first glance. A really important term that I heard frequently during my field work in the computer vision laboratory was "to make something work", meaning to make something running. This term was central to the everyday work of computer scientists and an inherent part of the research culture I experienced. Usually it was a problem that had to be solved. There was a certain task that had to be made to function. So this meant that most of the time the computer scientists dealt with things that were not functioning. However in the end, and that is what the computer scientist meant by "not functioning does not exist at all", the thing or problem or IPA they were dealing with had to function in any way. The character I am typing on my keyboard has to be displayed on the screen. If this task does not function the keyboard should not be for sale. What seems quite clear in the case of the keyboard might well be more complicated with other technologies or artefacts, for example in this case, Image Processing Algorithms.

What has 'functioning' actually got to do with 'seeing'? "Make something work" was, as I already said, a very popular phrase in the computer vision laboratory and therefore it can be treated as a guiding principle in everyday computer vision work. Let me have a look at a concrete example I already referred to in the previous chapter explaining how something is made to function in computer vision: Imagine a person just entering the room you are sitting in while reading this text. Suddenly, the person falls down and you witness it with all of your senses. Without any effort and reflection you see and recognise what happened: the person fell down and it does not take much effort for you to decide whether the person needs help, because he or she is injured or has even lost consciousness. In everyday situations we take being able to see and recognise such an event for granted, which is why we would usually not even mention that our perceptional apparatus functions. But if one has a closer look at human 'seeing' and perception there certainly are situations in which the 'functioning' of human perception is evaluated: for example, eyesight tests at an opthalmologist and the compensation of poor eyesight with spectacles, contact-lenses or even through laser surgery.

Now imagine how difficult it must be to teach a computer or IPA to detect a fall automatically. I will go into detail in just a moment. For now, I want to get straight to the question: How one can say whether the sociomaterial assemblage of automatic fall detection works (well, accurately and so on)? Or in other words: How can one say whether an automatic fall detection system sees and recognises a fall correctly? One preliminary answer is that the computer scientist and his or her respective scientific community are the primary evaluators and can be regarded as an "obligatory passage point" (Callon 1986). They define the criteria for evaluation. What is crucial about this is the understanding that these criteria are not immutable. They are temporary and of a fragile nature. 'Functioning', at least in the domain of Image Processing Algorithms is always conditional (e.g. it functions only in daytime) and probabilistic (e.g. it functions in 97% of the cases). It is a discursive practice (cf. Goodwin 1994: 606) about when, where and why a specific IPA functions.

The argument I wish to develop throughout this chapter is that what I call the 'Regime of Functionality', which could be characterised as an ordered and structured way of doing, making and presenting something as 'functional' and is a specific discursive practice that shapes the awareness of what counts as 'functional'. This 'Regime of Functionality' is deeply embedded in the current culture of computer vision and IPA research as I experienced in my field work and involvement with the field of computer vision.

Major events at which the ‘Regime of Functionality’ reigns supreme are demonstrations and presentations of systems equipped with IPAs. During my field work I witnessed and participated in several events at which IPAs were demonstrated to a wider public. The main concern of this chapter is to present my findings on demonstrations of computer vision and how these act as occasions at which the functioning of IPAs in the making is demonstrated and visions of their future use in greater sociotechnical systems are presented. I shall therefore present my empirical findings within the framework of general insights into IT-demonstrations and with Goffman's frontstage/backstage concept and connect it to the debate on changing norms and practices of academic, technoscientific work, ending the chapter with the debate on the politics of the algorithm.

First I shall present a closer look at the views and opinions of computers scientists on what functioning means for them and how they use and understand this expression in different contexts. I will also take a closer look at a powerful elaboration of ‘functioning’, namely ‘robust functioning’. I shall discuss and contrast these findings with those from the observation of demonstations that follow after my presentation of computer scientists´ views and opinions.

Definitions of ‘Functioning’

At the start of the group discussion about the meaning of ‘functioning’ that I set up, the participants, mostly junior computer scientists working in the same computer vision laboratory, tried to define what ‘functioning’ means from their points of view. As one can see there is a variation of definitions on different levels:

> Rafael: "Everything over and above coincidental, can be said to function." (General laughter) [93]

> Benjamin: "I would say something functions when it acts in the way it is supposed to act. Whoever developed the system or whatever had a certain intent and a certain goal of how it was supposed to act. If it really does, then one can say that it functions."[94]

What Rafael does, is to compare 'functioning' to coincidential functioning. He did not elaborate on his statement, because Benjamin immediately started talking, but one can guess what he intended to say with his amusing (because of the collective laughing afterwards) definition by starting a Gedankenexperiment. For example: an IPA has to detect human faces in 640x480 pixel images originating from a video stream. If this detection task is random, the face to detect could be anywhere on the 640x480 pixel screen. Considering how many pixels the face and the frame indicating that a face has been found (e.g. a circle or one of the commonly used green boxes) is made up of, there is a really low probability of detecting the face correctly. A slightly better result could be reached for example, if information was provided that faces in pictures are not usually found at the base of the image. In such a case, the detection algorithm could limit the search area and this would improve results, but still the probability of detecting and locating a face correctly would be very low. However, when taking Rafael´s definition seriously, this improvement would count as 'functioning', because it is better than coincidential.

Also in Benjamin´s definition, the small improvement in the face detection algorithm in my Gedankenexperiment would count as 'functioning' considering that the system designer intended the algorithm with a limited search area to work faster than the

[93] (Original Quotation/Translation by author) Rafael: "Alles, das besser als Zufall ist, das funktioniert." (Lachen)

[94] (Original Quotation /Translation by author) Benjamin: "Ich würd sagen funktionieren ist etwas verhält sich so wie es sich verhalten soll oder wie es von dem der das system oder was auch immer entwickelt hat, der hat ja eine bestimmte Absicht dahinter und hat ein bestimmtes Ziel, wie sich das verhalten soll und wenn es dann dieses Verhalten tatsächlich an den Tag legt dann kann man sagen, dass es funktioniert."

coincidental algorithm, because it had to search for a face on a smaller area. This approach to the discussion about what 'functioning' means, requires a clear statement at the beginning of the task that can be checked at the end of the task.

When all goals in this Gedankenexperiment have been reached, this would mean that the face detection algorithm working on a smaller area, works slightly better[95] and faster than the coincidental model. What actually happens when the advanced algorithm suddenly starts to work more slowly than the coincidental algorithm? A little later, another computer scientist, whom I shall name Oskar, brings his definition of 'functioning' into the discussion:

> Oskar: "Everything always functions until one gets a different result."[96]

With one unexpected, rogue result, everything could change again and the functioning system is history. This implies of course, that a level of functioning had been attained before. This again, requires an act of completion or of bringing something to an end by saying, "Yes, it works!" This act of completion can take place at very different stages in a process, so that the meaning of what counts as "working" is very fluid and blurred. It also depends on the use of the term 'functioning' or 'working' in different locations and situations.

Use of the Term 'Functioning'

Ideas of the future in innovation processes play a crucial role and visions drive technical and scientific activity (Borup et al. 2006). Thus, it is important to be aware of different places and activities, where people are talking about these ideas, promises and expectations of research. Borup et al. (2006: 292) note that there are quite contradictory expectations amongst people closely involved in scientific work: "when wearing a public entrepreneurial hat they might make strident claims about the promise of their

---

[95] In computer science the formulation "slightly better" would not be accepted, because it is too inexact and unclear. I would need to provide concrete performance benchmarks, such as time values.

[96] (Original Quote/Translation by author) Oskar: "Es funktioniert immer alles bis man ein anderes Ergebnis bekommt."

research", but "when among research peers, they will be much more cautious and equivocal, though publicly still committed to the promises associated with their field."

This insight connects very well with the use of the term 'functioning' in computer vision, as the following statements by computer scientists should show:

> Benjamin: "You just have certain criteria and with these criteria a function just might or might not fail; they vary then in this situation. If I work on something for my job (author´s note: in a commercial computer vision company) which is really going to be sold, then I have less scope for saying, "OK, it's not a tragedy if it fails" and say "OK, the criteria are not such a problem if it fails'. Though if you are developing something at Uni, where you know exactly which points have to be fulfilled but others are just 'nice to have', but don`t work out on time, then you´d say it functions anyway, because the criteria you applied to this system are not hard and fast; or sometimes, loose too."[97]

In his statement Benjamin distinguishes between his job ("für die Arbeit") and work at the university ("an der Uni") and notes that one has more room for manoeuver within the university and less within the job, because failing an assignment in a university setting is less crucial than in a job. I think I need to provide more background information here in order to fully understand Benjamin's statement.

The university computer vision laboratory where I based my participant observation had very close connections to a computer vision spin-off company at that time. In fact, the university lab and the spin-off company cooperated within research projects as two separated institutions and aside from that, many of the people working or studying at

[97] (Original Quotation/Translation by author) Benjamin: "Man hat halt dann gewisse Kriterien und diese Kriterien wo halt dann das funktionieren scheitern kann oder nicht, die variieren dann in dieser Situation. Wenn ich jetzt was für die Arbeit mache das dann tatsächlich verkauft wird dann ja hab ich dann halt weniger Spielraum zu sagen ok das ist nicht so tragisch wenn das nicht so hinhaut und sage ok das Kriterium ist nicht so tragisch wenn das nicht hinhaut, oder wenn man jetzt an der Uni etwas entwickelt, wo man genau weiss die und die Punkte müssen halt erfüllt sein, aber das andere ist halt nice to have aber geht sich halt zeitlich nicht immer aus, dann wird man da halt trotzdem sagen, es funktioniert, einfach weil die Kriterien, die man an dieses System anlegt ja nicht so fest sind; oder locker."

the university computer vision laboratory also worked for the other. So, when Benjamin talked about the job, he was referring to the company. When he talked about uni, he meant the university computer vision laboratory.

Where on the one hand, my observations confirmed this separation of the two sites, on the other, they also showed that the boundaries between university lab and spin-off company were blurred to a considerable extent. Both the university laboratory and the company generate and make use of synergies: enabled through both the intertwining of the two and a clear demarcation. Generally speaking in daily work life, no boundaries seemed to exist between university lab and spin-off company, but when speaking or presenting to a wider public, the demarcation of one to the other was often used to promote personal involvement in productive networks. In this case this meant having an academic research partner from the university on the one side, and on the other, a partner in private enterprise.

Benjamin's statement hints at the different meanings of 'functioning' used in different places and bound to this, the grade of completion that is required in order to count as 'functioning'. His statement illustrates the co-presence of 'functioning' and 'not-functioning' or 'failing'. Within the company you can only sell systems or products that really work. There is no conditional functioning – it just has to function! Within the university the use of the word 'functioning' is much more flexible because the system or product does not have to work in the same way as in private enterprise. One can say that it functions under certain conditions and would not otherwise. I shall elaborate on this point in the next paragraph when I present another text passage from the group discussion. I shall show the nuanced meaning of the term 'functioning' in scientific papers and then contrast this with two other meanings, namely 'advertising' in research output such as project reports and as a temporarily closed entity when technology is in action.

The Term ‘Functioning’ in Technoscientific Papers: Conditional and Probabilistic Functioning

In scientific papers, the tenor in the computer vision laboratory I visited seemed to be that there is no ‘functioning’- at least not in the sense that something just functions. There are carefully nuanced meanings of the term ‘functioning’. The following discussion of the first group discussion round verifies this:

> Ben: “In a (author´s note: scientific) paper - we are kind of careful I guess. I don´t know! Anyway, nobody writes in a paper that anything functions in general. I haven’t ever seen that.”[98]
>
> Oskar: “There you write the probability of it functioning is 80%.”[99]
>
> Ben: “Yes, exactly. There you write the probability of it functioning is 80% with these data. Maybe it functions better than another method. But in general, you never say it always functions.”[100]
>
> Greta: “Because there are always cases without a 100% probability of it functioning.”[101]

Following this conversation of computer scientists, I really had the impression that absolute ‘function’ does not exist. This is true and false at the same time. It is true, because there are always cases in which there is no 100% probability of something functioning. It is false because ‘functioning’ as a term in everyday usage, does exist. However, it exists in a sense that does not mean there is a 100% probability of something functioning but for example, an 80% probability. Thus, when we say something functions, we are actually always talking about ‘conditional functioning’. As

---

98 (Original Quotation/Translation by author) Ben: „Im Paper also ich weiß nicht, wird sind da eher vorsichtig glaub ich, also es schreibt niemand rein, dass irgendwas funktioniert im Allgemeinen, also hab ich noch nie gesehen…”

99 (Original Quotation /Translation by author) Oskar: „Da schreibst du es funktioniert zu 80%.“

100 (Original Quotation /Translation by author) Ben: „Ja genau da schreibst du es funktioniert bei den Daten zu 80%, es funktioniert besser als eine andere Methode vielleicht, aber du sagst nie, im allgemeinen es funktioniert einfach immer.“

101 (Original Quotation /Translation by author) Greta: „Weil es immer Fälle gibt wo es nicht 100%ig funktioniert.“

mentioned before, it is possible to say, for example that something functions with perfect lighting conditions. Even then, cases could probably be found that do not work using a specific algorithm or system, because different challenge has emerged.

In addition to this, functioning is always 'probabilistic functioning'. As the group discussion showed, in a scientific paper on computer vision one has to write "there is an 80% probability of it functioning" (e.g. face recognition algorithm XYZ does detect faces correctly in 80% of the cases). Even then, the conditions and settings have to be described in detail.

Most papers in computer vision are evaluation papers, meaning that the performance of an algorithm is presented. As described in Chapter Two, in a survey of Policy and Implementation Issues of Facial Recognition Technology, Introna and Nissenbaum asked among other things, "Does it actually work?" and elaborated on the evaluation of Facial Recognition Technology (Introna & Nissenbaum (2009: 21ff.).

They divided evaluation into three different types: technological, scenario, and operational. What I have described so far would be included in the technological evaluation type, because the performance of algorithms was tested. This means that such an evaluation is "normally performed under laboratory conditions using a standardized data set that was compiled in controlled conditions." (ibid.: 21) The main purpose and advantage of technological evaluations is the high degree of repeatability, but they are not designed to be evaluated under different conditions and settings. Scenario evaluations are designed to model real-world environments and populations and operational evaluations test systems in situ in their actual operational conditions (ibid.: 22). Introna and Nissenbaum note that ideally, FRT systems start with technological evaluation, followed by scenario evaluation, and finally by operational evaluation (ibid.: 21). This means for full understanding that if - in this case facial recognition - is said to be working, it has to have been tested not only in the lab, but also in situ in operational conditions.

The Term ‘Functioning’ as Advertising Strategy in Research Output (Proposals, Reports etc.)

Interestingly, and in contrast to what has just been said, I met the term ‘functioning’ during my fieldwork in a university computer vision laboratory many times. Outside a laboratory, in the media, we meet the term ‘functioning’ even more often as was demonstrated in Chapter Four. Here, I have to mention that it is not always the term ‘function’ itself that is used. Often, a more subtle way of getting the message across is employed to convey that a specific Image Processing Algorithm or computer vision technology really works. In Austrian conversational language this is usually worded with “es rennt” or “es läuft”, meaning that something is up and running, or functioning.

Before coming back to the group discussion, I would like to point to an Austrian newspaper article that was published in the newspaper *Kronenzeitung*, which has the widest circulation in Austria[102] and shows very well what is likely to happen once the term ‘functioning’ is used in the media: Science fact and science fiction are blurred.

---

[102] In 2009, the daily newspaper Kronen Zeitung reached 40.4% of the Austrian population (Source: http://www.media-analyse.at/studienPublicPresseTageszeitungTotal.do?year=2010&title=Tageszeitungen&subtitle=Total)

Seite 16 ÖSTERREICH Donnerstag, 10. Dezember 2009

Firma entwickelt neues Sicherheitssystem ● Computer erkennen Gesichter:

„Super-Kameras“ für Flughafen!

Sie können Gesichter erkennen, Personen durch Räume „verfolgen“, Identitäten prüfen – die Wiener Firma „x-pin.com“ arbeitet an einem neuen Sicherheitssystem für den Flughafen Wien-Schwechat, das schon 2010 zum Einsatz kommen könnte. Vorteil: Der neue „High-Tech-Computer“ besitzt eine Art künstliche Intelligenz.

VON MICHAEL POMMER

Datenschützer schlagen die Hände über den Köpfen zusammen – dabei sind Flughäfen weltweit schon immer Zonen gewesen, in denen uneingeschränkt überwacht und kontrolliert werden darf. Zu den typischen Kameras, die schon überall von den Decken hängen, kommen wohl schon bald neue Super-Geräte hinzu. Intelligente Kameras, die nicht stur auf alle Bewegungen zoomen, sondern Ereignisse und Personen erfassen, analysieren, auswerten und sogar Prognosen erstellen.

„Wenn ich den Flughafen betrete, wird mir vom System automatisch eine Nummer zugeteilt“, erklärt x-pin-Geschäftsführer Christian Raunegger sein Kiras-Projekt. Sofort werden offizielle Fahndungslisten verglichen. Aber das System kann noch mehr: Hat ein Passagier etwa beim Einchecken den Koffer stehen lassen, muss nicht extra ein teurer Entsorgungstrupp ausrücken – das „Super-System“ kann die Person auf dem Flughafengelände ausfindig machen und rasch informieren.

Massenansammlungen, Tumulte, offene Fenster. Der Computer erkennt atypische Ereignisse und schlägt Alarm. Raunegger: „Bisher müssen alle Flughafen-Videodaten sechs Monate gesichert werden. Wer oft verreist, ist so Dutzende Male im Jahr gespeichert. Mit dem Kiras-System werden nur noch auffällige Momente archiviert.“ Das dürfte Datenschützer ja vielleicht sogar beruhigen.

Das Projekt ist weltweit einzigartig und wird im zweiten Halbjahr 2010 schließlich abgeschlossen sein.

X-pin Geschäftsführer C. Raunegger

Foto: Andi Schiel

Die Video-Augen überwachen und analysieren jede Situation

Figure 4: **Newspaper Article "Super-Kameras" für Flughafen!**

This article from December 2009 describes the “super-cameras” mentioned beforehand in both the present tense and indicative form that are able to recognise faces, to track people throughout Vienna Airport or to monitor and analyse every situation. For example, “Super-cameras for the airport!”, “They can recognise faces.” Or “The video eyes monitor and analyse every situation.” It is also mentioned in the article that such a new security system “.... could already be deployed by 2010.” When reading the article the impression is created that the (technological super-camera) future (in this case 2010) is happening now (in this case December 2009) and therefore these “super-cameras” are virtually here. As far as I know there is nothing comparable to these “super-cameras” described in the article, in operation at Vienna Airport at the point of my writing. In addition, it is impossible to say to what extent this media article is a product of the journalist or of the public relations office of the computer vision company mentioned in the article.

More precise authorship can be ascertained with research output such as project proposals or reports. This research output is characterised by computer scientists in a different way to scientific papers, as the following excerpt of the group discussion shows:

> Ben: "I don't know - because in the scientific world, communication with the (author´s note: computer vision) community is pretty much through the means of papers. And to the outside world, these are maybe more like project proposals or project results and they are often just simplified and also a bit tweaked, I guess. Maybe things aren´t represented in the way they really are. One just says it functions like we planned, and..."[103]
>
> Rafael: "Here´s the perfect example: Institute "XYZ" has been writing for three years now that they have solved computer vision problem "ABC." We´ve been working on that project for three years now!" (General laughter)[104]
>
> Ben: "Yes, exactly."[105]
>
> Rafael: „Well, to the outside world not everything is shown as it really is."[106]
>
> Interviewer: "Well, where exactly do they say that? On their homepage?"[107]

---

[103] (Original Quotation/Translation by author) Ben: "Ich weiß nicht, weil in der wissenschaftlichen Welt da kommuniziert man halt mehr oder weniger die community mit papers. Und nach außen hin sind das vielleicht eher Projektanträge oder Projektresultate und die werden halt oft gerne vereinfacht dargestellt und bisschen beschönigt glaub ich, also da wird es vielleicht dann nicht so dargestellt wie es wirklich ist, da sagt man halt dann es funktioniert wie wir es vorgehabt haben und..."

[104] (Original Quotation /Translation by author) Rafael: „Da gibt es ein super Beispiel, das „Institut XYZ" (Anmk.: anonyimisiert) schreibt schon seit drei Jahren, dass sie das „computer vision Problem ABC" (Anmk.: anonymisiert) gelöst haben und wir arbeiten seit drei Jahren an dem Projekt." (allgemeines Lachen).

[105] (Original Quotation /Translation by author) Ben: „Genau ja."

[106] (Original Quotation /Translation by author) Rafael: „Also nach außen ist nicht immer das was wirklich ist."

Rafael: "Yes in promotional leaflets and so on. Well, three years ago they were already describing what we only integrated six months ago. But they already wrote three years ago that they could do that. Not, they are going to do that, but that they already can do that."[108]

Oskar: "Let´s get straight to the point: you don't know, if they had maybe already implemented similar stuff and maybe had solved the problem with it…"[109]

Rafael: "You seem to know all about it."[110]

Interviewer: "But this means, they have to have something, doesn´t it? Something so they can clarify; any reference so they can show, yes it already works, maybe to a limited extent and under certain conditions… so have they got something like that?"[111]

Rafael: "They already had that, sure, but at that time they only had a simple shape matcher and absolutely nothing that recognised the content of the shape, and in the leaflet they were already writing that they were doing the matching using the content etc. This means of course, as you saw, that it does construct correctly, because the problem is relatively simple in this case; but in reality it´s not, because of the stuff and

[107] (Original Quotation /Translation by author) Interviewer: „Also wo sagen sie das auf der Homepage oder…?"

[108] (Original Quotation /Translation by author) Rafael: „Ja in Werbefoldern usw. Also die haben auch schon vor drei Jahren das beschrieben, was wir jetzt vor einem halben Jahr eigentlich integriert haben. Und da haben sie aber schon vor 3 Jahren geschrieben, das können sie schon, nicht, das werden sie machen, sondern das können sie schon."

[109] (Original Quotation /Translation by author) Oskar: „Um es vorwegzunehmen: du weißt nicht ob sie vielleicht schon ähnliche Sachen schon implementiert gehabt haben und das vielleicht mit dem gelöst gehabt haben…"

[110] (Original Quotation /Translation by author) Rafael: „Das weißt du ziemlich genau."

[111] (Original Quotation /Translation by author) Interviewer: „Aber das heißt sie müssen ja irgendwas haben, oder? Wo sie verdeutlichen können, irgendeine Referenz wo sie dann zeigen können, ja das funktioniert ja schon, vielleicht eingeschränkt unter gewissen Bedingungen und so… also das haben sie dann?"

> how they had already decided to put it together. It was all much simpler. What they really had and what really worked."[112]

Very similar to my analysis of the newspaper article, the computer scientist Rafael describes an example, where another computer vision laboratory puts across that they have solved a computer vision problem, but in reality they have not actually done that. In reality, the "stuff" they really had was "much simpler" than what they said they had. The rhetorics of promoting their "stuff" seems crucial. This approach is very similar to the approach the computer vision company in the newspaper article took. Rafael's description spells this out perfectly: "But they already wrote three years ago that they could do that. Not, they are going to do that, but that they already can do that."

Writing in the present tense and with the indicative form seems to be widespread when speaking about ongoing and future research in computer vision and IPAs. My observations in the lab confirm this strategy. For example, I observed the preparation of a slide show about a project that had just recently been started. The main challenges of the project were formulated in the present tense and with the indicative form so that the impression was given that the research had already been done, with a perfect, ideal product as the outcome. In German as in the English language the present tense can also be used to describe activities in the near future (futurisches Präsens/present future). In doing so, time expressions (e.g. tonight, next week, in 2013) have to be used, otherwise the sentence does not have a future meaning. If one says, for example, "our face detection system works in 2013", this would always imply that it is only a matter of time and taken for granted that there will be a working face detection system in 2013.

---

[112] (Original Quotation /Translation by author) Rafael: „Das haben sie schon gehabt, sicher, aber sie haben damals nur einen reinen shape matcher gehabt, und überhaupt nichts irgendwas mit dem Inhalt und im folder ist aber trotzdem schon gestanden, dass sie über den Inhalt usw. das matching auch machen. Das heißt, natürlich du hast gesehen, das stellt es richtig zusammen, weil das Problem relativ einfach ist in dem Fall, und in Wirklichkeit aber nicht aufgrund von den Sachen, was sie bestimmt haben wie es zusammengesetzt wird. Also das Ding war viel einfacher. Das was sie wirklich gehabt haben und wirklich funktioniert hat."

We come closer to what I call the 'Regime of Functionality' when explaining this strategy of promoting specific research in a distinct time frame that brings possible future outcomes into the present. In the next group discussion, some basic players and contexts which have to do with this 'Regime of Functionality' are introduced:

> Oskar: "But with project proposals in general, also with final reports, you have to sell them. It's also about getting approval for a proposed project. If you were really honest about it, you wouldn´t write that there have been promising, preliminary results. But we learned not to do it in this way and after all it doesn't work; it probably would be unfavourable. Even though, in an honest or scientific way, the findings were maybe good or positive anyway. Well, we don't do that."[113]
>
> Interviewer: "Do you have an explanation for that? It seems to me, it goes without saying that you have to give a good account of yourself, but do you have an explanation why it is like that?"[114]
>
> Jonas: "You are actually advertising yourself. And as a university or whatever, you always want to show yourself in a good light to your funding body So that they´ll say, as it were, funding has been well spent and there was real outcome. And if you say that yes, we tried and we worked very hard at it, but in the end it didn't work out, or just doesn't work this way, then they get the impression that it was a waste of tax money, because then they´ll not say it works, but that it doesn´t work. Because in public opinion I don´t think there is much differentiation between the two.[115]

---

[113] (Original Quotation/Translation by author) Oskar: „Aber bei Projektanträgen allgemein auch bei den Abschlussberichten muss man das verkaufen. Es geht ja auch darum, dass du wieder Projektanträge bewilligt bekommen willst und dann wenn man dann so ehrlich wär und nicht mehr reinschreibt es waren gute Ansätze dabei, aber wir haben gelernt, dass wir es so nicht machen und es funktioniert nicht, wärs vielleicht schlecht, auch wenn es ehrlich und wissenschaftlich vielleicht eh gute oder positive Erkenntnisse wären, also so machen wir es nicht."

[114] (Original Quotation /Translation by author) Interviewer: „Habts ihr Erklärungen dafür wieso das so, das ist ja für mich das ist ja durchaus selbstverständlich, dass man sich gut verkauft sozusagen, aber habts ihr da Erklärungen dafür wieso das so ist."

[115] (Original Quotation /Translation by author) Jonas: „Das ist ja Werbung eigentlich für dich selber. Und du willst irgendwie als Uni oder was auch immer gut dastehen dem Fördergeber gegenüber so quasi dass man sagt, das Fördergeld wurde super verwendet und dabei ist wirklich etwas rausgekommen.

Ben: "Well, it's also about financing, or... because you have to acquire money, in the future too. You only write project proposals about how brilliant the group is and what great stuff they´re able to do and so on, as it were. It's all very superficial and probably exaggerated and actually isn't true."[116]

Jonas: "It's actually advertising."[117]

Ben: "Yes exactly.... it's more like advertising. You just have to do it this way."[118]

Lea: "Some aspects always works. You have really worked hard on it. You can always find something positive to show."[119]

Ben: "You just keep quiet about what didn't work..."[120]

Oskar: "If it doesn't work well, you just have to look for cases where it did work well."[121]

Jonas: "Exactly. This is important again for project reports and final reports. There, the things that worked out well are included."[122]

Und wenn du dann sagst ja wir haben es zwar probiert und wir haben wirklich intensiv daran gearbeitet, aber es ist halt so und so nichts geworden oder es funktioniert so nicht dann denken sie sich auch, ok das war Verschwendung vom Steuergeld, weil dann heißt es eben nicht es funktioniert, sondern es heißt es funktioniert nicht. Weil viel Differenzierung dazwischen gibt es glaub ich in der Öffentlichkeit nicht."

[116] (Original Quotation /Translation by author) Ben: „Also es geht auch um das Finanzielle, oder... weil man muss immer wieder Geld akquirieren auch in Zukunft dann wieder... und nur dann schreibt man einen Projektantrag quasi wie toll die Gruppe ist und was die nicht alles schon kann und so. Und das ist alles sehr oberflächlich und wahrscheinlich übertrieben, und stimmt so eigentlich nicht."

[117] (Original Quotation /Translation by author) Oskar Jonas: „Das ist Werbung eigentlich."

[118] (Original Quotation /Translation by author) Ben: "Ja genau... es ist dann mehr Werbung. Weil man es eben so machen muss."

[119] (Original Quotation /Translation by author) Lea: „Und irgendein Teil funktioniert immer. Man hat ja wirklich dran gearbeitet. Irgendwas Positives hast du immer vorzuweisen."

[120] (Original Quotation /Translation by author) Ben: „Man verschweigt halt lieber die Sachen, die dann nicht so funktioniert haben..."

[121] (Original Quotation /Translation by author) Oskar: „Wenn es nicht gut funktioniert muss man Fälle suchen wo es gut funktioniert."

This 'Regime of Functionality' seems to refer to something really enigmatic. People talk of truth and secrets, about supernatural powers and the unknown future, but in the end; is it all about the money? The story in a nutshell seems to be: In computer vision, researchers need to advertise themselves in the best possible light, or as part of an excellent research group with outstanding research output (and this strategy involves accentuating the good things, meaning the things that function and suppress the bad things that do not function so well) in order to get future funding from sponsors and funding bodies. In the case of the computer scientists I talked with, this funding equates with tax money, which is viewed as money that belongs to the public. So certain non-functioning things have to be kept secret, because the public seems to be unable to understand how complicated everything really is. This means that it does not make sense at all to tell the public that there is no strict demarcation between functioning and non-functioning and that functioning is necessarily, probabilistic, conditional and temporal.

Following my participant observation in and around the computer vision laboratory and the computer scientists' comments in the group discussion, I come to the conclusion that what I call the 'Regime of Functionality' fundamentally affects computer scientists' work and their scope of action. I observed this on several occasions in the lab, always when there was talk of "making something work," and it became especially clear and articulated when the computer scientist, Ben, said: "Yes exactly.... it's more like advertising. You just have to do it this way" When something like this has to be done, then it becomes clear that from this point of view there is no other way to choose, or to go. An invisible hand seems to determine what I describe as this 'Regime of Functionality' that prescribes what to do and how to act and that only renders visible to the outside whatever functions.

---

[122] (Original Quotation/Translation by author) Jonas: „Genau, das ist dann wieder für Projektberichte und Abschlussberichte wichtig. Da kommen halt die Dinge rein die gut funktionieren."

The Temporariness of 'Functioning': Technology in Action

Over the last three decades, computer scientists have seen a maturing of the field of computer vision and image processing. There is a growth of applications and an increasing number of software and hardware products such as digital cameras with a face detection function, on the market (Sonka, Hlavac & Boyle 2008). This requires in principal, that a certain technology or application works. Technology in operation requires that it functions, however this functioning is achieved and however this functioning effectively looks. The matter of technology in operation was not very prominent in the group discussion, but at one point it was mentioned by Oskar:

> Oskar: "It´s easy if you have industrial clients; everything always works till they say something doesn't work." (General laughter)[123]

Oskar's statement brings into play another characteristic of functioning: its temporariness. As long as technology in operation acts in the expected way, this counts as functioning. This condition is however, not immutable. It can change quickly. Obviously functioning, or what counts as functioning, can become not-functioning again. So, in order to speak about "real" functioning, computer scientsists add a very powerful adjective to functioning, namely the word 'robust'.

Better than just 'Functioning': ' Robust Functioning'

During the group discussion, I called attention to another aspect I had been able to observe in the laboratory. When preparing a presentation for a client (it was not clear if it was a client of the laboratory or of the spin-off company), an earlier version was reworked and in doing so, all imprecise terms were eliminated or changed. For example, the word "very" (German: "sehr") had been eliminated several times and where possible,

[123] (Original Quotation/Translation by author) Oskar: "Es ist einfach beim Auftraggeber wenn man Industrielle hat, dann funktioniert immer alles so lang bis die sagen da geht was nicht." (allgemeines Lachen)

changed to numbers or percentages. Numbers and percentages were also used instead of estimations (e.g. 90% of the picture instead of “very much of the picture”).

In many parts of the first draft of the presentation paper, the word “better” (German: “besser”) was used, for example to describe that something works “better” if done in this or that way. Following the advice of a senior researcher, the two PhD computer scientists working on the presentation paper changed “better” to the words “more robust” (German: “robuster”). As the following extract from the group discussion shows, “more robust” does actually mean the same as “better”, but in the context of computer vision this word has its own magic.

> Oskar: “If it works robustly, it works better.” (Laughing out loud)[124]
>
> Lea: “Reliable.”[125]
>
> Rafael: “This word is pure advertising, isn’t it?”[126]
>
> Ben: “Yes, it´s not really a… strictly speaking it’s not a scientific word, but it is something you use.”[127]
>
> Elias: “In principle you can say the program is robust and doesn’t crash, no matter what filter is used…”[128]
>
> Ben: “But we tend to use it with features and so on…”[129]
>
> Greta: “There are different criteria for robust or, like, that it doesn’t crash or if you have variations in the data, it can handle it. So, if everybody just used a blue whatever in our

---

[124] (Original Quotation/Translation by author) Oskar: „Wenn es robust funktioniert, funktioniert es besser.“ (Allgemeines lautes Auflachen).

[125] (Original Quotation /Translation by author) Lea: „So verlässlich.“

[126] (Original Quotation /Translation by author) Rafael: „Das ist ja reine Werbung, oder, dieses Wort.“

[127] (Original Quotation /Translation by author) Ben: „Ja das ist ja echt kein… streng genommen ist es kein wissenschaftliches Wort, das ist halt etwas das man verwendet.“

[128] (Original Quotation /Translation by author) Elias: „Ja prinzipiell kannst du sagen das Programm ist robust und stürzt nicht ab egal was für ein Filter…“

[129] (Original Quotation /Translation by author) Ben: „Aber wir verwenden es eher so mit features usw…“

data bank and then we only get a red one, then it goes haywire, and nothing works. That means it's hardly robust."[130]

Rafael: "This is classic: You take a normal algorithm, let's say any matching algorithm, then at a certain point you chuck in a RANSAC algorithm - that's just a different kind of algorithm. And then you write: I did exactly the same as the other guy did, but this time it´s robust. So, if he wrote stereo baseline matching or anything like that before, then you write robust baseline matching, because you have edged exactly one algorithm in. And that is pure advertising for yourself. [131]

Benjamin: "Maybe you could also say you can generalise it more, couldn't you? (General agreement) Or, that it works using a broader volume of data or input data."[132]

Oskar: "Or put simply it´s about, for example, car number plate recognition that only works when the sky is blue and the sun is shining. If somebody does it when it´s raining, there´s a thunderstorm and clouds… and if it still works, then it's robust."[133]

Ben: "Or more robust." (General laughter)[134]

---

[130] (Original Quotation /Translation by author) Greta: „Es gibt ja auch verschiedene Kriterien von robust, oder, eben sowas dass es eben nicht abstürzt oder dass es wenn du Varianzen in den Daten hast, dass es damit umgehen kann. Also wenn jetzt alle nur einen blauen verwendet haben für unsere Datenbank und dann kriegen wir eines mit rotem und dann geht es nur noch wwugutt (ein Kunstwort, das vermutlich auf einen deutlichen Anstieg hindeutet) und nichts mehr funktioniert, das ist halt dann wenig robust."

[131] (Original Quotation /Translation by author) Rafael: „Ein Klassiker was es bei uns auch gibt ist, du nimmst einen normalen Algorithmus her, irgendeinen matching Algorithmus, dann haust du an einer bestimmten Stelle RANSAC Algorithmus rein, das ist halt ein anderer Algorithmus, und dann schreibst du, ich hab genau das gleiche wie der andere gemacht, aber robust. Also, wenn er vorher geschrieben hat, stereo baseline matching oder sowas, dann schreibst du robust baseline matching, weil du genau einen Algorithmus dazwischen geschoben hast. Und das ist reine Werbung."

[132] (Original Quotation /Translation by author) Benjamin: „Vielleicht kann man auch sagen es ist besser generalisierbar, oder? (Allgemeine Zustimmung) Oder dass es halt auf einer breiteren Menge von Daten oder von Eingangsdaten funktioniert."

[133] (Original Quotation /Translation by author) Oskar: „Also einfach gesagt geht es zum Beispiel um eine Nummernerkennung bei Autos, die nur bei blauem Himmel und Sonnenschein funktioniert, macht einer was bei Regen, Gewitter, Wolken… und es geht auch noch, dann ist es robust."

‘Robust’ is a powerful add-on to ‘functioning’. Its use in computer vision is widespread, but its meaning can differ. The more common meaning is as a synonym for reliable as it is a more forceful word for ‘better’ and more suited to generalisation. In computer vision if anything is robust, that also seems to mean that it is prepared for coping with unexpected events occuring in the real world. Thus, robust functioning is more than just functioning and as such, it is widely used to advertise the success of a certain algorithm, program or system.

[134] (Original Quotation /Translation by author) Ben: „Oder robuster.“ (allgemeines Lachen).

The Demonstration of Functioning Image Processing Algorithms

Robust functioning does not necessarily mean that something works perfectly well. It is still a matter of concern to me when it can be said that something functions in a robust way and it is far from being clear what robust exactly means. In this regard, a really interesting occasion at which the robustness of Image Processing Algorithms is negotiated and tested are public, or semi-public demonstrations of IPAs. As I said before, during my fieldwork, I witnessed and participated in several occasions at which IPAs were also demonstrated to a wider public. In the following, I refer to STS literature dealing with IT demonstrations and presentations, both as information, and also for comparison with my own empirical observations of computer vision demonstrations.

IT Demonstrations and Presentations

Demonstrations of information technology (IT) are "occasions when an arrangement of computer hardware and software is presented in action as evidence for its worth" (Smith 2009: 449). Smith discusses the structure, role and status of IT-demonstrations, because while the demonstration of scientific experiments have been studied in detail (cf. Gooding, Pinch & Schaffer 1989), demonstrations of technology have received far less attention (Smith 2009: 450). Scientific experiments have had the purpose of promoting science to both business and government as the source of solutions to practical problems (ibid.: 452). Smith starts the analysis of IT-demonstrations with the notion "that a scientific demonstration is a reframing of laboratory work. That is, a demonstration frame constructs a presentable copy of the messy private experiment" (ibid.: 453). What is presented is an "idealized image of discovery" and "the scientific demonstrator is not really performing an original experiment but rather showing how it might be done" (ibid.: 453). Collins (1988), who analysed two television "public experiments" in the 1980s (rail containers holding nuclear waste would remain intact following a high-speed collision and anti-misting kerosene aerosol fuel could prevent the sudden occurrence of fire onboard a passenger aircraft), argued that both were deceptive because they were presented as experiments, but instead, were demonstrations in the sense that they were carefully designed with known outcomes that supported particular viewpoints in their respective public safety debates (cf. Smith 2009: 456f.). Smith shows, how IT-demonstrations attempt "to simulate a hypothetical future episode of a possible technology-in-practice, with the demonstrator playing the part of a user" (ibid.: 462). A really well educated user I may add. Bloomfield & Vurdubakis (2002) also see demonstrations as depictions of a future.

Before my main fieldwork period started, I had already informally attended two computer vision presentations, and I noticed that the presenters aimed to show what computer vision is capable of, by referring to a large number of images and videos. When watching these videos as an outsider, I really got the impression that they were showing real-life scenarios - technology that is already in operation - but as soon as some background information is available, it became clear that most of these images and videos are "only" test data where many of the lab researchers are even recognisable. This, of course, is also due to data protection issues and legal reasons because it is easier to get informed consent from laboratory staff than it is from people present in public space.

When browsing through one of these presentation slide shows, phrasings such as "Bildauswertung auf Erfolgskurs??" ("Is Image Processing tipped for Success??"), or "Videosensorik – hohe Erwartungshaltung an die Technik" ("Video Sensor Technology – great Expectations of Technology") can often be found. Here, the presentors were trying to establish a relationship between the technology presented (image processing, video sensor technology), great expectations, and the possibility of future success. Even though the computer vision presenter explicitly referred to the relation between the technology presented, great expectations and future success the assessment of this relationship is transferred and virtually outsourced to the audience by the means of careful phrasing (expressed through the question marks and the phrases such as 'great expectations of this technology').

Another concrete example of the role of IT demonstrations and presentations in connection with IPAs can be found in a contribution to the German technology journal

Technology Review (Heft 07/2011). The following quotation refers to the presentation of commercial face recognition software for mobile phones:

> „Wie das konkret aussehen kann, demonstrierte die US-Firma Viewdle in diesem Januar auf der Technikmesse CES in Las Vegas: Eine Handykamera nimmt eine Gruppe junger Damen auf, die Viewdle-Software denkt ein paar Sekunden nach, und schließlich blendet sie zu jedem Gesicht auf dem Sucher den dazugehörigen Namen ein. Außerdem durchstöbert sie soziale Netze wie Facebook und Twitter nach den Profilen der Abgebildeten. Wird sie fündig, zeigt sie die letzten Updates in einer Sprechblase über den Köpfen an. In der freien Wildbahn funktioniert das allerdings noch nicht – die Vorführung in Las Vegas beruhte auf einer eigens für die Show gebauten Demo-Version. Aber Viewdle hat von großen Firmen wie Blackberry-Hersteller RIM, Chipentwickler Qualcomm und der Elektronikmarkt-Kette BestBuy schon zehn Millionen Dollar eingesammelt, um daraus ein fertiges Produkt zu entwickeln.“ (Heuer 2011: 29)

Even if the software does not seem to work in real life scenarios yet (referred to as ‘In der freien Wildbahn’ in the article), but only worked while using a special demo-version[135], it demonstrated the possibility and plausibility of such technology in practice. Consequently, for the spectator and also for the reader of articles in the media about the demonstration it would seem to be only a matter of more investment and more development until this technology will also work in real life scenarios. This specific temporal framing of the technology as being very close to a real, saleable product within a self-evident, linear development path dissolves the categories of future and present.

Simakova (2010), who analysed organisational practices of technology launches and demonstrations in the IT industry, characterised the marketing practice of launching in terms of the production of ‘tellable stories’; meaning how organisations talk new technologies into existence. She described 'tellable stories' in terms of a narrative connecting particular attributes of technology constituencies inside and outside an organisation (ibid.: 554). Through ethnographic inquiry when participating in the activities before the launch of new RFID technology, she witnessed the tentativeness of

[135] You can see a similar demonstration video on Viewdle’s webpage [Aug 9, 2011]: http://www.viewdle.com/products/mobile/index.html

the launch and was able to deconstruct its definitive status and representative website. This investigation into the preparations usually hidden to outsiders, also challenged the impression that a technology launch is a singular event and the climax of a linear process leading to such an event (ibid.: 568).

IT presentations and demonstrations are essential (Smith 2009: 465), but there is still little known about their value. For example, it is unclear what different kinds of presentation and demonstration actually take place and for what purposes these different presentations and demonstrations are designed. One example of a specific type of presentation is certainly the respective organisation website. Also media articles can be regarded as a form of IT demonstration and presentation. The pertinent question is then how specific expectations and possibilities in a technology are translated into them.

What is also unclear are possible differences in this area between basic or applied research, development and production that need to receive further attention in the future.

In my area of research I can make use of Goffman's frontstage/backstage conception (Goffman 1959), applying it to the discussion about IT demonstrations and presentations. Goffman, in his classic book about *The Presentation of Self in Everyday Life* gives the example of teachers, whose behaviour can differ in classroom and staffroom. This means that the behaviour of people is dependent on the region where they are acting and performing. Goffman defines a region "as any place that is bounded to some degree by barriers to perception" (Goffman 1959: 66). As the example of teachers shows, their behaviour is not only dependent on the location or place alone, but the respective region is also defined by how it is constituted; who is there at what time. In the classroom during lessons there are usually pupils and one teacher present. In the staffroom, as a rule, there are no pupils present but there are other teachers and this status quo seems to have been protected for generations. But what is the impact of region—that is not always comparable with physical space—to the behaviour of people? Goffman explains,

> "… that when one's activity occurs in the presence of other persons, some aspects of the activity are expressively accentuated and other aspects, which might discredit the fostered impression, are suppressed. It is clear that accentuated facts make their appearance in what we have called a front region; it should be just as clear that there may be another region—a back region or backstage—where the suppressed facts make an appearance." (Goffman 1959: 69)

This means, according to Goffman, that in everyday life on the frontstage some facts may be accentuated and some may be suppressed, but on the backstage both accentuated and suppressed facts appear, including "vital secrets of a show" (ibid.: 70); and "show" in Goffman's thinking is the everyday appearance and interaction with other people in the front region. In different places this can vary, and the frontstage and backstage may be close together and connected; only divided by any spatial means of delimitation. In such situations where front and back region are adjacent, "…a performer out in front can receive backstage assistance while the performance is in progress and can interrupt his performance momentarily for brief periods of relaxation." (ibid.: 70). This points, on the one hand, to the ongoing interaction of front region and back region, but on the other hand, does also clearly demarcate the two regions from each other in Goffman's conception.

There are however, also some examples of what Goffman calls 'backstage difficulties', where the front and back can be close together and switch with each other from one second to the next. For example, in radio and television, "… back region tends to be defined as all places where the camera is not focussed at the moment or all places out of range of 'live' microphones." (ibid.:72). In such situations, everything out of camera sight or microphone range might be in a back region for television watchers or radio listeners, but it is a front region for studio guests. Goffman brings in the example of the announcer holding up a sponsor's product "at arm's length in front of the camera while he holds his nose with his other hand, his face being out of the picture, as a way of joking with his teammates." (ibid.: 72). When the camera suddenly sweeps towards the nose, 'backstage difficulty' has occurred. This example also refers to the interchangeability of regions. This means that "there are many regions which function at one time and in one sense as a front region and at another time and in another sense as a back region." (ibid.: 77) Front region and back region can also change in time, meaning that regions are time-dependent. In Goffman's words:

> "…a region that is thoroughly established as a front region for the regular performance of a particular routine often functions as a back region before and after each performance." (ibid.: 77).

Goffman gives the example of restaurants or stores a few minutes before these establishments open to the general public. Whereas the dining area of a restaurant suddenly changes from backstage to frontstage with the general opening, other areas of the restaurant might maintain their status as backstage, for example staff locker rooms. In this case, the backstage character is built into the room in a material way that defines them inescapably as a back region (ibid.:75). Next to this material character of regions, regions are also dependent to performativity:

> „…we must keep in mind that when we speak of front and back regions we speak from the reference point of a particular performance, and we speak of the function that the place happens to serve at that time for the given performance." (ibid.:77)

In my view, computer vision presentations and demonstrations could represent the frontstage of computer scientists' work, while the computer vision laboratory is more probably the backstage, in which computer scientists are usually among themselves. At least analytically, I conceptualise the inside of the lab as the backstage and what happens outside it as the front stage, whereas actions that usually take place inside the lab can temporarily also take place in protected areas outside it, for example at the installation site of a demonstration, as we will see in a moment. Vice versa, it is also possible for the actual backstage of the laboratory to become the frontstage, for example when a demonstration takes place inside the lab. Nevertheless, confronting practical action and backstage behaviour and language and comparing with computer vision presentations and demonstrations frontstage, might be a promising way to examine what Jasanoff and Kim described as "the understudied regions between imagination and action" (Jasanoff & Kim 2009: 123) and to understand how the functioning of IPAs is demonstrated to a wider public.

On the Frontstage of Computer Vision: Demonstrating Image Processing Algorithms in Action

In the last days of my field work I was able to participate in a computer vision demonstration. I did not only participate in the demonstration, in fact I was an active part of it, as I assisted the computer scientists from the lab with the installation and acted as a test subject later on. The demonstration was part of a university exhibition at the event of the presentation of an Austrian innovation award. For that purpose, a separate area was allocated to four different exhibitors and was opened to them two hours before opening to the public, in order to set up the stand. This installation process prior to the demonstration can be characterised, following Goffman's theory, as a temporary back region or backstage that with the opening to the public changes to a front region. It can be seen as situated within the transformation from backstage to frontstage.

Most time was spent during the backstage installation process finding the optimal collocation of the required equipment, especially the optimum balance between an ordinary network camera, an additional optical sensor (Microsoft Kinect) and a mattress. The mattress was necessary, because the demonstration showed an optical 'fall detection' system and thus, was supposed to prevent injuries from "test falls". The mattress also had the advantage that these "test falls" always had to be performed at this one location which was necessary for a correct detection of falls.

Kinect and network camera were connected to a laptop each and were installed on tripods at a height of about 2.5m and arranged at an ideal distance and angle to the mattress and tested in situ. This took quite a long time, because both Kinect and camera had to be carefully placed in order to represent the mattress correctly in the field of view. One advantage of this test site was that there was enough free space for arranging the equipment in the best way for supporting the functioning of the fall detection demonstration. In the first arrangement, a problem emerged for the two computer scientists involved in the installation process. Because the background which was the wall behind the installation was mostly grey, one of the two systems being demonstrated (the one with the network camera) looked as though it would fail because the designated test subject (me) was wearing mainly grey clothes that day. The problem was due to the difficulty of differentiating between grey background (wall) and grey foreground (me). This meant that for the camera and the connected IPA it was a problem of distinguishing me from the wall, because of the similar colours. Therefore the background wall was changed to white. In this case, this change was not too hard to achieve, because it was possible to change the position of the Kinect and camera on the tripods so that the background behind the mattress was a different wall and it was white. The other option was for me to change my clothes, but this turned out not to be necessary because the arrangement with the white wall had been achieved.

As soon as this “segmentation problem” was solved, another one emerged. Unfortunately the site of our demonstration was in front of a storeroom that was in use before the event. At the beginning of the installation process the mattress had to be removed repeatedly, because event equipment had to be put into the storeroom. The whole installation process had to be repeated again and again. Therefore, the setting and camera calibration took quite a long time, because it had to be rearranged several times, due to the marginal relocation of the mattress following staff interruptions, which changed the distances and angles between camera/visual sensor and mattress where the falls were to occur.

Unfortunately, or maybe fortunately, I accidentally slightly touched one of the tripods just at the very moment when everything had finally been set, so everything had to be recalibrated. After this, the tripod with the camera and visual sensor on it was removed to a safer place in order to prevent the necessity of another calibration during the demonstration following incautious acts of visitors or the autonomously moving robot from the next exhibition stand. Just in time for the opening, everything was set and ready for demonstration.

I was the main test subject in the demonstration, as I have already mentioned. Our two computer scientists were quite lucky to have me as I had already observed and been trained in falling the right way in the lab to activate the fall detection alarm, unlike any visitors to the demonstration as test subjects who may even have been dressed in white which would have raised the problem of segmentationagain. During preparation work in the laboratory one of the computer scientists had said that he should set a low parameter threshold for easier activation of the fall detection alarm. For the demonstration, it was preferable to risk false positive instead of false negative alarms. This meant, the possibility of the false detection of a non-fall was set higher than the possibility of not detecting a fall. This also meant that the number of possible false positive results (in which the subject did not actually fall, but a fall had been detected) were likely to occur more frequently. As an alternative to the worst case scenario of not being able to present a running system, the two computer scientists were advised by their supervisor to take a video to the demonstration that shows the abilities and proper functioning of the fall detection system.

Just before the exhibition opened and the invited public could witness the demonstration, the two computer scientists were quite worried about the possibility of their system failing, so everything was tested extensively before the first demonstration for a visitor. These worries were also due to the fact that it was the first time they were demonstrating the system to people outside their laboratory. At this point the development of the system was in its early stages and so, in the "back region" of the laboratory they had designed a special demonstration version with the main purpose of presenting an executable, which means in technical terms; a running system. One uncertainty of many, was how the system would react to unknown test subjects; visitors that would like to test the system for themselves. Until then, the presented system had only been tested on supervisors, other lab members and me, but not in a systematic way that would allow conclusions to be drawn about general and "robust" functioning dependent on user differences. It was also useful to have a "trained" test person like me for another reason: To be able to detect falls, the system working with the Kinect sensor had to be calibrated first. To do so, the test subject had to stand in front of the sensor in what one computer scientist called the 'Ψ' or "hands up" position to measure the basic proportions of the body and to activate the device, after which the detected person appeared like a skeleton on the screen. When a visitor wanted to try out the fall detection system, the calibration process was not possible, probably because this person was wearing a wide coat and so the usual body frame was not recognised during the "hands up" procedure as the nearly rectangular arrangement of upper arms and upper body was hidden by the wide coat. As a consequence the external test person had to take off the coat and the computer scientists calibrated again. So calibration, as well as fall detection worked in the end.

The site of interest, the mattress on the floor, was shown on a large screen so that visitors to the demonstration could witness both the "real" fall and the depiction of the fall on the screen. Every detected fall was shown to the visitors edged in red on the screen as soon as the fall had been detected. It happened quite a number of times that the computer scientists had to point out explicitly to the visitors that the red frame meant that a fall had been detected. This fact had frequently not been clear to the visitors. Therefore, the demonstration was not self-explanatory but needed guidance and explanation by the computer scientists. I also got the impression that people expected a little bit more, maybe something like loud acoustic alarms or other special effects; something spectacular. Nevertheless, on the frontstage of computer vision, by means of this demonstration it was conveyed to the public that something like visual fall detection does exist. It was demonstrated that such systems are still in their infancy, but that they already work. A person falling down was automatically detected by a system making use of a camera and another optical sensor. That this function was achieved under very fragile, stabilised, carefully designed, arranged and controlled conditions that were achieved backstage, before the demonstration started, was hidden to the visitor. This did not mean that the public was deceived or fooled, but that the very specific setting of the demonstration and also what I described as the 'Regime of Functionality' required the presentation of a functioning system that consisted of what Goffman calls "accentuated" facts.

As cited, Smith showed how IT demonstrations attempt "to simulate a hypothetical future episode of a possible technology-in-practice, with the demonstrator playing the part of a user" (Smith 2009: 462). In the case of the 'fall detection' demonstration only a very small part using accentuated components of a whole 'fall detection' system was presented in order to show that, in principle, it is possible and plausible to automatically detect a fall using the means of computer vision and image processing. A very attentive and well informed visitor of the demonstration would have realised that the detection was achieved by either a network camera or another optical sensor (Microsoft Kinect) that was connected to a laptop in the background, using special software that analysed the output of the cameras and the observed site at which the falls occured. The results of the detection were then presented on a screen that was connected to the laptop. The decison-making processes of the IPA was neither visible nor comprehensible for the visitor.

In addition to the demonstration of the technical component of 'fall detection' there were many visitor questions raised about the practical application and embedding of it into operative systems, or in other words, more generally speaking, there were questions raised about the significance of the technology. In this regard the computer scientists´ answers never challenged the 'functioning' of fall detection itself. They had no reason to do so as they just presented and demonstrated the functioning of the fall detector on the frontstage. This meant it was always assumed that the technical constituent of fall detection worked, even though the development of such a system was at the prototype stage and there were still many limitations, restrictions, and uncertainties, especially when implementing the basic algorithm in software; and consequently, the respective software in greater sociomaterial assemblages.

The computer scientists' comprehension of these questions usually moved in another direction. The realisation and implementation of a 'ready-to-use' product or system was presented in a very clear and well-elaborated way and as such it did not seem to be only a product of their imagination. What happened here went far beyond the pure technical constituent of the system that had been presented at the demonstration, because in order to make sense of the system, the computer scientists had to establish it as part of a ready made product that really did exist within a network of different actors and had been designed for a concrete purpose. In this case, the purpose of fall detection was to establish it within the framework of "Ambient Assisted Living", already mentioned in the previous chapter.

So, the fall detection system was presented as care technology and more concretely as emergency technology for elderly people, in order to enable, secure and facilitate their living in their own homes. As told during the demonstration the homes of the elderly could be equipped with fall detection sensors that would detect possible falls and send an alarm or notice to an outside person or organisation in order to call for help or assistance. What exactly was going to be sent and to whom was not yet clear and had still to be negotiated, but due to data protection issues it is likely that images could not be sent outside the home and could not even be saved locally. So, this device was presented as being privacy enhancing as no images were broadcast. This visual sensor approach has—in comparison to other fall detection or home emergency systems such as call button devices that have to be continuously carried on the body—the advantage that theoretically, in emergency situations the person who has fallen does not actively have to take action (e.g., press the call button), but the emergency is recognised even so and reported automatically.

The presented fall detection sensor for elderly people in ambient assisted living environments was embedded in the sociotechnical vision or concept of 'Smart Homes' (see Chapter Three) and it also ties in with, what recently was named telecare technologies (cf. Oudshoorn 2012). As distinct to telecare technologies such as devices monitoring blood sugar or blood pressure that are "aimed at monitoring and diagnosing a variety of chronic diseases at a distance" (ibid.: 122), "tele-emergency" technologies, such as optical fall detection are aimed at monitoring and diagnosing not chronic diseases, but singular, extraordinary events (e.g. falls) from a distance. This means that telecare technologies are interwoven into the daily routines of people and also need their cooperation (e.g. blood pressure has to be taken), whereas tele-emergency technologies only come to the fore in extraordinary, emergency situations. What telecare and tele-emergency technologies have in common is the importance of and dependency on place (ibid.). Oudshoorn showed "how places in which technologies are used affect how technologies enable or constrain human actions and identities" and "how the same technological device can do and mean different things in different places" (ibid.: 121). She notes that "sites such as the home are presented as 'tabula rasa' in which telecare devices can be introduced unproblematically" (Oudshoorn 2011). In her empirical research on German and Dutch telecare users and non-users, Oudshoorn showed how telecare devices reconfigured and transformed the home from a merely private place to a hybrid space of home and electronic outpost clinic, in which patients were expected to observe very precise schedules in order to keep the system running (ibid.: 129).

This brings me back to the fall detection demonstration and especially to the non-accentuated and supressed facts; all the uncertainties, limitations, restrictions, and special arrangements that did not appear on the frontstage during the demonstration, but that are crucial for an implementation at the imagined places of use; the homes of the elderly. During my field observations in the laboratory I witnessed a strong tendency among the computer scientists towards the view that elderly people are not willing to accept such systems in their homes. In this view, the home and, in this case, particularly the right to privacy inside the home was presented as anything but unproblematic. Nevertheless, it was part of the report on possible future uses during the demonstration to find out how the acceptance of such technologies could be achieved and how the development of these could go in the direction of protecting privacy[136]. This accentuation of acceptance and privacy issues did push other sociotechnical questions concerning the functioning of fall detection in the locations of use, into the background. I already wrote about major concerns about this. For example, the background/foreground segmentation problem that was solved in the demonstration of the 'fall detection' system by changing the background, but this procedure can hardly be influenced in elderly peoples´ homes. It is unimaginable that someone should always have to wear the same dark clothes and live in an apartment with dark walls and furniture keeping the lightning to a minimum, just in order to be able to detect if and when they fall.

As another example, the calibration problem occurred when the camera position was changed slightly, due to my clumsiness during the demonstration installation. The

[136] In this regard Suchman, Trigg & Blomberg (2002: 166) reported that for designers "…prototyping represents a strategy for 'uncovering' user needs…." From this perspective the 'prototype' is understood as a mediating artefact in designer-user interactions (ibid.: 168) that realises the involvement of (specific) user needs in technology design. In contrast to the process of uncovering user needs, the process to achieve (user) acceptance is a different one from my point of view, because it conceptualises the presented technology as "ready made" rather than still adaptable.

calibration of the system had to be repeated at the demonstration, so what would the situation be like in the home of an elderly person?

Occlusion is another problem. At the demonstration the carefully chosen site of interest (the mattress) was in full view of the camera, but what would the situation be like in private homes? In a home there is built-in, as well as moveable furniture and the messiness and diversity of living styles can easily disrupt the direct view of a camera onto possible places where falls can occur. The differentiation between background and foreground, calibration, and occlusion problems observed in the public demonstration are three examples of what could problematise the future implementation of a fall detection system in the homes of the elderly; or in homes in general. From my point of view based on my observations, it is to be expected that along with the implementation of optical fall detection sensors in private homes, private homes themselves need to change and somehow be adapted to the logic of such a system. Amongst other things this means sufficient illumination of all parts of the apartment, rearrangement of furniture and everyday objects, and a change in personal behaviour so that the configuration baseline of the system is not impacted negatively (as an example, my accidental displacing of the camera tripod). In brief, my conclusion is that place, technology, and human behaviour have to be synchronised for a functioning sociomaterial assemblage to be created and maintained. As an example, the proper functioning of this sociomaterial assemblage requires, amongst other things, for people to behave in a way that does not alter the camera calibration.

At this point the question has to be raised of whether human behaviour can appear in a way that image processing and behaviour pattern recognition algorithms can cope with sufficiently. My observations in the laboratory showed, as reported in the previous chapter, that in the case of the present fall detection system it is a matter of the differentiation between falling down and lying down. Similar to the questions raised with facial recognition systems, namely if there is enough variation among faces in order to differentiate among people, here, there is question of whether there is enough difference between critical falls that call for emergency action and intended actions similar to falling, such as lying down or doing exercises on the floor. Consequently, we need to think about the meaning and setting of false positive and false negative alarms and their outcome. Who is responsible then? How are potential algorithmic decisions—in this case deciding whether a critical fall has occured or not—to be assessed in legal terms? What might the legal status of IPAs be? I shall discuss these issues in the final and concluding chapter when reflecting on the politics of Image Processing Algorithms and especially about the shift from visual information sorting to visual information decision making.

Conclusions

The demonstration of the fall detection system was a great example to me of how technology is not only shaped by computer scientists, but also how a possible future society in which the respective technology is an integral part, might even be shaped by them when ideas of future uses are embedded in sociotechnical imaginaries. Exactly here, an interweaving of technology and society take place. This means that the material-semiotic configuration of a technical artefact and the technical process of producing does not only take place in the laboratory, but also in the social practice of telling stories about the future of this respective artefact or process, seen in my case, within the framework of a demonstration and presentation of a fall detection system. What is important to note in reference to my observations is the link and interdependence of these stories told of future uses, an understanding of the IPA and the system in which it is going to be employed, and the demonstration and appearance of the operative readiness of the technical system. These visions of future use could not exist without any real demonstration or proof of viability; and simultaneously, a system working perfectly, would be meaningless without such visions. Only the successful interplay of envisaged future uses and a demonstration of viability would seem to facilitate further realisation, development and use. This successful interplay involves what might be called a balance of power between what can be seen as a functioning, viable system, and the promises and expectations of future visions. Or, in other words, it is important to recognise "how wide the gap separating images from practices can become before an uncontrollable backlash is provoked" (Nowotny et al. 2001: 232).

Here, from a public understanding of science and technology perspective, the question arises to what extent an external observer who could potentially be affected, is able to differentiate on the one hand, between a demonstration in which the technology is presented as viable, because it accentuates the functioning aspects in a carefully designed setting promising future success and, on the other hand, the reality of technical uncertainties including the fact that functioning is always probabilistic, conditional and temporal. In this regard, my observations illustrate and connect to other findings. In short, these findings show that many of the technical uncertainties of bench and laboratory science are often invisible to the wider public (Borup et al. 2006: 272).

So why are these uncertainties invisible and often actively hidden away? Why does a conditionally viable system have to be presented to a wider public at a time when it is clear that more research still has to be done to make the system properly viable? As I already indicated, this what I call the 'Regime of Functionality' that is closely linked to what has been blamed for changing the norms and practices of academic, technoscientific work and 'new school' entrepreneurial scientists (cf. Lam 2010). I believe that a powerful guiding principle in the area of computer vision and IPA research into which I have been able to delve, is this very 'Regime of Functionality.' This is especially manifested - as was previously shown - in the demonstration to a wider public of IPAs and the systems into which they are integrated, bringing with it the danger of arousing great expectations more akin to science fiction because possible unknown and hypothetical future applications of IPAs and their fictional abilities are brought into the present as if they already exist in that specific form. As such, what was displayed in the computer vision demonstration is as much a fictional character as HAL 9000 from Kubrick's film and Clarke's novel *2001: A Space Odyssey* (see Chapter Three). A 'diegetic prototype' (Kirby 2011: 193ff.) that visibly demonstrated to a public audience, the utility and viability of the product (ibid.: 195) by accentuating what works, suppressing what does not work (yet) and by embedding the displayed system in meaningful tales of future uses. Thus, the performance of the fall detection demonstration was part of a dynamic sociomaterial "assemblage of interests, fantasies and practical actions" (Suchman, Trigg & Blomberg (2002: 175).

Temporality is a significant factor in respect of when and what kind of stories, fantasies, visions, promises and expectations are formulated about IPAs especially within the framework of the ongoing commodification of research. We can only assume that in the future, commodification of computer vision research will reduce and blur the timespan between present and future and simultanously also lessen and blur the difference between IPAs in the making and ready-made IPAs, as hard facts regarding time differences are missing. It can be expected that this trend towards a shortening and disappearance of a time lag will have significant influence on how societies relate to and trust in IPAs and their abilities. In this regard, from my point of view, it can be foreseen that a period of great expectations or even hype (cf. Bakker & Budde 2012) could be followed by a period of great disappointment as these very expectations could not be fulfilled. However it is also clear that such an increasingly obscure time lag will make it more difficult (in particular to outsiders or affected people) to judge whether an inherently opaque IPA is viable and true at a specific point in time.

In my view the 'Regime of Functionality' can be interpreted as a strategy and reaction of computer scientists to the shifting boundary between university and industry, between academia and business—as already described in this chapter—in their everyday working lives. However, in order to procure funding and secure resources for the future, and this means also the safeguarding of existing jobs or the creation of new job opportunities, the public has to be told and shown only the "good things" about their work in its early stages. This means to a great extent, accentuating favourable findings and in the case of computer vision, means showing and presenting what is (already) functioning more or less properly. Additionally, as indicated before, the strong connection of a university laboratory to a commercial company offers new job and business opportunities beside an academic career path. Lam (2010) presented a typology of scientists in the framework of university/industry ties emerging from in-depth individual interviews and an online questionnaire survey with UK-based scientists from five different disciplines: biology, medicine, physics, engineering, and computer science (ibid.: 312). She pointed out four different types: 'old school' traditionalists, hybrid traditionalists, entrepreneurial hybrids, and entrepreneurial scientists. Whereas old school traditionalists have the strong belief that academia and industry should be distinct from one another and for them, success should be pursued primarily within the academic arena, entrepreneurial scientists see the boundary between academia and industry as highly permeable and stress the fundamental importance of science/business collaboration. That this simple dichotomy fails in reality shows that the two hybrid categories are dominant: more than 70% of subjects in all disciplines can be allocated to the hybrid categories (ibid.: 317f.). What I witnessed during my field observations and partially described here, might be a combination of the two hybrid categories. This means that at the one hand, the commitment to the distinction between academia and industry which also includes a strong commitment to core scientific values, is achieved. On the other hand, in the course of a 'resource frame' (ibid.: 326) benefits of the extension from the solely scientific role to application and commercialisation following long years of basic research, are seized at (cf. Lam 2010: 325f.). At this point it has to be noted, that what I call the 'Regime of Functionality' as the basis for the securing of resources is not the only guiding principle in the everyday work of computer vision scientists. There are also various other organisational and individual, personal motivations that were not the focus of this analysis. One example in this context is the fun experienced by the mainly, (but not only) male computer scientists (cf. Kleif & Faulkner 2003) when giving (technological) things a playful try as described in the previous chapter. In my view, further research is certainly needed in this area, to explore not only individual perspectives and motivations in more detail, but also—seen in a more universal framework—the 'epistemic living spaces' (Felt 2009: 19 ff.) of (Austrian, European etc.) computer scientists working in the field of computer vision and IPA; research that nevertheless influence the ways computers are able to see and recognise.

Chapter Seven

Towards the Social Studies of Image Processing Algorithms (SIPA)

Computers are able to see. They have the ability to recognise: objects, people, faces, four to six facial expressions, specific (suspicious) behaviour patterns and falls, to name a few. Over the last few years, following my interests and my 'visiographic' strategy in researching computer vision and analysing in particular, Image Processing Algorithms (IPAs) and, based on an interdisciplinary, multi-perspective approach that brings together the fields of Science and Technology Studies (STS), Visual Culture Studies and Surveillance & Identification Studies, I would definitely affirm the statement: Computers are able to see and to recognise. Nevertheless, there is something very important missing in this statement that is crucial, but often disregarded or not mentioned: Computers are able to see and recognise in particular ways that are 'situated' and partial. If human vision is to be taken as a reference to which computer vision is to be compared—and this was the starting point for this analysis based expressly on Lucy Suchman's work on Human-Machine Reconfigurations (2007)—the realisation soon follows that human vision works in a more holistic and interactive way (cf. Bruce & Young 2011) but, similar to computer vision, human vision too is always 'situated' and partial (cf. Burri 2013). *Thus, the first basic insight to become aware of, is that both human and computer vision are fundamentally social, cultural, and political entities.* That means they both rely on diverse, multiple and changing societal negotiation and interpretation practices and while they are interconnected in many ways, they still differ significantly on several levels. For example, as might be expected, computer vision in its current state is rule-based. Thus, the impression might arise that it is also more predictable, objective and neutral than human vision although many results of IPAs are complex and opaque, making comprehensibility for humans more difficult. Especially when it comes to far-reaching, often binary decisions, made by IPAs it is important to acknowledge the sociocultural and political dimensions and the significance of these decisions that can be subsumed under the title of "*The Politics of Image Processing Algorithms*," one particular form of "*The Politics of Seeing and Knowing*."

It is essential to understand that IPA selections and decisions are based on specifically situated classification and standardisation practices that did not come into being artlessly and that rely on an objective, neutral, technical or natural foundation. As IPAs are fundamentally based on different forms of classification and standardisation, they pose—to use the words of Timmermans and Epstein—"sharp questions for democracy" (Timmermans & Epstein 2010: 70). This classification and standardisation "may (then) come to function as an alternative to expert authority" (ibid.: 71) and they might be contained as such "in rules and systems rather than in credentialed professionals" (ibid.). It is essential to note that all of this happens in the context of fundamental sociotechnical transformations that come along with the "grand narrative" (cf. Law 2008: 629) processes of digitalisation, computerised automatisation, and "smartisation" of devices, practices, and processes. Amongst other things, these phenomena of digitalisation, automatisation, and "smartisation" seem to bring with them, the reduction (or displacement) of human labour; they promise to create more security and safety; they seem to guarantee more economic efficiency through better and more objective decisions; they even pledge to provide more justice by acting more truthfully and neutrally. In short, digitalisation, automatisation, and "smartisation"—and as a fundamental part of these, IPAs—promise a better life for everybody and thus, a better society.

Recognising that more and more agency and authority and connected to this, great expectations, are attributed to IPAs that are only one specific form of automatisation, democratic societies are advised to discuss and reflect upon the sociopolitical distribution of responsibilities and power, especially among IPAs, the "smart" devices and automated systems they are part of, human operators and humans affected by these technologies. This discussion is inevitably also a debate on "what is good and desirable in the social world" (Jasanoff & Kim 2009: 122), because it sheds light on the (power) relations among various human and non-human actors and how they can or cannot live together. It reveals who is able to act in desired ways and who is suppressed in his or her way of living or acting. It shows who benefits from computers that see and therefore boosts their development or uses their capabilities, and who is affected adversely or even discriminated against through their use. It is clear that those constructing these "computers and machines that see" by developing and implementing IPAs, consciously or unconsciously exercise power, because they are able to decide what counts as relevant knowledge in every particular case (Forsythe 1993: 468). Thus, they are on the one hand in a position to decide and define what is real and what is true in the world, and on the other, they are simultaneously in a position to decide what is to be defined as desirable and undesirable, what is good and what is bad. It is then a way of "constructing uniformities across time and space through the generation of agreed-upon rules" (Timmermans & Epstein 2010: 71). The problem is that these "agreed-upon rules" are very particular and situation dependent and they might contain a wide array of tacit values and assumptions that represent the viewpoints of particular individuals. This especially is problematic when taking into account the technical authority attributed to a technological device or system as was the case with the "Automatic Toll Sticker Checks" (AVK) in operation on Austrian motorways referred to in Chapter Four, for example.

This thesis was written to provide a theoretically and empirically grounded basis for these important sociopolitical discussions and reflections. It analysed IPAs in order to explore human/computer vision relationships from different perspectives and angles and tried to follow these objects of interest to different places and sites. As such, it took a broad multi-perspective approach to cope with the highly complex, messy sociotechnical phenomenon of automatisation that is continuously in the making while simultaneously already making a difference. It elaborated on the fact that all attempts at giving computers and machines the ability to see, are in fact attempts at producing, processing and understanding (digital) images with the help of computer algorithms. Therefore, it made sense to understand the process of giving computers the ability to see as the sociomaterial process in which Image Processing Algorithms are developed, produced and implemented in devices or in larger systems; advertised, used, talked about, criticised, or configured. In short, processes in which IPAs are negotiated and formed at several sites and in several situations.

In what follows I will summarise the most important aspects and lessons learned, chapter by chapter, in order to bring them together and provide a starting point for drawing analytical conclusions. These conclusions are followed by the outline of a conceptual reflection framework for further analysis into the development of IPAs ("Social Studies of Image Processing Algorithms" [SIPA]).

As I elaborated in **Chapter Two**, based on the insights made in the fields of visual culture and surveillance and identification studies, human vision inevitably, is historically and culturally specific in all of its conceptions (cf. (Tomomitsu 2011; Kammerer 2008; Burri & Dumit 2008, Rövekamp 2004). This means human vision differs within time and from culture to culture. Meanings of entities to be observed are changing over time and they vary in different areas of the world. Who is interacting with whom, what is to be seen and known can bear very different meanings. What significance the event or object to be seen and observed has, is dependent on situated negotiation within a social practice. One such negotiation practice I explicitly referred to in Chapter Two was the historic case of Martin Guerre who had to be identified and recognised at court proceedings in 16th century France. The witnesses at the trial had to compare the appearance of a man who claimed to be Martin Guerre, to the picture of Martin Guerre in their imaginations as he had looked when he had left the place some years before. As unclear, and thus negotiable this recognition process was, so also are processes today, in which computers are part of the recognition process. The presence of seemingly refined facial recognition algorithms, the very image of technical sophistication and neutrality, does not close what Groebner named the threatening gap between appearance and description (Groebner 2001:21). As explained in Chapter Two, it is still a matter for complex interpretation and time-consuming human intervention in how far two patterns of appearance and description, between body and registered identity, or between human behaviour and pre-defined ground truth behaviour fit together. To sum up, it is a persisting process of negotiation that is taking place in sociocultural practices. The modes of these processes have changed, but the queries have remained the same. One of the central challenges in this regard is the question of visual expertise. It has become apparent that visual expertise is its own form of literacy and specialisation (cf. Burri & Dumit 2008: 302) and it has not been clear from the start who or what has this visual expertise. From this situation the question arises if and to what extent IPAs are, or could be positioned or perceived as visual experts. Referring back to the Martin Guerre case the question could be asked of whether it would have been possible to recognise the wrong Martin Guerre by the means of facial recognition or other IPA based technologies. What would have been different if such technology had been in use? This speculative question can only be answered adequately if society has a clear understanding of how much agency and authority is ascribed to the respective technology of facial recognition or similar IPA-based technology and how these are integrated within sociomaterial assemblages.

This makes it clear why it is important to understand how IPAs and devices or systems based on IPAs work, how they were made to work and what form of authority and visual expertise is attributed to them and by whom. This negotiation generally takes place in the academic and business fields of the (applied) computer sciences. It refers especially to the ways in which social order and social reality is inscribed into IPAs in computer vision laboratories. I exemplified this in **Chapter Three** with the example of automatic recognition of cows and what influence the selection of training images by the computer scientists can have on the ways cows are perceived by computers and subsequently also by humans that make use of this automatic recognition. What is real, and what is a real cow, is configured in the computer vision laboratory in such a case. The result is not an objective, universal view, but is situated and particular as is the case with human vision (cf. Burri 2013). In a familiar culture, cows might be adequately recognised as cows, but outside this culture, some kinds of cows might not be recognised as such because they differ too much from the prescribed standard template of the norm-cow within the program. So, *the second basic insight to take along from Chapter Three is that both computer vision and human vision are always situated and particular*

However, the meaning and social status (e.g. regarding visual expertise) of IPAs is not only a matter of negotiation in the field of the computer sciences, it is also negotiated within a broader background. One of the most important characters equipped with (black-boxed) IPAs is the "most famous computer that never was" (*The Guardian, June 2, 1997*): HAL 9000 from Kubrick's movie *2001: A Space Odyssey*. HAL 9000 is described as a cultural icon and "has come to serve as a leitmotif in the understanding of intelligent machines and the dangers associated with them" (Bloomfield 2003: 194). HAL 9000 and parodies of it, for example in the Simpsons episode *House of Whacks* (2001), to which I also referred in Chapter Three, mediate powerful visions and images of how future smart worlds with intelligent machines, of which IPAs are a vital part, could appear once they have been implemented and applied. In the case of the Simpsons´ smart 'Ultrahouse' I was able to show how close its conception is to visions of smart homes recently described by computer scientists and taken up in broader, socio-political discussions about smart futures. Such visions, whether they derive from popular culture or from the computer sciences, transport specific expectations and promises in the context of artificial intelligence and intelligent machines that influence and drive the development of IPAs and other relevant sensor technology to a considerable extent. What it comes down to, is that the imagery and visions of future, intelligent computers that can "see," are far beyond the current capabilities of IPAs and computer vision, because they present a more holistic, "human" version of vision. As such, it clearly shows the degree to which human and computer vision is interconnected, both continuously referring to each other. *So the third insight to underline, is that visions of the future influence the ways societal actors view, appropriate, and evaluate IPAs with all their capabilities and limitations. It can be clearly stated that the imagined capabilities are massively overestimated, while on the other hand, limitations are not taken into account in public understanding of IPAs. Thus, a wide array of promises and expectations is generated that cannot be fulfilled in these imagined ways.*

Concerning this, the location of the development and deployment of IPAs might play a significant role. Local differences and particularities have to be taken into account, instead of assuming that there is only one universal, worldwide procedure. This is why I also referred to the specific situation in Austria, because Austria's specific technopolitical identity as a "Gallic Village" when it comes to the introduction and development of new technologies (e.g. nuclear power plants) (cf. Felt 2013: 15) might influence the ways computers are being taught to see, or even lead to a ban on these efforts. During my field observations in Austria this situation was palpable, as seen in the efforts of computer scientists to cope with strict data protection regulations and to gain acceptance for their work and their imagined end products. This means that the specifically, Austrian techno-political identity has both enabled and restrained national development in computer vision. All observations made in this study have to be seen in this context, making it obvious that in other locations my observations could have resulted in different selections and conclusions because of different techno-political identities. This assumption has to be taken into account when reading and making conclusions from the following empirical chapters.

In **Chapter Four** I followed Nelly Oudshoorn (2003) and analysed the sociocultural testing of Image Processing Algorithms in newspaper articles and publicly available documents; another site where IPAs are being negotiated, discussed and seen within a wider framework. I concentrated on one of the first nationwide systems already in operation in Austria that is based on image processing, pattern recognition technology: the so-called 'Automatic Toll Sticker Checks' ("Automatische Vignettenkontrolle", in short: AVK) on Austrian motor- and expressways. A recurring "de-innovated" narrative in the media articles was that it is the camera which is at the centre of attention. The camera, not the IPA was positioned as the central actor of the system and it was also the camera that recognises the presence and validity of toll stickers, automatically. IPAs were completely neglected in the reports. Rather, they were blackboxed within the 'automatic' and 'innovative' camera. This blackboxing reinforced the view of an all-seeing, "magic" technological object, the automatic, innovative camera that is able to fish out any offenders from the cars driving on Austrian motorways. AVK was, with the exception of one single, critical article not contested in the media articles at all. On the contrary, it was described as being unproblematic, functional and familiar camera technology that made sense, especially in economic terms, by facilitating the collection of more toll and toll fines. Beyond that, the message sent to the readers was that it acts as a neutral moral agent in order to accomplish justice and fairness on Austrian motorways. This means AVK was positioned and reported on as the ultimate means of making everyone pay toll.

As the AVK system was mainly evaluated on its economic success by exclusively providing increased detection numbers and sales figures (though not providing lower numbers in some years), public understanding of this technology was led in the direction of full viability of the camera system while simultaneously presenting its economic value and its sense of justice. AVK was presented as a successful, ready-made, autonomous system, whereas the indispensible need for human intervention in the recognition process was mentioned only as a sideline. Additionally, a vast number of uncertainties that come with any image processing, pattern recognition technology such as biases, error rates, probabilities, uncertainties and false positive or false negative cases were not made an issue of in the media reports. Therefore it is quite clear that this account could lead to a widespread public understanding of "smart" camera technology generating high expectations that cannot be fulfilled, especially when it comes to more autonomous systems. For example, if we take the information presented in the answer to the first parliamentary questions, posed in parliament (atopq1; January 26, 2009) on 159 false positive cases in the initial period of AVK seriously, it means that without human intervention there would have been 159 car drivers that were wrongly detected by the system as toll sticker offenders. These 159 car drivers would have needed to prove their innocence to the authorities in order to avoid paying the unjust fines. Thus, in my view it should be the task of critical social scientists to make the public understanding of uncertainties a subject of the discussion in order to avoid disappointment and injustice in the future. The dominant message arising from how AVK was presented in the Austrian media is very close to what was said about CSI Forensic Science in Chapter Two: It is " ...easy, quick, routine and epistemologically very strong" (Ley, Jankowski & Brewer 2010: 13). In both cases this view leads to an asocial representation of IPAs, science, and technology in public and political discourse that underpins the so called "CSI-effect" (Collins & Evans 2012: 906): the exaggerated portrayal of science and technology to the public. *Thus, the fourth insight to be highlighted is that in media reports IPAs were blackboxed within more familiar devices (e.g. cameras) and as such, these devices were presented as asocial and acultural entities, which puts them into* *the position of uncontestable, neutral and objective moral agents in public understanding and discussion.*

In **Chapters Five and Six** exactly this dependence on society and culture and thus the various uncertainties of science, technology and IPAs that were widely missing and blackboxed in the media reports and in publicly available documents referred to in Chapter Four, were dealt with. In Chapter Six, in particular, I discussed what "functioning" means in the context of IPA development and deployment. "Making things run" or "work" and connected to it what I call a 'Regime of Functionality' was identified as being a constitutive practice in the computer vision laboratory in which I was a participating observer. But what it actually means if something is "running", "working" or "functioning" was far from being self-evident. Rather, it was recognised as a continuous discursive negotiation process also dependent on context, place and time. *So, the fifth insight to place emphasis upon, is that 'functioning' in the context of IPAs is always conditional (e.g. it functions only during daytime) and probabilistic (e.g. it functions in 97% of the cases). Moreover, as the saying "making things run" indicates, it is not ready- made and universally available, but a matter of a very specifically, situated "making" and negotiation procedure that is subsequently being blackboxed.*

One particularly interesting situation in which the "functioning" of IPAs was negotiated was a computer vision demonstration of an automated visual fall detection system. The system, still in its early stages, was carefully designed and installed to accentuate certain functioning aspects and suppress non-functioning ones, especially in regard to possible sites of application. The demonstrated system was presented within a framework of meaningful narratives about areas of future application. Through these discursive practices this half-baked system was already presented as a "functioning" system, or at least as being practically fully developed, while in reality much work was still needed to reach such an established status. As such, it was a 'diegetic prototype' (Kirby 2011: 193ff.).

This case was particularly interesting, because I was able to observe the whole production process in a computer vision laboratory from the start. In Chapter Five I presented my findings regarding the process of designing IPAs in computer vision laboratories. The sociotechnical construction of a ground truth, a term frequently used in computer vision that I consider to be a constituting sociomaterial element because it contributes significantly to what is real, defined or perceived as real, was of central importance. A ground truth defines the reference model for comparison with observed behaviour or an object of interest. Following the specifications of computer scientists, it predetermines how the respective behaviour or object of interest should appear, in order to be recognised as such. It works similarly to a reference image on a passport or in a facial recognition database or in any other individual identification technology to which the image of a specific person of interest is compared. *Thus, sixth insight to be noted is that the sociotechnical construction of the ground truth in computer vision laboratories standardises and defines what is perceived as real and true.* It has to be added here that this standardisation and definition is not neutral, objective, or universal, but is markedly selective, subjective, situated and particular.

In Chapter Five, I therefore followed the processes of how society, social order, and particular modes of reality and truth are inscribed into and manifested in the ground truth of three different cases where IPAs were used. What was demonstrated is that in contrast to the technical authority and neutrality often assumed, personal, subjective views that were negotiated in different sociotechnical constellations in and around computer vision laboratories were inscribed into the respective ground truth and thus, inscribed into the ability of the computer to see and recognise. In doing so, I was able to show its profoundly sociocultural character and how IPAs and computer vision are socially situated. The sociotechnical construction of a ground truth is the key area in which the analysis, perception and thus, the "truth" of IPAs is determined. This process constitutes the "experience-based" knowledge on which basis in further consequence, the visual world is perceived by IPAs and thus, potentially also by the people making use of, or being affected by IPAs and their rulings.

Image Processing Algorithms and how they are developed, designed, negotiated, and implemented in "sociomaterial assemblages" (Suchman 2008: 150ff.) were the focus of this multi-perspective explorative study. In what follows from my empirical findings, I shall describe the consequential trend away from visual information sorting towards more autonomous decision-making regarding this visual information and what the implications of this trend mean. Subsequently, I will comment on ethical, legal, and social aspects (ELSA) of IPAs, because these are widely missing in current debates about automatisation, smart CCTV or intelligent cameras. I will argue that such an involvement is a prerequisite and indispensible for future development. Finally, based on the explorations and findings of this study, I shall suggest a conceptual reflection framework for further sociotechnical analysis and development of IPAs. Referring and connecting to the "Social Studies of Scientific Imaging and Visualisation" (Burri & Dumit 2008) I shall call this attempt "Social Studies of Image Processing Algorithms" (SIPA).

From Visual Information Sorting to Visual Information Decision-Making

What is often referred to as "smart" technology can potentially change society; for example the sociomaterial assemblage of "Smart CCTV" in which IPAs are an essential and constituent part of. Especially here, the current trend is away from visual information sorting to visual information decision-making that is most likely to impact society and social order profoundly. This ongoing sociotechnical change brings with it a change in materiality and environments at possible and actual sites of operation that might not always be desirable for those affected. Indeed, the realisation of the necessary technology and making it work in a viable manner is a prerequisite for these systems. An examination of these processes of sociotechnical change show that many different entities are in flux. Existing "sociomaterial assemblages" (Suchman 2008: 150ff.) have been set in motion. It seems that nothing in particular and nothing as a whole remains the same if such a process is implemented. So if society, or those in relevant positions are willing to develop, implement or use IPAs, there should always be an awareness that further entities will also change. In addition to this, the question arises of whether affected people and entities are willing to change at all. From a democratic, political perspective the best case would be for everyone involved in this process to be willing to change, so that change can happen in a positive sense that could be termed technosocial progress. The worst case would be if the majority of those affected were not willing to change, but change happened anyway, because a powerful minority group was strong enough to force a change. However, if even only a minority group was not willing to change, the question could be asked of whether there were any alternatives to be considered so that the unwilling minority would also be taken seriously. So the decision for or against "smart" technology with integrated IPAs, or rather a specific sociomaterial assemblage of which IPAs are a part, is fundamentally a political consideration from the very beginning.

While it is generally agreed upon, both in the fields of computer vision and surveillance studies that privacy and data protection issues are important aspects to consider, the relationship between the "functioning" of IPAs and the modification of the sociomaterial assemblages they are integrated in, is widely neglected but would also be essential within this debate. This affects in particular, the materiality of the environments in which the visual sensors or cameras that deliver the input for IPAs are integrated. In what follows, referring to two of my main empirical cases; the automated toll sticker checks (AVK) and fall detection, I present a trend away from visual information sorting towards visual information decision-making and the implications of this trend. By doing so I shall show how the modification of the sociomaterial assemblages in which IPAs are integrated is imperative in the process of developing and implementing IPAs. As such, it is a strong argument for the involvement and participation of further contributors other than exclusively computer scientists in the process of designing IPAs and "smart" machines.

The system of automated toll sticker checks (AVK, cf. Chapter Four) which is supposed to recognise whether cars on Austrian motorways are equipped with an obligatory, valid toll sticker on the windscreen, is my example of the automated **sorting of visual information**. What is often referred to as an automatic (or even all-automatic) system is rather what should be called a semi-automatic system, because it became clear that human inspection was still needed for a final decision. In the case of the AVK system, images taken by high-speed cameras showing windscreens with suspected invalid or missing toll stickers, together with images of the car number plates are sent to an enforcement centre for further investigation. Only then can "compensatory toll claims" be sent to the car owner (whoever registered the car) by administrative mail. It has to be mentioned here that it is no surprise that the first systems in operation are centred on cars because, as was often argued by computer scientists during my fieldwork, it is much easier to automatically detect and recognise cars and their behaviour in comparison to humans and their behaviour. The argument being, that the behaviour of cars is more predictable, cars are easier to distinguish from their environment and they usually move in highly standardised settings such as on clearly marked lanes on motorways. Even so, when the toll sticker monitoring system in Austria was introduced in 2007, other relevant nonhuman actors such as the design of the toll sticker had to be changed in order to be more easily read/seen by the respective IPAs. Additionally, there was also a ban on tinted windows in order to make automated recognition possible and thus, improve the viability of the system. Although these changes were made to improve detection in the already highly standardised setting of motorways in Austria, there is a need to leave the final decision of whether a transgression had occurred to a human operator. Thus, the AVK system, and in particular the relevant IPA pre-sorts suspicious cases by saving the respective images as proof. Subsequently these cases are evaluated by human operators in an enforcement centre. This means the final and definite decision is left to the human operator. Of course, preliminary sorting is also a decision process in determining whether a car is suspected of not having a valid toll sticker. This does still impact the final decision to a considerable extent, as it narrows down the number of selected cases in a specific way, but the human decision is final, in contrast to the IPA decision. If in doubt, the human decision overrides the IPA decision. This example is only one amongst many where similar technology of a sorting nature is implemented and therefore can be seen as a decision aid or decision assistant. An example in public perception would be when seemingly fully automated results or matches of fingerprint and DNA analysis fundamentally need interpretation and intervention by skilful human experts (cf. Chapter Two).

While visual information sorting of invalid or missing toll stickers is already in operation on Austrian motorways, at the same time, algorithms that come closer to really **making an autonomous decision,** or a fully-automatic decision are in the making. Here, my empirical example of this kind of algorithm is in the case of fall detection algorithms (cf. Chapters Five & Six) that in extreme cases, could decide between life and death in a case where a critical fall took place, but which was not detected by the automatic system as was expected. It is clear that such a case brings with it the problem of responsibility. Two questions arising from this, are first: who or what is responsible for a possible error by the IPA, and second how could such an error be reconstructed. I will come back to these questions in the next section of this chapter when going into the ELSA aspects of IPAs. Beforehand, I will reflect on the problematic implementation of IPAs in existing sociomaterial assemblages such as private homes.

During the demonstration of a visual fall detection system which I participated in during my field work, as described in Chapter Six, I realised that the system being presented which was still in its infancy, was being shown to the interested public as a functioning system on the frontstage. It was implicitly advertised as a ready-made product, but backstage—and here I mean the lab work and the on-site installation preceding the public demonstration—it was still a very fragile system that had to be carefully designed and installed with an accentuation of certain functioning aspects that suppressed non functioning aspects, especially in regard to possible sites of application. These possible sites were seen especially in ambient assisted living (AAL) environments, e.g. in homes for the elderly, in order to detect critical incidents such as falls and subsequently to call emergency services.

One example of the suppressed aspects is what can be called the 'occlusion problem'. In the demonstration, the cameras used —or visual sensors, as they were sometimes referred to—were in direct and unconfined sight of the carefully chosen area of interest which was a mattress that had also been used in the lab beforehand when testing and experimenting with the system. When I thought about the system after the demonstration I asked myself what the situation would be like in real private homes: the imagined sites of application. Based on my observations at the site of the demonstration, I tried to imagine the situation in real private homes[137] in comparison to this site. In private homes there would be both built-in and moveable furniture, there would be untidiness and the diversity of living styles that could easily disrupt the direct sight of the camera onto possible scenes of accidents. Additionally, the places where falls could happen would not be limited to one specific, circumscribed area such as the mattress in the demonstration. Accidents can happen in each and every corner of an apartment. Of course, there could also be some seasonal or cultural variation. Think for example of a Christmas trees in a private apartment that occludes the field of vision of the camera.

This 'occlusion problem' is only one example for the challenging and problematic implementation of a visual fall detection system in the homes of the elderly and in fact in homes in general; homes that might be called "smart homes" on a more universal level. This example calls attention to possible fundamental limits of implementing IPAs in existing sociomaterial assemblages. Obviously, from a present-day perspective after having witnessed a fall detection demonstration and having insight into the implementation of facial recognition algorithms in standardised settings (cf. Introna & Nissenbaum 2009), along with the implementation of such an optical fall detection system in private homes, it is likely that many private homes would need to be configured and adapted to the inevitable visibility needs of such a system. This means that private homes in particular, the actual sites of operation and how they are designed and arranged, are vital parts of the sociomaterial assemblage of which IPAs are part. Amongst other things, this relates to sufficient illumination of all areas of the apartment, to the potential rearrangement of furniture and everyday objects, and the adoption of a certain behaviour in order not to impact the configuration camera baseline negatively (e.g. moving the camera during cleaning). In short, location,

[137] Here the question arises of how to imagine and design private homes, especially those for a specific social group, namely in this case, the elderly. When I think about those private homes I act in exactly the same way as the computer scientists when picturing elderly people falling. I refer to my own specific view of how private homes of the elderly look. In my case, the home of my own grandmother living in a 20th century detached house on the border between Germany and Austria.

technology, and human behaviour have to be synchronised so as to create and maintain a functioning sociotechnical system when it comes to the implementation of IPAs.

IPAs as 'Political Ordering Devices' and 'Colonisation Vehicles'

The question arises of whether affected people are aware of and willing to make the changes necessary for the successful implementation of, for example, an automatic fall detection system. The case of the automatic toll sticker monitoring checks on Austrian motorways, but also the case of face recognition that both work much better in highly standardised settings suggest that it is important not to underestimate the efforts that have to be taken to standardise and thus change environments and their materiality. This applies especially in private homes, where fall detection systems and other IPAs are planned. As such, **IPAs are 'ordering devices'** that clean up the everyday mess made by people and society—or to use Harry Collins´ words: that render extinct "troublesome cultural diversity" (cf. Collins 2010: 170)—that structure and order society and its socio-material organisation from their own very specific point of view. As such, following Winner (1980), they are not neutral ordering devices, but highly **'political ordering devices'**. They order society in a particular and specific, political way that was implicitly inscribed into them by computer scientists and operators during the process of developing, programming, and implementing. In this regard it is of great importance to make the domains of scrutiny or persons of interest visible, in order to be able to watch, track, and analyse them. Because IPAs are highly dependent on the visibility of their domains of scrutiny or persons of interest they co-produce visibilities. Meaning that once they are deployed they force their "allies" to remove all kinds of urban "caves" and hideaways in public, but also in private spaces. They make them break the anonymity of the mass, in order to pick out individual persons or objects, or they cause authorities to ban face covering in order to have free sight of faces and so on. An IPA can perform very well, but only until there are no images of the domain of scrutiny available: because the camera lense is blanketed, the persons of interest have covered their faces, or there is a piece of furniture or its shadow between camera and person. Then the IPA will not recognise the person or the event of interest. In this regard, IPAs simultaneously depend on and create "disciplinary spaces" (Foucault 1979), because

IPAs work only when the domain of scrutiny or person of interest is clearly visible and thus clearly locatable within a specific space. As IPAs do not allow much flexibility here, they are in Winner's terms "inherently political technologies" (Winner 1980: 128ff.), meaning that choosing IPAs means choosing visibilities, means choosing disciplinary spaces, means choosing a political system that allows these spaces.

This essential finding invites us to think about the necessary adaptation and standardisation of environments once IPAs have been implemented. IPAs necessarily seem to act like bulldozers that break and smooth the jungle thicket in order to cultivate fields in this formerly cluttered and inaccessible area. IPAs are then—to borrow the term 'colonisation' from the field of social ecology—**'colonisation vehicles'**. As much as the "colonisation of nature is the appropriation of parts of nature through society" (Bruckmeier 2013: 195), so too is the colonisation of existing sociomaterial assemblages, the appropriation of parts of society through IPAs. That means IPAs, understood as 'colonisation vehicles', modify sociomaterial urban landscapes in order to make use of these areas for specific observing actions such as recognising faces or facial expressions, detecting invalid toll stickers or critical human falls. Where there was a messy, sociomaterial urban landscape hidden from view before, there will necessarily be a clean, standardised, visible urban landscape afterwards once the 'colonisation vehicles' of IPAs have been deployed in this specific area. As a consequence, it might be the case that residents of these urban jungles fight the IPAs and the devices and systems they equip, as much as residents of the jungle fight against the bulldozers in order to save their homes which may seem cluttered but have been chosen by them. From the start, people generally and those affected should be informed about these fundamental interventions into their living environments. It is my view that principally, those affected should be put in a position to be able to participate in the discussions that decide about their living situation. Hence, the generally invisible, silent IPAs as integral parts of visual sensors or "Smart CCTV" systems are delusive: they appear smart and innocent but are in fact able to have wide-ranging sociomaterial effects. In order to accomplish visible, disciplinary spaces they need to devastate urban landscapes.

IPAs as Ground Truth Machines?: False Negative and False Positive Cases

Standardisation understood as the colonisation of messy, sociomaterial assemblages or urban landscapes seems to be a crucial step; even more so, assuming that the devices or systems including IPAs could act in more autonomous ways than in the case of sorting or decision-aid systems, or in devices such as the automated toll sticker checks. Depending on the implementation, the question arises if there is or necessarily must still be a human (expert, observer or supervisor) in the loop—even once the sociomaterial landscape has been fundamentally "colonized" by IPAs—who can evaluate and make use of the decision of the IPA. What are possible consequences? What is at stake becomes apparent when talking about false negative and false positive cases. These are the cases in which —if the numbers or rates are made public—what was perceived as true and real and what was perceived as untrue and unreal can be seen. Here it has to be noted that the concept of gathering false negative and false positive cases does always imply that there is one universal ground truth with which any domains of scrutiny are contrasted in order to evaluate accuracy. Thus, when talking about false negative and false positive cases in the context of IPAs the discussion can be whether and to what extent IPAs are (perceived as) "Ground Truth Machines" (cf. "Truth Machines" in Lynch et al. 2008) and what consequences come with such a status.

False negative cases, in which, for example, critical falls were not detected by an automated fall detection IPA, although critical falls actually occurred, are not taken into consideration and represented as they were just not recognised unless a human operator watched the visual material round the clock. While a false negative case in the example of the automated toll sticker checks only results in a possible loss of revenue—because the missing or invalid toll sticker was not recognised—it can result in the loss of human life in the fall detection example, because a critical fall was not detected and thus, no further emergency action was initialised. That means the decision not to send out an alarm, even though there was an emergency would generally have far further reaching consequences in the case of an (almost) autonomous fall detection system, than a false negative in the case of toll sticker detection. Which is why much more effort is needed to prevent false negative cases once we are confronted with autonomous IPAs. Here the question arises of how much rigour, what standards (e.g. in regard to false negative rates and how these are being evaluated and on what basis) and therefore how much transparency should be required from such future autonomous IPA systems. In further consequence, the question is how and by whom, reports of these standards should be released to the public, especially to users and affected people. This is a real (and currently unsolved) challenge as it is extraordinarily difficult and complex to determine false negative cases or false negative rates, especially in operational settings such as in the case of fall detection, because it takes great effort to collect operational sample data. One can critically ask how much effort, trial and error are effectively needed to analyse and determine false negative rates in the case of fall detection or of similar IPAs, especially in private homes. Without knowing or giving a final answer to this question, it is just an educated guess whether or not a project could fail (or be resisted) in spite of economic pressure to put the IPA and its respective device or system onto the market quickly. If an IPA is introduced onto the market too hastily without sufficient testing of false negative cases it becomes clear that consequences might be serious, not only for affected people but also, potentially for developers and distributors.

False positive cases include a different set of challenges. In the case of the automated preliminary sorting for toll sticker checks, false positive cases can be quite easily recognised by a human operator, because he or she is provided with image evidence data. In most cases, for a specialised human operator it is then not difficult to evaluate the validity or presence of a highly standardised toll sticker assuming the quality of the images is high. Even in doubtful cases, the decision can be made in favour of the client, as is reputedly done in the case of the Austrian AVK system. Automated, visual decision-making for fall detection is trickier. It is questionable whether image evidence sent to a human operator outside the home would even be possible due to privacy regulations. This applies to cases where saving image data for a specific period of time, or transferring this data from a private home to a specific place outside the home, is neither permitted nor desired. An operator could therefore not evaluate a scene on the basis of transmitted images seen on his or her computer that is probably far away from the site of application, as no images have been transmitted. Images that could give evidence of a specific event are eliminated from the system. If transmission had occurred, then in an emergency situation someone would have to visit the site where the images originated in order to see if there had really been a critical fall or if it was a false positive result. Compared to retrospective browsing through image evidence data in a central enforcement centre without any time pressure, as in the case of AVK, the immediate evaluation of critical fall detection at the site of the emergency needs much more effort in terms of time and costs that also have to be considered when planning the implementation of IPA-based, decision-making systems such as automated fall detection.

These considerations show that the successful implementation of IPAs is also dependent on the forms of existing sociomaterial assemblages or urban landscapes in which their deployment is planned. The more standardised, clearly structured, less cluttered and less private these existing, sociomaterial assemblages or urban landscapes are, the easier a successful implementation might be. As many sociomaterial assemblages are both profoundly cluttered and private, it must be pointed out that the implementation of IPAs and the devices they are part of in these cluttered, private assemblages is a pervasive interference and thus in the lives of people. Assuming that autonomously acting IPAs making decisions based on visual information are highly dependent on a clearly ordered, standardised, and "colonised" setting, the possible sites of implementation are limited considering the diverse and disordered ways of living in contemporary societies.

Following the discussion about false negative and false positive cases and rates, it becomes clear that the status of IPA devices or systems as possible "Ground Truth Machines" is very fragile, but nevertheless real. In order to become viable and true, they not only depend on the standardisation and colonisation of the settings in which they are implemented, but similar to other pattern recognition technologies such as fingerprinting or DNA profiling, they necessarily depend on the involvement, cooperation and interpretation of human evaluators or operators due to several uncertainties and restrictions that accompany IPAs. These uncertainties are intensified by the silent and implicit subjectiveness of IPAs. As was shown in Chapter Five, in contrast to the view that technical authority and neutrality are inscribed into the respective "Ground Truth" of an IPA (e.g. how a critical fall looks, what a cow looks like etc.), and thus inscribed in the ability of a computer to see and recognise correctly, what is inscribed is situation dependent, selective and subjective; views that have been negotiated in different sociotechnical constellations in and around computer vision laboratories. It has been stated that similar to human vision, the semantic processing of images by algorithms is a situated interpretative practice that is shaped by cultural traditions of seeing (cf. Burri 2012: 51) in the field of computer vision.

Also similar to fingerprinting and DNA profiling, a widespread public impression has arisen that blackboxed IPA devices and systems are infallible "Ground Truth Machines" that could overturn human perceptions or decisions such as eye witness testimony in court (cf. Lynch et al. 2008). There is however, a great difference to the expert professions of fingerprinting and DNA profiling that has the potential to debunk the "Ground Truth Machine" status of IPAs dealing with everyday domains of scrutiny. Here, from a human perspective, it is much easier to recognise when an IPA decision is profoundly wrong in mundane cases than it is in the highly specialised activities of fingerprinting or DNA profiling. This might be, because the "socially sensible" (Collins 2010: 123) human is an expert on "what everybody knows knowledge" (Forsythe 1993), or in Collins' terms, the "collective tacit knowledge" (ibid.: 11; 119) expert. By being confined within the diverse practices of everyday life, most humans seem to constantly and quasi-naturally update and adapt their ways of seeing and recognising in a non-formal way. Thus, in their own society or environment, in contrast to computers equipped with IPAs, most humans similar to the fictional computer HAL 9000 and its "Simpson's Ultrahouse 3000" version (cf. Chapter Three) seem able to cope with diverse, ambiguous, complex, and cluttered everyday situations very well in their perception. They are also able to differentiate and recognise subtle nuances in what they see. In this regard, IPAs really play the part of 'disclosing agents' (Suchman 2007: 226) that demonstrate how human vision can deal very well with the ordinary. For IPAs the mundane is the major challenge, because it can be said that IPAs have a fundamental problem with processing and interpreting diversity, ambiguity, situated actions (Suchman 1987, 2007) and local particularities.

This contradiction between the perceived status of IPAs as infallible "Ground Truth Machines" and their actual, limited and thus conditional and probabilistic status (cf. Chapter Six) of uncertainty, especially in everyday life situations, calls for a discussion of the ethical, legal, and social aspects and implications connected to the status and authority of IPAs.

Ethical, Legal, and Social Aspects of Image Processing Algorithms

In what follows I shall take a look at Image Processing Algorithms within the ELSA (Ethical, Legal, and Social Aspects) framework. ELSA has been, and is especially used in areas such as biotechnolgy, genomics, and nanotechnology in order to analyse the ethical, legal, and social issues raised by specific applications of these technologies. The Anglo-American and in this regard, the US version ELSI (Ethical, Legal, and Social Implications) in particular, has been perceived as more utilitarian in order to implement the respective technologies in an almost frictionless way (cf. Kemper 2010: 16f.). ELSA, conceptualised by the Continental European Social Sciences and Humanities in a more bottom-up way, leaves open the choice of 'aspects' and questions (ibid.: 17). Nevertheless, as is the case in Austria, ELSA has also been perceived as a means of facilitating the implementation of technology (ibid.: 19).

On the one hand I use the openness of the European approach but on the other, try to critically assess and challenge IPAs. Nevertheless, this procedure can help to facilitate the adequate and successful implementation of IPAs as it outspokenly questions aspects of friction.

One of the main socioethical questions connected to IPAs is the question of social justice and "fairness". Who benefits from IPAs and the sociomaterial assemblages of which they are a part and who does not? What are the advantages for some and how are these advantages related to the disadvantages of someone else? Altough there is a strong focus on the disadvantages, it is clear that both exist. Prainsack and Toom tried to emphasise this aspect by introducing the concept of *situated dis/empowerment* in the context of surveillance technologies, both to see and explain the oppressive/disempowering and the empowering aspects of surveillance (Prainsack & Toom 2010: 4) in order to explain its success or failure.

Here it has to be noted that the *situated dis/empowerment* of IPAs is hard to generalise as the word 'situated' indicates. This means, the constitution of the *dis/empowerment* of IPAs can be different from case to case, from time to time, from place to place. Which is why I shall proceed here in a situated manner, by focusing on one of my main empirical examples, the case of automated Fall Detection. As it is an entity in the making where there are not yet any operational settings available, thought about fall detection *dis/empowerment* is to be seen in the tradition of technology assessment with the added information of the empirical material already presented.

Empowerment

A best case scenario would be that a budget-friendly fall detection system based on IPAs that is installed in all the private homes of people wishing to have such a system, because they feel safer and want to live independently in their old age. As a side note, this logically assumes that a fall detecion system based on IPAs is the best way to achieve safety and independence. The installation does not impact homes to a considerable extent and they can proceed as usual in their daily life. So, nothing changes except that there are visual sensors installed in every room of the apartment and some modifications are made to the electronic infrastructure of the building that remain mainly invisible to the owner of the home. People living in the apartment can be sure that their privacy is respected and their data protected, because no outsider ever sees what happens inside or even outside (e.g. in the garden, on the balcony) the building. Image material is neither saved locally nor is it transmitted to a place outside the home. People know for sure that they are only being watched by a neutral, smart system that had been said to recognise automatically if they fall down in a critical way. This means that in the overall majority of the daily life situations the fall detection system passively operates in the background without creating any negative consequences to the people living in the place of operation. People do not need to change their behaviour in any way. They also do not need to cope for example, with internet or computer technology themselves. Instead they can rest on the sofa after lunch, they can place their Christmas trees in the same place it has been placed every year. Unless they fall down, they will not even notice the system.

Then suddenly, let us say about three years after installation, on a cold winter day the person living in this home falls down when getting out of a very hot shower, and stays lying on the floor of the steamy bathroom, very seriously hurt. Immediately, an alarm is sent to the ambulance emergency service. Some minutes later an ambulance crew enters the apartment with the key that was made available to them when the fall detection system was installed and they are able to save the life of the person who had the accident.

In such a case as described here, it is obvious that the affordable fall detection system was worth every penny it cost. It saved a life, and did not affect everyday life negatively beforehand. In addition to this, the relatives and friends of the respective home owner did not need to provide care on a regular basis and were able to live their lives in their own chosen ways. They also did not need to spend money on care. So overall, they might have saved money. Added to this, the emergency ambulance service also had an advantage from the fall detection system, as it was only called once when there was a real emergency. Finally, the private company selling the fall detection system also profited as it was able to do good business. While most actors in this whole scenario benefitted from the installation of the fall detection system or were empowered in their way of living, the hypothetical, usually female care attendants may have had disadvantages, because they would have lost many jobs in the home care sector.

Disempowerment

But what is the situation considering the insights from my empirical observations already elaborated upon? As it is likely that the installation of a visual fall detection system in the home of an elderly person does also affect and change the living environment to a considerable extent, the overall cost might be higher than expected. Meaning that this technology would be available to fewer people. Especially those with modest means could not afford the technology[138]. It could however, still be the case that the expected costs are below the costs of carers, because care is considered to be expensive. In such a case, it is likely that public or private health insurance institutions might choose fall detection over care, because it could save on overall costs, especially in times of economic crisis when the mantra of saving is omnipresent. In such a case, the question arises of whether people affected by a fall detection system installed in their premises, are willing to accept it instead of, or in addition to carers or caring relatives or friends. In the course of deciding for or against a fall detection system, people should be put in a position to get to know the capabilities, limitations and side effects of such systems. In contrast to the ideal installation described beforehand, it can be expected that the installation would impact everything in the home to a considerable extent, for example by the necessity for a rearranging and modernising of the furniture or by the installation of new brighter light sources. It is also questionable if there would be sufficient electronic infrastructure available in older houses, so that also in this regard, comprehensive installation would be necessary. During the installation process the question might also come up of how many and where exactly, cameras are to be installed in order to cover the whole area of the apartement for the protection of the inhabitant. Because some apartments might be full of nooks and crannies, be spacious and dimly lit, a large camera network would need to be installed. In such a case it is likely that selections would have to be made and also, due to privacy reasons some areas of the apartment might not be covered by cameras and automatic fall detection. The garden, corridors, bathrooms and lavatories might be some examples of places where cameras are, to say the least, highly problematic. In some areas of a room more cameras may need to be installed than might be expected. For example, if there is a large dining table in the middle of a room, at least four cameras would need to be installed on every side of the table in order to have free sight to a possible site of falling behind or below the table.

Considering that image material is neither saved locally nor transmitted to a place outside the home, because in Austria and the European Union, privacy and data

[138] Another possibility is vice versa: Those with modest means are forced to purchase the technology, because they are not able to afford human carers.

protection are fundamental rights and values, in the case of a detected fall, the emergency service is called to the apartment. Following my empirical observations it is likely that the detection threshold is set low in order to be sure that possible falls are recognised. In such a case it is likely that non-critical changes in position, such as lying down on the floor, because for example a coin fell to the floor and the inhabitant stooped in order to pick it up, might in some cases trigger the alarm. If these are recurring events, alarms might potentially not be taken seriously any more. Emergency operations are expensive and, as time goes by alternatives might be considered. For example, inhabitants would be advised to avoid certain behaviour. They would be instructed to avoid situations in which an alarm could be triggered. Apart from a feeling of insecurity, their everyday living is restricted to a considerable extent. Previously quasi-natural actions are questioned and in case of doubt, probably avoided. Another possibility of avoiding false alarms is to adjust the threshold setting of the fall detection system. Unless this happens soon after the installation this might entail additional costs, because a highly specialised IT expert needs to be called. Then, the threshold is set high in order to avoid false alarms. After a while the decision seems to be right, as there are no more false alarms.

But then suddenly, let us say about three years after these adjustments to the system, on a cold winter day the person falls when getting out of a very hot shower, and stays lying on the floor of the steamy bathroom, seriously hurt. However the upper body of the person who has fallen is up slightly, leaning against the wall. If there is really—but due to privacy reasons unlikely—a camera directed at the area in front of the shower, the event might not be detected, because steam led to poor visibility. Some minutes later when the steam has gone the prone position is also not detected, because the alarm has not been triggered due to the higher threshold. This is because the body, represented as a straight line is not sufficiently parallel to the plane of the floor (cf. Chapter Five). Thus, a critical fall has not been detected. A worst case would be that the casualty remains undetected on the floor of the bathroom for too long, because there are no carers or relatives looking after this person on a regular basis. Also a transmission of the image data of the scene in the bathroom to an operator or to relatives might have been useless, because the fall was not detected. The transmission of image data might have been useful in cases of false alarms in order to avoid emergency operations. Even if the image data with personal information is transmitted in an encoded form (e.g. in silhouette form), the intrusion on privacy might be disproportionate, because the event of, for example, picking up a coin from the floor is not a good enough reason for observing a person in their own, very private home. In addition the question comes up of how the affected person could know whether he or she is being observed at this moment of picking up the coin.

In such a worst case scenario, it is obvious that the fall detection system disempowered the affected people: Their familiar living environment had to be adapted to a considerable extent in order to install a fall detection system that was said to protect them by enabling them to live independently. They also had to adapt their behaviour to avoid false alarms, but in the case of a real emergency, the system did not detect the critical fall, because a false negative case had occurred. At first glance, only the company selling the system might have benefited, but depending on responsibilities it is likely that consequences would arise for the company once the affected customers would take legal action. In the following section, this issue of legal responsibility is discussed in more detail.

Legal Responsibilities

Along with more autonomous systems such as the fall detection system described, also legal responsibilities are subject to change. The main question arising in this context is who or what is, or should be made responsible? What about the distribution of liability and negligence (Beck 2009: 229f.) between humans and IPAs or computers? Are algorithms in any form and Image Processing Algorithms in particular, to be made responsible, in accordance with the law? According to Eisenberger (2011: 519), social responsibility is increasingly to algorithms, because they are gaining more and more selection abilities. But what happens if algorithmic selections and even decisions are at the heart of an autonomously operating system? What if an IPA like the one that is programmed to detect critical falls, fails? The *Royal Academy of Engineering* (2009) raised the question in this regard of how to deal with potential failures of autonomous systems: They asked if this "... could mean that such technologies are held back until they are believed perfect. But is this too strong a requirement?" (ibid.: 3).

It might be, because perfection is unachievable in total. It seems though, to be crucial that autonomous systems and their modes of operation—of which IPAs are an important part—are understood by society in order to manage possible risks and adverse effects. This raises questions about to what extent IPAs, or better, key figures, possibilities and limitations that go along with these algorithms have to be prominently specified and labelled by their vendors and distributors ("duty of declaration", cf. Eisenberger 2011: 521). In the case of IPAs such aspects could be the evaluation specification that is important for assessing the viability of an algorithm or a system. Was the system or IPA tested in a technical or scenario evaluation, or was the evaluation operational? What were the respective results (e.g. false positive, false negative rates)? The specification of the ground truth model is another aspect to be declared. Amongst other things this refers to what kind of and how many training images were used? How were and are the specific thresholds set that distinguish between one group and another, or between recognising whether an event has taken place or not, as the case may be? How dependent are IPAs on influencing factors such as weather conditions, population diversity, environments, lighting conditions and so on? In connection to this there should be a declaration or at least a discussion of what standards of technoscientific rigour are demanded from IPAs and who should define and control these standards. Here the question arises of who actually the (independent) experts for setting and controlling these standards are. Who should be involved in these processes, in what way?

Another aspect is in how far biases (e.g. higher recognition rates for specific groups or categories) have to be tested and declared before a system goes into operation (cf. "bias studies" in Introna & Wood 2004: 196). Or, once a system is in operation, in how far it has to be evaluated after a certain period of time in full operational conditions. Similar to the discussion of creating a 'research culture' in forensics and pattern identification disciplines[139], whose main reference point should be science and not law (an argument made by a number of academics, practitioners, and professional organisations, cf. Mnookin 2011), we should discuss what the creation of a 'research culture' in image processing and algorithms research could look like. In this regard it seems to be of great importance to apply a broad perspective to 'research culture', as in my view, the most critical point is reached when IPAs are brought outside academia to business and integrated in commercial products that affect a wide array of people. Here we are confronted with a well-known conflict of computer science to which I referred in Chapter Three when going into the (US) history of the computer sciences. The business sector of the computer sciences criticises that the algorithmical, mathematical orientation is too theoretical and not useful for the real world (cf. Ensmenger 2010: 134f.). Conversely, it might be necessary to apply exactly this mathematical orientation in order to avoid errors, biases, and failures.

It is my opinion that it is necessary to bring together the academic commitment and the business orientation of computer science. If we as a society wish to integrate machines or systems with decision-making agency that could be referred to as "autonomous systems" (cf. The Royal Academy of Engineering 2009) or "smart machines", we need to specify clearly on what basis the mostly binary decisions are made or not made by the integrated IPAs, and how these decisions can be explained, modified or suspended. This involves the demonstration and giving evidence for how abnormal patterns (and this refers to many different pattern recognition activities for which IPAs are designed; everything from an invalid toll sticker to a critical fall, to the recognition of a specific face, to suspiscious terrorist or criminal activity) can be clearly demarcated from all other (or almost all other) non-relevant patterns. What does 'clearly' demarcated mean and how do we process categories or phenomena that are not made for being clearly demarcated? In this regard it has to be emphasised that:

[139] Pattern identification can be defined as the association of a particular pattern or impression, such as fingerprints or shoeprints with a particular source (Mnookin et al. 2011: 730)

> "in highly complex situations, the breadth of human experience could give rise to better judgements than those made by a system 'programmed' by a narrow range of previous behaviour." (Royal Academy of Engineering 2009: 2)

In the context of Face Recognition Technologies (FRT) Introna & Nissenbaum noted that their

> "… view is that in order to achieve balanced ends, FRT must function as part of a intelligence and security infrastructure in which authorities have a clear and realistic vision of its capacities and role, as well as its political costs." (Introna & Nissenbaum 2009: 46)

For all practical implementation purposes, this asks for a careful distribution of responsibilities, power and agency of these IPA systems to computers and human users in order to manage possible risks and adverse effects. If such a system is proportional in legal terms, meaning that it must be shown that it is legitimate, suitable, necessary and resonable to achieve a specific aim, it should be implemented in a concrete context where the application is made-to-measure. In following Suchman (1987, 2007) *Workplace Studies* have demonstrated particularly, the importance of how technologies are being applied in situated actions, and how they can fail if they do not meet users' needs (Knoblauch & Heath 1999). A consequence is that such a system must support the complex sociomateriality of the specific work setting or of the specific living environment, in which it is going to be implemented in (Hughes et al. 2000). It has to be purpose-built and one-off (Introna & Nissenbaum 2009). This also means reflecting on implications should a system be employed in another place or at another point in time. Because "the same technological device can do and mean different things in different places" (Oudshoorn 2012: 121) the question has to be asked: What happens once IPAs travel in place or time and how does this affect society? At the present one can only think about these implications as it was done throughout this chapter. In the future, when more "smart" systems and devices based on IPAs colonise the world, further analyses are required to shed light on this important question. In the next section of this chapter, I suggest a conceptual reflection framework (SIPA) to give support to this question.

Social Studies of Image Processing Algorithms (SIPA)

As is mostly the case at the end of a specific research process or period, one result of this exploration of computers and their ability to see and particularly the analysis of Image Processing Algorithms is that more research is needed to further consider newly discovered and rediscovered questions and areas of concern that have not yet been analysed in detail. Thus, in what follows I suggest a conceptual reflection framework for the further analysis and development of IPAs in society and I also consider the question of how societies and the sciences could deal with IPAs in a responsible and reflective way of innovation?Because IPAs especially, as a crucial part of what is often referred to as smart or intelligent machines, are expected to be powerful actors and decision makers in the future, it is important to have such a conceptual reflection framework at hand that could guide their further empirical analysis and development.

Due to the fact that images are omnipresent when it comes to “computers with vision”, referring to and carrying on the “Social Studies of Scientific Imaging and Visualisation” (Burri & Dumit 2008) I refer to my suggested framework as the “Social Studies of Image Processing Algorithms” (SIPA).

SIPA is designed to provide a tool for studying and analysing Image Processing Algorithms, computer vision, and connected to them, relevant technologies or “smart” devices such as facial recognition or behaviour pattern analysis, from an interdisciplinary social science perspective, in order to understand and grasp these phenomena in all of their sociotechnical, cultural, political, ethical, and moral dimensions. As was shown throughout this study, the relationship between computers and their ability to see is a complex sociotechnical one. It has been established in particular through attempts to produce, process and understand (digital) images by means of computer algorithms. It is clear that the issue of computers and their abilitiy to see needs to be understood as a sociomaterial process in which IPAs are developed, produced and deployed in devices or in larger systems; advertised, used, talked about, criticised, and configured. In short, the processes in which IPAs are negotiated and formed in several sites and situations. In accordance with SIV, SIPA too, strongly refers to laboratory studies but it also goes beyond the techno-scientific laboratory. It follows

IPAs when they "leave" academic or technological territories, for example to the media, or to operational settings in other places and at other times. SIPA follows IPAs from production to implementation, to use or non-use in an object related manner which also means that it is not a restrictive framework, but encourages that specific methods used have to be adapted to a specific research object, to a specific place and time, and to a specific research question. Thus it can be stated that there is not only one single way to proceed, but that this process depends on what is important to know and why.

I understand SIPA as a specialisation and concretisation of SIV as it focuses particularly on Image Processing Algorithms and accordingly, also addresses the fields of professional computer vision, pattern recognition and image processing, explicitly. What SIPA offers is a sensitising concept (cf. Blumer 1986) and reflective perspective on a particular, complex, material-semiotic object of knowledge (Haraway 1997: 129)—the Image Processing Algorithm—that plays a silent leading role in the ongoing groundbreaking processes of computerisation, automatisation, and smartisation. SIPA is, as is every scientific concept or method, a political endeavour. Choosing SIPA means to choose a specific political practice that has the aim of approaching the phenomenon of Image Processing Algorithms from a particular point of view. It nevertheless has the aspiration of approaching IPAs from different perspectives in order to be able to handle its multiple dimensions.

Production of IPAs

The **production of images** is a basic requirement for the production of IPAs. Without images there are no IPAs and there is no computer vision possible. The production of images is the first level for analysis on the agenda of SIPA. In SIV, Burri and Dumit ask the question "how and by whom images are constructed by analyzing the practices, methods, technology, actors, and networks involved in the making of an image." (Burri & Dumit 2008: 300). They show that the production of images is dependent on a series of decisions concerning the machines, data, parameters, scale, resolution, and angles. These decisions and selections "do not depend on technical and professional standards alone but also on cultural and aesthetic conventions or individual preferences" (ibid.:

301). The production process of scientific images is far from being a neutral process and is shaped by sociotechnical negotiation with local variation also playing a role in the question of who is able to read images and who is allowed to read them. Visual expertise is its own form of literacy and specialisation (ibid.: 302).

SIPA builds on these insights and encourages more analysis of the role of images and image production in IPA development. It especially suggests analysing two types of images in the context of IPAs and their relation to each other in more detail. On the one hand this refers to *training images*. These influence and finally define the respective ground truth model. One can ask the following questions in order to analyse training image selection processes: Why are specific images chosen? How are these images constituted? From what sources do they come? In what way are they used to give evidence of an entity? On the other hand, it is important to have a look at *analysis images* that are compared with the ground truth template model once the IPA is in use. The question here is what the recording circumstances are: What influences the recording at the site of operation? Finally, the relation of *training images* and *analysis images* can be focused on: What are the differences between these two image sources and how might these differences influence the detection or recognition process?

Next to the production of images, the **production of the ground truth** ("ground truth studies") is the second analysis level that in my view, is the most significant in SIPA. That is, because the production of the ground truth can also be regarded as the production of the "Interpretation Template", or under specific circumstances as the production of a "Truth Template" that is the basis for all image interpretation done by an IPA. Referring to Chapter Five of this thesis, the question arises of what kind of knowledge or visual expertise was used in order to produce the respective ground truth. Is it more formalised, explicit or less formalised, tacit knowledge? Is it based on expert views or on everyday common sense? In this process, it is crucial to consider what aspects influence, characterise and are "inscribed" into the respective ground truth. Why were exactly these aspects chosen and of importance? How is proof given that the applied characteristics are real evidence for the specific domain of scrutiny? For example how can it be proved that the relationship between a straight line representing the human body and a plane representing the detected floor area indicates whether a person has fallen down? All in all, it should be comprehensible which specific and particular version of reality and truth has been transferred to and manifested in the respective ground truth. Once it has been formalised and determined, the question can be asked if there is still room for manoeuvre to either use the respective ground truth for image comparison, or to allow alternative (human) views (e.g. at another place or at another point in time). The analysis of the production of the ground truth is essential in understanding the political dimension of IPAs. It is key to the inscription of situated and particular (political) views into IPAs, and, depending on the assigned authority of the respective IPA, made durable and powerful.

The third level in studying IPA production processes is the **production of algorithms**. In order to translate the visual input material into the computer terms, it is necessary to apply mathematical methods. Thus, it can be asked what mathematical models or equations are used to formalise the domain of scrutiny? The SIPA researcher should follow transformation and reduction processes that take place in this matter. At this level, it is of importance to have a look at thresholds and how they are set. The question can be asked of how and why thresholds are decided on and set, in order to differentiate between different groups or different behaviour. How flexible or determining are these thresholds? The insights into the production of ground truth can be drawn upon when the question arises of what kind of knowledge or visual expertise was used.

Implementation of IPAs

IPAs cannot take effect without being connected to a hardware-software package. They need to be implemented in greater sociotechnical systems or into existing 'sociomateral assemblages' (Suchman 2008: 150ff.). IPAs not only need to adapt to other entities, but it might also be the case that the other entities need to adapt in the course of implementing IPAs. The first question is how and in what kind of sociomaterial assemblages IPAs are implemented. For what purpose (e.g. efficiency, authority, etc.) are they and the assemblages in which they are deployed, used or going to be used? What are the sociopolitical circumstances that led to the implementation? Were alternatives considered and if not why was this the case? If there is a possibility of following or reconstructing the implementation process, SIPA is concerned with how far there is a need to adapt, standardise and "colonise" the environments into which the IPA system is being implemented. For example, what are the total costs (also in non-financial, ethical, legal and social terms) of the implementation process? Here the question could also be asked of how people affected experience the changes that originate from IPA system implementation (e.g. automated toll sticker checks on Austrian motorways); a question not tackled empirically in this analysis. This touches also on the issue of data protection, privacy, transparency and participation and as such it is a core question for democracy. How far are fundamental rights such as privacy protected against IPA systems? How far are affected people involved in the implementation process? How are they informed about capabilities, limitations, uncertainties etc. that could influence their lives to a considerable extent?

On the analysis level of implementation, new or adapted forms of qualification have also to be considered: Are people working within or with IPA assemblages aware of the mode of operation? What is the relationship in power and authority between human operators and IPA selections or decisions? In accordance with the insights of this thesis it is clear that IPA systems operating on a semantic level cannot act fully autonomously, but must be integrated into social settings with professional staff who understand how the algorithms applied work. The more operators know about the ground truth in use with its error tolerances, thresholds and the reduction and standardisation of complexity, the better they are prepared to handle the technology and minimise possible risks such as false positive findings. Against this background, IPA systems can 'upskill' staff rather than the opposite (cf. Musik 2011). The assumption being that along with the implementation of such systems, operators have to be trained or learn on the job with practical experience in order to manage and work with IPAs. A reduction or even elimination of operators is unlikely, because human analytical and operational skill is still required and inevitable. As such this statement is in line with what Suchman, in reference to Ruth Schwartz Cowan (1983) notes: "…any labor-saving device both presupposes and generates new forms of human labor ."(Suchman 2007: 221).

Use and Non-Use of IPAs

Once IPA systems are in operation it makes sense to analyse the socio-technical constellations in situ, if access is available. That means SIPA is also concerned with the **use and impacts of IPA systems**. The main question is how IPAs and their sociomaterial assemblages are used in concrete applications. Also connecting to the level of implementation the question can be raised of whether all people affected are aware of the use and impact of the respective IPA system? How is awareness achieved in this regard, also as time goes by?

A very important analysis level is a **public understanding of IPA systems** that also could be re-formulated to a "Public Understanding of Smart Environments, Smart Devices, or Smart Machines." What understanding of a specific IPA system does the public have? What understanding is communicated by developers, operators and critics or by the media? Was this understanding also communicated in the phases prior to implementation? How was and how is this understanding used to promote or restrict further development or use? In contrast to understanding that has been communicated, the question of whether there are regular evaluations taking place should also be raised. If there are evaluations, it is of interest what kind of evaluation (e.g. economic, technical, operational) and by whom these are performed. How and why are the results of evaluations made public or kept secret?

In the course of evaluations, **bias studies** are an important means for the analysis of possible discrimination and new types of digital divide. As was referred to in Chapter Two, Introna and Wood demanded "bias studies" in the context of their study of the politics of face recognition technologies. They especially raised the question of what can be done to limit biases (Introna & Wood 2004: 195). As many biases seem to be inscribed into IPAs unconsciously, it is important at least to analyse biases once they are in operation. Because most IPA systems in operation are inaccessible, another possibility for gaining information about biases is an obligation to investigate biases and publicise the results of bias studies before a system affects people in a negative way.

An often neglected issue in STS and other connected areas are studies of non-use or resistance (cf. Oudshoorn & Pinch 2003: 17). Thus, it is important also to study the **non-use or resistance to IPAs**, because they are “a common aspect of the process of creating technological and social change” (Kline 2003: 51). So some of the relevant questions are: Where are IPAs not used or resisted? What are the reasons for this non-use or resistance? What practices of resistance occur and by whom are they performed? In this regard it might also be interesting to see what the alternatives to IPAs and IPA systems or devices are. Here on the level of non-use or resistance to IPAs it could be particularly important to carry out transnational or transcultural studies that compare different technopolitical cultures (cf. Hecht 2001) of non-use and resistance.

Towards responsible innovation: SIPA as a conceptual reflection framework for socio-technical solutions

As was already indicated earlier, SIPA does not only provide a conceptual framework for social scientists for the analysis of IPA production, development and either use or non-use. It also explicitly represents a reflection framework accompanying IPA research and development for computer scientists working on and with IPAs. This is not only a reaction to the claim of international science and technology to apply social and ethical considerations in research and development (cf. Schuurbiers 2011: 769), but is a claim for the real achievement of ‘sociotechnology’ (cf. Musik 2011: 351) in technoscientific practice and particularly in the design of IPAs. This means, as a consequence of social scientific research that is more cooperative than objectifying (Beaulieu 2010: 462) the conceptual and reflection framework of SIPA is a tool to bring together social scientists and computer scientists to reflect and work together in the specific research and innovation field, but also the business field of Image Processing Algorithms in order to achieve what could be subsumed under the term ‘responsible innovation’[140]. This cooperative work is not necessarily limited to computer and social scientists, it could also be possible to integrate (critical) artists working on and with IPAs such as those mentioned in Chapter Two. As such, SIPA encourages a specific form of participatory

[140] See Stilgoe, Owen & Macnaghten (2013) for a detailed overview on the different meanings of and the emergence of ‚responsible innovation.‘

design[141]. The involvement of other societal actors in IPA research and development might help to position computer vision onto a broader and more robust grounding. Because human vision is situated and particular (cf. Burri 2013) it is important to consider and make use of a great variety of situated and particular views that potentially contradict the situated and particular view of computer scientists. So, involving other people with other views could help to inscribe more diversity (and in this way, more democracy) into IPAs and thus, it could help to reduce—but never fully eliminate—influential semantic gaps. As such, SIPA can be seen as building upon but exceeding what was referred to as 'Midstream Modulation' (Fisher & Mahajan 2006) in order to enhance critical reflection in the laboratory (Schuurbiers 2011). 'Midstream Modulation' (MM) as a form of 'socio-technical integration research' (ibid.: 771)

> "is a means of incrementally influencing a technology during the "midstream" of its development trajectories. It thus asks *how* research is to be carried out, which is within the purview of engineering research, rather than *whether* a research project or product should be authorized, approved, or adopted, which is largely beyond the purview of engineering research. As an integral part of R&D activities, MM is a means by which research decisions might be monitored and broadened to take advantage of otherwise overlooked opportunities to weave societal factors into engineering decisions." (Fisher & Mahajan 2006: 3)

While Fisher and Mahajan's use of MM aimed to reflect critically on laboratory-based work, Schuurbiers (2011: 772) tried to enhance MM to reflect on the broader socio-ethical context of lab research. Here it is important to note how Schuurbiers comments on the relation between social scientists and laboratory practitioners (ibid.: 773), which is that the assumption underyling MM, in the context of SIPA, is not that computer scientists have a general 'reflective deficit' and social scientists are more reflective. Rather, as Schuurbiers suggests, it is the case that social scientists´ knowledge could complement natural scientists´ knowledge through interdisciplinary collaboration (ibid.). I would suggest going one step further. From the very beginning—and by the

[141] See Suchman (2007: 277f.) for more details on what participatory design is. In short, the guiding commitment of participatory design "…is to rethinking critically the relations between practices of professional design and the conditions and possibilities of information systems in use" (ibid.: 278).

very beginning, I mean the design level of research projects—sociotechnology should be carried out (cf. Musik 2011: 351). For research funding, this implies a promotion, in addition to basic research (e.g., on IPAs), of especially problem-centred instead of technology-centred research projects. Technology-centred projects make use of resources for the sake of developing one specific technology that is promoted as a ready-made solution for a pre-defined problem from first to last. In contrast, problem-centred projects would involve open inter and transdisciplinary engagement at the problem definition level, which would potentially lead to different comprehensive sociotechnological solutions. Of course, this procedure does not exclude what might be called —assuming that purely technical entities do exist—"technological" solutions, but it could avoid that asocial technological solutions are developed merely for the sake of economic growth or a seemingly sophisticated "smart touch" resulting in the need to adapt them laboriously to the messy sociomaterial assemblages and urban landscapes against the will and daily lives of the people living there.

Literature

Adey, Peter (2004): Secured and Sorted Mobilities: Examples from the Airport. In: *Surveillance & Society* 1(4): 500-519.

Akrich, Madeleine (1992): The De-Scription of Technical Objects. In: Bijker, Wiebe E. & Law, John (ed.): Shaping Technology/Building Society. Studies in Sociotechnical Change. The MIT Press, pp.205-224.

Amoore, Louise (2007): Vigilant Visualities: The Watchful Politics of the War on Terror. In: *Security Dialogue*, 38, 215-232.

Augusto, Juan Carlos & Nugent, Chris D. (eds.): Designing Smart Homes. The Role of Artificial Intelligence. Berlin, Heidelberg, NewYork; Springer.

Bakker, Sjoerd & Budde, Björn (2012): Technological hype and disappointment: lessons from the hydrogen and fuel cell case. In: *Technology Analysis & Strategic Management* 24 (6) 549-563.

Beaulieu, Anne (2010): Research Note: From co-location to co-presence: Shifts in the use of ethnography for the Study of knowledge. In: *Social Studies of Science* 40 (3): 453-470.

Beck, Susanne (2009): Grundlegende Fragen zum rechtlichen Umgang mit der Robotik. In: *Juristische Rundschau* 01/2009; 2009(6): 225-230.

Becker, Peter (2001): The Standardized Gaze: The Standardization of the Search Warrant in Nineteenth-Century Germany. In: Jane Caplan & John Torpey (ed.): Documenting Individual Identity. The Development of State Practices in the Modern World. Princeton/Oxford: Princeton University Press, 139–163.

Bloomfield, Brian & Vurdubakis, Theo (2002): The Vision Thing: Constructing Technology and the Future in Management Advice. In: Clark, T & Fincham, R (eds.): Critical Consulting: New Perspectives on the Management Advice Industry (Oxford: Blackwell Publishers. 115-129.

Bloomfield, Brian (2003): Narrating the future of intelligent machines. The role of science fiction in technological anticipation. In: Czarniawska, Barbara & Gagliardi, Pasquale (eds.): Narratives we organize by. Amsterdam: John Benjamins, 193-212.

Blumer, Herbert (1986): Symbolic Interactionism: Perspective and Method. University of California Press.

Boehm, Gottfried (1994) (Hrsg.): Was ist ein Bild? München.

Borck, Cornelius (2001): Die Unhintergehbarkeit des Bildschirms. In: Heintz & Huber (Hrsg.): Mit dem Auge denken. Springer. pp. 383 – 394.

Borup, Mads, Brown, Nik, Konrad, Kornelia & van Lente, Harro (2006): The Sociology of Expectations in Science and Technology. In: *Technology Analysis & Strategic Management* Vol. 18 No. 3&4, 285-298.

Bourdieu, Pierre (1977): Outline of a Theory of Practice. Cambridge University Press.

Bowker, Geoffrey C. & Star, Susan Leigh (2000): Sorting things out. Classification and its consequences. Cambridge & London.

Breckner, Roswitha (2003): Körper im Bild. Eine methodische Analyse am Beispiel einer Fotografie von Helmut Newton. In: *Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung (ZBBS)*, Heft 1: 33 - 60.

Brednow, Walter (1969): Von Lavater zu Darwin. Berlin: Akademie Verlag.

Bruce, Vicki & Young, Andrew W. (2011): Face Perception. Psychology Press.

Bruckmeier, Karl (2013): Natural Resource Use and Global Change. New Interdisciplinary Perspectives in Social Ecology. Basingstoke & New York: Palgrave Macmillan.

Bucchi, Massimiano (2004): Inside the laboratory. In: Science in Society. An Introduction to Social Studies of Science. New York: Routledge. S. 61-76.

Burri, Regula Valérie & Dumit, Jospeh (2008): Social Studies of Scientific Imaging and Visualization. In: Hackett et al. (ed.): The Handbook of Science and Technology Studies. Third Edition. MIT, 297–317.

Burri, Regula Valérie, Schubert, Cornelius & Strübing, Jörg (2011): Introduction: The Five Senses of Science. In: *Science, Technology & Innovation Studies* Vol.7, No.1, 3-7.

Burri, Regula Valérie (2012): Visual rationalities: Towards a sociology of images. In: *Current Sociology* Jan 2012 Vol. 60 Nr.1, pp.45-60.

Burri, Regula Valérie (2013): Visual Power in Action: Digital Images and the Shaping of Medical Practices. In: *Science as Culture* 22 (3): 367-387.

Busch, Lawrence (2011): Standards. Recipes for Reality. The MIT Press.

Callon, Michel (1986), Elements of a Sociology of translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. In: Law, John (Ed.): Power, Action and Belief: A New Sociology of Knowledge? London, Routledge: 196-233.

Caplan, Jane & Torpey, John (Hrsg.) (2001): Documenting Individual Identity. The Development of State Practices in the Modern World. Princeton.

Caplan, Jane (2001): "This or That Particular Person". Protocols of Identification in Nineteenth-Century Europe. In: Jane Caplan & John Torpey (ed.): Documenting Individual Identity. The Development of State Practices in the Modern World. Princeton/Oxford: Princeton University Press, 49–66.

Cole, Simon (2001): Suspect Identities. A History of Fingerprinting and Criminal Identification. Cambridge, MA.

Collins, Harry & Evans, Robert (2008): You cannot be serious! Public understanding of technology with special reference to „Hawk-Eye“. In: *Public Understanding of Science* 17: 283-308.

Collins, Harry & Evans, Robert (2012): Sport-decision aids and the "CSI-effect": Why cricket uses Hawk-Eye well and tennis uses it badly. In: *Public Understanding of Science* 21: 904-921.

Collins, Harry & Kusch, Martin (1998): The Shape of Actions. What Humans and Machines Can Do. The MIT Press.

Collins, Harry M. (1988): Public Experiments and Displays of Virtuosity: The Core-set revisited. In: *Social Studies of Science* 18 (4): 725-748.

Collins, Harry M. (1995): Science Studies and Machine Intelligence. In: Jasanoff, Sheila, Markle, Gerald, Petersen, James and Pinch, Trevor (Hrsg.): Handbook of Science and Technology Studies. Thousand Oaks, London, New Delhi: Sage. S. 286-301.

Collins, Harry M. (2010): Tacit and Explicit Knowlegde. University of Chicago Press.

Cooley, Charles Horton (1926): The Roots of Social Knowledge. In: *American Journal of Sociology* 32 (1): 59-79.

Cowan, Ruth Schwartz (1983): More work for mother: The ironies of household technology from the open hearth to the microwave. New York: Basic Books.

Daston, Lorraine (2001): Wunder, Beweise und Tatsachen: zur Geschichte der Rationalität. Frankfurt am Main: Fischer Taschenbuch.

Daston, Lorraine & Gallison, Peter (1992): The Image of Objectivity. In: *Representations* 40.

Daston, Lorraine & Gallison, Peter (2007): Objektivität. Frankfurt am Main: Suhrkamp.

Davis, Natalie Z. (1983): The Return of Martin Guerre, Cambridge/MA, London.

De Laet, Marianne & Mol, Annemarie (2000): The Zimbabew push pump: Mechanics of a fluid technology. In*: Social Studies of Science* 30(2): 225-263.

De Vries, Katja (2010): Identity, profiling algorithms and a world of ambient intelligence. In: *Ethics and Information Technology* 12/2010, S. 71-85.

Degele, Nina (2002): Einführung in die Techniksoziologie. Wilhelm Fink Verlag München.

Deleuze, Gilles (1992): Postscript on the Societies of Control.

Derrida, Jacques (1992): From "Shibboleth: For Paul Celan". In: Attridge, Derek (Ed.): Acts of literature. New York: Routledge, pp. 370–413.

Eisenberger, Iris (2011): Die Macht der Algorithmen. Der Verlust der Öffentlichkeit durch Personalisierung im Netz. In: *juridikum. Zeitschrift für Kritik/Recht/Gesellschaft* Nr.4/2011, S.517-522.

Ekman, Paul (1988): Gesichtsausdruck und Gefühl. 20 Jahre Forshcung von Paul Ekman. Paderborn: Junfermann.

Ekman, Paul (2007): Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life. New York: Owl.

Ensmenger, Nathan (2010): The Computer Boys Take Over. Computers, Programmers, and the Politics of Technical Expertise. The MIT Press.

FAO (2007): The State of the World's Animal Genetic Resources for Food and Agriculture. Edited by Barbara Rischkowsky & Dafydd Pilling. Rome.

Felder, Kay (2010): Gewichtige Erzählungen. Mediale Konstruktionen von Übergewicht zwischen genetischem Determinismus und Eigenverantwortlichkeit. University of Vienna: Master's thesis.

Felt, Ulrike (2007): Zukunftsszenarien als wissenschaftliche Ressource: Begegnungen zwischen Wissenschaft, Politik und Medien. In: Egloff, R., Folkers G. & Michel, M. (eds.): Archäologie der Zukunft. Zürich: Chronos. S.287-302.

Felt, Ulrike (2013): Keeping Technologies Out: Sociotechnical Imaginaries and the Formation of a National Technopolitical Identity; Pre-Print; Published by the Department of Social Studies of Science, University of Vienna, February 2013. Available at http://sciencestudies.univie.ac.at/publications

Felt, Ulrike (ed) (2009): Knowing and Living in Academic Research. Convergence and Heterogeneity in Research Cultures in the European Context. Prague: Institute of Sociology of the Academy of Sciences of the Czech Republic.

Felt, Ulrike, Fochler, Maximilian & Strassnig, Michael (2010): Experimente partizipativer ELSA-Forschung. Eine methodenpolitische Reflexion. In: Grießler, Erich & Rohracher, Harald: Genomforschung - Politik - Gesellschaft. Perspektiven auf ethische, rechtliche und soziale Aspekte der Genomforschung. Aus der Reihe: Österreichische Zeitschrift für Soziologie Sonderhefte Bd. 10. S.33-67

Fetterman, David M. (2010): Ethnography. Step-by-Step. Sage: 3rd edition.

Fisher, Erik & Mahajan, Roop L. (2006): Midstream modulation of nanotechnology research in an academic research laboratory. In: *Proceedings of the ASME international mechanical engineering congress and exposition—IMECE 2006*, November 5–10, Chicago.

Forsythe, Diana E. (1993): Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. In: *Social Studies of Science* 23 (3): 445-477.

Forsythe, Diana E. (2001): Studying those who study us. An Anthropologist in the world of Artificial Intelligence. Stanford University Press.

Foucault, Michel (1979): Discipline and Punish. The Birth of the Prison. Harmondsworth: Penguin Books, 195–228 (Panopticism).

Fox, Renee C. & Swazey, Judith (1992): Spare parts. Oxford University Press.

Frisby John P. & Stone, James V. (2010): Seeing. The Computational Approach to Biological Vison. Cambridge & London: The MIT Press.

Gates, Kelly (2010): The Tampa “Smart CCTV” Experiment. In: *Culture Unbound* 2: 67–89.

Gates, Kelly (2011): Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance. NYU Press.

Geertz, Clifford (1973): Thick Description: Toward an Interpretive Theory of Culture. In: The Interpretation of Cultures: Selected Essays. New York: Basic Books, pp. 3-30.

Gobo, Giampietro (2008): Doing Ethnography. London: Sage.

Goffman, Erving (1959): The Presentation of Self in Everyday Life. New York: Doubleday Anchor.

Gooding, David, Pinch, Trevor & Schaffer, Simon (ed.) (1989): The Uses of Experiment. Studies in the Natural Sciences. Cambridge University Press.

Gooding, David, Pinch, Trevor & Schaffer, Simon (eds.) (1989): The Uses of Experiment: Studies in the Natural Sciences. Cambridge University Press.

Goodwin, Charles (1994): Professional Vision. In: *American Anthropologist* 96 (3): 606-633.

Graham, Stephen & Wood, David M. (2003): Digitizing Surveillance: Categorization, Space, Inequality, In: *Critical Social Policy* 23 (2003); 227-248.

Groebner, Valentin (2001): Describing the Person, Reading the Signs in Late Medieval and Renaissance Europe. Identity Papers, Vested Figures, and the Limits of Identification, 1400–1600. In: Jane Caplan & John Torpey (ed.): Documenting Individual Identity. The Development of State Practices in the Modern World. Princeton/Oxford: Princeton University Press, 15–27.

Gross, Ralph, Shi, Jianbo & Cohn, Jeff (2001): Quo Vadis Face Recognition? http://www.face-rec.org/interesting-papers/General/gross_ralph_2001_4.pdf [03.02.2014]

Gruber, Stephan (2012): "Continuous Evidence": Vienna's Police and the Documentation of Identities, 1782–1848. In: Kristina Brišová & Jan Knejp (Hrsg.): Changes of Central Europe since the Middle of the 18th Century to Present. Contributions of the International Conference of Doctoral Candidates in Ostrava, May 25, 2011. Ostrava: University of Ostrava, pp. 105–115.

Gruber, Stephan, Meßner, Daniel & Musik, Christoph (2012): Personen identifizieren-Eine Geschichte von Störfällen. In: *Kriminologisches Journal* 44 (3): 219-224.

Gschrey, Raul (2012): "Der typische Deutsche" oder !Automatisierte Erkennung erfordert individuelle Charakteristika-sei durchschnittlich". In: *Kriminologisches Journal* 44 (3): 215-218.

Halfon, Saul (1998): Collecting, Testing, Convincing: Forensic DNA Experts in the Courts.In: *Social Studies of Science*, 28/5-6, pp. 801-827.

Hammersley, Martyn & Atkinson, Paul (2007): Ethnography. Principles in Practice. London and NY: Routledge. Third Edition.

Hanke, Christine (2006): „Ein klares Bild der Rassen?" Visualisierungstechniken der physischen Anthropologie um 1900. In: Heßler, Martina (Hrsg.): Konstruierte Sichtbarkeiten. Wissenschaft- und Technikbilder seit der Frühen euzeit. Wilhelm Fink. pp.241-261.

Haraway, Donna (1988): Situated Knowledges. The Science Question in Feminism and the Privilege of Partial Perspective. In: *Feminist Studies* 14 (3): 575-599.

Haraway, Donna (1991): Simians, cyborgs, and women: The reinvention of nature. New York: Routledge.

Haraway, Donna (1997): Modest_Witness @Second_Millenium. FemleMan_Meets_OncoMouseTM: Feminism and Technoscience. New York: Routledge.

Hargreaves, Ian, Lewis Justin & Speers, Tammy (2003): Towards a Better Map: Science, the Public and the Media. Swindon: Economic and Social Research Council.

Hecht, Gabrielle (2001): Technology, politics, and national identity in France. In: Allen, M.T. & Hecht, G. (eds.): Technologies of Power. Cambridge: MIT Press, pp. 253-293.

Helle, Horst Jürgen (1977): Verstehende Soziologie und Theorie der Symbolischen Interaktion, Stuttgart, Teubner.

Heßler, Martina (2006): Einleitung. Annährungen an Wissenschaftsbilder. In: Heßler (Hrsg.): Konstruierte Sichtbarkeiten. Wissenschaft- und Technikbilder seit der Frühen euzeit. Wilhelm Fink. pp.11-37.

Heuer, Steffen (2011): Nie mehr anonym. In: *Technology Review* 07/2001.

Higgs, Edward (2011): Identifying the English. A History of Personal Identification 1500 to the Present. London, New York.

Hine, Christine (2007): Multi-sited Ethnography as a Middle Range Methodology for Contemporary STS. In: *Science, Technology & Human Values* Vol. 32 Nr.6, 652-671.

Hughes, John, O`Brien, Jon, Rodden, Tom & Rouncefield, Mark (2000): Ethnography, communication and support for design. In: Luff, Paul, & Heath, Christian (eds.): Workplace Studies. Recovering Work Practice and Informing System Design. Cambridge University Press, 2000, pp.187–215.

Introna, Lucas D. & Wood, David (2004): Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems. In: *Surveillance & Society* 2 (2/3): 177–198.

Introna, Lucas D. and Nissenbaum, Helen F. (2009): Facial Recognition Technology: A Survey of Policy and Implementation Issues. Center for Catastrophe Preparedness and Response, New York University. Available at SSRN: http://ssrn.com/abstract=1437730

Ioannou, Spiros et al. (2007): Robust Feature Detection for Facial Expression Recognition. In: *EURASIP Journal on Image and Video Processing*, Volume 2007, Article ID 29081, 22 pages.

Irwin, Alan & Wynne, Brian (1996): Misunderstanding Science? The Public Reconstruction of Science and Technology. England: Cambridge University Press.

Jäger, Jens (2001): Photography. A Means of Surveillance? Judicial Photography. 1850 to 1900. In: *Crime, History & Societies* 5 (1): 27–51.

Jasanoff, Sheila & Kim, Sang-Hyun (2009): Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. In: *Minerva* 47: 119-146.

Jasanoff, Sheila (1998): The eye of everyman: Witnessing DNA in the Simpson Trial. In: *Social Studies of Science* 28: 713-740.

Jasanoff, Sheila (2004): States of knowledge. The co-production of science and social order. London/New York.

Johnson, Deborah G. & Wetmore, Jameson M. (2009): Technology and Society: Building our Sociotechnical Future. The MIT Press.

Kammerer, Dietmar (2008): Bilder der Überwachung. Frankfurt am Main: Suhrkamp.

Kemper, Oliver (2010): ELSA in der Genomforschung – eine vorübergehende Erscheinung? In: Grießler, Erich & Rohracher, Harald (ed.): Genomforschung – Politik – Gesellschaft. In: *Österreichische Zeitschrift für Soziologie* Sonderheft 10/2010.

Kirby, David A. (2003): Science Consultants, Fictional Films, and Scientific Practice. In: *Social Studies of Science* 33 (2): 231-268.

Kirby, David A. (2011): Lab Coats in Hollywood. Science, Scientists, and Cinema. MIT Press.

Kleif. T. & Faulkner, Wendy (2003):"I'm no athlete but I can make this thing dance"—men's pleasures in technology. In: *Science, Technology and Human Values* 28: 296–325.

Kline, Ronald (2003): Resisting Consumer Technology in Rural America: The Telephone and Electrification. In: Oudshoorn, Nelly & Pinch, Trevor (Eds.): How Users matter. The Co-Construction of Users and Technology. Cambridge & London: The MIT Press, pp. 51-66.

Knoblauch, Hubert & Heath, Christian (1999): Technologie, Interaktion und Organisation: Die Workplace Studies. In: *Swiss Journal of Sociology*, 25(2) (1999), 163–181.

Knoblauch, Hubert, Schnettler, Bernt, Raab, Jürgen & Soeffner, Hans-Georg (2006): Video Analysis: Methodology and Methods. Qualitative Audiovisual Data Analysis in Sociology, Frankfurt, Berlin, Bern, Bruxelles, N.Y., Oxford, Wien, Peter Lang, 2006.

Knorr Cetina, Karin (2001): Viskurse der Physik. In: Heintz, Bettina & Huber, Jörg (Hrsg.): Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Zürich: Ed. Voldemeer. S. 305-320.

Knorr-Cetina, Karin (1981): The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science. Oxford, N.Y., Toronto, Sydney, Paris, Frankfurt: Pergamon.

Knorr-Cetina, Karin (1984): Die Fabrikation von Erkenntnis: zur Anthropologie der Naturwissenschaft. Frankfurt am Main.

Knorr-Cetina, Karin (1995): Laboratory Studies. The Cultural Approach to the Study of Sceince. In: Jasanoff, Sheila, Markle, Gerald, Petersen, James and Pinch, Trevor (Hrsg.): Handbook of Science and Technology Studies. Thousand Oaks, London, New Delhi: Sage. S. 140-166.

Knuth, Donald E. (1968): The Art of Computer Programming. Vol.1: Fundamental Algorithms. Reading, MA: Addison-Wesley.

Krasmann, Susanne (2005): Mobilität: Videoüberwachung als Chiffre einer Gouvernementalität der Gegenwart. In: Hempel, Leon & Metelmann, Jörg (Hrsg.): Bild – Raum – Kontrolle. Videoüberwachung als Zeichen gesellschaftlichen Wandels. Frankfurt am Main: Suhrkamp.

Kropatsch, Walter G. (2008): History of Computer Vision. A Personal Perspective. http://www.icg.tugraz.at/Members/hgrabner/historyOfCV/kropatsch.pdf [03.02.2014]

Kruse, Corinna (2010a): Forensic Evidence: Materializing Bodies, materializing crimes. In: *European Journal of Women's Studies* 2010 (17): 363.

Kruse, Corinna (2010b): Producing Absolute Truth: CSI Science as Wishful Thinking. In: *American Anthropologist* 112 (1): 79–91.

Kuhn, Thomas S. (1962): The Structure of Scientific Revolutions. University of Chicago Press.

Lam, Alice (2010): From ‚Ivory Tower Traditionalists' to ‚Entrepreneurial Scientists'?: Academic Scientists in Fuzzy University-Industry Boundaries. In: *Social Studies of Science* 40 (2): 307-340.

Länger, Carolin (2002): Im Spiegel von Blindheit. Eine Kultursoziologie des Sehsinnes. Stuttgart: Lucius & Lucius.

Latour, Bruno & Woolgar, Steve (1979, 1986): Laboratory Life. The Construction of Scientific Facts. Princeton.

Latour, Bruno (1983): Give me a laboratory and I will raise the world. In: Knorr-Cetina K. & Mulkay M. (eds.): Science observed. Perspectives on the social study of science. London: Sage. 141-170.

Latour, Bruno (1987): Science in Action. How to Follow Scientists and Engineers through Society. Harvard University Press.

Latour, Bruno (1991): Technology is society made durable. In: Law, John (ed.): A sociology of monsters: essays on power, technology and domination. London & New York: Routledge, pp.103-131.

Latour, Bruno (1993): We have never been modern. Harvard University Press.

Latour, Bruno (1999): Pandora's hope: essays on the reality of science studies. Harvard University Press.

Latour, Bruno (2000): The Berlin key or how to do words with things. In: Graves-Brown, P.M. (Ed.): Matter, Materiality and Modern Culture. London & New York: Routledge, pp. 10-21.

Latour, Bruno (2005): Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press.

Latour, Bruno (2007): Eine neue Soziologie für eine neue Gesellschaft. Frankfurt am Main.

Law, John & Hassard, John (eds.) (1999): Actor Network Theory and After. Malden, MA: Blackwell

Law, John & Mol. Annemarie (2001): Situating Technoscience: an Inquiry into Spatialities. In: *Society and Space*, 19, 609-621.

Law, John & Urry, John (2004): Enacting the Social. Published by the Department of Sociology and the Centre for Science Studies, Lancaster University, UK, at http://www.lancs.ac.uk/fass/sociology/papers/law-urry-enacting-the-social.pdf

Law, John (2004): After Method. Mess in social science research. London & New York: Routledge.

Law, John (2008): On sociology and STS. In: *The Sociological Review* 56 (4): 623-649.

Lessig, Lawrence (1999): Code – and Other Laws of Cyberspace. New York: Basic Books.

Ley, Barbara L., Jankowski, Natalie, Brewer, Paul R. (2010): Investigating CSI: Portrayals of DNA testing on a forensic crime show and their potential effects. In: *Public Understanding of Science* Online First, doi: 10.1177/0963662510367571

Livingstone, David N. (2003): Putting Science in ist place: Geographies of Scientific Knowledge. Chicago, London: The University of Chicago Press.

LSE (2005): The Identity Project: An Assessment of the UK Identity Cards Bill & Its Implications. Interim Report (London: London School of Economics, 2005).

Lynch, Michael (1990): The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. In: Lynch, Michael & Woolgar, Steve (ed.): Representation in Scientific Practice. The MIT Press, pp.153–186.

Lynch, Michael, Cole, Simon A., McNally, Ruth, & Jordan, Kathleen (2008): Truth Machine: The Contentious History of DNA Fingerprinting. University of Chicago Press.

Lyon, David (2003): Surveillance as Social Sorting. Privacy, Risk, and Digital Discrimination. Abingdon.

Lyon, David (2007a): Surveillance, Security and Social Sorting: Emerging Research Priorities. In: *International Criminal Justice Review* 17 (3): 161–170.

Lyon, David (2007b): Surveillance Studies. An Overview. Cambridge.

Magnet, Shoshana Amielle (2011): When Biometrics Fail. Gender, Race, and the Technology of Identity. Duke University Press.

Manghani, Sunil (2012): Anonymous: The Dilemma of the Mask. http://futurezone.at/english/anonymous-the-dilemma-of-the-mask/24.579.305 [03.02.2014]

Meßner, Daniel (forthcoming): Registrieren, dokumentieren und wiedererkennen – Die Einführung biometrischer Identifizierungstechniken um 1900. Universität Wien: Dissertation.

Michael, Mike (2006): Technoscience and Everyday Life: The Complex Simplicities of the Mundane. Maidenhead: Open University Press.

Mitchell, William J. T. (2001): The Reconfigured Eye: Visual Truth in the Post-photographic Era. The MIT Press.

Mitchell, William J.T. (2002): Showing seeing: a critique of visual culture. In: *Journal Of Visual Culture* 1(2): 165-181.

Mitchell, William J.T. (2005): What do Pictures Want? the Lives and Loves of Images. University of Chicago Press, Chicago.

Mitchell. William J. (1994): Picture Theory: Essays on Verbal and Visual Representation. University of Chicago Press.

Mnookin, Jennifer, Cole, Simon A., Itiel Dror, Barry A.J. Fisher, Max M. Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, Michael D. Risinger, Norah Rudin, Jay Siegel, and David A. Stoney (2011): The Need for a Research Culture in the Forensic Sciences, In: 58 UCLA Law Review 725 (2011).

Möllers, Norma & Jens Hälterlein (2012): Privacy Issues in Public Discourse. The Case of „Smart“ CCTV in Germany. In: Innovation. The European Journal of Social Science Research. DOI:10.1080/13511610.2013.723396

Monahan, Torin (2010): Surveillance in the Time of Insecurity. New Brunswick, New Jersey, and London: Rutgers University Press.

Moses, Yael, Adini, Yael & Ullman, Shimon (1997): Face Recognition: The Problem of Compensating for Changes in Illumination Direction. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 19.7 (1997): 721-732.

Mraczny, Sebastian (2012): Blindheit und die kulturelle Produktion visuellen Sinns. In: Zeitschrift für Österreichische Soziologie (2012) 37:185-201.

Musik, Christoph (2009): Die Sehnsucht, das Innere des Menschen in seinem Äußeren zu erkennen. Von der Pysiognomik zur automatischen Gesichtsausdruckerkennung. Universität Wien: Masters thesis.

Musik, Christoph (2011): The thinking eye is only half the story: High-level semantic video surveillance. In: *Information Polity* 16 (4): 339-353.

Norris, Clive & Armstrong, Garry (1999): The Maximum Surveillance Society: The Rise Of CCTV. Oxford.

Norris, Clive (2002): From Personal to Digital: CCTV, the Panopticon and the Technological Mediation of Suspicion and Social Control, In: Surveillance as Social Sorting. D. Lyon (ed.), London, Routledge, 2002, pp.249-281.

Norris, Clive (2003): From personal to digital CCTV, the panopticon, and the technological mediation of suspicion and social control, In: Surveillance as Social Sorting. Privacy, Risk and Digital Discrimination. D. Lyon (ed.): London and New York, Routledge, 2003.

Nowotny, Helga, Scott, Peter & Gibbons, Michael (2001): Re-thinking Science—Knowledge and the Public in an Age of Uncertainty. Cambridge, UK, Polity Press, 2001), pp. 232.

Orlikowski, Wanda J. (2007): Sociomaterial Practices: Exploring Technology at Work. In: *Organization Studies* 28 (09): 1435-1448.

Oudshoorn, Nelly & Pinch, Trevor (2003): How Users and Non-Users Matter. In: Oudshoorn, Nelly & Pinch, Trevor (Eds.): How Users matter. The Co-Construction of Users and Technology. Cambridge & London: The MIT Press, pp. 1-25.

Oudshoorn, Nelly (2003): The Male Pill. A Biography of a Technology in the Making. Duke University Press.

Oudshoorn, Nelly (2011): Telecare Technologies and the Transformation of Health Care. Houndmills, Basingstoke, New York: Palgrave Macmillan.

Oudshoorn, Nelly (2012): How Places Matter. Telecare Technologies and the Changing Spatial Dimensions of Healthcare. In: *Social Studies of Science*, 42 (1): 121-142.

Person, Jutta (2005): Der pathographische Blick: Physiognomik, Atavismustheorien und Kulturkritik 1870-1930. Königshausen & Neumann.

Picard, Rosalind W. (2001): Building HAL: Computers that sense, recognize, and respond to human emotion, in: Human Vision and Electronic Imaging VI, part of IS&T/SPIE9s Photonics West, Society of Photo-Optical Instrumentation Engineers.

Pinch, Trevor (1993): "Testing - One, Two, Three... Testing!": Toward a Sociology of Testing. *Science, Technology, & Human Values*, Vol. 18, No. 1, , pp. 25-41.

Pinch, Trevor J. & Bijker, Wiebe E. (1987): The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit of Each Other. In: Bijker, Wiebe E., Hughes, Thomas P. & Pinch, Trevor J. (ed.): The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology. The MIT Press.

Prainsack, Barbara & Kitzberger, Martin (2009): DNA Behind Bars: Other Ways of Knowing Forensic DNA Technologies. In: *Social Studies of Science* 39 (1), 51–79.

Prainsack, Barbara & Toom, Victor (2010): The Prüm Regime. Situated Dis/Empowerment in Transnational DNA Profile Exchange. In: *British Journal of Criminology* 50, 1117–1135.

Prainsack, Barbara (2010): Partners in crime: the use of forensic DNA technologies in Austria. In: Hindmarsh, Richard & Prainsack, Barbara (eds.): Genetic Suspects. Global Governance of Forensic DNA Profiling and Databasing. Cambridge University Press, pp.153-174.

Rammert, Werner & Schubert, Cornelius (ed) (2006): Technografie. Zur Mikrosoziologie der Technik. Frankfurt/New York.

Rammert, Werner (2007): Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie, Wiesbaden, VS Verlag, 2007.

Richter Rudolf (2002): Verstehende Soziologie. Wien: Facultas.

Rosenfeld, Azriel (1969): Picture Processing by Computer. In: *ACM Computing Surveys (CSUR)* Volume 1 Issue 3, pp.147-176.

Rövekamp, Elke (2004): Das unheimliche Sehen - Das Unheimliche sehen. Zur Psychodynamik des Blicks. FU Berlin: Dissertation.

Sacks, Harvey (1978): Notes on Police Assessment of Moral Character'. In: Van Maanen, J. & Manning, P.K. (ed.): Policing: a view from the street. New York, Random House.

Schatzki, Theodore R., Knorr-Cetina, Karin & von Savigny, Eike (2001): The Practice Turn in Contemporary Theory. New York: Routledge.

Schmölders, Claudia (1997): Das Vorurteil im Leibe: eine Einführung in die Physiognomik. Berlin: Akademie Verlag.

Schnettler, Bernt & Pötzsch, Frederik S. (2007): Visuelles Wissen. Erscheint in: Rainer Schützeichel (Hrsg.): Handbuch Wissenssoziologie und Wissensforschung, Konstanz: UVK. pp.472 – 484.

Schulz-Schaeffer, Ingo (2000): Sozialtheorie der Technik. Frankfurt/Main.

Schuurbiers, Dan (2011): What happens in the Lab: Applying Midstream Modulation to Enhance Critical Reflection in the Laboratory. In: *Science and Engineering Ethics* (2011) 17:769-788.

Shapiro, Linda G. (1983): Computer Vision Systems: Past, Present, and Future. NATO ASI Series Volume 4, 1983, pp 199-237.

Sheller, Mimi & Urry, John (2006): The new mobilities paradigm. In: *Environment and Planning*; 38; 207-226.

Silverstone, Roger & Hirsch, Eric (Eds.) (1992): Consuming Technologies: Media and information in domestic spaces. London/New York: Routledge.

Simakova, Elena (2010): RFID 'Theatre of the Proof': Product launch and technology demonstration as corporate practices. In: *Social Studies of Science* 40/4, 549-576.

Simon, Herbert (1969): The Sciences of the Artificial. The MIT Press.

Sismondo, Sergio (2010): An Introduction to Science and Technology Studies. 2nd ed. Blackwell Publishing.

Smith, Wally (2009): Theatre of Use: A Frame Analysis of Information Technology Demonstrations. In: *Social Studies of Science* 39: 449-480.

Sonka, Milan, Hlavac. Voclav, Boyle, Roger (2008): Image Processing, Analysis, and Machine Vision. CL-Engineering. 3rd edition.

Stacy, M. (2003): “Tampa Police Eliminate Facial-Recognition System”, *Atlanta Journal-Constitution*.

Stork, David G. (1997): HAL’s Legacy: 2001’s Computer as Dream and Reality. The MIT Press.

Sturken, Marita & Cartwright, Lisa (2001): Practices of Looking: An Introduction to Visual Culture. Oxford University Press.

Suchman, Lucy (1987): Plans and Situated Actions. The problem of human machine communication. Cambridge Cambridge University Press.

Suchman, Lucy (2007): Human-Machine Reconfigurations. Plans and Situated Actions. 2nd Edition. Cambridge University Press.

Suchman, Lucy (2008): Feminist STS and the Sciences of the Artifcial. In: E. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman (eds.): The Handbook of Science and Technology Studies, Third Edition. Cambridge & London: MIT. S.139-164.

Suchman, Lucy, Trigg, Randall & Blomberg, Jeanette (2002): Working artefacts: ethnomethods of the prototype. In: *British Journal of Sociology* 53 (2): 163-179.

Suchman, Lucy, Danyi, Endre, & Watts, Laura (2008). Relocating Innovation: places and material practises of future-making. Centre for Science Studies, Lancaster, UK.

Surette, Ray (2005): The Thinking Eye. Pros and cons of second generation CCTV surveillance systems, In: *Policing: An International Journal of Police Strategies & Management* 28/1 (2005), 152-173.

Swoboda, Gudrun (2002): Lavaters Linienspiele. Techniken der Illustration und Verfahren graphischer Bildbearbeitung in einer physiognomischen Studiensammlung des 18. Jahrhunderts. Wien: Dissertation.

The Royal Academy of Engineering (2009): Autonomous Systems: Social, Legal and Ethical Issues. Copy available: www.raeng.org.uk/autonomoussystems

Timmermans, Stefan & Epstein, Steven (2010): A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. In: *Annual Review of Sociology* 2010. 36:69–89.

Tomomitsu, Jennifer (2011): Tinkering with the object. Investigating Manipulation and enactments of ‘seeing’ in scientific imaging practice. PhD thesis Lancaster University.

Traweek, Sharon (1988): Beamtimes and Lifetimes. The World of High Energy Physicists. London: Harvard University Press

Turaga, P., Chellappa,R., Subrahmanian, V.S. and Udrea, O. (2008): Machine Recognition of Human Activities: A Survey, CSVT 18(11) (2008), 1473-1488.

Tutton, Richard (2011): Promising Pessimism: Reading the Futures to be Avoided in Biotech. In: *Social Studies of Science*, 1-19.

Van der Ploeg, Irma (2011): Normative Assumptions in Biometrics: On Bodily Differences and Automated Classifications. In: Simone van der Hoof & Marga M. Groothuis: Innovating Government. Normative, Policy and Technological Dimensions of Modern Government. The Hague: TMC Asser Press, 29–40.

Verbeek, Peter-Paul (2011): Moralizing Technology. Understanding and Designing the Morality of Things. The University of Chicago Press.

Weyer, Johannes (2006): Modes of Governance of Hybrid Systems. The Mid-Air Collision at Ueberlingen and the Impact of Smart Technology. In: *Science, Technology & Innovation Studies* Vol. 2, Nov. 2006, 127-149.

Wimmer, Matthias et al. (2008): Tailoring Model-based Techniques to Facial Expression Interpretation. achi, pp.303-308, First International Conference on Advances in Computer-Human Interaction, 2008.

Winner, Langdon (1980): Do Artefacts Have Politics? In: Daedalus 109: 121-136.

Wyatt, Sally (2008): Technological Determinism is Dead; Long Live Technological Determinism. In: Hackett et al. (ed.): The Handbook of Science and Technology Studies. Third Edition. MIT, 165-180.

Appendix

ASFINAG eMail information about AVK (May 2012)

Zum Einsatz der „Automatischen Vignettenkontrolle (AVK) können wir Ihnen gerne folgende Informationen übermitteln:

- Erste AVK-Anlage ging im Dezember 2007 in Betrieb, **derzeit sind insgesamt fünf solcher mobilen Anlagen in Betrieb**

- **AVK – aus Verkehrssicherheitsgründen und als unterstützende Maßnahme zu manuellen Kontrollen**
  Die AVK ist ein digitales Kamerasystem und ist als ergänzende bzw. unterstützende Kontrollmaßnahme zur manuellen Vignettenkontrolle durch die Mautaufsicht zu sehen. Dort, wo aus Gründen der Verkehrssicherheit keine Ausleitung möglich ist, kommt die AVK zum Einsatz (z.B. auf mehrspurigen Autobahnen in Ballungsräumen, Autobahnen ohne Pannenstreifen, etc.). Sie dient damit dem erhöhten Sicherheitsaspekt zugunsten unsere Kunden, aber auch der Mautaufsichts-Mitarbeiter. Die AVK-Anlagen sind kein Ersatz für die manuelle Kontrolle durch die Mautaufsicht.
  Mit der AVK erfolgt eine stichprobenartige Überprüfung, mit häufiger (in etwa wöchentlicher) Versetzung der Anlagen.

- **AVK – zur Steigerung der Vignettenmoral**
  Vignettenmoral ist an und für sich mit rund **98%** schon sehr hoch - die AVK dient der weiteren Steigerung der Vignettenmoral, vor allem i.S. unserer Verantwortung wie im Gesetz vorgesehen: **Faire Behandlung aller Autobahnbenutzer, so dass nicht einige zu Lasten der Gesamtheit das Netz ohne Vignette benutzen.**

- **Datenschutz**
  Diese Technologie kommt bundesweit an immer wieder wechselnden Standorten am gesamten Autobahnen- und Schnellstraßennetz zum Einsatz. Erfasst und geahndet werden nur jene Fahrzeuge, wo nachweislich ein Vergehen gegen die Mautpflicht in Österreich vorliegt. Aufgenommen werden ein Überblicksbild mit Kennzeichen sowie ein Detailbild von der Windschutzscheibe. Der Kamerawinkel ist dazu so eingestellt, dass keine Gesichter von Lenker und Beifahrer erkennbar sind. Die erfassten Daten werden zur Kontrolle auch manuell nachbearbeitet. Es gilt: im Zweifel für den Kunden! Für die AVK besteht auch eine den Vorgaben

des Datenschutzgesetzes entsprechende Grundlage und die Datenanwendung wurde vor Inbetriebnahme beim Datenverarbeitungsregister ordnungsgemäß angemeldet und registriert.

Table of Articles (Chapter Four)

Article ID	*Date*	*Nature of Article*	*Source of Article*	*Original German Title of Article*
a01	02.05.2007	Radio/Online	Ö1 Frühjournal	Testbetrieb automatische Vignettenkontrolle
a02	29.08.2007	Newspaper	Der Standard	Mit Kameras zur "intelligenten Straße"
a03	13.11.2007	Newspaper	Kleine Zeitung	Automatische Kontrolle
a04	11.12.2007	Newspaper	Tiroler Tageszeitung	Eine Asfinag West wäre ein Schuss nach hinten
a05	12.12.2007	Magazine	News	Die erste automatische Vignetten-Kontrolle: Pickerlsünder sind nun komplett chancenlos
a06	12.12.2007	Online	ORF Wien	Kameras starten heute mit Vignettenkontrolle
a07	12.12.2007	Newspaper	Presse	Elektronisches Auge kontrolliert Vignette bei jedem Auto
a08	13.12.2007	Newspaper	Neue Kärntner	Vignettenkontrolle jetzt automatisiert
a09	13.12.2007	Newspaper	Presse	Asfinag startet automatische Vignettenkontrolle
a10	13.12.2007	Newspaper	Salzburger Nachrichten	Mautpreller im Visier
a11	13.12.2007	Newspaper	Der Standard	Länderschau - Automatische Vignettenkontrolle
a12	13.12.2007	Newspaper	Tiroler Tageszeitung	Automatische Kontrolle der Mautvignette startet
a13	13.12.2007	Newspaper	Vorarlberger Nachrichten	Jagd auf Vignettensünder
a14	19.12.2007	Press Release	ARBÖ	ARBÖ: Vignette heuer in zwei verschiedenen Layouts - beide gültig
a15	20.12.2007	Newspaper	Vorarlberger Nachrichten	Blickpunkt
a16	20.12.2007	Newspaper	Der Standard	Länderschau - Unterschiedliche Mautvignetten
a17	20.12.2007	Newspaper	Neues Volksblatt	Vignette 2008 mit zwei gültigen Layouts
a18	21.12.2007	Newspaper	Kronen Zeitung	Verwirrung um Vignette
a19	25.12.2007	Newspaper	Kleine Zeitung	Chefebene wird schlanker
a20	13.01.2008	Newspaper	Tiroler Tageszeitung	Zweite Vignetten-Variante erleichtert Kontrolle
a21	18.01.2008	Newspaper	Vorarlberger Nachrichten	Keine Chance für Mautsünder
a22	09.02.2008	Newspaper	Leipziger Volkszeitung	Österreich führt mobile Mautkontrolle ein

a23	13.02.2008	Newspaper	Express	Ösi Kontrollen
a24	23.02.2008	Newspaper	Abendzeitung München	Hightech-Jagd auf die deutschen Maut-Muffel
pq1	27.11.2008	Parliamentary Question	Harald Vilimsky (FPÖ)	vollautomatische Vignettenkontrolle (278/J)
atopq1	26.01.2009	Answer to Parliamentary Question	Doris Bures (SPÖ), Bundesministerin für Verkehr, Innovation und Technologie	Anfragebeantwortung 333/AB zu 278/J
a25	19.05.2009	Newspaper	Tiroler Tageszeitung	Mautmuffel im Westen
a26	25.11.2009	Newspaper	Oberösterreichische Nachrichten	Pro Tag werden in Österreich 270 Vignettensünder erwischt
a27	05.01.2010	Newspaper	Kurier	Mautpreller im Rückzug: Nur zwei von 100 schwindeln
a28	08.07.2010	Newspaper	Wiener Zeitung	Neuer Rekord bei Jagd auf Vignetten-Sünder
pq2	09.07.2010	Parliamentary Question	Mayerhofer, Vilimsky (FPÖ)	Vignette - Qualität und Bestrafung trotz vorhandener Vignette (6177/J)
atopq2	09.09.2010	Answer to Parliamentary Question	Doris Bures (SPÖ), Bundesministerin für Verkehr, Innovation und Technologie	Anfragebeantwortung 6113/AB zu 6177/J
a29	11.10.2010	Newspaper	Kurier	Neue Vignettenkontrolle, bessere Räumung
a30	01.02.2011	Newspaper	Tiroler Tageszeitung	Ohne Vignette droht Strafe
a31	02.02.2011	Newspaper	Salzburger Nachrichten	Keine Schonfrist für Vignettenmuffel
a32	09.02.2011	Newspaper	Kronen Zeitung	Autodieb fuhr ohne Vignette: ASFINAG straft Wagenbesitzer
a33	23.02.2011	Newspaper	Kleine Zeitung	Auf der Jagd nach den Vignetten-Sündern
a34	18.03.2011	Newspaper	Tiroler Tageszeitung	So wenig LKW wie 2004, Geldmaschine Amras
a35	20.06.2011	Press Release	EFKON AG	EFKON LIEFERT ASFINAG AUTOMATISCHES VIGNETTEN-KONTROLLSYSTEM
a36	21.06.2011	Newspaper	Neue Vorarlberger Tageszeitung	Weitere automatische Systeme zur Vignetten-Kontrolle
a37	27.06.2011	Newspaper	Neue Kärntner	Vignettenkontrollen auch bei voller Fahrt

a38	28.06.2011	Newspaper	Kronen Zeitung	Kampf gegen Mautpreller mit mobiler Vignetten-Kontrolle
a39	05.07.2011	Online	heise online	LKW Maut: Ohne Auftrag keine Lieferung
a40	23.11.2011	Newspaper	Oberösterreichische Nachrichten	Pro Tag werden im Schnitt 329 Vignettensünder erwischt
a41	31.01.2012	Newspaper	Tiroler Tageszeitung	Frist endet: Maut-Sheriffs nehmen Fahrt auf
a42	31.01.2012	Newspaper	Neue Kärntner	Um Mitternacht folgt Petrol auf Mango
a43	16.03.2012	Newspaper	Wiener Zeitung	Neuer Rekord an Vignetten-Sündern
a44	16.03.2012	Online	ORF News	Neuer Rekord an Vignettensündern
a45	17.03.2012	Newspaper	Oberösterreichische Nachrichten	Rekord an Vignetten-Sündern
a46	26.03.2012	Newspaper	Der Standard	Asfinag will 2,8 Milliarden bei Straßenbau sparen
a47	23.10.2012	Newspaper	Kleine Zeitung	Straße spricht mit dem Autofahrer

German Abstract

Alle Versuche, Maschinen und Computern die Fähigkeit des Sehen beizubringen, sind Versuche, digitale Bilder herzustellen, zu bearbeiten und vor allem ihre Inhalte zu verstehen. Zu diesem Zweck ist es zwingend notwendig, Bildverarbeitungsalgorithmen zu entwickeln und anzuwenden. Bildverarbeitungsalgorithmen werden zu einflussreichen politischen und gesellschaftlichen Akteuren und Entscheidungsträgern. Deshalb ist es wichtig, ein tiefgehendes Verständnis davon zu erreichen, wie genau diese Algorithmen Bilder erzeugen, bearbeiten und vor allem semantisch interpretieren.

“Computers and the Ability to See” basiert auf einem interdisziplinärem Zugang, welcher die akademischen Felder der Wissenschafts- und Technikforschung (STS), der visuellen Kulturstudien und der Überwachungs- und Identifizierungsstudien verbindet. Es ist insbesondere inspiriert von Lucy Suchmans Arbeit zu ‘Human-Machine Reconfigurations’ (Suchman 2007) und dem visuellen STS Zugang der ‘Social Studies of Scientific Imaging and Visualization’ (Burri & Dumit 2008). Die Dissertation schreibt sich somit in die theoretischen Rahmen des (feministischen) Posthumanismus und der materiellen Semiotik ein. Damit verbunden ist die Entscheidung, die konkreten Praktiken von nichtmenschlichen Entitäten und ihren spezifischen Handlungsfähigkeiten empirisch zu untersuchen (vgl. (Suchman 2007: 1).

Die empirische Analyse von Bildverarbeitungsalgorithmen bettet sich ein in die grundlegenden soziotechnischen Transformationsprozesse, die mit den Begriffen Überwachungsgesellschaft (hier insbesondere das Phänomen der “intelligenten” Videoüberwachung), Digitalisierung, Automatisierung und “Smartisierung” von gesellschaftlichen Praktiken, Artefakten und Geräten zusammengefasst werden können. Auf dieser Grundlage erforschte die Dissertation Mensch-Computer (Re-)Konfigurationen, indem sie die Ausverhandlung und Entwicklung mit Fokus auf die politische und gesellschaftliche Signifikanz von Bildverarbeitungsalgorithmen in unterschiedlichen Situationen und Umgebungen von den Laboren der Bildverarbeitung bis hin zu den Medien in den Blick nahm. Die Forschung folgte unter Einbeziehung eines breiten Methodenspektrums der qualitativen Sozialforschung (Teilnehmende

Beobachtung, Gruppendiskussionen, Interviews, Dokumentenanalyse) einer 'visiographischen' Strategie und entwickelt darauf aufbauend in den Schlussfolgerungen den konzeptuellen Reflektionsrahmen der "Social Studies of Image Processing Algorithms" (SIPA). Dadurch leistet die Arbeit einen wichtigen Beitrag zu der Frage, wie Gesellschaft und Wissenschaft mit Bildverarbeitungsalgorithmen in ihrer Funktion als 'politische Ordnungsapparate' in einem verantwortlichen Weg der Innovation umgehen können. Dabei ermutigt SIPA explizit die Zusammenarbeit von Sozial- und ComputerwissenschaftlerInnen sowie die Einbeziehung weiterer gesellschaftlicher Akteure wie zum Beispiel KünstlerInnen. SIPA beinhaltet also auch Fragen und Ebenen, die sich mit der Steuerung, Regulierung und mit ethischen, rechtlichen und gesellschaftlichen Aspekten von Bildverarbeitungsalgorithmen auseinandersetzen.

English Abstract

It is a basic requirement of all attempts to configure machines and computers with the ability to see, that these are in fact, attempts to produce, process and understand digital images by means of computer algorithms. Those becoming powerful social actors and decision makers, it is important to understand exactly, the production, processing, and interpretation of digital images by algorithms where the semantic interpretation element is central.

"Computers and the Ability to See" is based on an interdisciplinary, multiperspective approach that is framed by the academic fields of Science and Technology Studies (STS), Visual Culture Studies and Surveillance & Identification Studies. It especially is inspired by Lucy Suchman's work on 'Human-Machine Reconfigurations' (Suchman 2007) and the Visual STS approach of the 'Social Studies of Scientific Imaging and Visualization' (Burri & Dumit 2008). This links to what could be summarised as the theoretical frames of (feminist) post-humanism and material-semiotics, and connected to it, to the commitment "to empirical investigations of the concrete practices" of nonhuman entities and their specific agencies (Suchman 2007: 1).

The most relevant sociotechnical transformation processes that framed the empirical analysis with computer vision and more specifically with Image Processing Algorithms (IPAs) are what could be condensed in the "grand narrative" (cf. Law 2008: 629) terms of surveillance society (especially what often is referred to as Smart CCTV or intelligent video surveillance) as well as the digitalisation, automatisation, and "smartisation" of social practices, artefacts and devices. On these grounds, the thesis explored 'Human-Computer Vision (Re-) Configurations' by analysing the negotiation and the development, and by focusing on the political and social significance of Image Processing Algorithms in different sites from the computer vision laboratory to the news media. In doing so, the research followed a 'visiographic' strategy that applied a wide array of qualitative methods (participant observation, group discussions, interviews, document analysis).

In the conclusions the thesis discusses the question how societies and the sciences could deal with the 'political ordering devices' IPAs in a responsible and reflective way of innovation. In this regard it suggests the "Social Studies of Image Processing Algorithms" (SIPA), a conceptual and reflective framework for the further development and analysis of IPAs, encouraging social scientists, artists and computer scientists to reflect and work together in the specific research and innovation field, but also the business field of computer vision. The SIPA scheme also covers questions of governance, regulation, and ELSA (ethical, legal, and social aspects).

Curriculum Vitae

Name	Christoph Musik, Bakk.phil. MA
Date and Place of Birth	30.09.1983 in Rosenheim (Bavaria, Germany)

Educational Background

10/2010-2014 Doctoral Studies in the Social Sciences (Science and Technology Studies) at the Department of Science and Technology Studies, University of Vienna with stays as visiting Phd:

- 03/2011-07/2011 Centre for Science Studies, Lancaster University, UK
- 04/2013-06/2013 Institut für Kriminologische Sozialforschung, Universität Hamburg, Germany

10/2007–10/2009 Master in Sociology (MA), Department of Sociology, University of Vienna

10/2004–08/2007 Bachelor in Sociology (Bakk.phil), Department of Sociology, University of Vienna

Collaboration in Research Projects & Fellowships

03/2014-08/2014 Reader in Visual Sociology at the Department of Sociology/University

- SS 2014 VO+SE Visual Surveillance – Bilder der Überwachung (with Roswitha Breckner & Robert Rothmann)

10/2013-02/2014 Project manager at EDUCULT – Denken und Handeln im Kulturbereich

- Evaluation of the theatre project „13 Kisten“ (BRASILIEN.13 caixas)

10/2010-10/2013 Recipient of a DOC-team-fellowship of the Austrian Academy of Sciences at the Department of Social Studies of Science and Technology, University of Vienna (http://www.identifizierung.org)

08/2009-05/2013 ‚small | world | ART | project' – a participatory art project (http://www.smallworld.at), with Helene A. Musik;

‚small | world | ART | exhibition', Kubus EXPORT, Vienna 11. - 26.05.2013

03/2012-02/2013 Reader in STS at the Department of Science and Technology Studies, University of Vienna

- WS 2012/2013 UK ‚Technologie und Gesellschaft'
- SS 2012 SE 'The Same Person'. Past, Present, and Future of Identification Practices and Techniques (with Stephan Gruber & Daniel Meßner)

09/2009-09/2010 Project employee at the Institute for Advanced Studies Vienna

- ‚tripleB ID - Identifikation von Bedrohungszenarien in Banken durch Bildanalyse' (KIRAS security research scheme)
- ‚Networked miniSPOT - On the Spot Ereigniserkennung mit low-cost Minikameramodulen und Kommunikation über robuste Netzwerke der Gebäudeautomation' (KIRAS security research scheme)

09/2007–08/2009 Fellow at the Institute for Advanced Studies Vienna in the research group ‚equi' (Employment – Qualification – Innovation)

Selected Publications

MUSIK, Christoph (2012): The thinking eye is only half the story: High-level semantic video surveillance. In: Webster, C. William R. / Töpfer, Eric / Klauser, Francisco R. / Raab, Charles D. (eds.) (2012): Video Surveillance – Practices and Policies in Europe. Vol. 18 of Innovation and the Public Sector. Amsterdam, Berlin, Tokyo, Washington D.C.: IOS Press. pp. 37-51.

GRUBER, Stephan/MEßNER, Daniel/MUSIK, Christoph (2012): Personen identifizieren – Eine Geschichte von Störfallen. *Kriminologisches Journal*, Heft 3 (2012), S. 219-224.

MUSIK, Christoph (2011): The thinking eye is only half the story: High-level semantic video surveillance. *Information Polity* 16/4: 339-353.

MUSIK, Christoph & VOGTENHUBER, Stefan (2011): Soziale Implikationen automatischer Videoüberwachung. Sozialwissenschaftliche Erkenntnisse aus dem Projekt TripleB-ID. IHS Projektbericht, Wien. Im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie (BMVIT) und der Österreichischen Forschungsförderungsgesellschaft mbH (FFG).

BLAUENSTEINER, Philipp/KAMPEL Martin/MUSIK, Christoph/VOGTENHUBER, Stefan (2010): A Socio-Technical Approach for Event Detection in Security Critical Infrastructure, accepted at Intl. Workshop on Socially Intelligent Surveillance and Monitoring (SISM 2010) in

conjunction with IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA, June 2010

MUSIK, Christoph (2009): Die Sehnsucht, das Innere des Menschen in seinem Äußeren zu erkennen. Von der Physiognomik zur automatischen Gesichtsausdruckserkennung. Master thesis Universität Wien 2009.

Peer Reviewer

Public Understanding of Science (SAGE)