OULU 2014
A 622

ACTA UNIVERSITATIS OULUENSIS

A
SCIENTIAE RERUM
NATURALIUM

Florian Matusek

SELECTIVE PRIVACY PROTECTION FOR VIDEO SURVEILLANCE

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 622

FLORIAN MATUSEK

SELECTIVE PRIVACY PROTECTION FOR VIDEO SURVEILLANCE

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in OP-sali (Auditorium L10), Linnanmaa, on 7 May 2014, at 12 noon

UNIVERSITY OF OULU, OULU 2014

Acta Univ. Oul. A 622, 2014

Supervised by
Docent Raija Halonen
Doctor Reda Reda

Reviewed by
Professor Thomas P. Keenan
Professor Jari Veijalainen

Opponent
Professor Jari Porras

ISBN 978-952-62-0414-7 (Paperback)
ISBN 978-952-62-0415-4 (PDF)

ISSN 0355-3191 (Printed)
ISSN 1796-220X (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2014

Matusek, Florian, Selective privacy protection for video surveillance.
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering
Acta Univ. Oul. A 622, 2014
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

An unparalleled surge in video surveillance has occurred in recent years, due to some tragic events such as terror attacks, bank robberies and the activities of organized crime. Video surveillance technology has advanced significantly, which has even enabled the automatic tracking of individuals. However, in the opinion of the public the increase in security has brought about a decrease in personal privacy. Through video surveillance citizens could be monitored more easily than ever before, thus considerably intruding into their personal privacy. It was assumed that security and privacy in video surveillance was a zero-sum game in which citizens were forced to choose one over the other.

This study was based on the belief that this notion is false. It was assumed that it can be possible to keep personal privacy while guaranteeing the utmost security. A solution to this issue was sought using Hevner's design science research guidelines and design science research cycles. A video surveillance system was designed and constructed that would protect the personal privacy of uninvolved individuals under surveillance while still providing a high level of security, namely the Privacy Enhancing Video Surveillance system PEVS. PEVS protected the privacy of individuals by automatically scrambling the image regions where people were present in video streams. If a criminal act should take place, it was possible, with the proper authorization, to selectively unscramble the data of individuals of interest to analyze the situation. This enabled to analyze the situation without intruding into the privacy of uninvolved people on the one hand, while on the other hand using the data as evidence of possible criminal activity. Hence, the privacy of individuals was protected while maintaining the same level of security.

PEVS provided the first technology-based video surveillance solution, which showed only relevant individuals in the image while leaving the identity of everyone else unrevealed. Therefore, the main contribution of this thesis was the construction of a novel approach to video surveillance systems, capable of selectively protecting the privacy of individuals. This included introducing an architecture for a privacy preserving video surveillance system, which consisted of several sub-constructs. These included storage techniques for privacy data and shadow detection and segmentation methods, which increased the accuracy and speed of previous methods. Further, novel security and privacy metrics for video surveillance were introduced. The overall system was a significant improvement over the existing knowledge base that has thus far seen only first steps to selective privacy protection but has failed to provide a complete system.

Keywords: CCTV, computer vision, data protection, privacy, video surveillance

Matusek, Florian, Valikoiva yksityisyyden suojaus videovalvonnassa.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähkötekniikan tiedekunta
Acta Univ. Oul. A 622, 2014
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Videovalvonnassa on tapahtunut viime vuosina merkittävää kasvua johtuen järkyttävistä tapahtumista kuten terrori-iskut, pankkiryöstöt ja järjestäytyneen rikollisuuden toimet. Videovalvontateknologia on kehittynyt merkittävästi mahdollistaen jopa yksittäisten ihmisten automaattisen seurannan. Turvallisuuden lisääntymisen katsotaan kuitenkin vähentäneen yksityisyyttä. Videovalvonnan avulla ihmisiä pystytään seuraamaan helpommin kuin koskaan aikaisemmin tunkeutuen täten heidän yksityisyytensä alueelle. On oletettu, että turvallisuus ja yksityisyys videovalvonnassa on nollasummapeliä, jossa kansalaisten on valittava yksityisyyden ja turvallisuuden välillä.

Tämä tutkimus perustuu olettamukseen, että edellä esitetty ei pidä paikkaansa, vaan että on mahdollista suojata yksityisyys samalla taaten täysi turvallisuus. Ratkaisua tähän ongelmaan etsittiin suunnittelutieteellisen tutkimuksen avulla. Työssä suunniteltiin ja toteutettiin videovalvontajärjestelmä PEVS (Privacy Enhancing Video Surveillance system), joka suojaa valvonnanalaisten sivullisten yksityisyyttä ja siitä huolimatta tuottaa korkean turvallisuustason.. PEVS suojaa henkilöiden yksityisyyttä salaamalla automaattisesti videoaineistosta ne kuva-alat, joissa esiintyy ihmisiä. Mikäli laitonta toimintaa havaittaisiin, olisi riittävillä käyttöoikeuksilla mahdollista purkaa salaus mielenkiinnon kohteena olevien henkilöiden kohdalta tilanteen analysoimiseksi. Tämä mahdollisti yhtäältä puuttumattomuuden sivullisten yksityisyyteen ja toisaalta tiedon käyttämisen todistusaineistona mahdollisen rikoksen tutkimisessa. Tällä järjestelmällä yksityisyys oli mahdollista suojata samanaikaisesti, kun turvallisuudesta huolehdittiin.

PEVS mahdollisti ensimmäistä kertaa maailmassa videovalvonnan, joka näyttää vain relevantit henkilöt jättäen muiden henkilöllisyyden paljastamatta. Sen takia tämän tutkimuksen merkittävin kontribuutio oli uudenlaisen lähestymistavan kehittäminen videovalvontaan, joka kykenee valikoivasti suojelemaan ihmisten yksityisyyttä. Tämä ratkaisu sisältää yksityisyyden suojaavan, useita rakenneosia sisältävän videovalvontajärjestelmäarkkitehtuurin esittelyn. Rakenneosiin kuuluu yksityisen tiedon tallennusmenetelmiä ja varjontunnistus- ja segmentointimetodeja, jotka paransivat aiemmin käytettyjen metodien tarkkuutta ja nopeutta. Lisäksi esiteltiin uudenlainen turvallisuus- ja yksityisyysmetriikka videovalvonnalle. Toteutettu järjestelmä on huomattava lisäys nykytietämykseen, jossa yksityisyyden suojan osalta on otettu vasta ensiaskelia ja joka ei mahdollista kattavaa järjestelmää.

Asiasanat: konenäkö, tiedon suojaus, valvontakamera, videovalvonta, yksityisyys

They that can give up essential liberty to obtain a little temporary safety deserve neither liberty nor safety.

Benjamin Franklin, Historical Review of Pennsylvania, 1759.

Preface

This dissertation is the result of long hours of research and work that would not have been possible without the help of some people very dear to me.

First, I would like to thank my main supervisor Raija Halonen whose enormous patience, motivation and support greatly contributed to making this dissertation a reality. Further, I would like to thank Olli Martikainen for introducing me to the world of research and also for his patience in teaching me in intense workshops on how to work scientifically.

I also would like to thank Reda Reda, who was not only my second supervisor but who also is and will continue to be a mentor and dear friend to me. Without his support, energy and never-ending belief that I will make it, I would have achieved nothing near the result we see today. For this and everything else he has done for me in the past years I would like to express my special gratitude.

I am further indebted to my dissertation follow-up-group: Kari Kaila, Christoph Egger and Lukas Juen, who supported me after writing the thesis.

Further, I thank Stephan Sutor and Klemens Kraus, two very dear friends of mine and also my fellow co-founders of KiwiSecurity. This company the three of us have started is an amazing adventure and I firmly believe that we will lead it to great success in the future. It is nearly impossible to find two people with such tremendous knowledge and energy combined with the business acumen, which will enable us to conquer the world.

My special thanks goes to Smaranda Elena Corbeanu for her patience, support and advice during the long hours of work in the final spurt of this endeavor.

Further, I would like to thank my brother Severin Matusek, who provided together with Smaranda, input on Foucault and the Panopticon.

Last but not least my deepest gratitude goes to the people who have laid the foundation for everything I do: my parents. Without their ongoing support and wisdom to never settle for less than the best, fight for what you believe in and get out and create the next Big Thing, I probably would have stayed at home and never would have risked anything.

Thanks to all of you for being part of my life.

Abbreviations and definitions

ADR	Action Design Research
CSA	Critical Surveillance Area
CTMC	Continuous-Time Markov Chains
HOGs	Histograms of Oriented Gradients
ICT	Information and Communication Technologies
KPI	Key Performance Indicator
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
PbD	Privacy by Design
PbP	Pixel-by-Pixel (shadow detection)
PETS	Performance Evaluation of Tracking and Surveillance
PEVS	Privacy Enhancing Video Surveillance
PoD	Privacy on Demand
PQL	Privacy Quality Level
PTZ	Pan Tilt Zoom
QoP	Quality of Performance
ROI	Region of Interest (in video surveillance)
TB	Texture-based (shadow detection)
VS	Video Surveillance
VSCQ	Video Surveillance Coverage Quality
VSCR	Video Surveillance Coverage Redundancy

Artifact and compression artifact

Artifact in this dissertation refers to an IT artifact as used in Design Science Research, e.g. by Hevner *et al.* (2004). In this dissertation it thus represents a constructed system or parts thereof. Contrary to that, compression artifact refers to a noticeable distortion in an image, which is similar to image noise.

Exaptation

Exaptation is a term used in evolution science to describe a change of the function of a trait during evolution. E.g. feathers were originally developed for heat regulation and later adapted for flight. In Design Science Research, exaptation is defined as a category of research contributions where existing solutions are extended to new problems.

False positives

False positives are detections that do not correspond to a correct detection result (e.g. if an object is detected that is actually a shadow), as opposed to true positives.

False negatives

False negatives are counted if no detection occurred, even though it should have (e.g. if an object is present in the scene that was not detected), as opposed to true negatives.

Incident

Incidents are defined as events, which should be prevented, which might threaten security and which should trigger a reaction by security personnel. In the context of this dissertation what constitutes an incident depends on the definition by the organization using the proposed system. For example, an incident could be defined as a small event such as theft of a product in a shop or could include major events such as terrorist attacks.

Image segmentation

Image segmentation is the task of segmenting an image into several regions, which subsequently can be classified and tracked.

Mean Time To Failure (MTTF)

Reliability is defined as the probability of a system being up and running throughout an interval without system level repair and is expressed through the Mean Time To Failure. According to the International Telecommunication (ITU-T) recommendation E.800, reliability is defined as the ability of the system or an item to perform a required function under given conditions for a given time interval.

Mean Time To Repair (MTTR)

Mean Time To Repair is a value describing the maintainability of a system and is the average time required to repair failed components. MTTR consists of time for failure detection, alarm notification, dispatch, repair/replacement and reboot time.

Pixel-by-pixel shadow detection (PbP)

PbP shadow detection applies the shadow detection rules to each pixel separately.

Privacy

Privacy in this dissertation is defined as the condition of a person of being free from being observed. Image data in connection to privacy is understood in accordance to the 1995 EU directive on the protection of individuals (European Parliament 1995). For further information see the definition on privacy and identity information.

Privacy information / identity information

Privacy information or identity information describes attributes of a person, which make him or her identifiable. In general, this includes not only the appearance of a person in a video frame but specific movements, behavior and gait as well as voice and odor. However, in this study, privacy and identity information are defined as physical appearance, which makes a person identifiable in a single frame, i.e. the face, body and clothes, but not behavior, gait, voice or odor, which might be part of a separate study. Identity information does not encompass everything that Heikkinen *et al.* (2004) summarize as personal information, including needs, requirements and actions, but only information, which helps to identify a specific person.

Scrambling / pixelization

In this dissertation the term “scrambling” refers to obscuring a part of an image, which afterwards is not discernible anymore. Pixelization is a technique used in image and video editing whereby the resolution of parts or all of an image is reduced.

Security

In this dissertation the term security refers to a physical state of being free from danger or threat. Security includes preventing acts of crime such as, but not limited to, theft, armed robbery, suicide attacks, riots, terrorist attacks or other acts that might endanger human life. While similar, security does not include safety, which describes threats not caused intentionally such as plane crashes, fires or other accidents. Even though the presented results of this study could help improve safety as well, it was not a focus of this study.

Texture-based shadow detection (TB)

Texture-based shadow detection checks larger texture patches (usually between 7×7 to 9×9 pixel windows). Texture-based shadow detection reduces noise in the image. However, it is only feasible if the background is textured. Asphalt roads, for example, do not contain enough texture information to support reliable TB shadow detection.

Uninvolved individuals / persons

Uninvolved individuals or persons are defined as all persons present in an area under video surveillance, which are not involved in a given incident as defined by the user of a security system. Hence, uninvolved individuals in this dissertation represent all persons in a video stream whose privacy should be protected.

Contents

1 Introduction

Due to a series of terror attacks at the beginning of the 21st century and increasing criminal activity in recent years, a growing number of video surveillance cameras were deployed world-wide (Porikli *et al.* 2013). Today, video surveillance networks are being built that cover entire cities and consist of up to one million cameras (Bankok-Post September 6th 2012). Moreover, video surveillance is increasingly deployed as a sensor in fields beyond security, such as traffic monitoring, parking, safety in hospitals, elderly care and retail (Cumming & Johan 2013, Tang *et al.* 2013, Zhang *et al.* 2013). Similar to the Internet, where, as Greenwald and MacAskill (2013) revealed, almost everything is analyzed and stored, data in video surveillance can be kept for a very long time as well. If everything we do can be analyzed and be collected, the privacy of people under surveillance is being threatened (Such *et al.* 2013). This concern is further reflected in the opinion of the general public. About 90% of all adults in the United States are concerned about their privacy with 26% feeling that they have lost a lot of their privacy already (Taylor 2003). This controversy between security and privacy was the main motivation of this work, which aimed at providing a technological solution to this problem.

1.1 Purpose

Demand for security and video surveillance has constantly increased in recent years. Accordingly, video surveillance technology has seen rapid advancement in recent years, with algorithms analyzing behavior of people under surveillance automatically (Wadhwa 2012, Cristania *et al.* 2013). However, these technological advancements have been simultaneously eroding the personal privacy of people even more than they had before. Personal privacy is a valuable asset to most persons and one that should be protected as much as possible. It is an asset that is also most cherished when it is no longer in existence. First attempts have been made to regulate video surveillance use through legal frameworks and to better protect the privacy of people. However, only few technological solutions to this issue have been proposed, with none of them providing adequate privacy protection in real-world scenarios. In this study, a technological solution to the challenge of protecting the privacy of people under video surveillance was sought and found. The purpose of this study was to provide a contribution to this issue and to solve the main research question, which was formulated as follows:

> *"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"*

The research question was formulated after interviews with groups of people, who were influenced by video surveillance, for example at their work place, and a thorough study of the existing knowledge base. Both the interviews and the study of the knowledge base revealed that there was a strong need for a video surveillance system providing adequate privacy protection, which should not diminish overall security of a site and that no system fulfilling all requirements had been built before.

This research question represented the demand for a system that allowed protecting the privacy of innocent individuals while providing security. Security in this context was defined as physical security of an area under surveillance and it was assumed that a video surveillance system would support security personnel to provide physical security. Thus, security personnel should be enabled to use video surveillance as a tool to provide security while providing means to protect the privacy of persons who are not of interest. However, a privacy preserving system would not by itself decide whose privacy should be protected and whose identity should be revealed. This decision would be up to a human observer and to compliance regulation in an organization. Hence, such a system should be embedded in appropriate organizational processes to be effective.

1.2 Motivation

Privacy always becomes a more prominent issue when its existence is being threatened, recently due to new developments in information processing systems. Hence, it would be reasonable to think that at the same time as surveillance by the government increased, privacy concerns increased as well. Many aspects, which are associated with privacy today, such as liberty, freedom and property, were discussed in philosophy for centuries. However, the concept of privacy itself did not emerge until the end of the 19th century with the first recognized article about privacy and "the right to be left alone" by Warren and Brandeis (1890). At the time of its writing, several new inventions were making a discourse about this topic necessary. Photography was established already for some years and the telephone had already been invented twenty years prior. Moving pictures started to emerge. The yellow press tried to find out every detail about celebrities, started stalking them and reporting information about their private lives.

In the 19th century, English philosopher Jeremy Bentham invented the concept of the Panopticon (Foucault 1975). It was the basic concept of a structure for institutions such as hospitals, schools and, most importantly, prisons. In the center of the round structure a tower was placed that allowed the person in the tower to see what was happening around him. In a circle around the tower a set of rooms or cells were constructed. A person in the cell could only see straight outside and could not see the people in neighboring cells. Through the central placement of the tower, the person in the tower was able to see everybody but, by means of clever lighting, the people in the cells could not see the guard. This way, they never knew if they were being watched or not, thus creating a kind of psychological self-discipline. Bentham originally wanted to use the concepts to increase efficiency in factories: on the one hand personnel cost of guards would be reduced and on the other hand workers would be more disciplined.

Foucault (1975) uses the concept of the Panopticon as an example to describe his discipline society. In his view, the Panopticon is a metaphor for the current status of our society. It is the pure implementation of un-personalized power in a state. In contrast to Bentham he does not use a small number of people controlling the Panopticon. In our society it becomes multi-dimensional with every individual being controlled by different powers with different interests.

Orwell (1949), in his famous novel *1984*, paints a picture of a society that is completely controlled by a dictator called Big Brother that turns out to be a name for the ruling power. In the fictional state of Oceania the society is under perfect and total surveillance via the use of thousands of telescreens, microphones and other technology. Every person in the state is aware that he or she is under control, not least because of omnipresent posters that remind people “Big Brother is watching you”.

If we look at the situation today, did any of the visions of Bentham, Foucault or Orwell become true? Have we become, through new technology and possibilities, a transparent society? Especially since events such as 9/11 triggered a global increase of security and surveillance, visions like these do not sound so fictional anymore. Data of individuals is collected, stored and shared like never before, as some recent revelations suggest (Greenwald & MacAskill 2013). Supposedly to protect us, photos, fingerprints and credit card data are stored and we are completely scanned before we are allowed to board an airplane.

Without a doubt security has become more important in our contemporary world. Whether this view is subjective as a result of reports from the media or if our world is actually becoming less secure, is not important at this point. Since 9/11 security demands have risen and technology is developed to increase security and if the technological possibilities are there, they are used. The trend to more security in public spaces cannot be stopped. This is why applied ethics should not block and build up walls but rather try to influence the development to ethically acceptable applications. The main argument against video surveillance is not using cameras themselves, for example to view live images, but the loss of privacy if the image of persons is captured in a video stream, digitized and stored to be searched later on. However, this argument cannot be compared with traditional norms. There has been a significant change in the last twenty years towards freer information sharing. Humans and their values are adapting to new situations.

Pro and cons for video surveillance are evident: On the one hand it may help to increase our security; it is used as a deterrent for criminals and delivers evidence to investigate and solve crimes. On the other hand the privacy of citizens is constantly invaded and there is a fear that the gathered information can be abused. It is one of the basic characteristics of technology that it has advantages as well as disadvantages. Every technological advancement, be it the clock in the 13th century or the railway in the 19th century, has broadened our possibilities.

To understand the arguments against video surveillance one should look at the importance of privacy. What exactly changes in our behavior when we are filmed in public with our knowledge? What are the dangers of storing information about us? The danger to freedom and autonomy is there because the person under surveillance knows less than the person watching him or her. Her expectations of others, how they might act towards her, are based on incorrect or incomplete information. She does not know if she is watched or not, i.e. if someone is currently looking at the video surveillance feed or not. This is why privacy is important, because in order to be free and autonomous we need control over our self-portrayal. Of course, what needs to be protected and what is private is a parameter than can change and it varies between cultures. A person living in a big city might share information differently than someone from a village. Europeans might deal differently with intimacy than Asians. Anybody who lives in modern society must become comfortable with coexisting with many people in a small space and the subsequent relinquishment of some privacy rights. Independent of how much a society values privacy it is an asset that needs to be protected.

The relationship between information processing systems and privacy is an uneven one. Information processing continues to develop at a fast pace, driven by numerous interests by industry, producers, the stock market and, not the least, the consumers. Privacy however, has a much weaker lobby. Today, information processing develops at a much faster pace than measures to protect the privacy of people that is being compromised by rapidly advancing technology. Taylor (2003) revealed that already about 90% of all adults in the United States are concerned about their privacy and 26% feel they have lost a significant part already. Considering the level of privacy intrusion video surveillance brings and large-scale deployments of cameras today, it is about time to bring the protection of privacy to the same level. Finding technological solutions to the problem of privacy intrusion in video surveillance was the motivation for this work, namely providing privacy without compromising security.

First approaches to find technological solutions have been made before. The first generation of privacy protection systems masked entire image regions, such as desks of employees or entrances, which were defined by fixed image coordinates (Wickramasuriya *et al.* 2004). This proved to be very ineffective, especially as soon as people started moving out of masked areas (see Fig. 1). An advantage of this technique was its simplicity and its lack of significant performance requirements, which is why it is still used as a standard feature in many video surveillance cameras on the market today.

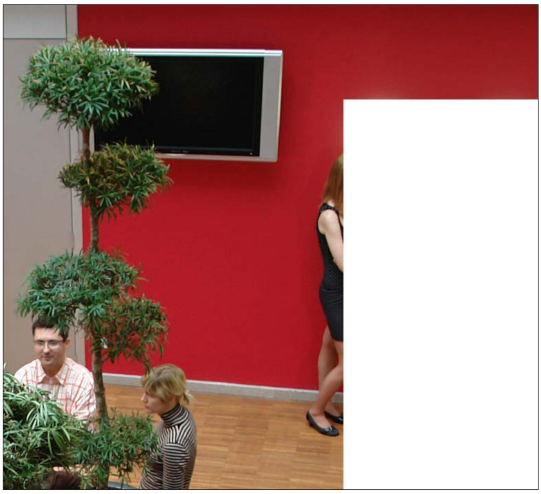


Fig. 1. First generation of privacy protection systems (simulated image).

The second generation of systems allowed masking of moving objects, thus achieving privacy protection in video surveillance more efficiently (Dufaux & Ebrahimi 2006, Chattopadhyay & Boult 2007). Fig. 2 illustrates this concept. As can be seen, the masked area was no longer defined by specific image coordinates but by pixels of moving objects in the image. Shadows cast by individuals were masked just like the individuals themselves.

However, this method had several disadvantages: First, all movement was masked, including background movement, shadows and highlights. This often resulted in the masking of most of the image, even though only a small portion was actual identity information. This rendered privacy protection algorithms unfeasible since only a small part of the image could be seen clearly. However, identity information, and only identity information, should be masked in the image in order to protect privacy efficiently. Robust background subtraction with accurate shadow detection would have to be the basis for any future privacy protection system.

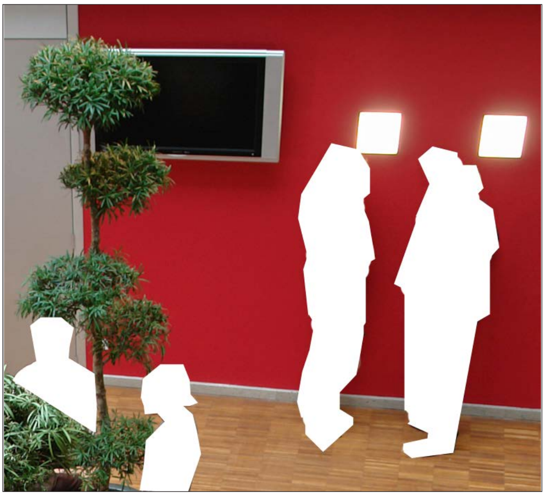


Fig. 2. Second generation of privacy protection systems (simulated image).

Second, it was not possible to distinguish between individuals. Thus, either everyone or no one in the image could be masked. However, in many applications it would be desirable to unscramble only specific people in the image. For example, if the original unmasked video material of a covert operation had to be used in court, covert agents would be seen and their identity revealed. In shops it would be beneficial to the privacy of employees if only those who committed theft were unscrambled and not everybody else in the image as well. In many situations it would be desirable that only those who triggered an alarm would be unscrambled. This showed that it was critical for the protection of privacy of people under surveillance to be able to unscramble only relevant individuals in the image and that new systems fulfilling this requirement were needed. This sentiment was further reflected in the research of the environment, as elaborated in Section 1.5.

The debate between privacy and security has been framed incorrectly as a zero-sum game in which we are forced to choose one over the other. But protecting privacy does not have to decrease security; it merely involves adequate oversight, regulation and appropriate technological solutions. Solove (2011) states that our current views of privacy and security are based on mistaken views about what it means to protect privacy and the costs and benefits of protection. We have to start a paradigm transformation with four main pillars:

- Privacy is essential to freedom
- Privacy *and* security are needed simultaneously and the gain of one is not the result of a trade-off of the other
- Privacy and security is a positive sum, not a zero-sum
- Apply privacy by design (PbD): PbD should be the new foundation principle, as stated by Cavoukian (2013). With the increasing complexity and interconnectedness of information technologies, nothing short of building privacy right into system design should suffice. Accordingly, PbD was developed in order to describe the principle of implementing privacy proactively into systems themselves. As Langheinrich (2001) points out, future systems that are built on the principle of PbD should focus on seven areas: notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, access and recourse.

In conclusion, three developments led to this work:

- Crime and terrorist attacks led to an increase in video surveillance cameras watching our daily life
- This invasion was, to a large extent, at the cost of personal privacy
- Privacy and security are still considered as a trade-off

Even though intensive research within video surveillance has been done, research in privacy protection in video surveillance is still at the beginning of its development and leaves many areas of improvement. If privacy is considered in today´s video surveillance systems, it is mostly viewed as a network security and encryption problem (Chang *et al.* 2012) and not as a problem of viewing private information itself.

1.3 Prior research

In recent years, computer vision methods, which aim to protect the privacy of people and objects in video surveillance (Cavallaro 2007) have emerged. The basic approach to achieve privacy protection in video surveillance is to extract or hide identity information present in an image, such as the identity of a person, before security personnel views it. Hence, the privacy of uninvolved individuals is protected.

Accordingly, the study of prior research was divided into three different areas of interest: Security and privacy in ICT in general, detection of regions of interest for privacy protection (e.g. people, faces, motion) and privacy preserving techniques in video surveillance. Prior research showed that privacy is a much debated issue in ICT systems today and plays an important role in critical applications (Buscher *et al.* 2013). Security and privacy are especially important in areas where sensitive data is handled, such as the healthcare sector (Appari & Johnson 2010, Alemán *et al.* 2013, Caine & Hanania 2013). With the emergence of cloud applications, security and privacy in the datacenter become an area of importance as well, as summarized by Hamouda (2012). Together with technology itself, it is important that guidelines and policies are employed as well (Johnson *et al.* 2010b).

Person tracking is an active research field with various sub-fields, which were relevant for this study. In order to track an object, background subtraction is first applied to detect foreground pixels (Bharti 2013). Next, image segmentation segments all foreground pixels into objects (Li *et al.* 2012, Caleiro *et al.* 2013). Finally, objects are associated in subsequent frames in order to allow tracking of an object through a scene (Kim *et al.* 2013). Further, face detection and recognition methods could be used to detect regions of interest (Badii & Einig 2012). Recognition of faces allows the recognition of certain individuals as well (Ahonen *et al.* 2004).

Privacy preserving methods in video surveillance are still an emerging field. However, a number of works have been published that deal with the issue. Cavallaro (2007) gives a clear overview over different privacy preservation methods. There are several types of approaches. Some authors suggest scrambling methods that are reversible in case of an incident (Martínez-Ballesté *et al.* 2013, Paruchuri *et al.* 2013), while others scramble permanently (Saini *et al.* 2012). Furthermore, storage of privacy information in video surveillance is a research area of interest. Baaziz *et al.* (2007) presents a method for efficient and secure storage of privacy information. There is also a trend to be able to scramble only individuals in an image, which requires some kind of identification such as face recognition or RFID tags (Wickramasuriya *et al.* 2004, Chen *et al.* 2007). Bringing privacy protection directly to the camera is another direction researchers are following. For example, Winkler and Rinner (2013) present a smart camera with on-board privacy protection.

The study of prior research revealed that while some research in privacy preserving video surveillance technology has been done, there remain some open issues. Few works focus on selectively unscrambling specific individuals in the image. Those that do take this issue into account base this selection on identification criteria such as face recognition and RFID tags. This has the disadvantage that protected people need be registered first, limiting the application of the method. Additionally, from a data protection point of view, it is questionable if it is feasible to protect the identity of a person by identifying this person first. Furthermore, the performance of methods such as person tracking or private data storage was limited and could not be used in a complete system where many other tasks need to be performed. There was also no guidance on how a privacy preserving video surveillance system could be designed from an architectural point of view. Finally, there was a lack of approaches to measure security and privacy in a given situation, which is important if a privacy preserving system is employed in real-world applications.

1.4 Research methods

For this study, the seven design science research guidelines, proposed by Hevner *et al.* (2004) were followed. They represent a checklist of seven items that need to be followed in order to create meaningful research results. Further, the design science research cycles by Hevner (2007) were applied in order to ensure that results of the research fulfill a need by the environment and build on existing knowledge. The research cycles represent loops in research, for example between requirements of the environment and the construction of the artifact or between the build and evaluate process. These loops are repeated until the optimal result is found. Research was performed in the following steps: first, corresponding to Hevner's (2007) relevance cycle, requirements of the environment were investigated. This cycle was repeated during the research to acquire feedback by the environment if defined requirements were met by the system. This information was gathered through a survey with three different groups in companies employing video surveillance: Employees without security responsibility, security chiefs and users of security systems (e.g. security guards). 37 people answered this survey, which provided the basis for formulating research questions. Second, the main research question was formulated with several sub-questions according to parts of the construct to be built. Third, prior research in the area of privacy protection in video surveillance was investigated in the existing knowledge base, corresponding to Hevner's (2007) rigor cycle. Relevant publications were found by searching in scientific databases, by following citations in relevant works and by existing knowledge of the field. This was a constant process followed during the study to not only rely on existing knowledge at the start but to benefit from new developments during the research. Fourth, the construct with all sub-constructs was built according to Hevner's (2007) design cycle. Each sub-construct was evaluated, resulting in a build-evaluate loop until a satisfying solution could be found. Finally, the whole construct was evaluated and discussed. Apart from empirical evaluation and case studies, the complete system was presented to the same groups, who provided the requirements of the environment in order to check if their requirements were met.

1.5 Requirements and environment

Before formulating research questions and constructing an artifact, requirements of the environment were gathered. These were gathered by surveying three different groups of people, which were influenced by video surveillance systems:

- Employees without security responsibilities
  (henceforth abbreviated to "employees"),
- users of security systems
  (e.g. security guards, henceforth abbreviated to "users") and
- chiefs of security (i.e. security responsibility on management level).

In total, 37 individuals of companies that employ video surveillance responded to the survey. The results revealed that the group of employees valued their personal privacy the most. They did also feel that security is important but there was a strong demand that their privacy would be better protected. Users and security chiefs on the other hand did not feel as strong about privacy protection as employees but stated that they believed they could provide the same level of security if they did not know the identity of a person during an incident. However, they stressed the importance of identifying a person after an incident on demand. These were the two main insights of these interviews: protection of privacy was very important to people but in order to provide security it should be possible to reveal the identity of a person in case of an incident.

1.6 Research questions

Based on user requirements, the following main research question, which was answered in this work, was formulated:

> *“How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?”*

While working with this main question, components of the system were identified and it was decided to add additional sub-questions to cover sub-components of the system. Many other questions could be researched, however these questions were the most important that required answers in order to construct a system that works in real-world scenarios and provides actual benefits to the customers:

1. “*What existing methods can be used to achieve selective privacy protection?*”

   Before constructing new methods, existing knowledge should be researched.

2. “*What kind of metrics exist to measure video surveillance security and privacy coverage in a* given *area?*”

   Privacy protection does not only require adequate technological solutions but appropriate implementation already during the planning phase when installing a new system. Metrics can ensure that an appropriate level of privacy is provided.

3. “*How much can the performance and accuracy of background subtraction and shadow detection methods be increased?*”

Background subtraction, together with shadow detection, are an integral part of the system to be developed to detect which parts of the image could identify a person.

4. "*How can the accuracy and performance* of *discrimination between objects be increased?*"

   After background subtraction, segmentation methods cluster foreground pixels into objects or people, which subsequently can be associated and tracked.

5. "*How can the accuracy of* segmentation *in mean shift based methods be improved?*"

   Mean shift based clustering and segmentation methods are known to be fast and applicable in real-time applications but not as reliable as other methods. Hence, basing segmentation on such methods and increasing their accuracy is a viable approach for segmenting people.

6. "*What storage methods for privacy* sensitive *video data should be used to store privacy information to ensure maximum privacy protection and security*?"

   Storage of private information is a sensitive and critical topic. Storage should be secure and efficient.

1.7 Research contributions

This study brought one main contribution to the existing knowledge base: the construction of a video surveillance system that allowed for selective protection of the privacy of individuals, namely the Privacy Enhancing Video Surveillance system or PEVS. Using PEVS, the video images of all people in a video stream could be scrambled to protect their privacy. In case of an incident, authorized personnel could choose to unscramble selected suspects in the image, without having to unscramble any uninvolved individuals. Hence, the privacy of all uninvolved individuals in the video was protected. Each part of the system was built with a focus on robustness, performance and speed. No overall video surveillance system, which could selectively protect the privacy of individuals, had been proposed before. Hence, this was a significant contribution to the knowledge base. Further, several contributions could be made in sub-constructs, such as shadow detection and segmentation algorithms. In whole, five main contributions could be attributed to this work: a system that provided privacy protection to selected individuals, the overall system with all its components, the PEVS system architecture, improvements to algorithms of sub-components and novel security and privacy metrics. Table 1 further summarizes the results of this dissertation as suggested by March and Storey (2008).

Table 1. Formalized results of this dissertation.

Item	Description
Problem	Video surveillance systems become ubiquitous but privacy of individuals is threatened. Current systems do not protect the privacy of people adequately.
IT Artifact (Solution)	Construction of artifacts resulting in an overall privacy enhancing video surveillance system (PEVS).
Evaluation Method	Measuring performance of each sub-component as well as the overall system. Using case studies to show usage in real-applications as well as surveys to verify that defined requirements were met.

1.8 Structure

This dissertation is structured as follows. Chapter 2 presents prior research in the area of privacy protection in video surveillance as well other connected areas. In Chapter 3 the methodology used during the research as well as research questions are discussed. Chapter 4 introduces the construct built and presents each part of the overall system while in Chapter 5 the empirical and case study based evaluation of the construct is presented. In Chapter 6 the results of the construct are discussed and an outlook for the future is provided. Finally, a conclusion in Chapter 7 concludes the dissertation and summarizes the study.

2 Prior research

Privacy issues in video surveillance is a much debated topic, with studies in various cities and countries, such as London (Kroener 2013), Canada (McPhail *et al.* 2013) and Chicago (Schwartz 2013). In this chapter, prior research from a technical point of view for privacy preserving video surveillance, with different aspects of the topic, is introduced. The content of prior research works is summarized as well as advantages and drawbacks given. This chapter is divided into four different parts. First, it is presented how security and privacy is dealt with in the ICT field in general. Second, an overview of methods to detect areas to be privacy protected is given. Third, privacy preserving techniques in video surveillance, both embedded and server-based, are investigated. Finally, a conclusion summarizes the current state-of-the-art in the areas presented.

2.1 Security and privacy in ICT

The relationship between security and privacy spans over the whole field of ICT (Stahl 2007, Noordin 2013). Most prominent examples nowadays include privacy issues in social networks, as investigated by Stutzman *et al.* (2013) and van der Velden and El Emam (2013). Further, with the advance of smart phones, security on mobile devices becomes critical (Portokalidis *et al.* 2010). A way to more secure mobile communication could be direct peer-to-peer communication, as implemented by Porras *et al.* (2004). Some even argue that ICT systems are inherently insecure and should thus not be used for critical applications at all (Ondrisek 2008). Stahl (2007) argues that privacy has been identified as one of the major ethical issues in ICT from the early days of the debate on computer and information ethics. Privacy always plays an important role in critical applications, such as emergency management systems, as pointed out by Buscher *et al.* (2013). Hence, before researching the topic in the specific area of video surveillance, security and privacy in ICT in general is investigated.

Security and privacy are not only a technical issue in ICT, but an issue of how ICT systems are implemented in an overall security management process (Kraus 2010) and how ICT projects are implemented in general (Kruger *et al.* 2006). Halonen and Paavilainen (2005) and Martikainen and Halonen (2011) show how problems in the implementation in ICT projects can be identified and potentially be avoided. Initiatives such as the common criteria for information technology security evaluation (CCMB 2012) support the quest for secure ICT systems.

Security and privacy in the healthcare sector is of growing importance and a critical area in ICT where privacy breaches have to be prevented (Appari & Johnson 2010). Appari and Johnson (2010) give an in-depth overview over the state-of-the-art of security and privacy in the healthcare sector. They show that considering the United States of America, virtually all parts of the healthcare sector will be supported by ICT in the near future. This could help the USA to save more than $81B annually by moving to electronic medical records. Thus, they argue, security and privacy will become a globally vital issue in this area. Alemán *et al.* (2013) give an in-depth overview over e-health publications. Furthermore, patients increasingly demand more control over their data and how it is used (Caine and Hanania 2013). The work of Miller and Tucker (2009) shows that the effect of government regulation of patient data for privacy protection hinders ICT adoption in the healthcare sector. They continue that, for example, by prohibiting networking between different hospitals, ICT adoption is reduced by 25%. Moreover, they show that prohibitions reduced compatibility of ICT systems implemented in different hospitals.

With the increased use of cloud services, privacy issues have growing relevance in datacenters as well. This is especially relevant since web providers store very personal information of users, which are requested from them when signing up (Heikkinen *et al.* 2004). Xiao and Xiao (2013) give an overview over security and privacy issues in the cloud and identify the most important privacy attributes of a cloud offering: confidentiality, integrity, availability, accountability and privacy-preservability. Hamouda (2012) gives a good overview over security and privacy issues in cloud computing as well. Ullah *et al.* (2013) point out challenges and best practices of cloud computing while Li *et al.* (2013) show how to ensure secure storage of health data in the cloud. Yu *et al.* (2013) show how privacy policies can automatically be enforced in a Platform as a Service.

Legal frameworks and guidelines are just as important as technology when it comes to security and privacy (Earp *et al.* 2002). Johnson *et al.* (2010b) investigate existing guidelines for issuing security and privacy policies and suggested new guidelines. Their evaluation is based on a template-based policy authoring method that was introduced in an earlier work (Johnson *et al.* 2010a). In this method, policies are created by a policy author who uses templates created by a template author, who again uses policy elements issued by the policy element author. In addition to the existing guidelines, Johnson *et al.* (2010b) suggest three new ones: support appropriate limitation of expressivity, i.e. limit the choices the user has when issuing a policy in a way that policies cannot contradict themselves; communicate risk and threats, i.e. communicate to the user the consequences of their actions; and provide access to metadata, i.e. providing the policy author with all information about the meaning of terms he might not be familiar with.

Besides policies, laws and regulations are an important tool for providing and enforcing security and privacy (Breaux & Anton 2008). However, it is often not clear which regulation should be used at which point of time. Compagna *et al.* (2009) investigate methodologies on when to use which method for providing security and privacy in the health care sector using a real-world example. Herrmann (2007) gives a guide how to measure security and privacy and resilience with defined metrics.

Reeder *et al.* (2010) study how users use ICT systems, in this case SPARCLE, that use natural language to define policies and tried to identify problems. They find that most problems arise due to "group ambiguity", i.e. that composite terms are used to describe a group. However, it is not clear what this group describes. The second biggest problem is terminology mismatch, i.e. that multiple terms are used to describe the same thing. Similarly, five different problematic areas are identified by Reeder *et al.* (2010): group ambiguity, terminology mismatch, negative rule, missing element and rule conflict. Subsequently, they propose possible solutions for each of these challenges.

Considering this background, privacy preserving technology for video surveillance can be built while considering a secure and robust architecture around it (Kraus *et al.* 2009). Table 2 summarizes publications on security and privacy presented in this section.

Table 2. Summary of publications on security and privacy.

Description	Reference
Privacy and ethics in Islamic ICT	Noordin (2013)
How privacy and information disclosure evolved in Facebook	Stutzman *et al.* (2013)
A study on how teenage patients deal with privacy	van der Velden & El Emam (2013)
How to handle privacy data in emergency management systems	Buscher *et al.* (2013)
A study on security and privacy in e-health	Alemán *et al.* (2013)
Demand by patients for better control over their private data	Caine and Hanania (2013)
Review of security and privacy issues in cloud computing	Xiao and Xiao (2013)
Security, privacy and portability challenges in cloud computing	Ullah *et al.* (2013)
Secure sharing of private e-health data in cloud computing	Li *et al.* (2013)
Enforcing privacy policies with a Platform as a Service	Yu *et al.* (2013)
How to provide security and privacy in cloud computing	Hamouda (2012)
A tool to analyze the output and benefits of ICT projects	Martikainen and Halonen (2011)
Overview over security and privacy in the healthcare sector	Appari and Johnson (2010)
A security management process for large scale systems	Kraus (2010)
Policy authoring for security and privacy policies	Johnson *et al.* (2010b)
Policy templates for policy authoring	Johnson *et al.* (2010a)
Usability challenges in security and privacy policy-authoring interfaces	Reeder *et al.* (2010)
Quantifying the effect of privacy regulation in electronic medical records	Miller and Tucker (2009)
Methodological support for assessing security and privacy requirements	Compagna *et al.* (2009)
A dynamic architecture for large scale video surveillance systems	Kraus *et al.* (2009)
Security in electronic voting systems	Ondrisek (2008)
Methodology to derive access rights and obligations from legal regulations	Breaux and Anton (2008)
Security and privacy metrics with focus on regulatory compliance	Herrmann (2007)
A framework to evaluate ICT security awareness	Kruger *et al.* (2006)
A hierarchical model to model information system failures	Halonen and Paavilainen (2005)
A framework to examine privacy management practices	Earp *et al.* (2002)

2.2 Detecting areas to protect

This section presents methods to detect a region of interest in an image that subsequently can be scrambled to protect the privacy of people. Two main categories are investigated: person tracking and background subtraction and face detection and recognition.

2.2.1 Person tracking and background subtraction

Robustly tracking people in video surveillance is an on-going research topic with challenges especially in crowded scenes (Shu *et al.* 2012, Shitrit *et al.* 2013). Tracking usually involves first detecting movement in the image, called background subtraction (Huang *et al.* 2013, Noh & Jeon 2013) to decide which part of this movement is of interest. Bharti (2013) provides a satisfactory overview of current developments. In this step often highlights and shadows are detected as well (Nadimi and Bhanu 2004), which are then filtered out of the image in order to reduce unwanted detection. Shadows can be classified into different categories, depending on how they were generated (Prati *et al.* 2001). First, a distinction can be made depending on movement. Shadows, which are cast by static objects, such as buildings or parking cars, are called static shadows, while shadows, which are generated by moving objects, are called dynamic or moving shadows (Prati *et al.* 2001). Second, shadows can be distinguished by the surface they appear on. The part of an object that is not illuminated by light sources is called self-shadow. As elaborated by Prati *et al.* (2001), cast shadows are shadows that are cast on a surface by an object that blocks light from the light source. In this work moving cast shadows are relevant since they present a problem to algorithms relying on motion information.

Next, the image is segmented into regions that belong together. Several general-purpose segmentation algorithms have been developed in recent years. One technique is based on density or color histograms where peaks and valleys in the histogram distribution are used to locate clusters (Caleiro *et al.* 2013). Apart from this technique, other algorithms have been proposed including level-set (Li *et al.* 2012), graph partitioning (Shi & Malik 2000) and watershed (Beucher 1992, Zhang *et al.* 2012).

After segmentation, association between segmented objects in subsequent frames is performed (Siebel 2003, Kim *et al.* 2013). Tracking methods can be distinguished whether they are tracking objects within a single video stream or whether they are using multiple cameras for tracking. Zervos (2013), for example, uses face recognition while Bouma *et al.* (2013) use appearance for multi-camera tracking. There are also the following different approaches to tracking.

Blob-based tracking

A blob is defined and identified as a region of interest. Blob-based tracking generates blobs by subtracting the image of the background from a scene and then threshold the resulting difference image (Rosales & Sclaroff 1998). There are two main classes of blob detectors: differential methods based on derivative expressions and methods based on local extremes in the intensity landscape. These operators are also referred to as interest point operators or interest region operators. This approach is very reliable but can only be used with static cameras where the background can be subtracted.

Kernel-based tracking

Kernel based tracking methods (Comaniciu *et al.* 2003), such as mean shift (Fukunaga & Hostetler 1975, Beleznai *et al.* 2004, Beyan & Temizel 2012), are often used because of their computational simplicity and speed. This technique is an iterative localization procedure based on the maximization of similarity measure (Beleznai *et al.* 2004). Mean shift clustering is applied to background-differenced image sequences. This is also an efficient and reliable technique that could be utilized for further applications.

Template-based tracking

A detection can further be achieved by matching templates (Jurie & Dhome 2001) or using texture patches (Leung & Malik 2001). This is a straightforward algorithm where a template of a given object is created and the matching region closest to the template has to be found in each frame. If a template does not match an object perfectly this approach can be unreliable.

Color-based tracking

Color-based tracking algorithms use color information of an object to find it in different frames (Bradski 1998). The drawback of this approach is that colors can change considerably in different images and especially on different cameras. To counteract this, color normalization has to be performed.

Feature-based tracking

Feature-based methods can deliver accurate results but are computationally complex (Zhou *et al.* 2009). Feature-based methods describe each object using a set of features, such as SIFT (Gao *et al.* 2012), local binary patterns (Heikkilä *et al.* 2009, Pietikäinen *et al.* 2011), BRIEF (Demiröz *et al.* 2012) or HoGs (Dalai *et al.* 2005, Descamps *et al.* 2012), which are sought in consecutive frames. This way an association between objects in different frames is achieved. The most popular feature based tracker is the Kanade–Lucas–Tomasi feature tracker, which tries to find features that could be more efficiently used for tracking (Shi & Tomasi 1994).

Model-based tracking

Model-based tracking algorithms are based on finding a parameterized model that approximates the shape of the object that is to be tracked (Siebel 2003).

Summary of tracking methods

For each situation the right kind of tracker has to be selected very carefully since nowadays no general tracker algorithm is available. For example, if the type of object to track is known in advance, such as people or faces, this information can also be used for certain types of trackers. Model-based trackers learn a model of the object from a big set of images of the type of the object offline. In situations where a static camera is present, trackers, which use background subtraction, can be used. There, a statistic model of the background is learned and subtracted from the current frame. Thus, foreground objects can be extracted and concurrently tracked. Tracking accuracy and performance is usually measured by comparing tracking results (object location & ID) to a set of ground truth data (Kavasidis *et al.* 2012). Tracking can further be used for higher-level reasoning to understand intention and behavior of individuals (Izadinia *et al.* 2013).

In order to achieve better accuracy, tracking methods can be tailored to applications in crowded scenes (Kratz & Nishino 2012), tracking of pedestrians (Cai & Pietikäinen 2011), airborne tracking (Yilmaz *et al.* 2003) or for mobile applications (Räty 2008, Sutor *et al.* 2008b). Mazzon and Cavallaro (2013) use pre-existing knowledge about the scene and provide a map of the surroundings to the algorithm. Building on the assumption that humans tend to walk towards entrances and exits they can improve tracking results in multi-camera tracking applications. Tracking results can further be improved by using multiple information sources, additional to video. For example, Yang (2013) proposes a system that uses laser scans in combination with visual tracking methods.

2.2.2 Face detection and recognition

The second approach to detect regions of interest for privacy masking, which is most prominent in literature, is to detect or recognize faces in videos (Li & Jain 2011). Since people are identified the easiest through their face, masking this region is the most straightforward approach for privacy protection (Badii & Einig 2012). Bradski (1998) uses face detection for tracking of the face. Ahonen *et al.* (2004) propose a face recognition method using local binary patterns, while Bowyer (2004) specifically use it for privacy applications. Badii and Einig (2012) argue that if using face detection as criteria for privacy protection, the right balance between intelligibility and privacy protection needs to be chosen. They propose a method using pixelization and edge detection. Face detection and recognition can never be 100% accurate. Thus, it is even more important to consider protection against face spoofing (Määttä *et al.* 2011) to decide if an area should be privacy protected or not. Table 3 summarizes publications on person tracking presented in this section.

Table 3. Summary of publications on person tracking.

Description	Reference
Tracking using appearance cues over long sequences	Shitrit *et al.* (2013)
Background subtraction using spatio-temporal analysis	Huang *et al.* (2013)
Background subtraction using multiple cues	Noh and Jeon (2013)
Review of background subtraction techniques	Bharti (2013)
Review of color spaces and segmentation methods for object detection	Caleiro *et al.* (2013)
Predicting object tracks using velocity-space reasoning	Kim *et al.* (2013)
Face recognition for multi-camera tracking	Zervos (2013)
Multi-camera tracking in a shopping mall	Bouma *et al.* (2013)
Level set method for semi-automatic segmentation	Li *et al.* (2012)
Segmentation using independent component analysis and watershed	Zhang *et al.* (2012)
Evaluating multi-target tracking for activity recognition	Izadinia *et al.* (2013)
Combining laser detection and video analysis for multi-target tracking	Yang (2013)
Multi-camera tracking exploiting camera site knowledge	Mazzon and Cavallaro (2013)
Using face detection for achieving intelligibility in privacy protection	Badii and Einig (2012)

Description	Reference
Handling occlusions by using part-based tracking-by-detection	Shu *et al.* (2012)
Adaptive mean-shift for multi object tracking	Beyan and Temizel (2012)
Feature-based tracking on omnidirectional cameras	Demiröz *et al.* (2012)
Person detection for indoor scenarios	Descamps *et al.* 2012)
Person tracking using spatio-temporal motion patterns	Kratz & Nishino (2012)
Object tracking using particle filtering	Gao *et al.* (2012)
A tool for generating ground truth data for detection and tracking	Kavasidis *et al.* (2012)
Handbook for face recognition	Li and Jain (2011)
Local binary patterns for still images	Pietikäinen *et al.* (2011)
Person re-identification using self-similarity	Cai & Pietikäinen (2011)
Scale invariant feature transform (SIFT) for object tracking	Zhou *et al.* (2009)
Using center-symmetric local binary patterns for ROI description	Heikkilä *et al.* (2009), Pietikäinen *et al.* (2011)
A video surveillance camera, independent of power and network connections	Sutor *et al.* (2008b)
A prototype of a mobile, intelligent video surveillance system	Räty (2008)
Using histograms of oriented gradients for person detection	Dalai *et al.* (2005)
Shadow detection using color and intensity	Nadimi and Bhanu (2004)
Using mean shift for person detection in groups	Beleznai *et al.* (2004)
Using local binary patterns for face recognition	Ahonen *et al.* (2004)
Investigating privacy concerns in face recognition	Bowyer (2004)
Object tracking in airborne thermal cameras	Yilmaz *et al.* (2003)
Implementing a complete person tracking system	Siebel (2003)
Kernel-based object tracking	Comaniciu *et al.* (2003)
Evaluation of shadow detection methods	Prati *et al.* (2001)
A framework for object tracking in videos	Jurie and Dhome (2001)
A model to recognize the visual appearance of materials	Leung and Malik (2001)
Modeling background with an adaptive mixture of Gaussians	Stauffer and Grimson (1999)
Combining low-level & mid-level information for person tracking	Rosales and Sclaroff (1998)
Face tracking using mean shift	Bradski (1998)
A method good feature selection for tracking	Shi and Tomasi (1994)
Using watershed for image segmentation	Beucher (1992)
Mean shift for pattern recognition	Fukunaga & Hostetler (1975)

2.3 Privacy preserving video surveillance

In this section prior research with a focus on privacy protection in video surveillance is analyzed. Two types of systems are differentiated: server-based systems, which are built to run on a central server component, and embedded systems, which are built to be run directly on smart cameras. Cameras that feature enough processing power to perform tasks in addition to image capturing are typically called smart cameras (Beer *et al.* 2009). The distinction between server-based and embedded is made since performance on smart cameras is limited and hence it is harder to achieve satisfactory results on such embedded systems.

2.3.1 Server-based privacy preserving video surveillance

Systems, which are built to be run on a central server component are described as server-based (Rinner & Wolf 2008). In this section, approaches to server-based privacy protection systems in video surveillance are investigated. Cavallaro (2004) separates the data stream from the surveillance cameras into two classes:

- Personal data (for example faces and vehicle license plates)
- Behavioral data (movement information)

According to Cavallaro (2004), the notion of different classes of data streams is based on the fact that for identifying an individual and for understanding the scene, different information is required. This data can be handled differently, depending on authorization. Personal data is automatically masked. However, only motion information is used. Background noise is accounted for but there is no selective privacy protection. Hence, different approaches exist on which regions of the image should be scrambled and how. Further, Cavallaro (2007) gives a clear overview over privacy protection systems in video surveillance and how they work. Using a privacy preserving video surveillance system the police force can maintain security levels while protecting the privacy of those being filmed.

Paruchuri *et al.* (2013) and Martínez-Ballesté *et al.* (2013) propose scrambling methods that can be reversed if required. This saves storage space by storing only one stream rather than two separate streams for unscrambled and scrambled data. Korshunov and Ebrahimi (2013) state that privacy protection can be provided by warping all faces in a video stream. However, with this approach, people can still be identified by humans from other visual traits. Peng *et al.* (2013) propose to use the H.264 standard to scramble specific regions of interest in an image.

Sohn *et al.* (2011) propose a system that scrambles privacy sensitive regions (faces) in JPEG XR compressed video streams. Using JPEG XR has several advantages, one of them being the high quality of the images. By using a secret key in the encoding phase, the original image regions can be decoded if the secret key is known. The choice on how image parts to be scrambled are chosen also impacts the quality of the detection. Saini *et al.* (2012) face the issue that most detection methods to detect regions to be privacy protected are unreliable and propose a method that provides satisfactory results even if detectors are inaccurate. Zhang *et al.* (2010) on the other hand combine privacy protection with person tracking and face detection. This way, the faces of people can selectively be scrambled or unscrambled.

Kitahara *et al.* (2004) present a system to pixelate identity information in mobile cameras by detecting faces. They use multiple cameras to understand a 3D position and to detect faces. Unfortunately, this approach is currently not feasible in real-world scenarios where data from multiple cameras is seldomly available. Upmanyu *et al.* (2009) propose a privacy preserving video analysis system that does not merely preserve privacy merely for the end user but for the internal system. They continue that the input image is split in several parts and sent to different servers for computation. No server knows the complete input image but only performs a task on its dedicated part. Finally, Upmanyu *et al.* (2009) only merge the results, ensuring that at no point one component gains knowledge about the complete input image.

Baaziz *et al.* (2007) on the other hand face the issue of storage of privacy information. They present a privacy protecting video surveillance system, which provides efficient storage of data, providing means for privacy protection and secure methods to protect the video data from malicious attacks. It uses watermarking to identify original images and to detect manipulation. However, it can only mask movement in the scene and thus does not provide selective privacy protection. Zhang *et al.* (2005) go a step further by storing privacy-related data as a watermark directly in the video stream, which can only be retrieved using a secret key. Similarly, Choi *et al.* (2011) use data from an encrypted H.264 stream to recover private data in case of authorized access. Boult (2005) shows the possibility of using invertible cryptographic methods to scramble private data in a video. Only with access to a decryption key can the original data be decrypted.

Further, Boult and Woodworth (2008) take the issue of privacy protection in biometrics in general into account and propose using Biotopes, which are revocable tokens based on biometric features that preserve the privacy of users. Brassil (2005) aims at giving people under surveillance more power over what is stored about them. He proposes a system that allows users to decide if videos of themselves can be distributed using their mobile communication devices. Similarly, Barhm *et al.* (2011) present a system where users can set their individual level of desired privacy. The more knowledge about the environment where a system is used is known, the better people can be identified. Chen *et al.* (2007) present a system for hospitals that could mask individuals and unmask others, based on a training set. The system is trained with images from the hospital staff. The system then only shows the outlines of patients and shows staff un-obscured. Wickramasuriya *et al.* (2004) propose a similar system based on standard sensor technology such as RFIDs.

Another noteworthy approach to privacy protection is to edit identity information in a way that humans can still identify the individuals but automatic algorithms cannot (2005). This technique is called de-identifying. Newton *et al.* (2005) present possibilities to de-identify images of faces in such ways that face recognition software cannot identify the individuals anymore but they still remain recognizable by humans. Dufaux (2010) provides a framework to measure the effectiveness of privacy preserving methods in video surveillance and suggests improvements. Korshunov *et al.* (2012) provide a further evaluation on the effectiveness of privacy preserving methods. Table 4 summarizes publications on server-based privacy preserving video surveillance, which were discussed in this section.

Table 4. Summary of publications on server-based privacy preserving systems.

Description	Reference
Storing private information in a modified video stream	Paruchuri *et al.* (2013)
Using a protection stream to protect & unprotect a video stream	Martínez-Ballesté *et al.* (2013)
Obfuscating faces by warping	Korshunov and Ebrahimi (2013)
Encrypting private information in H.264 streams by chaos and selective encoding	Peng *et al.* (2013)
Evaluating the effectiveness of different scrambling methods	Korshunov *et al.* (2012)
Approach for a more secure encryption of H.264 video	Choi *et al.* (2011)
A method for setting individual privacy settings in video surveillance	Barhm *et al.* (2011)
A framework for evaluating scrambling methods	Dufaux (2010)
Management of and deployment of smart camera networks	Beer *et al.* (2009)
Protecting privacy by splitting video frames into random sub-frames	Upmanyu *et al.* (2009)
An introduction to distributed smart cameras	Rinner and Wolf (2008)
Introducing biotopes for protecting privacy in biometrics	Boult and Woodworth (2008)
Storage and privacy preserving techniques in video surveillance	Baaziz *et al.* (2007)
Using learned data to scramble persons in videos	Chen *et al.* (2007)
Privacy protection by hiding private by watermarking and encryption	Yabuta *et al.* (2005)
Hiding privacy information in a video stream as watermark	Zhang *et al.* (2005)
Cryptographic invertible scrambling of movement in video	Boult (2005)

Description	Reference
Setting privacy information for video surveillance on mobile phones	Brassil (2005)
De-identifying faces to hinder automatic face recognition	Newton *et al.* (2005)
Protecting privacy in mobile cameras	Kitahara *et al.* (2004)
Using RFID to identify persons to scramble or hide in media spaces	Wickramasuriya *et al.* (2004)

2.3.2 Embedded privacy preserving video surveillance

From a data protection point of view it is most desirable to protect the privacy of individuals as close to the image capture source as possible, hence directly on the camera (Rinner & Wolf 2008, Winkler & Rinner 2013). This section focuses on methods, which are performed directly on a smart camera.

Winkler and Rinner (2010a) present an extensive overview over privacy oriented video surveillance systems, arguing that these should fulfill the fundamental requirements Integrity, Authenticity, Freshness and Timestamping, Confidentiality, Access Authorization and Availability. They measure each proposed method in the literature against these requirements. The work also mentions user involvement as one of the most important issues in privacy oriented video surveillance. The user should know what kind of privacy protection is present and, ideally, be able to control which information is shown. Most of the systems that have this feature, however, rely on separate handheld devices.

Further, Winkler and Rinner (2010b) and Winkler and Rinner (2012) present a smart camera with privacy protection performed already on-board. In contrast to other systems, this system contributes research from trusted computing. It does not only focus on masking people in the video stream, but provides the following attributes as well: Integrity (by signing each frame), Authenticity (signing of frames), Confidentiality (by masking people) and Access Control (by providing different privacy levels with different access control options). It sends all the information of the video stream over the network but encrypts the private parts of the image separately. Only the public, non-sensitive video is accessible by everybody. In order to prevent copying or spoofing code on the camera, a multi-step trusted boot procedure checks in various steps if the integrity of the system has been compromised. Next steps in their research are bringing trusted components even closer to the sensor by either hard-wiring security processes to the sensor itself or by providing a dedicated chip. This way, a part of the camera can still be available to developers on an open Linux-based platform while still providing security and privacy. Senior *et al.* (2003) provide a system that offers different levels of privacy already on a smart camera, such as silhouette, movement, ID and appearance.

On the other hand, Chattopadhyay and Boult (2007) present privacy protection embedded in a DSP processor of a smart camera. In contrast to others, they focus mainly on the hardware aspects, arguing that DSP processors have all ranges of speeds and sizes and are optimized to leave enough CPU speed for the OS and processing. They detect ROIs depending on faces, skin or motion and then mask these areas using JPEG compression. It is further possible to consider privacy aspects during the encoding stage of a video stream. Yabuta *et al.* (2005) present a method that hides privacy information inside a JPEG stream itself. They propose that during JPEG compression privacy data (pixels form moving objects) is encrypted and watermarked separately. Yabuta *et al.* (2005) argue that this way, only one stream has to be stored to provide privacy protection. Usually, a private and a masked stream are stored. Using the right key, the private parts of the video stream can be restored. This system requires fixed cameras. However, it is not stated by Yabuta *et al.* (2005) if this is implemented directly on the camera or not. Fleck and Straßer (2010) present a distributed system that automatically detects pre-defined critical events on the smart camera in a geo-referenced world-model and protects the privacy of people by masking them directly in the smart camera. If a person falls, an image symbolizing this incident is shown. This application is developed to be used in assisted living scenarios to detect if an individual is falling or not while preserving their privacy. Table 5 summarizes publications on embedded privacy preserving video surveillance systems, which were discussed in this section.

Table 5. Summary of publications on embedded privacy preserving systems.

Description	Reference
Integrating privacy and security in the video sensor	Winkler and Rinner (2013)
Interacting with video surveillance cameras using mobile tools	Winkler and Rinner (2012)
A privacy preserving smart camera using trusted computing concepts	Winkler and Rinner (2010a)
Integrity, protection, authentication and confidentiality on a smart camera	Winkler and Rinner (2010b)
Distributed automatic privacy protection on smart cameras for assisted living	Fleck and Straßer (2010)
Implementing invertible cryptographic obscuration on a smart camera	Chattopadhyay and Boult (2007)
A smart camera with different privacy abstraction options	Senior *et al.* (2003)

2.4 Concluding remarks

The presented prior research in privacy preserving video surveillance methods described in Section 2.3 is analyzed based on the criteria defined in Table 6.

Table 6. Criteria used to evaluate privacy preserving methods.

Criteria	Description
Complete system	Practical applicability in real-world scenarios can only be provided if proposed methods are embedded into complete video surveillance systems. This poses special challenges due to system performance and interconnections between system components.
Real-time application	Performance optimizations are necessary to provide a system, which is applicable in real-world scenarios where real-time performance is a requirement.
Load balancing	Computational complexity highly depends on image content. By balancing the load between different image tasks, which in most cases feature different amounts of complexity, performance can be increased significantly.
Storage optimization	The amount of data that has to be stored on the smart camera should be minimized while still preserving the original data.
Shadow detection	In order to increase the accuracy of scrambling methods shadow detection has to be an integral part of a privacy preserving video surveillance system.
Selective privacy protection	The ability to scramble/unscramble specific individuals will be a must-have feature of privacy preserving VS in the future.

Table 7 shows results of the analysis of prior research based on these criteria. Only publications that propose new privacy preserving methods are listed, no evaluation studies or new frameworks. A “No” in a column states that this publication does not take the aspect in question into account at all.

Table 7. Overview of the study of publications against privacy protection criteria.

Publications / Criteria	Complete	Real-time	Load Balance	Storage Optim.	Shadow Detection	Selective Privacy
Paruchuri *et al.* (2013)	Yes	-	No	No	No	No
Martínez-Ballesté *et al.* (2013)	No	Yes	No	No	No	No
Korshunov and Ebrahimi (2013)	No	Yes	No	No	No	No
Peng *et al.* (2013)	No	Yes	No	No	No	No
Winkler and Rinner (2013)	No	Yes	Yes	No	No	No
Winkler and Rinner (2012)	Yes	Yes	No	No	No	No
Barhm *et al.* (2011)	Yes	No	No	No	No	No
Winkler and Rinner (2010b)	Yes	Yes	No	No	No	No
Fleck & Straßer (2010)	No	-	-	No	-	No
Upmanyu *et al.* (2009)	No	Yes	No	-	No	-
Boult and Woodworth (2008)	No	-	No	No	-	No
Baaziz *et al.* (2007)	No	No	No	Yes	No	No
Chen *et al.* (2007)	Yes	Yes	-	No	No	Yes[1]
Chattopadhyay & Boult (2007)	Yes	Yes	No	No	-	No
Yabuta *et al.* (2005)	No	No	No	No	No	No
Zhang *et al.* (2005)	No	No	No	Yes	No	No
Boult (2005)	No	Yes[2]	No	No	No	No
Brassil (2005)	No	No	No	Yes	No	No
Kitahara *et al.* (2004)	Yes	Yes	No	No	No	No
Wickramasuriya *et al.* (2004)	No	Yes	No	No	No	Yes[3]
Senior *et al.* (2003)	No	Yes	No	No	No	No

[1] Yes, but only in a limited setting. The algorithm has to be trained with a limited set of people (hospital personnel).

[2] Yes, the use of Biotopes increases performance.

[3] Yes, but persons have to wear RFID or other tags.

The study of prior research revealed that while some research in privacy preserving video surveillance technology has been done, there were some open issues:

- Selective privacy protection: few works focus on selectively unscrambling specific individuals in the image. Those that do take this issue into account base this selection on identification criteria such as face recognition and RFID tags. This has the disadvantage that protected individuals need to be enrolled first, limiting the application of the method. Furthermore, from a data protection point of view, it is questionable if it is feasible to protect the identity of a person by identifying this person first.

- Algorithm performance: the performance of methods such as person tracking or private data storage was limited and could not be used in a complete system where many other tasks need to be performed.
- Architecture: there was further no guidance on how a privacy preserving video surveillance system could be designed from an architectural point of view.
- Metrics: there was a lack of approaches to measure security and privacy in a given situation, which is important if a privacy preserving system is employed in real-world applications.

3 Research methodology

Before starting research one has to define a clear research methodology in order to empirically answer defined research questions. This work was based on one main research question and a number of sub-questions, which are outlined in this chapter. Further, a research methodology based on design science research is presented. Many works have been published on research methodology in information science (IS), investigating advantages and disadvantages of each method (Walls *et al.* 1992, March & Smith 1995, Markus *et al.* 2002, Hevner *et al.* 2004, Järvinen 2004a, Järvinen 2004b, March & Storey 2008). Research for this work was done using design science research principles, using the seven design science research guidelines by Hevner *et al.* (2004) as well as the design science research cycles by Hevner (2007).

3.1 Design Science

Video surveillance systems in general and intelligent video surveillance systems in particular are part of the category of information systems. Hence, when deciding on the methodology to use in this thesis, one common to information science research should be chosen. Hevner *et al.* (2004) outline that research in information systems science is characterized by two paradigms: behavioral science and design science. Behavioral science seeks to develop and verify theories that explain human or organizational behavior. It is based on natural science research methods and works in a similar way by postulating theories (principles, laws) and trying to verify them. By understanding human or organizational behavior better and gathering information on the interactions between people, technology and organizations, the way information systems are used can be improved and the efficiency of an organization increased. Design science on the other hand has an engineering approach, which tries to solve problems by building new information system artifacts (Markus *et al.* 2002). Aken (2004) argues that in management research design science should be employed, instead of explanatory sciences such as physics.

Design science has its roots in the studies of the artificial (Simon 1996). Simon (1996) defines artificial as “produced by art rather than the natural” and as the product of engineering and design. Hence, design science is a problem solving process, with the principle that knowledge and solutions to problems can be found by building an artifact. It aims at creating new ways to increase the efficiency and effectiveness of analysis, design, implementation, management and use of information systems (Denning 1996). Benbasat and Zmud (1999) argue that empirical IS research should be "implementable,...synthesize an existing body of research,…[or] stimulate critical thinking". Designing advances in this area is difficult however, since established theory is often insufficient with information technology being applied in areas, where it was previously unthinkable (Markus *et al.* 2002). When working with design science one has to be aware of the dichotomy of design (Hevner *et al.* 2004). Design is both the process of creating something new as well as a product (artifact) resulting from this process. Design is both a verb and a noun. The process of designing an artifact is a series of expert activities. The created artifact is then evaluated, which provides feedback for another round in this build-and-evaluate loop.

In order to support researchers in this loop and yield the highest quality results, Hevner *et al.* (2004) define seven design science research guidelines that were followed in this work in order to achieve meaningful research results.

These cover building a meaningful artifact (Guideline 1) that solves a concrete problem (Guideline 2). Thorough evaluation of the artifact is necessary in order to achieve a purposeful result (Guideline 3) and a relevant research contribution (Guideline 4). Research methodology and rigor are similarly important (Guideline 5) and require the search for an effective artifact to the problem (Guideline 6). Finally, after the design process is complete, it is important to properly communicate the results, both to technical as well as management audiences (Guideline 7).

Table 8. Seven design science research guidelines (Hevner *et al.* 2004).

Guideline	Description
1: Design as an artifact	The result of design science research has to be an artifact in the form of a construct, a model or an instantiation.
2: Problem relevance	The artifact has to be the solution to a concrete and relevant business problem.
3: Design evaluation	The quality, utility and efficacy of an artifact must be proven and demonstrated with rigorous evaluation.
4: Research contributions	The contribution to the knowledge base has to be clear and verifiable.
5: Research rigor	Design science relies on rigorously applied research methodology.
6: Design as a search process	All available means have to be used to search for the solution of a given problem.
7: Communication of research	Research results have to be communicated both to technical as well as management audiences.

Sein *et al.* (2011) argue that common design science research approaches focus solely on the creation of a technological artifact but do not pay enough attention to evaluation in the context the artifact should be used in and the business context the research problem arose from. They propose to apply action research to design science, thus defining Action Design Research (ADR) to overcome this gap. Action research aims to link theory with practice by combining theory generation with researcher intervention. ADR focuses on building, intervention and evaluation of an artifact and thus takes the theoretical creation of an artifact into account but also its influence on users and its ongoing use in context.

Sein *et al.* (2011) define four different stages with several principles for ADR (see Fig. 3): Problem formulation, building, intervention and evaluation, reflection and learning and formalization of learning. Stage 1, problem formulation, represents the process of defining a research problem. The problem might be articulated by different sources, e.g. practitioners, end-users or researchers. It consists of two principles. First, Practice-Inspired research views field problems as knowledge creation opportunities, as opposed to theoretical puzzles. Second, Theory-Ingrained Artifact stipulates that any artifact created and evaluated is informed by theories. Stage 2 consists of building and evaluating the artifact. This stage includes intervention with the environment the problem formulation came from. It incorporates three principles. First, Reciprocal Shaping stipulates that in a recursive process, IT artifact and organizational context are influencing and changing each other. Second, Mutually Influential Roles suggests that project members are constantly learning from each other. Third, Authentic and Concurrent Evaluation states that, as opposed to other theories, ADR views evaluation not as a separate stage, which is performed after an artifact is built but as an inherent process, which should be performed during the creation of an artifact. Stage 3 is a continuous stage, parallel to the other two stages. In this stage, everything should constantly be reflected upon and knowledge gained should be used to influence the artifact building process. It consists of one principle: Guided Emergence. This principle suggests that the artifact created does not only reflect the preliminary design but also "its ongoing shaping by organizational use, perspectives, and participants" (Sein *et al.* 2011). In the research presented in this dissertation Design Science Research cycles, as presented in Section 3.2 with a practical approach as suggested by ADR was applied.

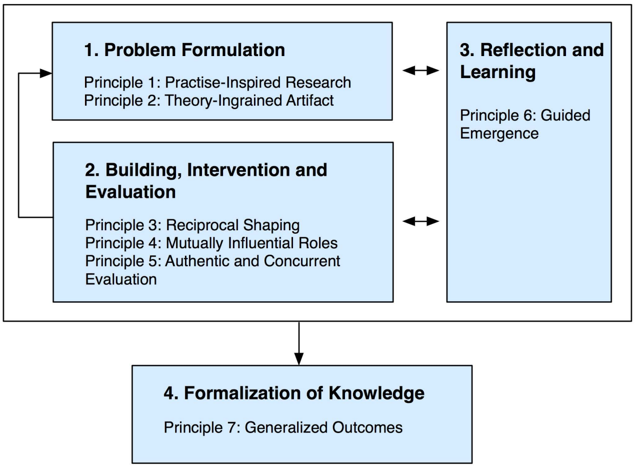

1. Problem Formulation
Principle 1: Practise-Inspired Research
Principle 2: Theory-Ingrained Artifact
2. Building, Intervention and Evaluation
Principle 3: Reciprocal Shaping
Principle 4: Mutually Influential Roles
Principle 5: Authentic and Concurrent Evaluation
3. Reflection and Learning
Principle 6: Guided Emergence
4. Formalization of Knowledge
Principle 7: Generalized Outcomes

Fig. 3. Action Design Research stages as defined by Sein *et al.* (2011).

The knowledge contribution framework, proposed by Gregor and Hevner (2013), divides possible knowledge contribution in design science research into four quadrants, depending on their solution maturity (y-axis) and application domain maturity (x-axis) (see Fig. 4). Works in DSR can be evaluated against this contribution framework to establish where an evaluated work fits into. Contributions can either be routine design, an improvement, an invention or an exaptation. If the solution as well as the application domain maturity is high, a routine design, with no major knowledge contribution, is present. Low solution but high application domain maturity define an improvement while low solution as well as application domain maturity define an invention. Finally, an exaptation is characterized by high solution but low application domain maturity. Gregor and Hevner (2013) state that inventions are very rare. If the contribution is an exaptation, existing solutions are applied to new problems or fields.

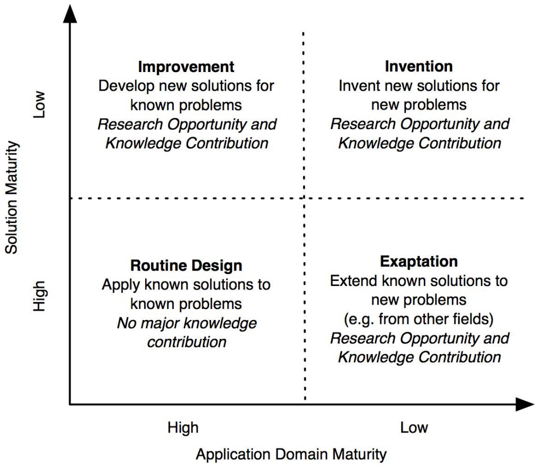

Improvement
Develop new solutions for known problems
Research Opportunity and Knowledge Contribution
Invention
Invent new solutions for new problems
Research Opportunity and Knowledge Contribution
Routine Design
Apply known solutions to known problems
No major knowledge contribution
Exaptation
Extend known solutions to new problems (e.g. from other fields)
Research Opportunity and Knowledge Contribution
Solution Maturity
Low
High
High
Low
Application Domain Maturity

Fig. 4. Knowledge Contribution Framework (Gregor & Hevner 2013).

3.2 Design science research cycles

Hevner (2007) and Hevner *et al.* (2004) define three Design Science Research Cycles to be used in information systems research. This approach features three research cycles in the areas of the research environment, design science research and knowledge base (see Fig. 5).

The relevance cycle is performed as a bridge between the environment and design science research and gathers user requirements and performs field-testing. The design cycle represents the core activities in design science research of building and testing artifacts and their evaluation. The rigor cycle bridges design science research and knowledge base and connects artifact design with all available information, including experience and the state-of-the-art. All cycles are repeated until desired results are achieved.

These cycles can be found in the present research. Hevner's research cycles are used for the overall work, researching user requirements, prior research and then design, as well as for each sub-artifact that is developed.

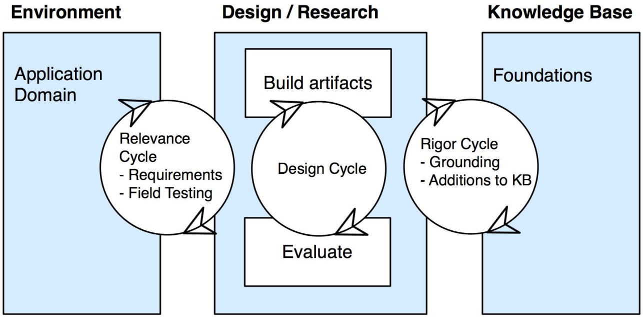


Fig. 5. Simplified depiction of (Hevner 2007) Design Science Research Cycles.

3.3 Addressing the research problem

Research was performed in several phases, which all built on each other. Each phase delivers the prerequisites for the next phase and is part of one of the cycles as shown in Fig. 5:

- Researching user requirements of the environment (Relevance Cycle)
- Formulating research questions (Design Cycle)
- Researching prior research in the knowledge base (Rigor Cycle)
- Constructing the artifact and sub-artifacts; see Fig. 6 (Design Cycle)
- Evaluation and discussion (Relevance Cycle & Rigor Cycle)

3.3.1 Researching user requirements

First, requirements were gathered to understand the need of the environment. In order to find out how important the issue of privacy protection in video surveillance is and which features are demanded of a video surveillance system that supports privacy protection, 37 persons of companies employing video surveillance were interviewed:

- Employees without security responsibilities (abbreviated to “employees”),
- users of security systems (e.g. security guards, abbreviated to “users”) and
- chiefs of security (i.e. security responsibility on management level).

The questions covered, among others, the following questions, which were to be answered on a Likert-type scale of one to ten, with one meaning "not important" and ten "very important":

1. *Do you feel more secure since video surveillance is used at your work place?*
2. *If you had the choice, would you install video surveillance at your work place?*
3. *Do you feel that video surveillance data might be used in an inappropriate way?*
4. *Do you think that by using video surveillance your personal freedom is restricted?*
5. *How important is personal privacy to you?*
6. *If available, would you appreciate privacy preserving video surveillance technology at your work place?*
7. *Can you provide the level of security as today if you can see what a person does, without knowing the identity of the person?*
8. *In order to provide the same level of security as today is it important to you to have the possibility to "unlock" identity information in case of an incident?*
9. *Are there different levels of authorization in your organization that allow the revelation of different levels of identity information, based on the authorization of the user?*

The average answers to these questions, separated by group, are shown in Table 9.

Table 9. Results of the interviews (average answers for each group).

Question	1.	2.	3.	4.	5.	6.	7.	8.	9.
Security chief	9	9.5	3	2	7	4	8.5	9.5	5
User	9	9	5	4.5	7	4.5	-	-	-
Employee	5	6	7	7	7.5	9	-	-	-

These results revealed that while employees consider video surveillance an appropriate tool to increase security, personal privacy is very important to them. They feel that increasing video surveillance usage creates uneasiness. If one does not control the technology, one cannot be sure how it is used. Hence, there was a strong demand for privacy preserving technology, which can assure employees that their privacy is respected. This further reflects the public discussion and opinion of video surveillance in the German region. The results from the two other groups, security technology users and chiefs of security, revealed that they can imagine offering the same security if they do not know the identity of individuals *during the incident*. However, for them it is vital to provide security to be able to reveal the identity of a person in case of an incident. These were the two main insights of these interviews: privacy protection is very important to people but in order to provide security it should be possible to reveal the identity of a person in case of an incident.

3.3.2 Formulating research questions

The results of the interviews revealed requirements for a privacy preserving video surveillance system and led to the main research question:

> *"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"*

Security was deemed important in the sense of physical protection of property and life. It was further assumed that a video surveillance system would support security personnel to provide physical security. At the same time, especially employees demanded that their privacy would be protected. However, a system, which would protect the privacy of persons while providing security would not decide by itself whose privacy should be protected. This decision would depend on regulation in the organization, which would use the system.

Studying the prior research revealed that while some sub-components necessary for a selective privacy preserving system exist, most of the components are currently too slow and unreliable to be used in a complete system, which could be used in a real-world scenario.

When working with this main question, components of the system were identified and it was decided to add additional sub-questions to cover sub-components of the system. Many other questions could be studied, however these questions were the most important to answer to develop a system that would work in real-world scenarios and would provide actual benefits to the customer:

1. "*What existing methods can be used to achieve selective privacy protection?*"

   In order to protect the privacy of uninvolved individuals while still making relevant people visible, privacy protection has to be performed for individuals selectively. An example for an application for this feature is an office building, where the privacy of employees needs to be protected while suspicious strangers should be clearly visible. Further examples include the protection of public places, high security areas and VIP areas. In order to achieve selective privacy protection, people have to be separated and identified in an image.

2. "*What kind of performance measurement metrics exist to measure video surveillance coverage in a given area and can these be improved?*"

   In order to set up a video surveillance system that aims at protecting the privacy of individuals, privacy considerations have to go beyond technology itself and have to include privacy by design already during the planning phase. This means planning the deployment of video surveillance cameras to maximize security while minimizing the amount of cameras used. In order to achieve this in practice, guidelines are needed for system integrators and planners concerning how to set up camera layouts with defined levels of security and privacy.

3. "*How much can the performance and accuracy of background subtraction and shadow detection methods be increased?*"

   Before segmenting and tracking people in a video image (sub-question 1), foreground objects have to be separated from the background (background subtraction). While background subtraction is a reasonably well researched field, there remains room for improvement and there are few good and fast shadow detection methods. These are critical since shadows have similar features than objects, e.g. their shape and movement, but should not be detected as foreground since this would lead to false tracking results. As with all methods in this work, in order to be performed in real-time, they have to be performed in a fast and efficient manner. Hence, this sub-question is important before segmentation and tracking methods can be researched in the next step.

4. "*How can the accuracy and performance of discrimination between objects be increased?*"

   Object segmentation is a step before the actual tracking of objects in an image can be performed. With segmentation, different objects, which have been detected as foreground using background subtraction, can be distinguished and subsequently tracked. Current object segmentation methods are very performance intensive and inaccurate. If they are inaccurate, people whose privacy should be protected might be unmasked. Further, if used on smart cameras, segmentation methods should be optimized for performing on limited hardware resources.

5. "*How can the accuracy of segmentation in mean shift based methods be improved?*"

   After object segmentation, tracking methods can use this output to track objects or people in the image. Mean shift based methods are known to be very fast segmentation algorithms, however accuracy is not satisfactory compared with other, slower methods. Therefore, using mean shift as a basis and researching methods to increase its accuracy is the task for this sub-question.

6. "*What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?*"

   For data protection and performance reasons it is critical to decide where to store identity information. From a data protection perspective, it should be stored before it leaves the camera or be transferred in a highly encrypted way. From a security perspective, scenarios such as when smart cameras are stolen have to be considered. From a technical point of view, limited storage resources on smart cameras have to be considered. Furthermore, different storage methods have to be taken into account, which range from storing two video streams (one privacy protected, one original) to just keeping metadata about the identity of a person.

3.3.3 Prior research

After the research questions were clearly defined, a review of the prior research was performed to understand the available knowledge base (see Chapter 2 for a detailed discussion of this topic). This defined what tasks still have to be fulfilled in order to build an overall selective, privacy-protecting video surveillance system and hence to answer the main research question.

3.3.4 Constructing the main artifact and sub-artifacts

The construct was divided into different components in order to reflect the different parts that needed to be researched, built and/or enhanced. The defined sub-components were:

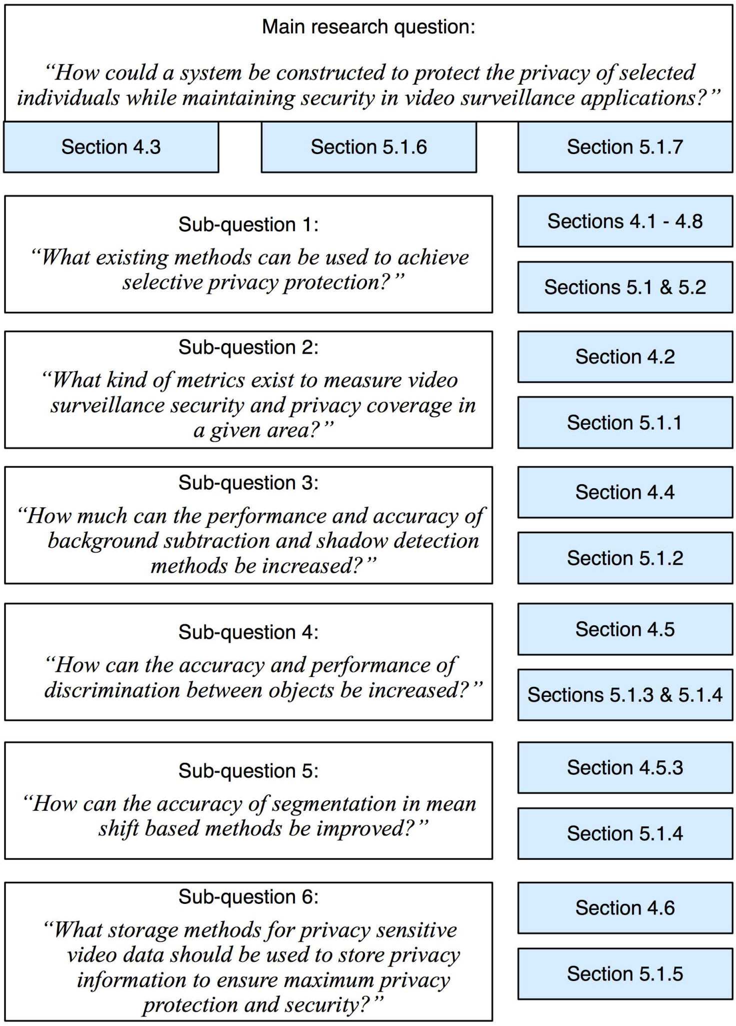
- Security and privacy metrics for video surveillance
- System architecture of the system to be developed
- Object tracking, including
    - Background subtraction and shadow detection
    - Image segmentation
    - Object association
- Storage of identity information

Fig. 6 shows how the main research question and the sub-questions are answered in each of these sub-components. Mostly, each sub-component represents the answer to one sub-question. However, sub-components might answer multiple questions at once. This figure is shown at the beginning of each chapter to clearly illustrate what questions are answered.

For each of these components existing methods were researched. If a method existed that did not yet meet the requirements, the aim was to extend it to fit to the requirements. If no method existed (e.g. security and privacy metrics), it was newly developed based on experience in similar fields or problems. While the final evaluation was done in a separate step, each component was already evaluated separately and compared to prior methods. This evaluation further led to a build-evaluate-build loop, as described by Hevners design cycles (see Section 3.1).

3.3.5 Evaluation of the construct and discussion

After all components were built, the whole construct was evaluated and compared with other systems in order to get an understanding what the contribution to the knowledge base was. The prototype of the final construct was shown to selected customers, who were part of the interviews during the requirements phase, to gather feedback and input in the form of a survey. During an in-depth discussion, the results of each component were described as well as ideas for future improvements were given. These ideas will lead to new research in this area in the future.



Main research question:

“How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?”

Section 4.3 — Section 5.1.6 — Section 5.1.7

Sub-question 1:
“What existing methods can be used to achieve selective privacy protection?”
Sections 4.1 - 4.8
Sections 5.1 & 5.2

Sub-question 2:
“What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?”
Section 4.2
Section 5.1.1

Sub-question 3:
“How much can the performance and accuracy of background subtraction and shadow detection methods be increased?”
Section 4.4
Section 5.1.2

Sub-question 4:
“How can the accuracy and performance of discrimination between objects be increased?”
Section 4.5
Sections 5.1.3 & 5.1.4

Sub-question 5:
“How can the accuracy of segmentation in mean shift based methods be improved?”
Section 4.5.3
Section 5.1.4

Sub-question 6:
“What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?”
Section 4.6
Section 5.1.5

Fig. 6. Depiction where the main research question and each sub-question (left) is answered (right).

4 Privacy enhancing video surveillance

This chapter describes the construct, which was built as a result of understanding prior research and the current state-of-the-art of privacy and video surveillance as well as of researching the environment. This chapter first introduces the overall system functionality before describing each part of the developed construct.

4.1 Introduction

The aim of the construct was to develop a privacy enhancing video surveillance system, which would scramble identity information in video surveillance data while allowing authorized personnel to selectively unscramble individuals. As elaborated in Section 1.2, the first two generations of privacy enhancing video surveillance systems only protect the privacy of people in unsatisfactory ways. The first generation scrambles complete image regions (Fig. 1), thus not providing any protection for people moving out of the regions. The second generation scrambles all movement in the image (Fig. 2), thus not providing the possibility to unscramble a video of a suspect without compromising the privacy of all other individuals in the image.

The current study introduced the third generation of privacy protection systems: selective privacy protection. Selective privacy protection aimed to remedy the shortcomings of previous generations by using tracking and matching algorithms to identify individuals in the video. Using this information, authorized users could choose to unmask only offending individuals without compromising the privacy of possible bystanders. Selective privacy protection increased the complexity of previous privacy protection generations by adding domain problems relevant to object tracking. Apart from association problems one of the main issues of automatic privacy protection were lighting changes and shadows. Both conditions occur frequently in outdoor scenarios resulting in large areas of the video being privacy protected. Accordingly, these considerations were part of the design process of the proposed system.

Fig. 7 shows an example of a scene where individuals are scrambled but two selected individuals are not (simulated image). By detecting, tracking and identifying specific individuals, scrambling and unscrambling could be performed in a flexible way so that no more people than necessary had to be unscrambled. This new system aimed at solving the drawbacks of the second generation and provided a number of further advantages. Using a privacy protection system of the third generation, authorized users were able to unscramble only offending individuals without compromising the privacy of possible bystanders. This unscrambling process could be implemented securely by using personal chip cards and asymmetric encryption.

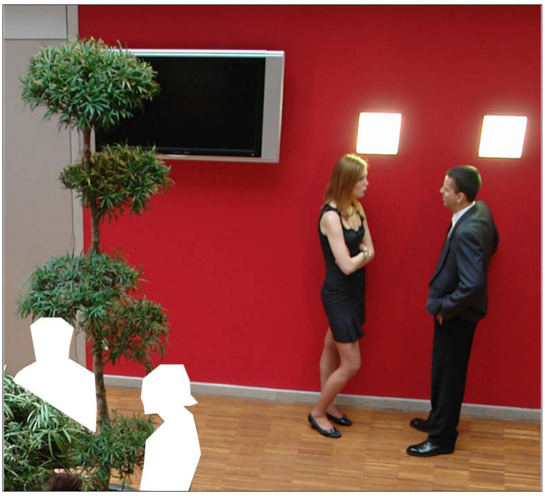


Fig. 7. Selective privacy protection (simulated image).

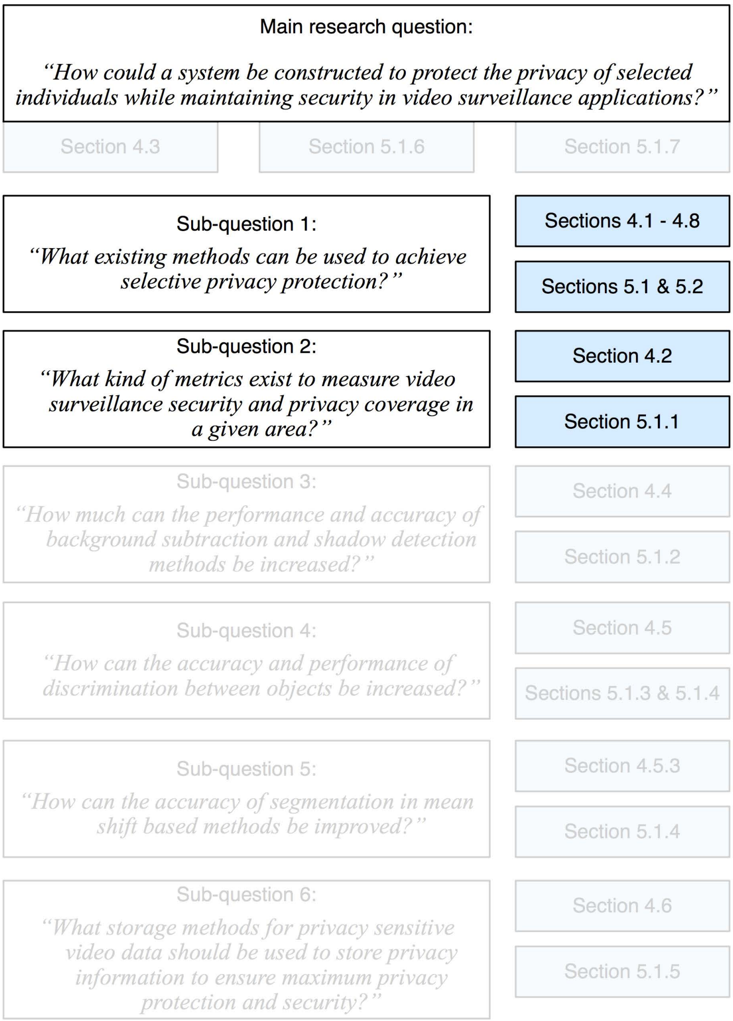
In order to build this new generation of privacy protection systems the following was necessary:

- An overall video surveillance architecture that was built from ground up to provide privacy protection ("privacy by design") and provides a flexible framework for identification technologies
- Robust tracking methods, which reliably track individuals to provide object information for a masking module
- A robust background subtraction method with advanced shadow detection and elimination methods as well as image segmentation
- Secure and optimized storage of critical privacy information

Building such a system has resulted in the system presented in this work. In the following sections the individual sub-constructs of this system are introduced.

4.2 Security & privacy metrics

This section presents metrics for video surveillance deployments in order to provide a measure on how secure and privacy protected a defined area under surveillance is. As such, this section answers sub-questions one and two of the research questions (see Fig. 8).



Main research question:

"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"

Section 4.3

Section 5.1.6

Section 5.1.7

Sub-question 1:

"What existing methods can be used to achieve selective privacy protection?"

Sections 4.1 - 4.8

Sections 5.1 & 5.2

Sub-question 2:

"What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?"

Section 4.2

Section 5.1.1

Sub-question 3:

"How much can the performance and accuracy of background subtraction and shadow detection methods be increased?"

Section 4.4

Section 5.1.2

Sub-question 4:

"How can the accuracy and performance of discrimination between objects be increased?"

Section 4.5

Sections 5.1.3 & 5.1.4

Sub-question 5:

"How can the accuracy of segmentation in mean shift based methods be improved?"

Section 4.5.3

Section 5.1.4

Sub-question 6:

"What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?"

Section 4.6

Section 5.1.5

Fig. 8. Placement of this section for the answering of the research question.

4.2.1 Visual performance parameters

In order to standardize the visual coverage of the video surveillance systems novel metric parameters are proposed (Sutor *et al.* 2008a). Those parameters were indispensable for design and engineering of large-scale video surveillance systems, especially when protecting critical infrastructure.

Visual Surveillance Coverage Quality (VSCQ)

To design and determine the required configuration of the video surveillance system for critical infrastructure, the position, distribution and number of required cameras must be carefully determined. The system must be able to provide a maximum coverage of the surveillance focus and must guarantee at least a minimum coverage level even if one or more cameras fail. To measure this coverage level, the following performance parameter was introduced. In this thesis Visual Surveillance Coverage Quality (VSCQ) is described as a "metric to describe the quality of a video surveillance system measuring the redundancy of the visual coverage of the video surveillance cameras to the area under surveillance".

VSCQ is defined in different coverage levels. The higher the coverage level, the more parts of the system can fail without losing situational awareness. A higher coverage further results in fewer blind spots due to occlusions, because multiple cameras observe the same area. This was especially important when using automated video analytics, where video streams would be automatically analyzed to detect critical events. Especially for determining the position of a person and tracking him a higher number of cameras for one scene would improve results significantly. This was important when new tracking methods were introduced. With a high VSCQ level video streams can be analyzed better, resulting in more options, including tracking people in crowds and tracking more people per area. The proposed metrics could further simulate attack scenarios showing which attack has which impact on the performance of a video surveillance system. Not necessarily entire rooms were covered with the same level - in an art gallery the walls could be covered with a higher and the middle of the rooms with a lower level; on an airport it might be the opposite.

In the following passages, five coverage levels are presented. The illustrations shown in Fig. 9 represented a 2D projection of a 3D scene. Actual image processing computation was done in a virtual 3D space. CSA stands for Critical Surveillance Area, i.e. the area of interest.

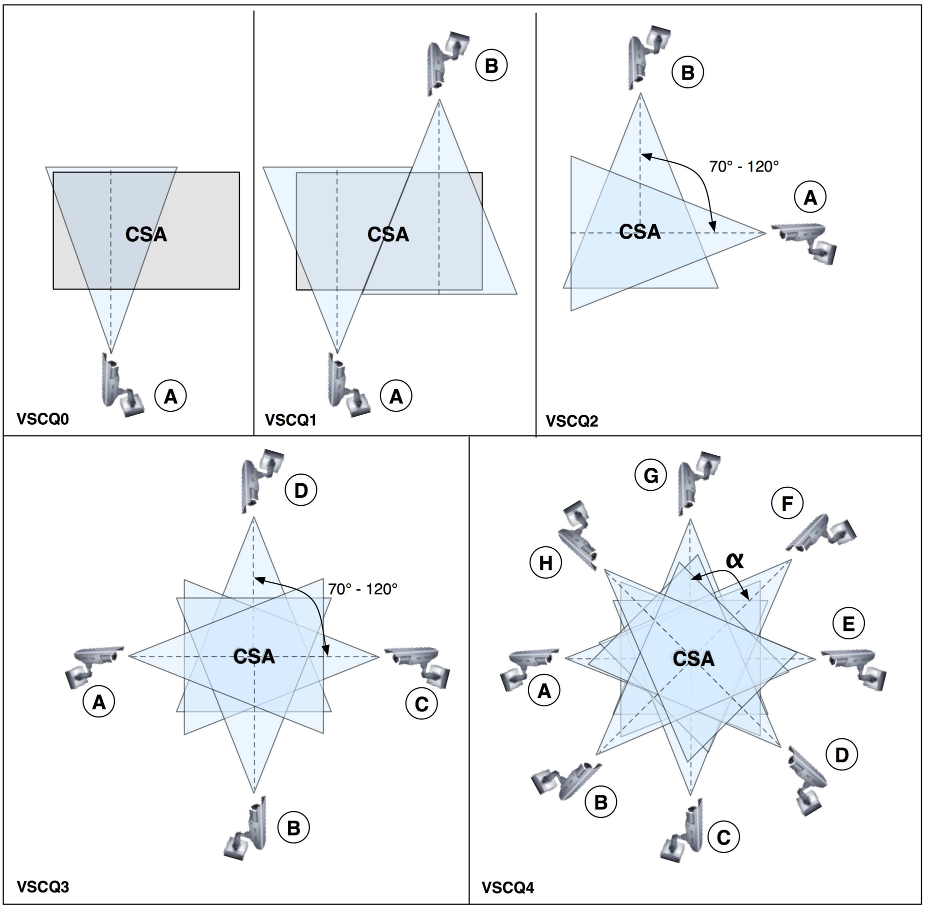


Fig. 9. Different types of video surveillance coverage quality levels.

VSCQ0: Part coverage

As shown in Fig. 9, the lowest coverage level represented the most basic coverage with one camera observing a given area. Naturally, VSCQ0 was the least secure level; if the camera would fail there would be no backup to take over the tasks of the camera and the area previously observed by the camera was not seen. VSCQ0 should not be used in secure areas; however, it might be suitable for less frequented areas that don’t contain critical objects. Due to occlusions or other obstructions it might not always be possible to track people and analyze behavior consistently using automated video analysis software.

VSCQ1: Full coverage

VSCQ1 required two or more cameras to cover the observed area but only with minor overlapping (see Fig. 9). For automated video analysis this configuration had the disadvantage of making it hard to handover the tracked person to other cameras due to the minimal overlapping areas, especially given a high density of people.

VSCQ2: Two dimension coverage

As shown in Fig. 9, VSCQ2 required at least two cameras observing the same area with an angle between their viewpoint axes of 70 and 120 degrees (ideally the angle would be 90 degrees). This way a calculation of the position of a person was more accurate than in VSCQ1 because one of the cameras would have a side view of the person at all times; this also meant that occlusion handling was arguably better due to placement on a different axis. The main disadvantage of this level was the limitation of visual identification of camera tampering.

VSCQ3: Two dimension double side coverage

As depicted in Fig. 9, VSCQ3 required an area to be observed from two different sides redundantly. This level should be used in areas where important places, such as at an airport, had to be protected. The placement on each side of the object allowed visual conformation of the current state of the object even in heavily crowded scenes. If one of the cameras failed, the quality of the system was reduced to VSCQ2 with an additional camera; this had minimal implication to video analytics because coverage was still two-dimensional and allowed separation of occluded individuals. Due to the placement of the cameras, visual tampering detection was also possible in this level.

VSCQ4: X point surround coverage

VSCQ4 required more than double-sided coverage of an area (see Fig. 9). When using this level the first step was to determine the camera count to be used, which was assumed to be x, where x had to be greater than four. The cameras were placed in a way that the angles between the cameras were evenly distributed; e.g. eight cameras were used to cover an important area so the placement of the cameras followed a 45° pattern. Hence, the angle α (see Fig. 9) was calculated as follows:

$$p_x = \begin{cases} 1 & \text{if } (c_{pre} - c) < T_c \text{ and } (i_{pre} - i) > T_i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

For x greater than five, a loss of a single camera did not decrease the surveillance coverage due to an increase in redundancy, which was the main advantage of this level. In addition, visual tampering detection was possible. A quality improvement may also be present but the more cameras were added the less the actual improvement due to increasingly overlapping views.

4.2.2 Extending the VSCQ

Adding cameras to the system could extend the VSCQ levels with devices providing additional capabilities. These capabilities included top-down cameras to help flow analysis as well as person separation, an extra camera that is solely used to increase the level of redundancy or a PTZ camera to allow high definition videos in the designated areas. Such additions to the coverage level were denoted by adding G for a ground camera, E for an extra camera or P for a PTZ camera to the VSCQ level. In some conditions an extra static or PTZ camera could be used to replace a malfunctioning camera, which guards against a VSCQ level loss but the extended functionality offered by the addition was lost temporarily.

4.2.3 Visual Surveillance Coverage Redundancy (VSCR)

To ensure the desired coverage quality even in case of failure a level of redundancy was introduced, which was placed in four categories that offer increasingly stronger insurances against component failure, e.g. by incorporating parallel redundant cameras. The redundancy requirements were defined as shown in Table 10.

For r0 to r2 the number of the redundancy level was also an indicator of the quality of redundancy, the special case r3 which was only valid for VSCQ4 was an exception as the number of camera views is decreased with a failed camera.

Table 10. Description of the different redundancy level and requirements.

Level	Description
r0	No redundancy. Failure of a single camera may result in a loss of video surveillance in this area.
r1	Failure of a camera means a reduction in VSCQ level but will never result in a total loss of video surveillance.
r2	Any single camera can fail without a loss in the VSCQ level. For example, this can be achieved by adding additional parallel cameras.
r3	This is only applicable for VSCQ4: Failure of a single camera will result in a slight loss of quality because of the highly overlapping camera views, however the redundancy level is only decreased if the camera count reaches four or less.

4.2.4 VSCQ grammar

The term “grammar” was used in this context in the sense that it was a set of formation rules of a formal language. The grammar used to describe the Video Surveillance Coverage Quality VSCQ was defined as follows, where the non-terminal Ω is the starting symbol. A non-terminal is a term, which can be substituted by another symbol, according to the given VSCQ grammar.

$$\begin{aligned} \Omega &= VSCQ[0-4]-r[0-3]\ \ \Xi\ \ \Psi\ \ \mathrm{Z} \\ \Xi &= +G-r[0\,|\,2] \\ \Psi &= +[0-9]^{+}E \\ \mathrm{Z} &= +[0-9]^{+}P \end{aligned} \tag{2}$$

Using this grammar the coverage quality of any given scene could be described, entering the corresponding values into the above VSCQ grammar definition. Table 11 shows examples on the usage of the grammar.

Table 11. Examples on how to use the VSCQ grammar.

Grammar	Description
VSCQ2−r0	Full coverage without redundancy, with VSCQ level 2.
VSCQ2−r1+G−r0+1P	Full coverage with redundancy level one; an extra ground camera that is not used redundantly and a PTZ camera.
VSCQ4−r3+G−r0+2E+3P	Surround coverage with redundancy level 3 and two extra cameras and three PTZ-cameras and not redundant top-down cameras. This could be an example for a station concourse.
VSCQ2−r2+G−r2	Full coverage with two redundant cameras. In addition a redundant ground view is added.
VSCQ3−r2+3E	Two-dimensional double side coverage with four redundant (i.e. eight) and with three extra cameras.

4.2.5 Privacy Quality Levels (PQL)

Similar to VSCQ, PQL defined what kind of privacy protection for a given area is present. This way, individuals under surveillance knew at any time how much their privacy was intruded upon. This information could be provided to these people by means of signs in the area under surveillance. Naturally, in the event of an emergency or critical situation, authorized personnel could access all original data using PEVS. Three different levels were defined.

PQL0: No privacy protection

This was the first privacy protection level. No privacy protection was provided, all detected information was recorded and stored uncensored and unencrypted.

PQL1: Part privacy protection

This was the middle privacy protection level. Access to recorded material was restricted, live streams could be viewed by every person having access to the system. However, a restricted secure access rights system was used. In addition, data was encrypted. For scrambling of personal information, a low-level scrambling technique was required, which scrambled movement but kept certain information about a person, such as color of their clothes.

PQL2: Full privacy protection

This was the highest level of privacy protection. All identity and privacy information in the video was automatically made unidentifiable. Access to live and recorded data was restricted by a secure certificate system. Moreover, recorded data was encrypted using this certificate. Access policies and an advanced user rights management system were employed. For scrambling of personal information, high-level scrambling had to be chosen, which allowed no possibility to get information about a person. Hence, all movement, facial information and color of individuals needed to be scrambled.

Integration into VSCQ grammar

PQLs could be incorporated in the VSCQ grammar similarly. Taking the examples shown in Table 11, the grammar was extended to VSCQ2-r0 / PQL1 or VSCQ2-r1+G-r0+1P / PQL0. PQLs included information, which can be provided to security and law authorities, whenever requested, about privacy protection precautions being in place. Furthermore, PQLs are extra information for data protection agencies, which authorize video surveillance installations.

4.2.6 Engineering and dimensioning of the system parameters

The requirements for any security system included quality-based factors such as coverage or response time. However, just as important was the reliability of a security system. Given that it is impossible to know in advance when threatening events might occur, a reliability approaching 100% was aspired. Components were classified as follows:

- Fallible components, such as cameras, switches, processing units, routers and repeaters. For every component the MTTF can be evaluated.
- Infallible components, such as cables and other components without electronic parts. These are components that just fail in case of destruction. For these components it is not possible to calculate a MTTF.

In order to achieve close to 100% reliability, fallible components had to be used redundantly. To ensure a certain VSCQ level even in the worst case the right dimensioning and positioning had to be chosen. First of all the VSCQ level that was necessary to guarantee the right level of physical security had to be evaluated.

As discussed in the case studies later on the necessary VSCQ levels depended on and can vary by the usage of an area. This level was called threshold VSCQ level. At this threshold VSCQ, a minimum level of the physical security must be guaranteed. Some extra features might not work properly and more than usual false alarms are generated. This means, whatever happened, the whole situation was still under control by bundling human resources.

Next, the expected reliability (MTTF) of every single fallible component and the underlying fallible components was required. These underlying components were, depending on the system architecture, the power supplies, the network and primary image processing units. In other words it was any part of the system that could cause a single camera not to deliver pictures. The reliability of a single camera was composed by their underlying fallible components and an extra factor for destruction or vandalism. This vandalism factor and the MTTR depended on the placement of the camera. For instance, an indoor camera in an airport four meters above the ground could be hardly destroyed by vandalism but would be difficult to be repaired during usual airport operation. Now the availability for every single camera could be calculated from the MTTF, the MTTR and the vandalism factor.

The last step in the preparation period of evaluating the right dimension of the redundancy level for a video surveillance system was to define the VSCQ level that was favored in a best-case situation, a situation where all components were working correctly. This level was called desired VSCQ level. Here, all desired extra features could be guaranteed and the rate of false alarms was low enough that there were free human resources. There were two conditions to be fulfilled:

1. The dimensioning had to be considered in a way that the threshold VSCQ level could be guaranteed in a worst-case scenario. Such a scenario could be the loss of a camera and the loss of a network component or the loss of a single camera at the same time as a whole network cable leg was destroyed, either by sabotage or by vandalism. This corresponded with a failure of a fallible component and the loss of an infallible component or the simultaneous failure of multiple fallible components but not more that 50% of all cameras. Here, connections between single cameras such as network components and power circuits had to be taken into account. Special attention in this worst-case scenario had to be given to commonly used infallible components.

2. The second dimensioning rule guaranteed the desired VSCQ level in a “normal” trouble scenario caused by normal component availability. Such a disaster scenario was defined by the loss of any single fallible component. This event might be caused by a lost camera connection due to network failure or by a broken or destroyed camera.

The redundancy level had to be chosen in a way that these rules are always fulfilled. However, the availability was falling with the age of a component.

4.3 Privacy enhancing video surveillance system architecture

In this section the overall system architecture of a novel privacy preserving system is presented: the Privacy Enhancing Video Surveillance architecture called PEVS (Matusek 2012). This section thus helps answer the main research question and parts of sub-question one, since a new system architecture unifies all system components, which are later described as separate sub-constructs (see Fig. 10):

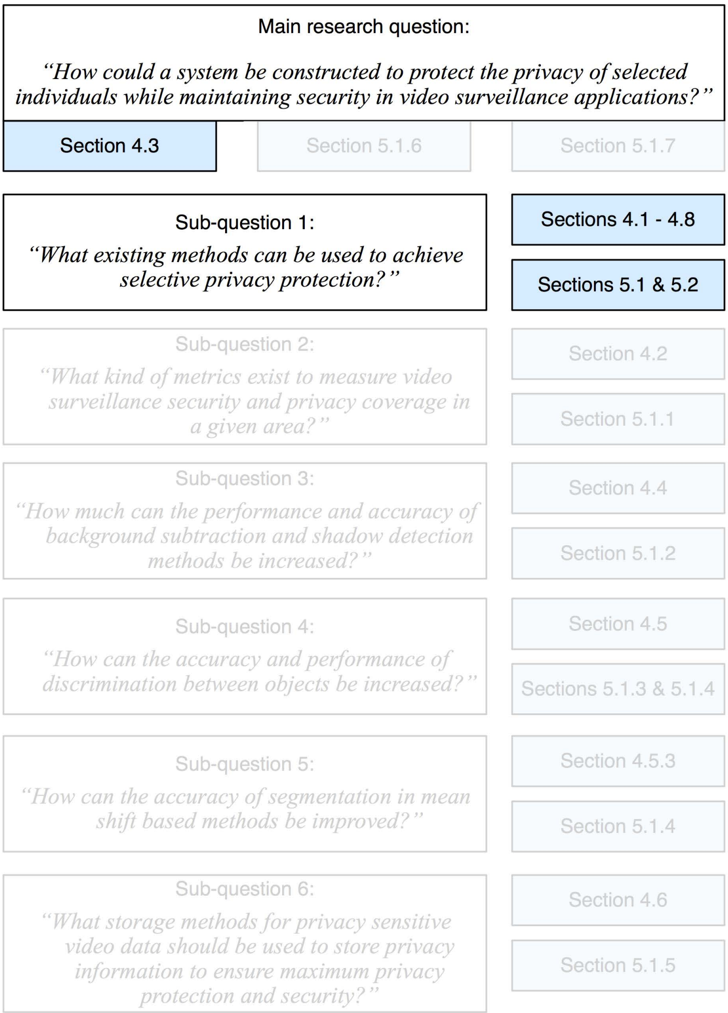

Main research question:
"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"
Section 4.3
Section 5.1.6
Section 5.1.7
Sub-question 1:
"What existing methods can be used to achieve selective privacy protection?"
Sections 4.1 - 4.8
Sections 5.1 & 5.2
Sub-question 2:
"What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?"
Section 4.2
Section 5.1.1
Sub-question 3:
"How much can the performance and accuracy of background subtraction and shadow detection methods be increased?"
Section 4.4
Section 5.1.2
Sub-question 4:
"How can the accuracy and performance of discrimination between objects be increased?"
Section 4.5
Sections 5.1.3 & 5.1.4
Sub-question 5:
"How can the accuracy of segmentation in mean shift based methods be improved?"
Section 4.5.3
Section 5.1.4
Sub-question 6:
"What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?"
Section 4.6
Section 5.1.5

Fig. 10. Placement of this section for the answering of the research question.

4.3.1 Video surveillance architecture without privacy protection

An intelligent video surveillance system might be constructed as depicted in Fig. 11 (parts, which were dealt with in the research are shaded). The PEVS system expands this concept by adding privacy protection mechanisms to parts of this architecture (Matusek 2012). Each part of this system in described in the following paragraphs.

Input / sensors

The type of input sensors was determined according to the task to be executed. This varied from video, binary sensors, acoustic, ultrasonic, radar, infrared, visual, ultraviolet and X-rays or a combination of multi-sensor types.

Pre- and data processing

Preprocessing included individual calibration of the different sensors categories, noise filtering and data abstraction. Additionally, filters to detect and remove snow and rain as well as algorithms to stabilize the image and rectify it were used. In addition, preprocessing provided central synchronizing of all sensors and sensor categories.

Data fusion

In data fusion, object tracking was performed and combined with all available sensor information. This was followed by statistical, threat and situation analysis. Data fusion in next generation video surveillance systems was not only concentrating on extraction of video metadata but also increasingly focusing on the process of metadata consolidation.

Data storage and retrieval

High performance storage and retrieval technologies were available commercially. However, it was rather difficult to find a technology providing privacy enhancement while providing the same level of performance. In Section 4.6 a privacy-enhanced video surveillance algorithm is presented. This algorithm posed a number of challenges to storage optimization, query performance, security management, access control and performance management.

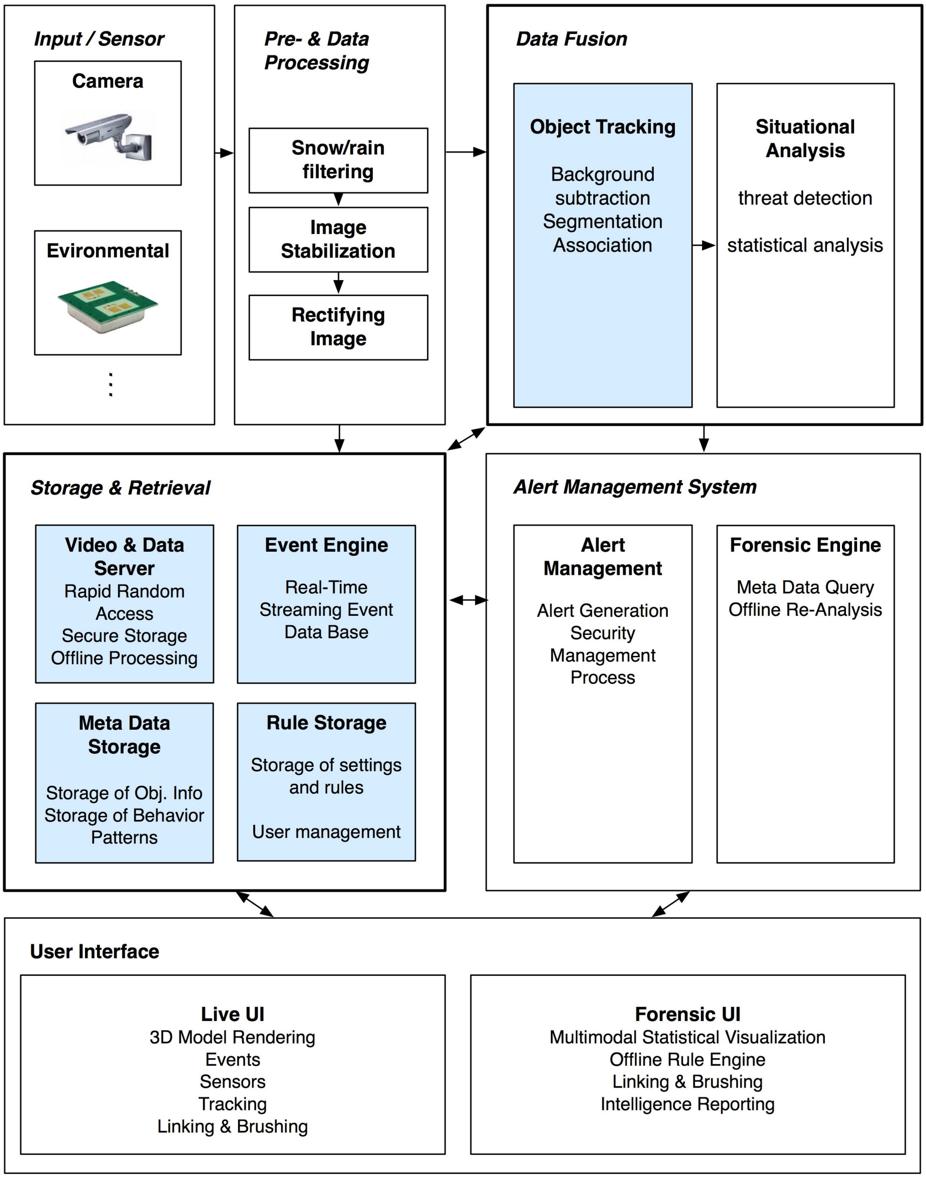


Fig. 11. Typical video surveillance system processing architecture.

Alert management system

The alert management system was closely tied to the user interface, providing all necessary data the user might need. It consisted of an alert management engine and a forensic engine. Alert management generated and stored all alerts the situational analysis engine detected. It further implemented different security management processes for different situations in order to deliver alerts in a way the security personnel would use them.

Output

Output included presenting the information in a user interface to the user. It played closely together with the alert management and the storage and retrieval systems. The live UI provided live video feeds to the user while the forensic UI could access stored data for reanalysis.

4.3.2 The PEVS architecture

The video surveillance architecture presented in Fig. 11 was taken as a basis for PEVS, which included only parts of the overall architecture. PEVS included the parts of the video surveillance architecture that are framed with a black border in Fig. 11. All other parts could be used as well but were not part of PEVS. Fig. 12 shows an overview of the PEVS system, including all parts of the system, which are elaborated upon in the following section.

Input & preprocessing

The input to PEVS was image data in various formats (e.g. H.264, MPEG4, MJPEG) and sent either from an image sensor of a camera or from another system, such as a video management system. Before this data was fed into PEVS, the image was preprocessed to be optimized for image processing. This included rectifying and stabilizing the image as well as reducing noise caused by environmental conditions such as snow or rain. The preprocessed image was then sent to PEVS where it was processed by a number of processes.

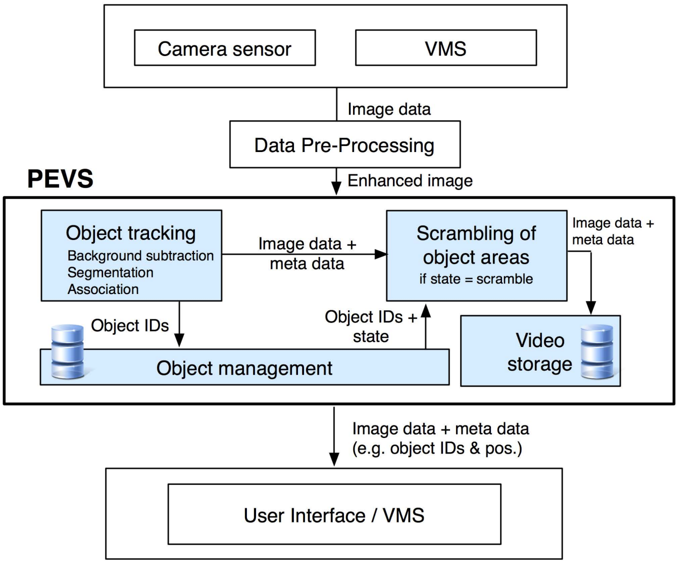


Fig. 12. Overview of the Privacy Enhancing Video Surveillance architecture.

Object tracking

First, tracking was performed, which included background subtraction, shadow detection, segmentation and association between different objects. Tracking assigned object IDs to each object found. These objects were tracked through a number of video frames and associated to one another. The tracking module sent the image and metadata such as the object IDs to the scrambling module.

Scrambling

The scrambling module acquired the state from object management and depending on that either scrambled the area in the image where a specified object was present or not. The scrambling module further sent the image and scrambling metadata to video storage as well as an output of PEVS.

Object management

An object management database stored all object IDs, their position in each frame and their state, which could have two values (either "scramble" or "clear"). This was important for deciding which image areas should be scrambled and which not. By default, each object had the state "scramble". If a user decided to unscramble a person, the state of an object in object management changed from "scramble" to "clear".

Use case

The output of PEVS was sent to a video management system (VMS). In the user interface of the VMS, the user could decide which objects should be scrambled and which not by clicking on the corresponding object. Alternatively, using external means for identifying a person (e.g. through biometric face recognition or RFID tags), this decision could be done automatically by matching an object ID with a person database. Fig. 13 shows an architectural view on the system where the process of requesting a video stream with an individual person unscrambled is shown. If the state of an object, either through a user decision or through automatic recognition, was changed to "clear", this information was sent to object management where the state of the object was changed in the defined time frame. By default, image streams were scrambled live as they were recorded. If the state of an object was changed retrospectively, e.g. if a user decided that in a recorder video stream a specific person should be unscrambled, PEVS requested the recorded image data including metadata from the storage archive and scrambled only image regions of objects with a "scramble" state.

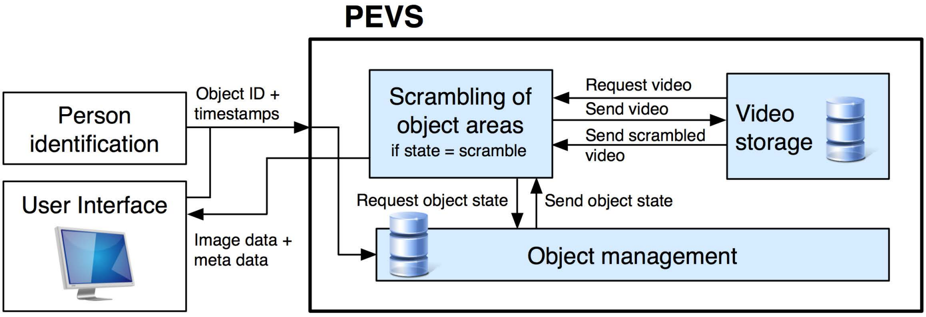


Fig. 13. Architectural view of possibilities to unscramble a person.

PEVS was designed to be platform independent so different implementations of PEVS could run on different platforms. For example, the complete process (excluding the user interface) could be implemented to be run directly on a smart camera. There were also hybrid approaches, where part of the system would be implemented in a smart camera and one part would be implemented centrally. Fig. 14 shows a possible example, where tracking and scrambling as well as storage of identity information (e.g. image regions where people are seen) could be done in smart cameras while visualization and storage of camera images with identity information removed was done centrally. Where identity information is stored depends on the storage technique chosen (see Section 4.6 for details). The cameras could send alarms, scrambled video feeds and identity information (using correct authorization). Additionally, mobile devices could access scrambled as well as original video feeds. Wireless access to the control center was managed over 3G, LTE, WiFi or WiMAX networks. Since identity information was only stored in the cameras themselves, no user could access this information without proper authorization. A typical user was only allowed to view scrambled video information. In case of a critical incident and a user with appropriate authorization could request the original feed with identity information. The request was sent from a standalone user interface or from a mobile device to the control center. Using the authorization, the control center in turn requested the identity information from the correct smart camera. If the authorization was correct, the camera sent the image regions with identity information to the control center. This information was passed on to a data fusion module that combined this information with the scrambled video frames in the database and sent the original video feed to the control center, which passed them on to the user.

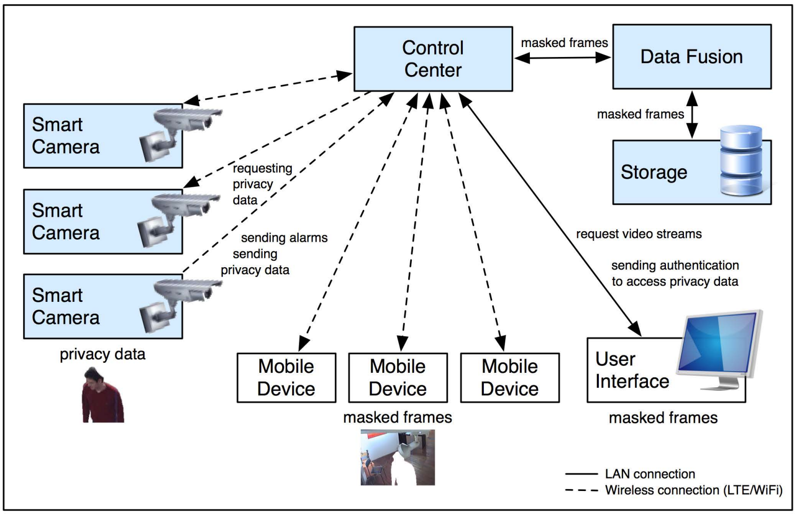


Fig. 14. PEVS usage with smart cameras and mobile devices.

4.4 Background subtraction and shadow detection

This section describes the first part of the object tracking module, as shown in Fig. 12. It further helps answer research sub-question three (see Fig. 15).

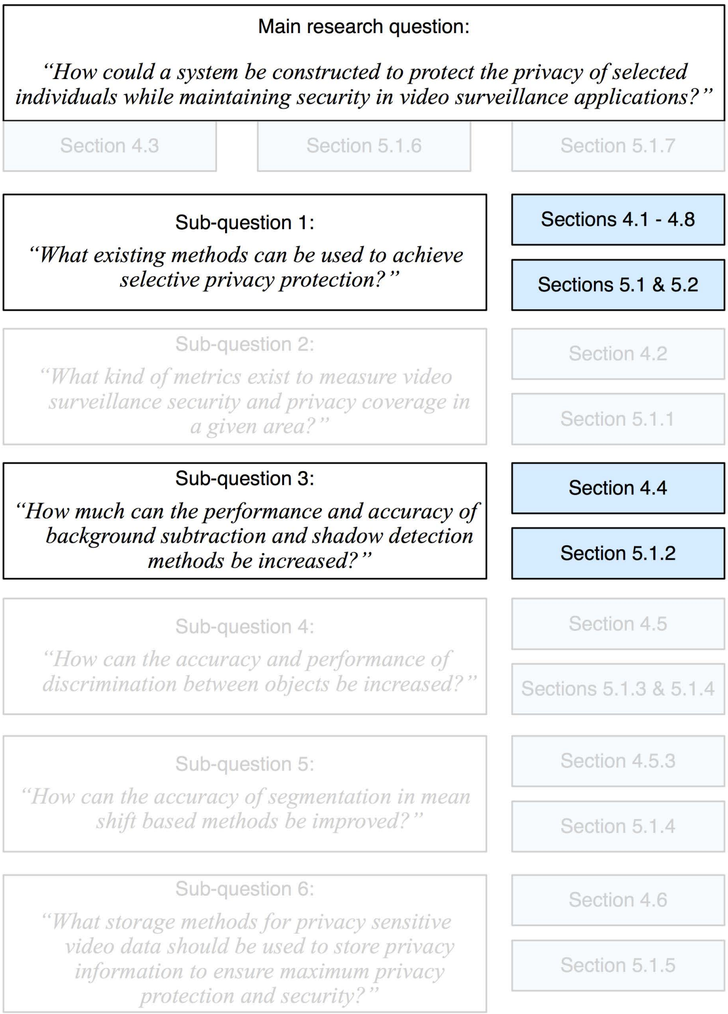

Main research question:
"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"
Section 4.3
Section 5.1.6
Section 5.1.7
Sub-question 1:
"What existing methods can be used to achieve selective privacy protection?"
Sections 4.1 - 4.8
Sections 5.1 & 5.2
Sub-question 2:
"What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?"
Section 4.2
Section 5.1.1
Sub-question 3:
"How much can the performance and accuracy of background subtraction and shadow detection methods be increased?"
Section 4.4
Section 5.1.2
Sub-question 4:
"How can the accuracy and performance of discrimination between objects be increased?"
Section 4.5
Sections 5.1.3 & 5.1.4
Sub-question 5:
"How can the accuracy of segmentation in mean shift based methods be improved?"
Section 4.5.3
Section 5.1.4
Sub-question 6:
"What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?"
Section 4.6
Section 5.1.5

Fig. 15. Placement of this section for the answering of the research question.

4.4.1 Introduction

The first step in the object tracking module was extracting motion information using background subtraction. Background subtraction detects movement in an image by subtracting background pixels from the current frame and thus provided metadata for image segmentation and association. Background subtraction detects all pixels that differ from background pixels, which are learned over time. Hence, shadows are detected as well. It was undesirable to use metadata with shadows since shadows do not contain critical identity information of a person and thus should not be scrambled. Hence, it was important to employ robust shadow detection after background subtraction to deliver accurate object information in PEVS (Matusek *et al.* 2008).

4.4.2 Scrambling and shadow detection

The goal for PEVS was to scramble 100% of identity relevant information in the image while minimizing the number of pixels scrambled, which did not contain identity information (see Section 4.8.2). Shadow detection ensured that pixels that were not part of a person or an object were not masked. However, a shadow detection algorithm had to perform the critical balance between detecting all shadow pixels while not unmasking any private information. Fig. 16 shows a typical situation with two individuals with no shadow detection, demonstrating the effect of shadows on privacy scrambling. All movement in the image was masked. Here, shadows were classified as part of the person. All masked pixels amounted to approximately 19.800 pixels. Roughly 6.500 of the pixels were shadow pixels, which did not need to be scrambled in order to preserve privacy. With a shadow detection algorithm these pixels would not have been masked and more scene information could have been kept.

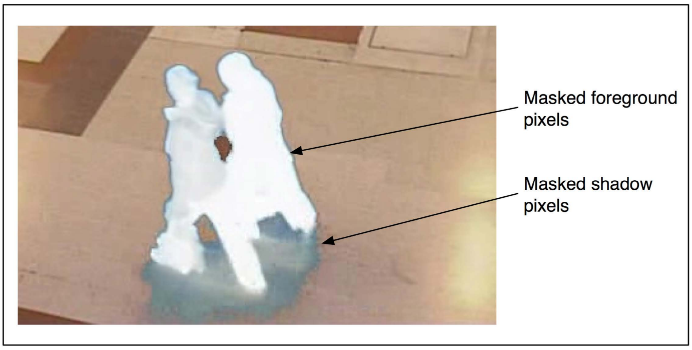


Fig. 16. Two masked individuals with shadows.

4.4.3 Shadow detection with low footprint

Since PEVS represented a complete system that should be deployed in real-world applications and needed to perform in real-time, efficient algorithms were sought. Rather than using complex shadow detection methods, several established methods were combined in order to create a highly accurate and efficient shadow detection method. More specifically, a mixture between methods developed by Horprasert *et al.* (1999) and Nadimi and Bhanu (2004) was implemented.

Video analysis processing queue

Fig. 17 shows the object tracking process used in PEVS. The input to the object tracking module was represented by the current image of a video. It was transferred to the background subtraction module, which separated foreground from background. Further, shadow detection was performed on the image with the background removed. Next, detected image regions were segmented and clustered (see Sections 4.5.2 and 4.5.3). Further, association of segmented objects ("tracking") on the unscrambled image region was performed. Using this object information, those objects that should not be seen were scrambled. This was based on user choice or automatic detection (see Section 4.3).

Scrambling could be performed in several ways, including dividing all pixels to be scrambled into blocks and setting the color value to the average color value of a block to a chosen color or by inverting the pixel values. The original, unmasked, image regions were encrypted and stored locally in internal memory. Next the shadow detection method is discussed.

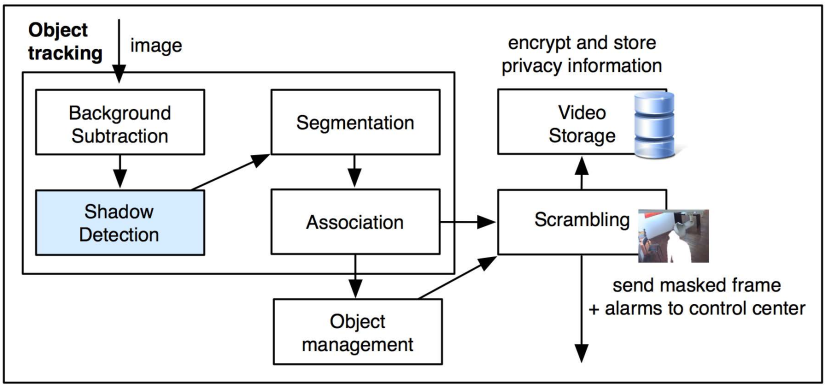


Fig. 17. Object tracking process in PEVS.

Shadow detection and removal

For the implemented shadow detection method a novel combination of shadow detection methods introduced by Horprasert *et al.* (1999) and Nadimi and Bhanu (2004) was used. The algorithm is elaborated upon and every stage of the shadow detection process is explained in the following sections, while the shadow detection pipeline is shown in Fig. 18.

Background subtraction

As Fig. 18 shows, input to the shadow detection pipeline is provided in the form of an original and a binary image, resulting from background subtraction. In this first stage, motion in the image was detected. The output of this stage was a binary mask with moving pixels marked. The result of this stage could be used to reduce shadow detection to just areas that are detected as motion and that therefore could cause problems later on. Further, this pre-selection of image regions reduced computational complexity and increased the performance of the overall system.

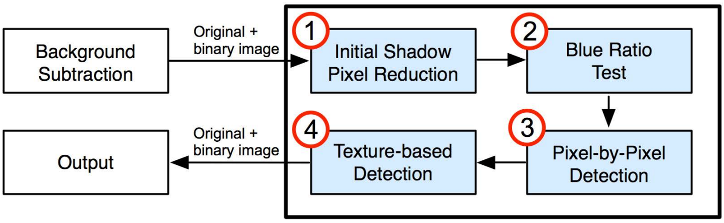


Fig. 18. Shadow detection pipeline.

The background subtraction implementation was based on the method “Mixture of Gaussians” proposed by Stauffer and Grimson (1999). This method was chosen because of its ability to model background motion, thus being robust against repeating background motion such as waving trees or flags. This way, false positives during motion detection could be reduced and a good model of the foreground on which to perform shadow detection was produced. A drawback of this method was its speed. Since it models each pixel with at least three distributions this algorithm was slower than simple background subtraction. However, accuracy was an important factor and since background subtraction provides the basis for all following processing steps, the input data to the shadow detection pipeline had to be as accurate as possible. This is why processing power was sacrificed in exchange for more accurate results. According to Stauffer and Grimson (1999), this method for motion detection used not only one single Gaussian to represent the background, but a mixture of Gaussians. In this approach multiple Gaussians can represent the background, based on their persistence, variance and a threshold T. T is a measure of how much of the data should be accounted as background (Stauffer & Grimson 1999).

Initial shadow pixel reduction (intensity check)

In the first stage of the shadow detection pipeline (see Fig. 18), pixels that were candidates for shadow pixels were removed. It was assumed that pixels on a detected surface could not be shadows if they have a higher intensity than the actual background. So if a pixel had a higher intensity than the background it was either a highlight, which was also detected as moving foreground, or an actual foreground object but no shadow. These pixels were left out from the object mask and did not have to be checked again.

Blue ratio test

Similar to the previous stage, the blue ratio test (Fig. 18) reduced the number of potential shadow pixels by using a condition on the pixels. It exploits the observation that shadow pixels, which fall on neutral or gray surfaces, such as asphalt roads, tend to be bluer in tone (Nadimi & Bhanu 2004). Therefore, pixels with a bluish chromaticity value were pre-selected to be shadow pixels.

Pixel-by-pixel shadow detection (PbP)

In order to improve the results of the detection, two different detection methods were used simultaneously: pixel-by-pixel (PbP) and texture-based detection (TB). If a pixel was classified by either of the methods as a shadow pixel, it received a confidence rating of 0.5. If both methods classified it as shadow, it accumulated to 1. Only pixels that had a rating of 1 were removed at the end of the process.

In the PbP shadow detection option, a pixel was classified as a shadow if the pixel had similar chromaticity but lower brightness than the background. Since at this stage background subtraction was already performed, the goal was to delete shadow pixels that were incorrectly detected as foreground motion. It was assumed that all pixels, which were no shadow pixels but were detected as foreground pixels were actual pixels of foreground objects. A pixel was classified as a shadow pixel if it had a similar chromaticity but a significantly lower intensity than the background. Because of color variation in the images of the sequence, chromaticity and intensity values could vary within a small range and thus required the introduction of thresholds. If the change of the intensity was over a certain threshold and the change of the chromaticity under a certain threshold, a pixel was classified as a shadow pixel, as shown in the following equation:

$$p_x = \begin{cases} 1 \text{ if } (c_{pre} - c) < T_c \text{ and } (i_{pre} - i) > T_i \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

where:

P_x	Current pixel
c	Current chromaticity value
i	Current intensity value
c_{pre}	Mean of past chromaticity of non-shadow pixels
i_{pre}	Mean of past intensity values of non-shadow pixels

T_c	Chromaticity threshold
T_i	Intensity threshold

If the first condition was met, a shadow pixel was found and it was marked as such. Otherwise the pixel kept its classification.

Texture-based shadow detection (TB)

Because TB detection treated a number of pixels as a patch, noise resulting from camera sensors, bad image compression or wrong motion detection could be avoided. Also, if shadow detection was performed, it was likely that shadows on foreground objects that had a similar color than the background, were also detected as cast shadows and therefore deleted. If a texture patch was used, the pattern of the background and the foreground object had to match in order to result in false shadow detection.

One disadvantage of this method was that it did not work properly on background surfaces that were not highly textured. In typical surveillance scenarios such textures might be a property of asphalt streets, pavements or white walls. Consequently, this method worked well on surfaces such as grass or brick walls, which had many distinctive features. The classification rule was similar as in the pixel-based approach, with the difference that new thresholds were used:

$$S = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} I(p,q) \tag{4}$$

where:

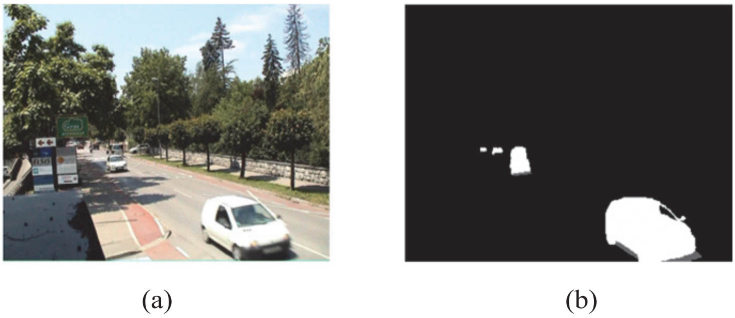
p_x	Current pixel
c	Current chromaticity value
i	Current intensity value
c_{pre}	Mean of past chromaticity of non-shadow pixels
i_{pre}	Mean of past intensity values of non-shadow pixels
T_{tx_c}	Chromaticity threshold
T_{tx_i}	Intensity threshold

This rule was checked against every pixel value in the current texture patch, which was usually a 7×7 window that was moved over the image. If all pixels in the patch met condition (4) the texture-patch was declared to be a shadow patch.

Otherwise it was left as is. The results were blocks in the form of the patch window.

Output

The output of the algorithm was a shadow pixel mask that could easily be removed. This left just the foreground objects of the scene, which in the next step were segmented into objects and subsequently tracked PEVS. Fig. 19 shows an example of output of background subtraction and shadow detection. Fig. 19 (a) shows the original image and Fig. 19 (b) the foreground mask with foreground pixels marked white and shadows pixels marked gray.



(a) (b)

Fig. 19. Example output of background subtraction and shadow detection.

4.5 Image segmentation & object association

This section presents segmentation methods in order to distinguish between different objects as well as association of an object in consecutive frames (tracking). It thus answers research sub-questions four and five (see Fig. 20).

4.5.1 Introduction

Since image segmentation and clustering methods work best using domain knowledge, an approach to image segmentation was sought that could deliver accurate results while minimizing computational complexity. Hence, two segmentation methods with a focus on performance were developed and optimized for different situations. Based on the environment, either one could be switched on or off or both could be used if enough hardware resources were available.

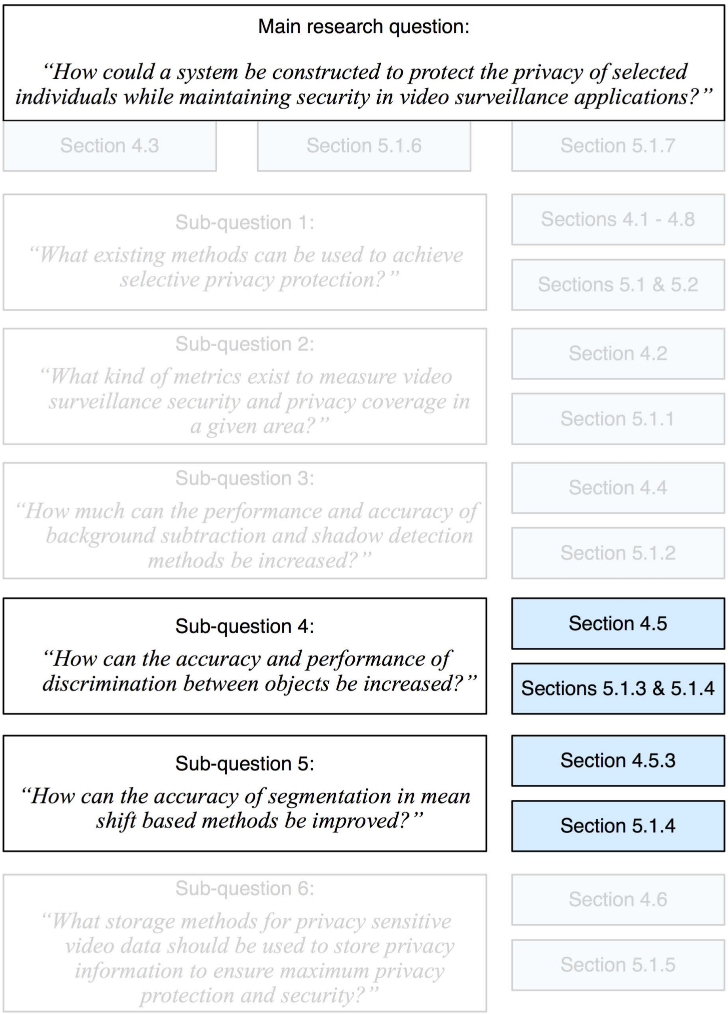

Main research question:
"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"
Section 4.3
Section 5.1.6
Section 5.1.7
Sub-question 1:
"What existing methods can be used to achieve selective privacy protection?"
Sections 4.1 - 4.8
Sections 5.1 & 5.2
Sub-question 2:
"What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?"
Section 4.2
Section 5.1.1
Sub-question 3:
"How much can the performance and accuracy of background subtraction and shadow detection methods be increased?"
Section 4.4
Section 5.1.2
Sub-question 4:
"How can the accuracy and performance of discrimination between objects be increased?"
Section 4.5
Sections 5.1.3 & 5.1.4
Sub-question 5:
"How can the accuracy of segmentation in mean shift based methods be improved?"
Section 4.5.3
Section 5.1.4
Sub-question 6:
"What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?"
Section 4.6
Section 5.1.5

Fig. 20. Placement of this section for the answering of the research question.

4.5.2 Hot-spot blob image segmentation

This section presents the first image segmentation algorithm, which used a block-based method to reduce image resolution, while maintaining all relevant information and in turn down-scaled the problem complexity and processing performance (Kraus *et al.* 2008, Matusek *et al.* 2010).

In this work, the term "block" is used in a very general way and stands for a certain image area. It can range from a single pixel to a square or even rectangular image part containing multiple pixels. The block size should be chosen to be significantly smaller than the expected object size in order to ensure sufficient resolution for analysis and tracking. A block size of 8x8 pixels was found to be the optimal trade-off between loss of resolution and computing performance as determined by empirical tests. Furthermore, the block size was kept constant within the image and during the process. Each block was represented by the data shown in Table 12.

Table 12. Data describing a block.

Parameter	Description
I	The index of the block. It held a unique position within the image similar to an index in a one-dimensional array.
S_b	The state of the block. It influenced the algorithmic behavior and could change throughout the algorithm.
W_b	The weight of the block corresponding to the integrated intensity within the block's area.
A_b	The covered image area in units of blocks. It starts with one and is incremented for blocks with certain states as the blob size increases.
R_b	The reference to another block. It can link one block to another block.

For the unprocessed image, a block started with: S_b = unassigned, $W_b = 0$, $A_b = 1$, R_b = "no reference". This was the preprocessing stage. Throughout the stages of the algorithm S_b could change to one of the following states: *irrelevant*, *relevant*, *assigned*, *center*, *joined center* and *junction*, where all states but relevant ones were possible final states (see Fig. 21). When the background subtraction module presented in Section 4.4 designated a block as background, this block was no longer relevant for the algorithm and thus labeled as irrelevant. On the other hand, if the background model flagged the block as foreground, it was tagged as relevant. Only relevant blocks were considered for further calculations and could either become center blocks if a certain amount of neighboring blocks had the correct state, which was an indication that the location of the block might be part of a new blob within the image, or assigned if the block was in close proximity to another block that belonged to a center.

Furthermore, blocks that connected areas of different assignments would be labeled as junction. Finally, the different parts connected by junctions could be bridged or separated according to certain rules derived from their characteristics forming a bigger blob or splitting blobs into smaller segments.

The algorithm was performed in stages numbered from one to six (see Fig. 12). A preprocessing stage was also introduced, which reset any information contained in the blocks used in a previous frame. This allowed the minimization of allocations, which improved the performance. Fig. 22 shows block states of each of the stages (from left-to-right and top-to-bottom): first, the original image, then the output of the background model, the classification of blocks into relevant (white) and irrelevant (black) blocks, labeling centers (green) and associated (red) blocks, labeling junctions (blue), conversion of centers to joined centers (yellow), cancellation of shadow blocks (gray), the final bounding boxes of the objects.

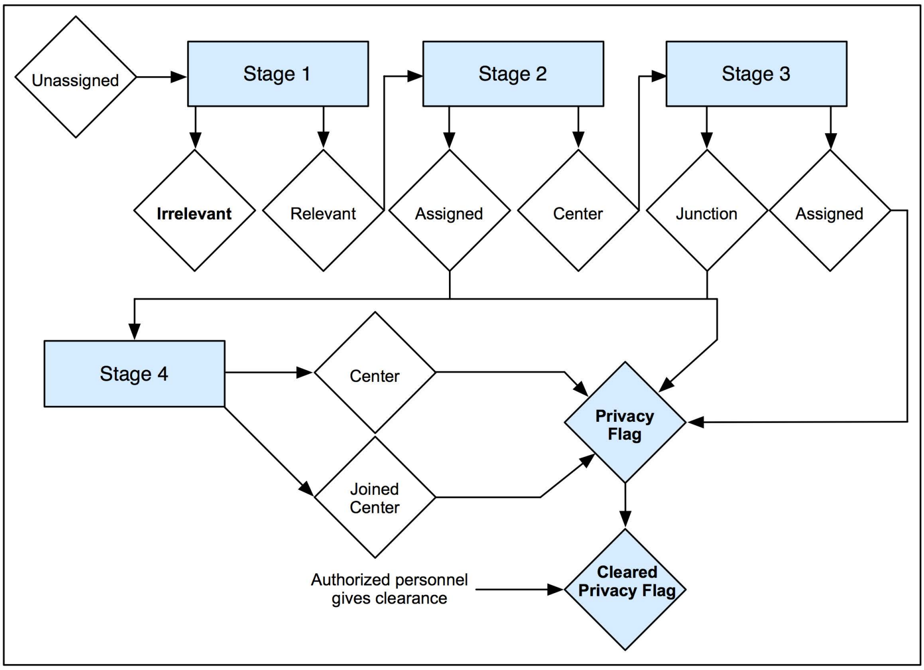


Fig. 21. Algorithmic stages of hot-spot blob image segmentation.

Stage 1

The algorithm started by calculating the integral sum of intensities of all blocks (SoI) deemed relevant by the background model, placing the SoI into the W_b

variable for each block and building a list for these blocks. The list was sorted by W_b where the highest W_b was the first element, the second element was the second highest and so on. If W_b was below a certain threshold t_i a block was completely discarded and set S_b = irrelevant. Therefore, the list only contained blocks with S_b = relevant.

Stage 2

For each block in the list, starting with the first, S_b was checked. If $S_b \neq$ relevant, it meant that the block had already been assigned to a center and did not need to be processed in this step. Otherwise the block was processed and all block states in the neighborhood were checked.

The neighborhood was a possible design parameter of the algorithm and could include only the adjacent blocks (as implemented in the current work) or also blocks further away. Depending on the implementation, the algorithm did an iterative check of how many neighbors were found with S_b = relevant. In the first iteration it checked if there was a block within the image where all neighbors were in relevant state. If this held true, the block was labeled as center (S_b = center) and all neighboring block states were changed to S_b = associated. Furthermore, W_b of every associated neighbor was added to the weight of the center block W_c.

If one or more blocks were found to already be associated, the algorithm proceeded by finding all corresponding centers and associated the current block with the center with the highest W_c. This corresponded to setting S_b = associated, storing the centers address in R_b and adding W_b of the current block to W_c.

Should the first iteration yield no centers at all (and therefore no associations as well) the number of neighbors needed to form a center decreased and the iterative search for centers and associations continued until all blocks were either center or associated.

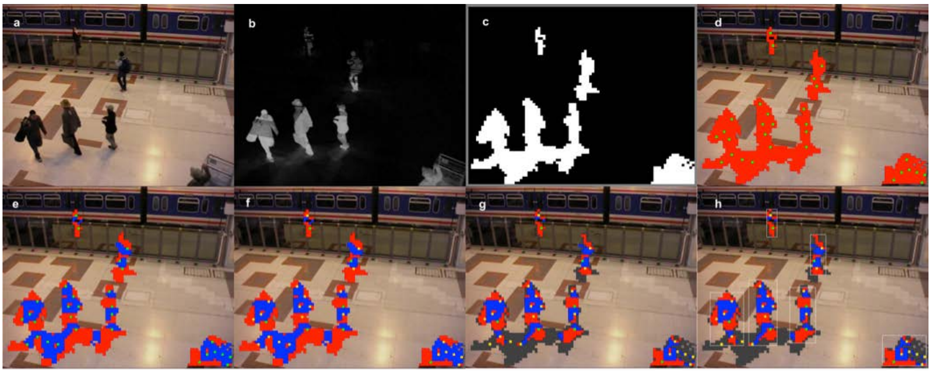


Fig. 22. Block states for the different algorithmic stages.

Stage 3

After labeling and associating the blocks, possible borderlines (junctions) between the regions of different centers had to be found. The list containing the relevant blocks was traversed once more and all blocks that were in a similar state and had one or more blocks with different center references in their neighborhood were marked as junction. In order to manage the weight of the junctions a junction object was introduced. This junction object held references $R_{j,1}$, $R_{j,2}$ to two center blocks, a weight W_j and an area A_j. The junction objects were identified by the two references and stored in a list. If a block was part of a junction, the list of junctions was iterated to find the corresponding object. If no corresponding references were found in the list, a new junction was created with $W_j = 0$ and $A_j = 0$ and appended to the list. In either case the weight and area of the current block was added to the values of the junction.

Stage 4

After finishing the search for blocks being part of a junction, the list of junctions was sorted according to W_j. If the junction was found to be of relevance (e.g. by comparing to a threshold t_j or by analyzing the balance of weights of the two centers with respect to the junction weight), the centers were to be joined. In this case the state of the center with less weight (the weak center) was changed to joined center and the reference was updated to point to the second center (the strong center), which effectively merged the two centers in an efficient way. Now, the final center of a blob could be found simply by traversing the center reference chain from any block until the reference didn't change anymore.

Stage 5

A new object called "blob" was introduced, which essentially held the relevant data of one segmented region within the image. A blob object consisted of the data described in Table 13.

Table 13. Data describing a blob object.

Value	Description
L_{blob}	List of blocks belonging to it (and sharing the same center)
C_{blob}	Final center of the blob
BB_{blob}	Coordinates of the final bounding box (left, right, lower, upper border)
W_{blob}	Total weight of the blob
A_{blob}	The total area of the blob

One last time the list of blocks was traversed to create one blob per center and to store all associated blocks in the reference list.

Stage 6

Due to static occlusions within the scene or object parts with very similar color to the background image, an object could be split into two or more blobs. To avoid this unwanted behavior, a simple model-fitting algorithm based on the shape of a human approximated by a rectangle was implemented. The dimensions of the model were manually calibrated at three distinct positions in the image and interpolated in between for every other position (barycentric interpolation). As the head (or top) regions of the objects were the most stable areas (generally fixed with respect to the object's center and mostly free from shadows) the list of blobs was sorted according to the blobs' y-coordinate starting with the uppermost blobs (low y-coordinate). A rectangular shaped acceptance area was positioned with congruent upper border to the bounding box BB_{blob}. Furthermore, the acceptance area was placed horizontally with an offset to the center of BB_{blob}. The offset depended on the perspective of the scene, which yielded shear/rotation of the objects and the size of the acceptance area. The offset value was calibrated by hand at the left and right border of the scene and linearly interpolated between these positions.

If any other blob had ample overlap with the acceptance area, this blob was joined to the “accepting” blob and then deleted from the list of blobs. Ample overlap was given if k_o percent of the blob’s bounding box was within the acceptance area ($k_o = 50\%$ was chosen in the current implementation).

Postprocessing stage

A final filtering of the remaining blobs was performed. Two strategies were applied that could lead to deletion of a block: first, if the size of blobs was much smaller than the size of a human estimated by the model and second, if an approximated width/height ratio was larger than 1. The more the computed ratio and size differed from the constraints, the lower the confidence rating; candidates with a confidence rating below a threshold t_c were removed. After this stage the remaining blobs could be visualized with a rectangular outline.

The blocks remaining after the postprocessing stage represented the relevant foreground areas. In the best case foreground areas were regions of movement including individuals.

Shadow Elimination

After calculation of the background model most of the shadows were already removed, as presented in Section 4.4. Using the information calculated for this segmentation algorithm, remaining shadows could be removed as well, without additional computational effort. The block density is defined by

$$d_b = W_b / A_b \tag{5}$$

where the area was measured in units of blocks, it was compared to the density of the center

$$d_c = W_c / A_c \tag{6}$$

of every block within a blob. If

$$d_b > d_c \cdot k_d \tag{7}$$

where k_d is a constant factor (0.95 in this implementation), the block’s coordinates were used to update the bounding box to accommodate this block.

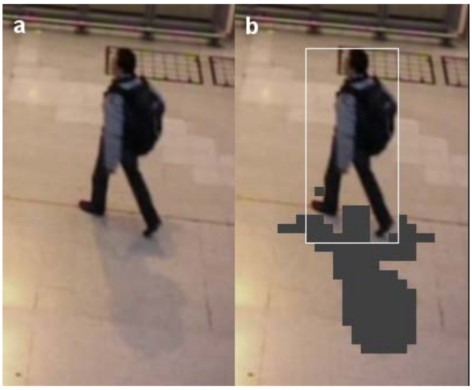


Fig. 23. Shadow detection using already calculated values.

The same applied if the maximum intensity value within the block was higher than d_c divided by the number of pixels of a block. This ensured that those blocks, which hold small but bright details, were not labeled as shadows (e.g. at object borders or within small objects). Fig. 23 shows the effect of this mechanism. Fig. 23 (a) shows the original image while Fig. 23 (b) shows detected shadow pixels (gray blocks) and the blob bounding box (white). This mechanism significantly reduced the remaining shadows while keeping objects with a generally low density in the difference image. After processing all blocks in this way, the bounding box was defined for this blob. This procedure offered the advantage of using the already computed values also needed for the main algorithm, which resulted in an easily implementable and efficient way to detect shadows.

4.5.3 Segmentation using mean shift clustering

This section presents a fast clustering algorithm, which represents the second segmentation algorithm proposed for the PEVS system (Sutor *et al.* 2008c, Matusek 2011).

Introduction

Clustering is a widespread task in pattern recognition and image processing. Mean shift has become one of the most popular clustering algorithms (Fukunaga

& Hostetler 1975). For the special case of video surveillance, a very efficient approach for clustering difference images for detecting and tracking people is proposed by Beleznai *et al.* (2004). This could be accelerated dramatically by performing all calculations on integral images (Viola & Jones 2001). However, this approach limits the mean shift calculation to a uniform kernel, which reduces the flexibility of this algorithm. In this section, exponential integral kernels are introduced which allowed mean shift to be calculated on integral images with weighted non-uniform kernels. This brought the benefit of very efficient calculation and the advantage of weighted clustering eliminating outliers and improving overall robustness.

In this work, mean shift (Yizong & Cheng 1995) was applied to background subtraction described in Section 4.4. This resulted in an input for the algorithm as shown in Fig. 24 (right). These background-subtracted images were then clustered to detect foreground objects, which were to be obfuscated. These corresponded to the brighter pixels in the image while the background remained dark.

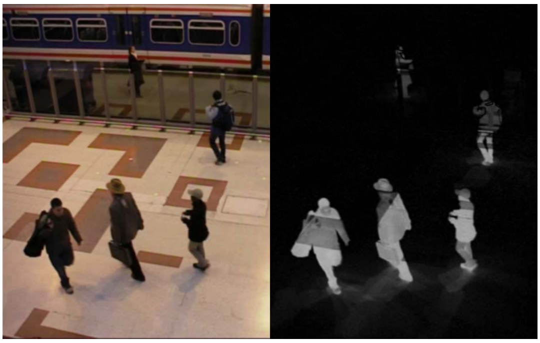


Fig. 24. Original (left) and background-differenced frame (right).

The mean shift clustering procedure on background-differenced images was carried out in four steps: generating seed points, determining the mean shift vector, generating converging path and grouping paths.

Generating seed points

Seed points were generated around local maxima in the difference image. Around every seed point, an area of interest was generated. This area was usually chosen to be rectangular for computational complexity; it could just as well be circular or elliptical according to the chosen clustering algorithm. The size and shape of this area were important tunable parameters, usually set to the approximate size of the object to be detected and tracked. For this area a weight function was defined to give different weights to the pixels in further calculations. This area with its weight function was referred to as a “kernel”. In the case that all weights were the same (constant) this was called a uniform kernel, otherwise a non-uniform kernel. Note that in this work the term kernel is used as a term for a weighting function that can be applied for mean shift calculation.

Determining the mean shift vector

On every kernel area a vector pointing towards the highest density point was determined. This point corresponded to the brightest spot with respect to the background-differenced sequences. This vector was called the mean shift vector. In the case of a uniform kernel the mean shift vector pointed towards the center of gravity that was not necessarily inside an area of high density, which could be troublesome in some cases as the following sections show.

Generating converging path

The kernel was set to the point the mean shift vector pointed to and the whole procedure started over until the displacement fell below a certain threshold or a maximum number of iterations. Usually convergence was reached within a few iterations. These consecutive points, starting from the seed point to the point of termination, formed the mean shift convergence path.

Grouping paths

All paths converging towards the same mode were sought and grouped. It was significant to note that in practice displacements of a few pixels might occur due to limited kernel support and rounding errors. The grouped seed points formed the bounding box of an object. Due to the mean shift procedure, holes in the difference image were bridged. This resulted in the clustering of objects that were approximately the size of the kernel.

Summary of the clustering procedure

The choice of kernels might differ according to the task or scene to which the mean shift clustering was applied. A simple uniform kernel was much faster to calculate when using integral images, while a non-uniform kernel, e.g. a Gaussian kernel, was less prone to outliers. However, it was not possible to calculate the mean shift procedure on the efficient integral images.

In the following section a non-uniform integral kernel is presented, which demonstrated the advantages of a weighted kernel and the efficiency of integral image calculations.

Calculation of mean shift using a uniform kernel

Assuming a rectangular region of interest and a uniform kernel, the mean shift vector was calculated by first summing up all pixel values in the region. For a kernel with coordinates (*x,y*) as top-left corner, the sum s was calculated on the Image I as

$$s = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} I(p,q) \tag{8}$$

with w and h as the width and height of this area respectively. Further, the x-weighted area sum s_x was calculated as

$$s_x = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} x \cdot I(p,q) \tag{9}$$

and the y-weighted area sum s_y was calculated as

$$s_y = \sum_{p=x}^{x+w} \sum_{q=y}^{y+h} y \cdot I(p,q) \tag{10}$$

The mean shift vector coordinates (x_{new}, y_{new}), which represented the coordinates of the next point in the mean shift convergence path, were thus given by

$$x_{new} = \frac{s_x}{s} \tag{11}$$

$$y_{new} = \frac{s_y}{s} \tag{12}$$

Integral images

Usually, the computationally expensive summations had to be calculated for every single seed point in the image and subsequently needed to be iterated several times. However, this computationally expensive process could be severely accelerated. A speed boost of up to a factor of 30 had been measured (Beleznai *et al.* 2004).

Three images were pre-calculated on the incoming difference image. First, the integral image was calculated as

$$\mathrm{int}(x,y) = \sum_{p=0}^{x}\sum_{q=0}^{y} I(p,q) \tag{13}$$

The corresponding value at (x,y) was the sum of all gray values in the image area $[(0,0);(x,y)]$, which was calculated as

$$s = \mathrm{sum}(ax,ay,bx,by) = \mathrm{int}(bx,by) - \mathrm{int}(ax,by) - \mathrm{int}(bx,ay) + \mathrm{int}(ax,ay) \tag{14}$$

Here, the advantage was that the sum of all pixels in the desired area $[(ax,ay)–(bx,by)]$ was simply calculated by four additions and subtractions. Fig. 25 exemplifies this. The *x*- and *y*-weighted integral images were calculated the same way, for the *x*-weighted area sum int_x:

$$\mathrm{int}_x(x,y) = \sum_{p=0}^{x}\sum_{q=0}^{y} p \cdot I(p,q) \tag{15}$$

Similarly for the y-weighted area sum int_y:

$$\mathrm{int}_y(x,y) = \sum_{p=0}^{x}\sum_{q=0}^{y} q \cdot I(p,q) \tag{16}$$

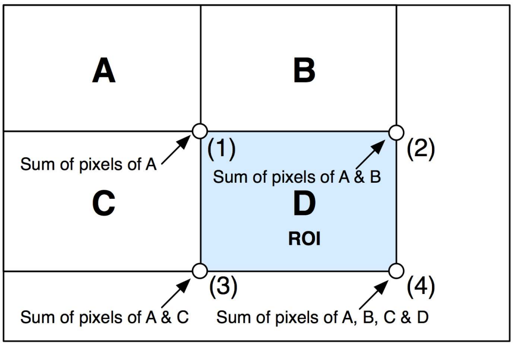


Fig. 25. The area sum of D, computed as: (4) + (1) – (2) – (3).

Non-uniform integral kernels

Non-uniform kernels weighted pixel values depending on their location, i.e. giving pixels closer to the kernel center a higher weight than pixels further away. This was useful since pixels in the center of an object had a higher probability of belonging to the object itself. Nevertheless, it was still desirable to use the efficient integral images as a data structure. Accordingly, the following problem occurred: when considering a uniform kernel, pixel values were not weighted (or merely weighted by a constant factor). However, if a higher weight closer to the kernel center was desired, it would be possible to split the kernel in half and weigh each side linearly as depicted in Fig. 26. Note that this is only discussed for the vertical case, the horizontal case was analogous.

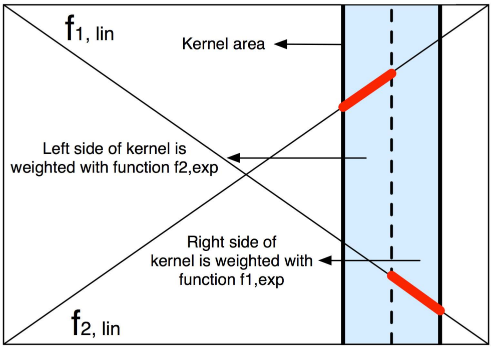

f1, lin
Kernel area
Left side of kernel is weighted with function f2,exp
Right side of kernel is weighted with function f1,exp
f2, lin

Fig. 26. Constructing a linear kernel resulting in asymmetrical weights.

Inside the kernel, the pixel values on the left side were weighted with a linear monotonous growing function $f_{2,\mathrm{lin}}$; The pixel values on the right side were weighted with a linear monotonous falling function $f_{1,\mathrm{lin}}$. Constructing the linear weight function W(p) from these two functions, weighting became asymmetrical. The following equations show this problem for W(p) as a linear function:

$$
\begin{aligned}
&x_{\mathrm{new}} = \frac{S_x}{S} = \frac{\sum\limits_{p=x}^{x+w}\sum\limits_{q=y}^{y+h} y \cdot I_{\mathrm{weighted}}(p,q)}{\sum\limits_{p=x}^{x+w}\sum\limits_{q=y}^{y+h} I_{\mathrm{weighted}}(p,q)} = \\
&\frac{\sum\limits_{p=x}^{x+w/2}\sum\limits_{q=y}^{y+h/2} y \cdot I_{\mathrm{weighted\_left}}(p,q)}{\sum\limits_{p=x}^{x+w/2}\sum\limits_{q=y}^{y+h/2} I_{\mathrm{weighted\_left}}(p,q)} + \frac{\sum\limits_{p=x+w/2}^{x+w}\sum\limits_{q=y+h/2}^{y+h} y \cdot I_{\mathrm{weighted\_right}}(p,q)}{\sum\limits_{p=x+w/2}^{x+b}\sum\limits_{q=y+h/2}^{y+h} I_{\mathrm{weighted\_right}}(p,q)}
\end{aligned}
\qquad (17)
$$

$$
\begin{aligned}
I_{\mathrm{weighted\_left}}(p,q) &= I(p,q) \cdot W_{\mathrm{left}}(p) \\
I_{\mathrm{weighted\_right}}(p,q) &= I(p,q) \cdot W_{\mathrm{right}}(p) \\
W_{\mathrm{left}}(p) &= p \\
W_{\mathrm{right}}(p) &= B - p
\end{aligned}
\qquad (18)
$$

When solving these equations, it could be seen that the weights were not symmetrical. Hence, linear weighting functions could not be applied.

The exponential integral kernel

A self-similar kernel function needed to be found in order to avoid the symmetry problem that was shown when using linear functions. This was a function that fulfilled

$$W(p) = C \cdot W(p + a) \qquad (19)$$

where C was a constant factor that would be averaged out and a was constant shift in *x*-direction. The group of functions that fulfilled this condition was the class of exponential functions as the following shows:

$$\begin{aligned}
&W(p) = q^{x} \\
&q^{x} = C \cdot q^{x+a} \\
&q^{x} = C \cdot q^{x} \cdot q^{a} \\
&C = \frac{1}{q^{a}}
\end{aligned} \tag{20}$$

Hence, every exponential function, as exemplified in Fig. 27 was suitable for constructing a weighting function that could be applied to integral images to calculate the mean shift vector, i.e. an integral kernel. The weighting function was constructed of two exponential functions $f_{1,exp}$ and $f_{2,exp}$, which had the property of being self-similar and hence guaranteed a symmetrical weighting relative to the kernel center.

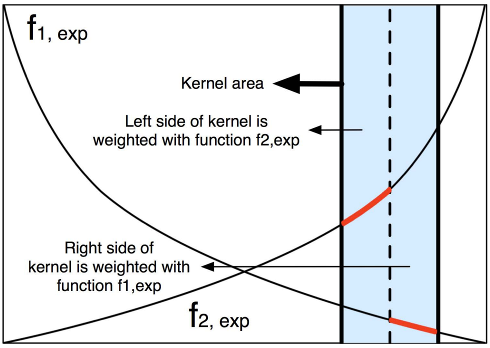

f1, exp
Kernel area
Left side of kernel is weighted with function f2,exp
Right side of kernel is weighted with function f1,exp
f2, exp

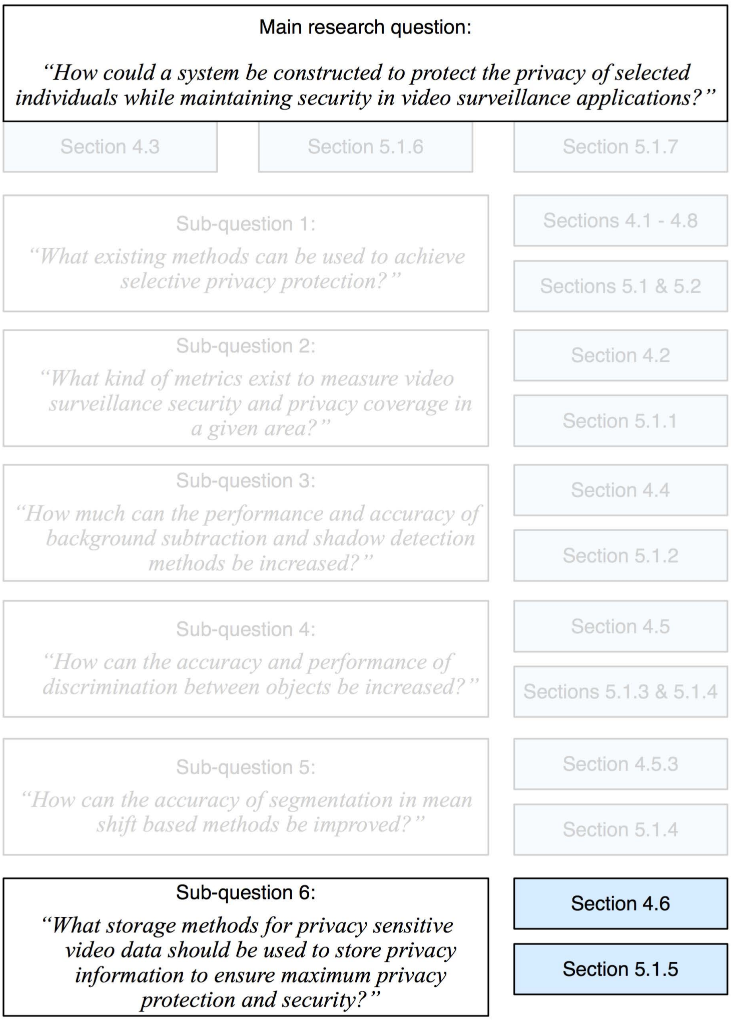
Fig. 27. A weighting function constructed from two exponential functions.

4.5.4 Object association

After objects had successfully been segmented and detected, in order to track them over a period of time, they had to be associated with objects from previous frames. By finding a match between an object in the current frame with an object in the previous frame, the tracking of the trajectory of this object through a video stream could be achieved. There are a number of possibilities to associate objects ranging from linear matching to Kalman predictions (Kalman 1960). Object association was not part of the research presented in this thesis. However, there is a multitude of excellent works on object association available within the current knowledge base (Schulz *et al.* 2001, Piva *et al.* 2005, Xing *et al.* 2009).

4.6 Storage of identity information

This section proposes and analyzes different storage techniques for identity information in the PEVS system. Further, three different approaches regarding the place privacy relevant data was stored, are presented. This section answers research sub-question six (see Fig. 28).



Main research question:

“How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?”

Section 4.3

Section 5.1.6

Section 5.1.7

Sub-question 1:

“What existing methods can be used to achieve selective privacy protection?”

Sections 4.1 - 4.8

Sections 5.1 & 5.2

Sub-question 2:

“What kind of metrics exist to measure video surveillance security and privacy coverage in a given area?”

Section 4.2

Section 5.1.1

Sub-question 3:

“How much can the performance and accuracy of background subtraction and shadow detection methods be increased?”

Section 4.4

Section 5.1.2

Sub-question 4:

“How can the accuracy and performance of discrimination between objects be increased?”

Section 4.5

Sections 5.1.3 & 5.1.4

Sub-question 5:

“How can the accuracy of segmentation in mean shift based methods be improved?”

Section 4.5.3

Section 5.1.4

Sub-question 6:

“What storage methods for privacy sensitive video data should be used to store privacy information to ensure maximum privacy protection and security?”

Section 4.6

Section 5.1.5

Fig. 28. Placement of this section for the answering of the research question.

4.6.1 Introduction

When privacy relevant information was scrambled in a video stream and it was desired to keep the original video information for later use, storage issues of this sensitive information had to be considered. When it came to storage of private information, such as the face or image of a person, there was a tradeoff between high security, storage capacity and performance. This section presents different storage techniques that were developed with respect to these parameters (Matusek & Reda 2008). Furthermore, three different approaches were proposed regarding to where privacy relevant data was stored. Each one of these methods provided a different level of security depending on the hardware to be used.

4.6.2 Levels of privacy

In order to increase privacy of individuals in video surveillance, one could delete or mask privacy relevant data. There were two techniques for storing identity information with a different level of privacy protection. The more the rights of the individual were protected, the less efficient the storage process became. One had to find the optimal tradeoff between these two reciprocally proportional parameters. First, masking the face of a person made it hard to identify while still preserving information about the movement. However, in this case a person could be identified by the color and nature of their clothes. Second, masking the whole body of a person made identification impossible. However, this resulted in decreased security since some movements could not be observed. Different ways of masking could also provide a different level of privacy. As shown in Section 2.3, ways of masking include scrambling (Dufaux & Ebrahimi 2006), de-identifying (Newton *et al.* 2005) and pixelization (Kitahara *et al.* 2004). While the methods of masking did not influence storage consumption, they influenced CPU usage. The amount of masking performed on a person, however, had a direct effect on storage consumption. As shown in 4.6.4 and 5.1.5, these effects could be extensive.

4.6.3 Storage techniques

Three storage techniques were proposed to ensure secure storage of privacy relevant data (S1 to S3).

Masked and original frame storage (S1)

This technique stored two complete frames, the original and the frame with the individuals masked. This resulted in better performance since the frame did not need to be re-encoded but also in higher storage consumption (see Section 5.1.5). Fig. 29 shows this setup. If the user interface requested a masked frame, the application server fetched the complete frame from storage and did not have to re-encode it. In this case access control to the database had to be of very high quality, as the original frame stored an identity in the legal sense, which is subject to a great deal of protection both on a European and on a national level. The proof of a water-tight life-cycle management for the original frames was the basic prerequisite for any implementation. The advantage of this technique was that no re-encoding was necessary however storage consumption was high since two full video streams were stored.

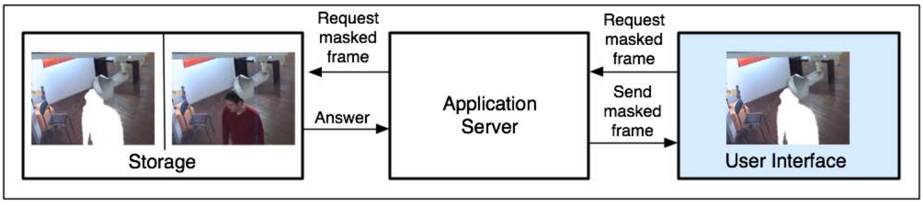


Fig. 29. Masked and original frame storage.

Metadata and frame storage (S2)

This technique stored metadata about the privacy relevant area in an XML format. Additionally, the original frame was stored. However, the metadata could only contain geometric shapes, which approximate the shape of the person. This results in a larger masked area than necessary. Theoretically, a pixel-based description could be done in XML as well, however in this case there is no storage benefit compared to a pixel description. As shown in Fig. 30, the frame was masked in the application server. If the user interface requested a masked frame, the application server fetched the original frame and the metadata and re-encoded the frame to create a masked frame.

The same issues as in S1 for the original frames applied. In addition the XML metadata needed to be especially protected, because otherwise it would have been possible to manipulate the geometric shapes to uncover the individual that was monitored. The advantage of this technique was that it required low storage consumption but it masked more image area than necessary since only geometric shapes were used to describe a private image area.

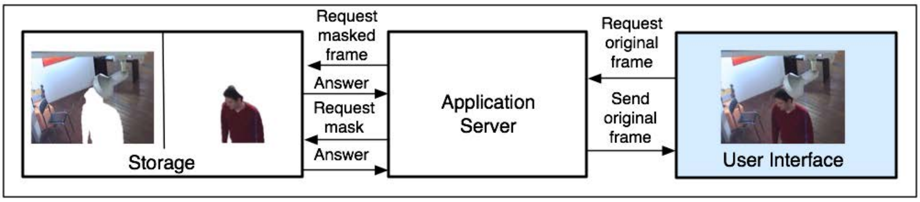


Fig. 30. Metadata and frame storage.

Mask and masked frame storage (S3)

This technique stored the masked frame and the mask separately (see Fig. 31). If the user interface requested a masked frame, the application server fetched the frame and sent it without re-encoding to the user interface. If the user interface requested the original frame, the application server fetched the masked frame and the mask and combined them to re-create the original frame. In this case, re-encoding was necessary. The mask should not be externally reachable at all, which simplifies access control. The advantages of this technique were low storage consumption and higher security since private data could be stored in a different, secure location. The disadvantage was that re-encoding of the video stream was necessary when private data needed to be accessed.

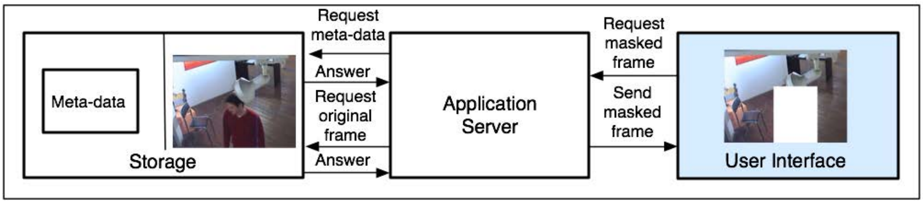


Fig. 31. Mask and masked frame storage.

Summary of storage techniques

Table 14 summarizes the advantages and disadvantages of each storage technique proposed.

Table 14. Advantages and disadvantages of each storage technique.

Storage technique	Advantage	Disadvantage
Masked and Original Frame Storage (S1)	No re-encoding necessary	High storage consumption
Metadata and Frame Storage (S2)	Low storage consumption	Masks more than necessary
Mask and Masked Frame Storage (S3)	Low storage consumption, high security	Re-encoding necessary when private data is accessed

4.6.4 Storage policies

Depending on where the privacy relevant data was stored in the system, different levels of security could be provided. The earlier the privacy relevant data was removed from the frame, the better the privacy protection. Figure 32 shows three possible configurations for storing privacy data. The shaded box in Figure 32 indicates how much of the system and its communication is secure. The set-ups shown in Figure 32 are elaborated below.

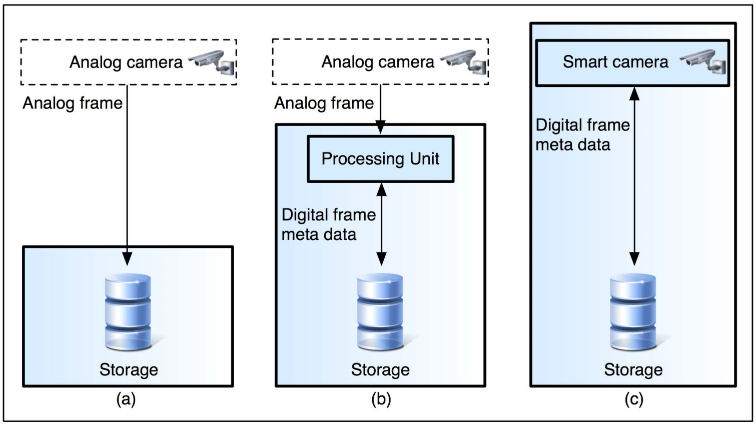


Figure 32. Three options to store information in the system.

Central storage

Figure 32 (a) shows an analog setup where the analog video stream was transmitted to storage. This was a central, server-based storage approach. It could be used if analog cameras were set-up and the video stream was sent directly to a central location where it was digitized and stored. This approach provided the least privacy protection since the video stream was sent including all privacy relevant data (see Table 15). In the central storage privacy data was cut out and saved separately according to the techniques outlined in 4.6.3.

A hot standby database in the form of a geo-redundant backup was necessary. The questions of how these identical copies were synchronized and where the change of transport encryption to storage encryption takes place were two examples of the evaluation one had to make in order to determine the actual security risks within the system.

Hybrid storage

Figure 32 (b) shows a hybrid approach where privacy data was masked in a processing unit close to the camera. This unit could be a DVR or an embedded box incorporating a DSP, which could perform tasks such as encoding and separating a stream. In this way, privacy relevant data was separated from the rest of the stream close to the camera. However, the communication between the analog camera and the processing unit contained the full video stream and was therefore insecure in a privacy sense (see Table 15). In addition, it meant that the full information was available at the site and thus was vulnerable to a central attack.

Smart camera storage

Figure 32 (c) shows a smart camera approach where privacy data was separated on the camera. This approach provided the highest security, since privacy relevant data did not have to be transmitted over the network and could be stored on the camera directly in a highly encrypted way. Only if identity information were requested with correct authentication, this data would be sent over an encrypted network connection (e.g. using SSL or a VPN tunnel). It required a network camera capable of encoding and separating video streams. If these hardware requirements could be met, this approach was preferable over the others (see Table 15). However, if the security of the smart camera could not be guaranteed, it would be insecure to store data on the camera itself. In this case the complete video stream could be sent over an encrypted connection and be stored centrally as well.

Summary of storage policies

Table 15 shows advantages as well as disadvantages of each storage policy.

Table 15. Advantages and disadvantages of storage policies.

Policy	Advantage	Disadvantage
Central storage	Works on legacy infrastructure	Insecure, little privacy protection
Hybrid storage	Legacy infrastructure privacy protection on site	Insecure comm. between analog camera and DVR
Smart camera storage	Secure overall system, best privacy protection	Smart camera, new infrastructure required

4.7 Smart camera collaboration

When using smart cameras for processing video surveillance data, limited resources have to be considered. In order to increase the performance of a system embedded on smart cameras, two methods, which took the current hardware limitations into account, were proposed (Matusek *et al.* 2009).

4.7.1 Load balancing

In Fig. 33 a proposed performance load balancing method is shown, which made use of the processing power of all available neighboring smart cameras. Once a smart camera was put up, it automatically detected and allocated its neighbors and determined which smart cameras could be used to share computer vision tasks. These tasks included segmentation and object association and all pixel-based, and thus performance-intense, image-processing tasks (e.g. background subtraction). If the required performance of a smart camera CPU exceeded 90% of its nominal performance value, task components would automatically be assigned to the nearest under-loaded smart cameras, thus ensuring a real-time high performance execution without bottlenecks, especially in critical situations.

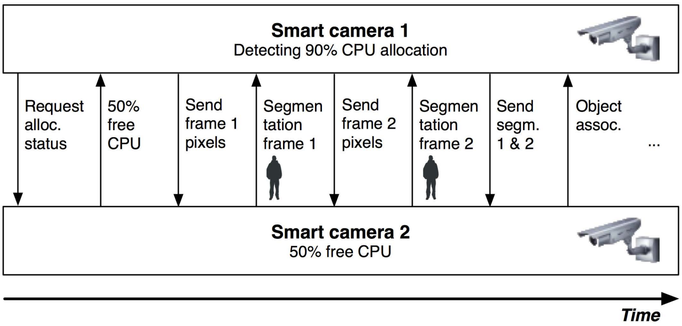


Fig. 33. Performance load balancing between different smart cameras.

4.7.2 Collaboration between smart cameras

Several computer vision tasks could benefit from the available load balancing process between different smart cameras. Two different applications were proposed for this process:

- Multi-camera tracking & detection: Smart cameras could handover information about an object to neighboring smart cameras (e.g. features, color, size, trajectory, outline, pixel values) in order for the next smart camera to find the object again (similar to Fig. 33). This increased the accuracy and performance of the video analysis.
- Fixed smart cameras could send alarms to neighboring smart cameras. This way, movable smart cameras (PTZs) could zoom in to a person, thus delivering high resolution shots of a suspect, allowing the finding of a criminal to be more feasible and increase the chance of success.

4.7.3 Dynamic property adaptation

In order to increase the accuracy of employed computer vision methods and at the same time decrease bandwidth requirements while using all available data, a dynamic property adaptation method was proposed. Using this method, the properties of the smart camera could dynamically be set, depending on the application. As outlined in 4.2.1, where quality levels for different applications were defined, different applications required different settings. Face recognition, for instance, would require high resolution but a low frame rate. The same was true for all recognition tasks. On the other hand, tracking required a low resolution but a high number of frames per second in order to successfully track objects and their motion through a scene. Dynamic property adaptation adapted the properties of each smart camera to correspond to the desired output. This way, for instance, the full resolution of a smart camera could be used while not exceeding available bandwidth resources since the frame rate was reduced at the same time.

4.8 Scrambling and unscrambling of individuals

The final component lacking for the overall system was the scrambling and unscrambling of individuals. This was a straightforward approach in choosing different methods to making identification of a person to be scrambled impossible.

4.8.1 Scrambling methods

There were different possibilities to scramble a defined area in an image. These could be varied according to the privacy protection level desired. These methods were divided into two categories.

Low-level privacy protection

Low-level scrambling included scrambling techniques, which masked all personal information but in a way that information about a person could still be kept. These included dividing the area to be scrambled into blocks and assigning the average color value of each block to it. This way, information about the color of the clothes a person was wearing was kept. Furthermore, “ghosting” could be employed, which again divided the area into blocks but assigned to each block a color value of the corresponding background block but with an intensity value of the foreground image. A further form of scrambling was inverting the pixel values of each pixel (essentially creating a negative) and thus making the person difficult to identify (see Fig. 34 (b)). This was the weakest form of scrambling since all pixel values could be computed back to their original values by merely inverting them again.

High-level privacy protection

In order to provide maximum privacy protection, a minimum of information about a person should be kept. This minimum might be the movement itself, the general shape or the size of a person. One method to provide such protection was to color all foreground pixels in one defined color. This way, no color information about the clothes of a person was kept. However, the way a person moved was still recognizable and thus would provide personal information about the person. A method to avoid this problem was pseudonymisation (see Fig. 34 (c)). This meant that for each person in the image an avatar was chosen and rendered over the area of the person in the image. This was done independent of the person's age, gender or physical appearance. All areas that were not identified as a person were scrambled using one of the other techniques. This way, it was impossible to identify a person by the color of their clothes, the way they moved or by the shape of their bodies. Pseudonymisation was the method with the highest level of privacy protection.

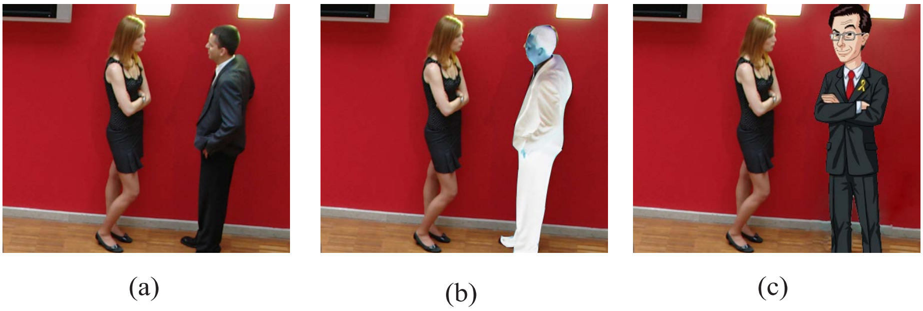


Fig. 34. Original (a), inverted (b) and pseudonymised (c) individuals (simulated image). Avatar © Yahoo! Blog.

4.8.2 Scrambling of movement

By default, all movement in the image was scrambled. The reason for this was that in case a person was not detected by the tracking module, it had to be avoided that this person becomes visible. However, as a result of this, background noise was scrambled as well. In order to scramble all movement in the image by default, the output from background subtraction, as described in Section 4.4, was used, as shown in Fig. 19. Each of the foreground pixels, except marked shadow pixels, was scrambled by the chosen scrambling method. Fig. 35 shows a video surveillance scenario from a live trial on a parking area, where through PEVS all movement, i.e. persons, in the scene was scrambled. The white rectangle represented the result of the segmentation of different persons while the number in the rectangles indicated the object ID, which was tracked through the scene.

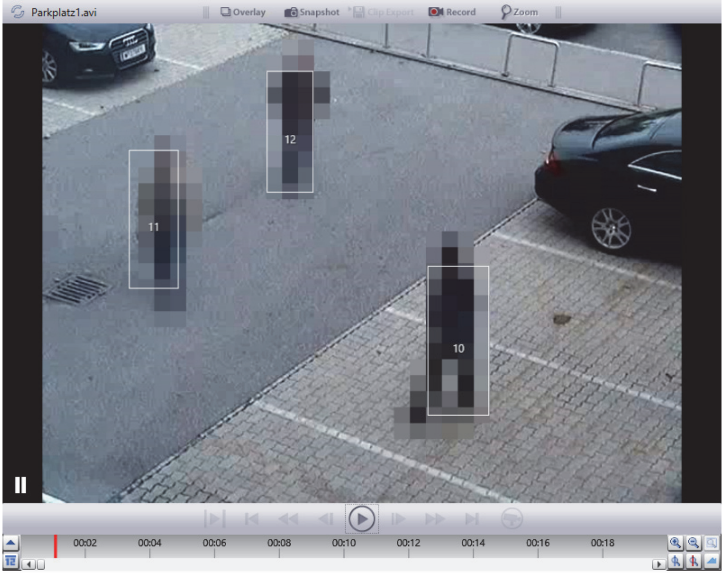


Fig. 35. Video surveillance scenario with all people scrambled.

4.8.3 Unscrambling individuals

If the user, or automatic identification, as explained in Section 4.3.2, chose a specific individual to be unscrambled (typically by clicking on the person in the user interface), the state of the chosen object ID was set to “clear” in object management (see Fig. 12). For each frame processed, the corresponding image area of each object with a “clear” state was exempted from scrambling. This way, all movement in an image was scrambled except marked individuals. This could be performed live or forensically. The typical use case involved people to be unscrambled retrospectively, after a certain event occurred and video data was to be used for an investigation. Fig. 36 shows the result of this process, where one person (object ID “10”) in the scene shown in Fig. 35 was selected to be unscrambled. As can be seen, all people in the scene except object 10 remained scrambled and thus their privacy was protected. This way, privacy is provided on demand (PoD) and can be switched on or off, depending on correct authorization.

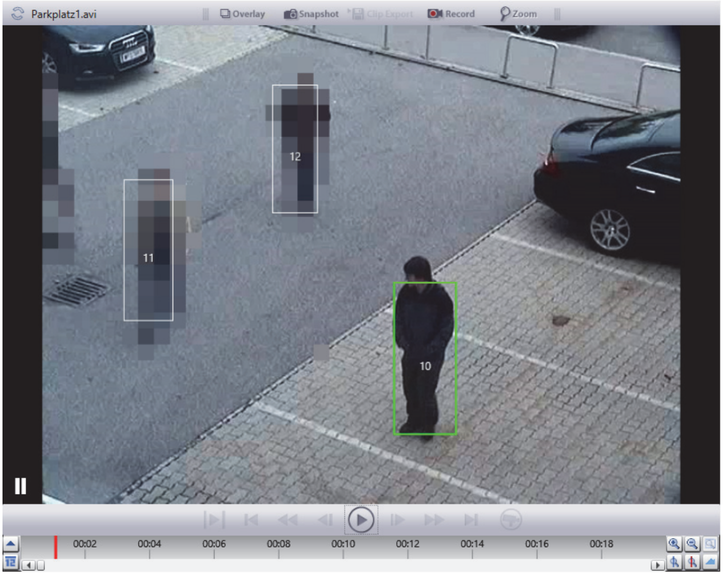


Fig. 36. Video surveillance scenario with one person unscrambled.

5 Evaluation

Research guideline 3 (“design evaluation”) by Hevner *et al.* (2004) states that “the utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.” (p. 83). The evaluation in this work followed this guideline in various ways. First, as Hevner (2007) suggests, each part of the system was evaluated during the build processes in a continuous build-evaluate loop (design cycle). Thus, results of the evaluation could be implemented into the build process immediately and influence the construct in a significant way. Furthermore, the overall system was evaluated when finished and checked against the environment (relevance cycle). This field-testing enabled reviewing whether or not the overall construct matched the requirements of the users, which had been gathered before artifact construction. This cycle further influenced the construct itself, as it could be adapted to the users’ needs. For example, user interface elements could be changed to be more comfortable for the user. The evaluation was divided into two main parts, performance evaluation and survey-based evaluation.

5.1 Performance evaluation

Performance evaluation was aimed at evaluating the performance of the each sub-component and the overall system in order to arrive at a measureable criterion for establishing the contribution of this work. Performance evaluation was performed using two methods of evaluation:

- Empirical: measuring of performance in comparison to established methods and to established ways of using a system.
- Case study: using case studies to show benefits of a system and establishing its use in real-world scenarios, as suggested by Soy (1997).

In the following sections both techniques for evaluation are described first for sub-components of the system and second for the overall system.

5.1.1 Security & privacy metrics case study

In order to evaluate the security and privacy metrics presented in Section 4.2, they were applied in a case study at one of the largest international airports in Europe. The complete system at this airport consisted of more than 2,000 video surveillance cameras. The metrics developed were applied in collaboration with a systems integrator on site to critical areas at the airport to ensure high VSCQ. After the case study, the system integrator was interviewed with open questions to gather feedback on the impact of the quality levels. The airport consisted of different sections where different VSCQ levels and PQL were required. The following sections were considered high security sectors where VSCQ levels were applied:

- Check-in
- Security Check
- Gates
- Restricted Areas
- Parking lots
- Airfield

Accordingly, the redundancy level was kept above r1 within those sections in order to guard against failure.

Check-in area

The check-in area was often highly crowded. This caused people to occlude each other and objects such as pieces of luggage. This caused automatic tracking of people from a single viewpoint to be error prone. Hence, a VSCQ level of at least 3 was necessary; level 4 was desirable to assure that each object could be seen from different viewpoints. Additional G-cameras would further assist computer vision analysis of the cameras streams when the area was crowded. Additional P-cameras could be utilized to automatically generate close-up shots of interesting areas.

In order to fulfill the first dimensioning rule VSCQ3 had to be guaranteed, for the second dimensioning rule VSCQ4 was necessary. This required the use of more than three camera pairs in both horizontal dimensions, which meant VSCQ4. The cameras had to be bound to independent infallible components and resources like networks and power supplies. That meant VSCQ4-r3. A loss of even half of the cameras still resulted in VSCQ3. The ground view camera was redundant: +G-r2. Privacy protection was necessary since this was an area to which anyone had access. This resulted in: VSCQ4-r3+G-r2 / PQL2.

Security check

The security check only required VSCQ1 because people were controlled in lines. Additional G-cameras were added for flow-control, i.e. detecting individuals moving in the wrong direction. Since the security check was a high security area, where security personnel needed to react fast in case of an incident, a minimum of privacy protection was to be used. VSCQ1-r1+G10 / PQL1 was applied.

Gates

As in the check-in area a VSCQ level of at least 3 was required to ensure situational awareness. P-cameras were additionally deployed to allow close-up views. Minimum privacy protection was to be used. This resulted in VSCQ3-r1+25P / PQL1.

Restricted areas

Looking at restricted areas and corridors, multi-coverage might not be necessary because it was sufficient to detect an intrusion where crowding would not occur. However, it was desirable to maintain partial overlapping of camera views, to ensure seamless tracking of individuals. Since this was a non-public area, no privacy protection was necessary. This corresponded to VSCQ1-r1 / PQL0.

Parking lots

In outdoor places, such as parking lots, it was especially desirable to add P-cameras to gain close-up shots of license plates or faces, especially when the distance of the cameras to the objects of interest was quite large. Otherwise, a VSCQ level of at least 3 was kept assuming all cameras cover a longer distance than in indoor-situations. This was a public place that anyone could access. Hence, maximum privacy protection was necessary. This resulted in VSCQ3-r2+10P / PQL2.

Airfield

On the airfield a VSCQ level of 4 was required since the viewing distance was long. Furthermore, additional cameras were added around the perimeters of the airport to allow reliable intrusion detection along the entire airfield as well as P-cameras for zoomed views. Since this was a restricted, high security and non-public area, no privacy protection for working personnel was employed. Here VSCQ4−r3+30E+10P / PQL0 was chosen.

Results

After the VSCQ levels were applied, the system integrator was interviewed on the impact and change of the application. The following questions were asked:

- *Have you used a similar measure for surveillance coverage quality and privacy protection before?*
- *How would you assess the impact of the implementation of the VSCQ levels?*
- *How would you assess the impact of the implementation of the privacy quality levels?*
- *Will you implement the levels in the future in your work?*

The answers to the questions revealed that this kind of metric for video surveillance was not known before the case study and was unheard of in the industry. Privacy protection in video surveillance was not known in general. The impact of both levels was deemed very valuable with three main advantages:

- Reduced time for security planning: due to a better understanding where and what type of camera and how many cameras must be placed already in the planning phase, security planning could be completed much faster. Additionally, it would be easier to discuss security and privacy requirements for an area with the customer already during the planning phase as well.
- Reduced cost: first, due to reduced time for security planning, overall projects would become less costly for the customer. Second, since the amount of cameras for each area is optimized, no unnecessary cameras would be installed. This would again reduce both hardware and labor costs.
- Better protection: both the system integrator as well as the customer could be sure that all defined areas are secured as planned and privacy protection is implemented where needed.

The system integrator stated that he would use the metric in the future in larger projects, where the value of such a metric would be significantly higher than in smaller installations, which are less complex. The application of the VSCQ levels resulted in a better understanding of the camera setup on site. It was valuable for the system integrator to receive measureable values for specific areas on how much protection was currently provided and how this should be changed in the future. It showed that this metric could be an important tool for security planning and system deployment, which eventually would increase security and privacy as well as decrease cost by showing optimal camera usage.

5.1.2 Background subtraction and shadow detection

In Section 4.4 an improved shadow detection method is presented. To evaluate the method, metrics which were proposed by Prati *et al.* (2001) were used. Accordingly, results from the presented method could be compared to shadow detection algorithms evaluated in Prati *et al.* (2001), which use the same input sequences. Prati *et al.* (2001) define two metrics, which show the quality of the shadow detection algorithms, namely shadow detection rate η and shadow discrimination rate ξ . The first one corresponds to minimizing false negatives (FN), i.e. the shadow points classified as non-shadow points. The second rate corresponds to minimizing the false positives (FP) rate, i.e. non-shadow points detected as shadows. Those two rates were calculated as follows:

$$\eta = \frac{\mathrm{TP_S}}{\mathrm{TP_S} + \mathrm{FN_S}}; \varepsilon = \frac{\overline{\mathrm{TP_F}}}{\mathrm{TP_F} + \mathrm{FN_F}} \qquad (21)$$

where TP and FN are true positives and false negatives respectively. The subscript S stands for shadow and F for foreground. $\mathrm{TP_F}$ is the number of ground-truth points of the foreground objects minus the number of points detected as shadows, but belonging to foreground objects.

Results

Figure 37 shows an example frame from the results of pixel-wise shadow detection in a street (scene 1). From left to right, the original frame (a), ground truth (b) and the actual shadow detection result (c) are shown. White pixels represent detected foreground, gray pixels detected shadows and likewise black pixels detected background. Due to high quality input, apart from four spots in the picture, no noise was produced.

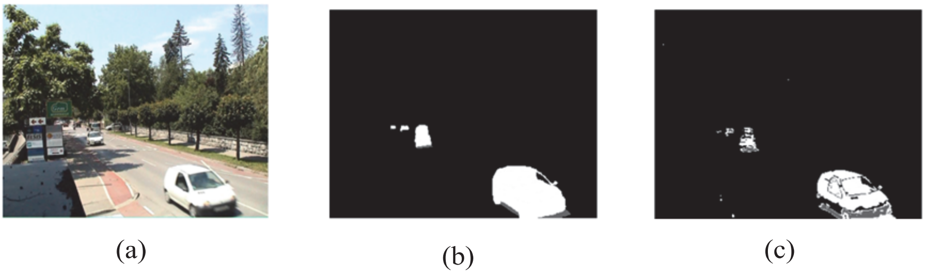


Figure 37. Shadow detection results from scene 1.

In Figure 37 shadows under the car in front were detected, however also parts of the front of the car were detected. This was caused by the dark area at this part of the car that had the same color value as detected shadows. Shadows under the car further back were also detected. Shadow detection rate η and shadow discrimination rate ξ are shown in Table 16 and Fig. 38. Results without post-processing and with post-processing are given.

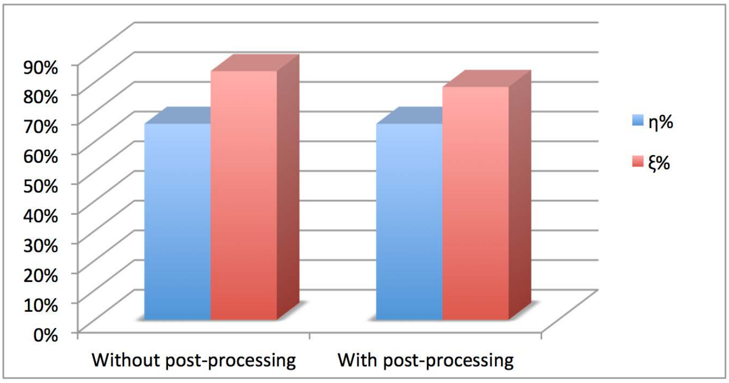


Fig. 38. Results of measurements of scene 1 (street).

In Fig. 39 an example frame of the results of the shadow detection in scene 2 is presented. White pixels represent foreground, gray pixels shadows and black pixels background. The image resolution of the input frames in this scene was low (320×240 pixels). Compression artifacts were the reason that noise was produced. As in scene 1, the original frame (a), ground truth (b) and the shadow detection result (c) are shown. White pixels represent foreground, gray pixel shadows and black pixel background.

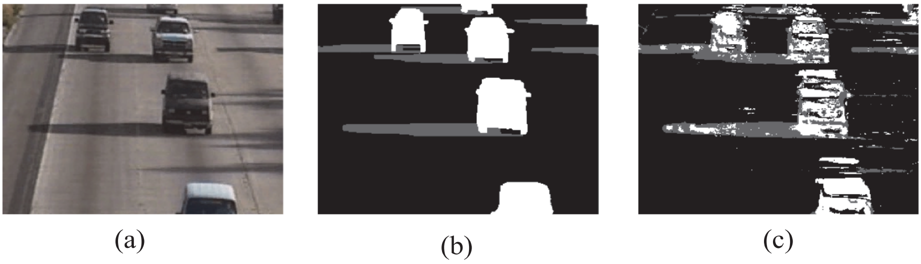


Fig. 39. Shadow detection results from scene 2.

While most parts of the cast shadows were detected, so were parts of the objects. Specifically, dark cars posed a problem due to their color value, which was similar to color values of shadows. In Table 16 and Fig. 40 shadow detection rate η and shadow discrimination rate ξ from scene 2 are shown. Also, results from comparable shadow detection algorithms (SNP, statistical non-parametric, and SP, statistical parametric), which yielded the highest shadow detection and shadow discrimination rate on this scene in Prati *et al.* (2001), are given.

The shadow detection rate without post-processing was slightly lower than with the SNP algorithm (81.07% compared to 81.59%), however the shadow discrimination rate was higher (66% compared to 63.76%). As in scene 1 (see Table 16) the shadow detection rate was higher with post-processing while the shadow discrimination rate decreased. As can be seen with post-processing, the method scored a significantly higher shadow detection rate than both SNP and SP.

Table 16. Shadow detection (η) and discrimination rates (ξ) for the scenes.

Technique	Scene 1		Scene 2	
	η%	ξ%	η%	ξ%
Without post-processing	66%	83.72%	81.07%	65.66%
With post-processing	66%	78.40%	85.06%	60.84%
Statistical non-parametric	n/a	n/a	81.59%	63.76%
Statistical parametric	n/a	n/a	59.59%	84.70%

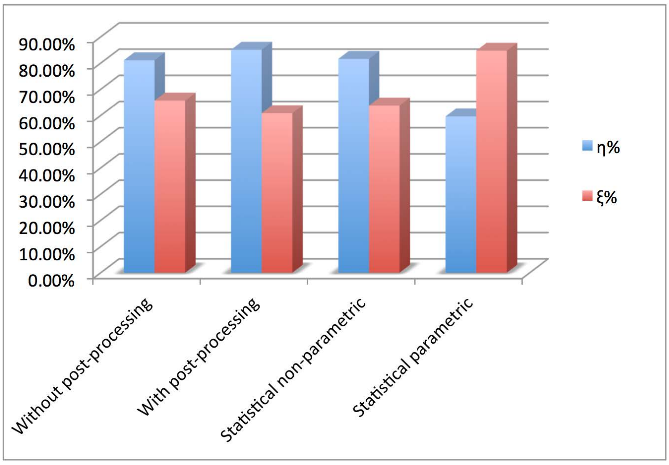


Fig. 40. Results of measurements of scene 2 (highway).

5.1.3 Blob-merging segmentation evaluation

To evaluate the segmentation method using the blob-merging approach presented in Section 4.5.2, the PETS 2006 (Ferryman & Crowley 2006) data set was chosen as it shows typical situations in video surveillance including problems such as shadows, reflections and occlusions. To provide a useful measure of the performance of the algorithm, each single frame was checked by hand for false positives. The checks were performed beginning with frame 349 (initialization of background model ended at this point) until frame 2,224 (see Fig. 41). After tuning the parameters 202 false positives were found in frame 1,875 of the sequence (see Table 17 for details). This corresponded to a true positive rate of 89.2%. Most parameters of the algorithm were not depending on absolute quantities and thus should be relatively independent on the chosen test sequence for achieving best results.

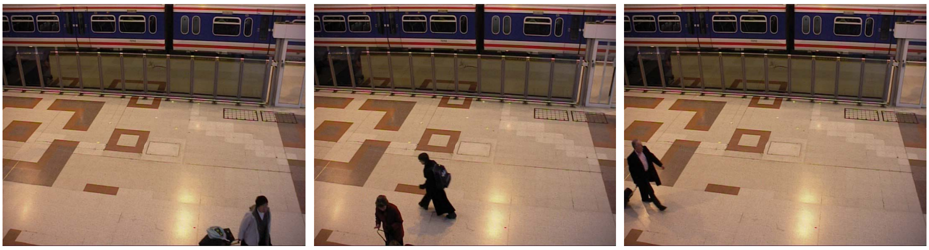


Fig. 41. Frames 349, 1,684 and 2,224 of the PETS sequence.

Table 17 shows these results and the types of errors, which cumulated to 202 false positives. 44% percent of all errors where “object not found” errors, which was attributable to occlusions with static objects in the scene that were in front of relevant objects and covered a large part of them. The “shadow interpreted as object” errors (40%) came from the constraint that low intensity objects were not removed from the scene, as this would lead to more “object not found” errors. Therefore, all shadows that were separated from their originator and were big enough in size were interpreted as objects. This further showed the importance of shadow detection algorithms as presented in Section 4.4.

Table 17. Detailed distribution of errors in the PETS test sequence.

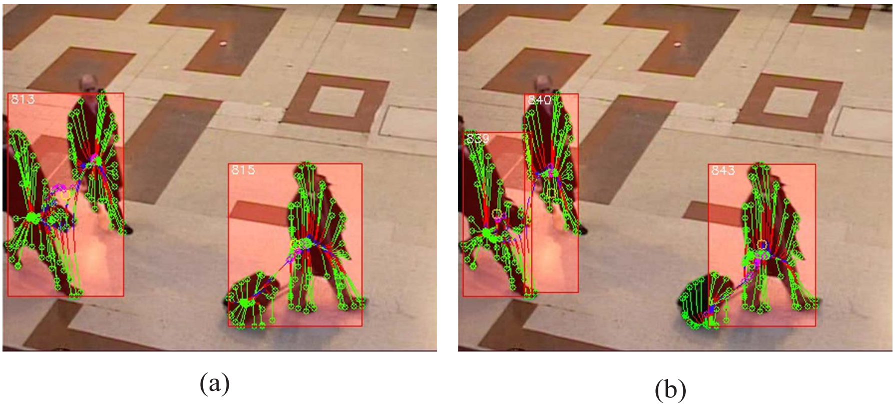
Error	Count	Percent
Object not found (too small)	89	44%
Shadow interpreted as object	80	40%
Split object	31	15%
Object too large	2	1%
TOTAL	202	100%

The “split object” errors (15%) arose from unwanted separations of junctions within an object (often due to low intensity areas in the difference image). On the contrary the “object too large” errors (1%) originated from unwanted bridging to objects. The algorithm needed a maximum computation time of 47.48 ms for about 1,100 relevant blocks (8x8 pixels per block) present in the image with a resolution of 720x576. The computation time of the algorithm for the whole sequence (3,021 frames) was 5.063 seconds, which corresponded to 1.655 milliseconds on average per frame. The tests were performed on a 2.13 GHz Intel Core 2 Duo machine with 1GB of RAM.

5.1.4 Mean shift segmentation evaluation

The proposed method was implemented and applied to various test sequences. In Fig. 42 a snapshot of a single frame of a PETS 2006 sequence is illustrated comparing the presented mean shift clustering on background-differenced images using a uniform kernel and the proposed constructed exponential kernel. The mean shift convergence paths are shown and bounding boxes are superimposed for each separately detected object, hence for each cluster center.

The applied kernels, no matter if weighted or not, needed to be rectangular to make use of the integral image data structure approximating the human contour outliers. These were bound to overlap when humans were in close proximity. This often caused mean shift to converge towards the "wrong" object. This phenomenon can be seen in Fig. 42 (a). The two individuals on the left were merged into one cluster, while they were separated in Fig. 42 (b) where weighting towards the kernel center was applied.



(a)

(b)

Fig. 42. Experimental results for the proposed mean shift procedure.

In some situations a slower convergence towards the detected cluster centers could be observed, however, further calculations and tuning need to be done to fully explore the power of this approach. Finally, due to the constructed kernel, the calculation speed of the weighted kernel in comparison to the uniform kernel was insignificantly higher, but the memory requirement was increased by a factor of four.

5.1.5 Storage techniques case studies

To evaluate the storage techniques presented in Section 4.6, a case study as well as measurements between the different storage techniques proposed was performed.

Case study

The following two case studies were considered before deciding on a concrete case study:

- Case studies that would warrant the storage and retrieval of original frames with all identity information preserved. These would be case studies that had a direct connection to public safety and security of military institutions (e.g. ammunition storage), where recognition of adversaries and reaction to recognized behavior patterns needed to be immediate.
- Second were case studies that did not warrant the storage and retrieval of the original frames. These were commercial case studies like the monitoring of shopping malls, where the storage of complete identities would not be allowed or in public institutions where very private areas (such as toilets) were under surveillance and where it would be sufficient to fetch privacy information only in the event of an incident.

In both types of case studies one of the crucial KPIs would be whether an analysis needs to be done in real-time or not. This would be reflected in the selection of the hardware (in-memory database or not), the storage network type (distributed database or centralized) and the data handling (data preprocessing or not).

In order to test the storage method chosen during the evaluation phase in a real-world scenario, a test setup was installed at an international airport, which represented the first type of case study considered. Since this was a public institution in a high security area it was chosen to test privacy protecting measures. Privacy concerns were especially at hand in areas where people do not wish to be recorded on video (e.g. near toilets). Since high security was also imperative eight smart cameras were chosen. Storage technique 3 (S3) was chosen in order to limit the amount of storage space needed while still masking only privacy relevant data, keeping security to a maximum. The smart cameras sent the privacy protected (masked) video surveillance images to storage while saving the privacy relevant data (the mask) encrypted locally, according to

Figure 32 (c). If not requested, the privacy relevant data was never transferred over the network.

Fig. 43 shows the setup in this case study. Eight smart cameras units were used, which stored privacy information directly on the camera. The smart cameras stored the privacy data on their local storage while sending the masked frames to central storage. Only on request from the application server with appropriate authentication would the cameras send privacy data. When the user interface requested a certain frame it also had to send the authentication data of the user. The application server then decided what the user was allowed to see. If the user had the right to see the original data, it fetched the masked frame from storage and requested the mask from the corresponding camera. The camera sent the mask back to the application server, which in turn stitched masked frame and mask together to send it to the user interface. Using this technique, storage consumption was kept to a minimum and privacy protection and security to a maximum. The test lasted for one week with eight cameras with a resolution of 640x480 pixels, recording 15 frames per second. This amounted to 9,072,000 frames per week and camera and to 72,576,000 frames in total. Storage space amounted to 333GB for the masked frames and only 2GB for the masks.

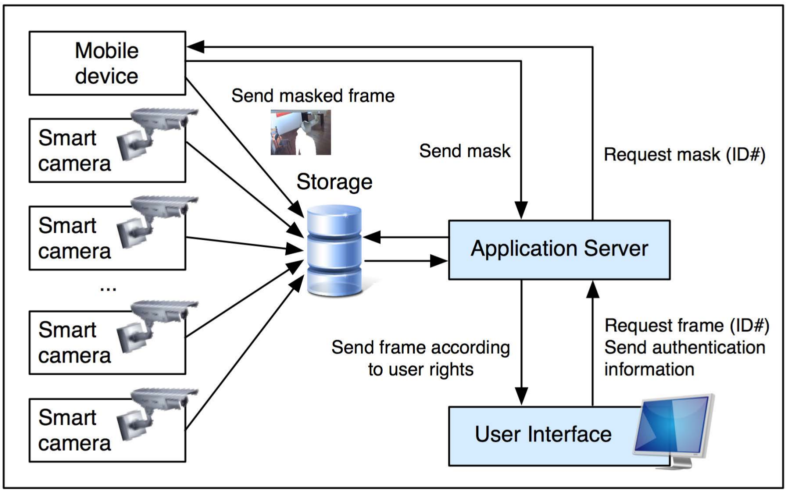


Fig. 43. Setup for the case study at an international airport.

Results of measurements

In order to compare the different storage techniques proposed in Section 4.6.3, performance and storage consumption for each frame were evaluated. Table 18 and Table 19 show the results of encoding in respect to time and storage consumption. Encoding was tested using an Intel Core 2 Duo CPU with 1.86GHz, 1GB of RAM and running Windows XP SP2. A typical video surveillance video was chosen with three to four people in the field of view. As shown in Table 18 the machine took 47 seconds for 1,074 frames, resulting in 44 milliseconds per frame in average. The storage of a video file containing 1,074 frames was 4.90MB, which results in 4.70KB per frame in average. If JPEG sequences were saved, one frame was approximately 105KB (see Table 19).

According to these results, different storage techniques were compared and evaluated. Considering a camera setup where a video surveillance system is recording 24 hours with each storage technique, the CPU and storage consumption shown in Table 20 was obtained.

Table 18. Encoding time.

Parameter	Value
Number of frames	1,074
Encode Time	47 sec.
Avg. time / frame	44 msec.

Table 19. Memory requirements.

Parameter	Value
Compressed file size (MB)	4,9
Amount frames	1,074
Avg. KB / frame	4.7
Avg. single JPG frame (KB)	105

Table 20. Comparison between the three different storage techniques.

Technique	Frames	Time to dec. (s)	Storage (GB)
Masked and Original Frame Storage (S1)	1,296,000	0	11.8
Metadata and Frame Storage (S2)	1,296,000	57	5.9
Mask and Masked Frame Storage (S3)	1,296,000	57	7.9

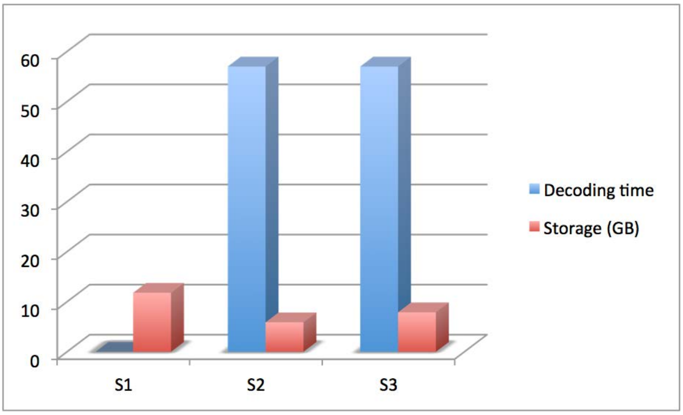


Fig. 44. Results of comparison between different storage techniques.

As can be seen in Table 20 and Fig. 44, storage technique 1 (S1) took up significantly more storage space than the other methods (11.8GB) but required no time for re-encoding. Storage technique 2 (S2) and S3 took up 5.9GB and 7.9GB respectively while required the same time for re-encoding. However, S3 added the extra benefit of masking only the desired parts of the image, thus not ignoring any information and keeping security to a maximum. A trade-off between encoding time and storage therefore had to be made. In a real-world scenario, as can be seen in the case study in the beginning of this section, storage technique 3 (S3) would be the technique of choice because it provided maximum security while protecting privacy.

5.1.6 PEVS architecture empirical evaluation

One of the main goals of the current research of developing a privacy protection system for video surveillance was that it only affects the performance of the overall system minimally compared to state-of-the-art privacy protection systems. In order to evaluate this, performance parameters of the developed overall PEVS system were checked against a state-of-the-art video management system (Milestone XProtect Enterprise Edition) and a state-of-the-art privacy protection system (KiwiVision Privacy Protector 2.2).

System set-up

In order to achieve objective results, all the systems evaluated were tested with the same set-up. Calculation speed was measured in frames per second. The set-up provided the following conditions:

- One FullHD (1080p) video surveillance camera
- One 8-core high performance server
- Video surveillance scenarios, where the number of people in the room varies (one to nine people simultaneously), from the PETS 2007 data set (see Fig. 45)
- The task for the privacy protection systems was to protect the privacy of one VIP in the room
- Calculation speed was measured in frames per second

Fig. 45. Examples of the test sequence used.

Measurements

As can be seen in Table 21 and Fig. 46, performance with PEVS was significantly higher than compared to current privacy protection systems (all movement masked, see Fig. 2). In Fig. 46 the x-axis shows results for different numbers of people in the room while the y-axis gives frames per seconds. The relative performance increased the more people were present in the room, since only VIPs had to be masked. Naturally, some performance was lost due to the tracking process, which was dependent on the amount of individuals to be tracked. However, tracking was performed on feature points and image blocks and masking was performed on a pixel basis. This was why a performance increase could be achieved using tracking to separate only the person that needed to be masked instead of masking all people in the image. Since no calculations except decoding had to be done on the image in the traditional video management system Milestone, performance for this system was linear. Performance for systems without privacy protection would always be better than if calculations would have to be performed. As can be seen, a significant performance increase compared to state-of-the-art privacy protection systems could be achieved using PEVS.

Table 21. Calculation speed of different systems (frames per second).

Number of people	1	2	3	4	5	6	7	8	9
w/o privacy protection (Milestone)	576	576	576	576	576	576	576	576	576
State-of-the-art PP (KiwiVision)	400	180	120	90	72	60	51	45	40
PEVS	380	361	343	326	310	294	279	265	252

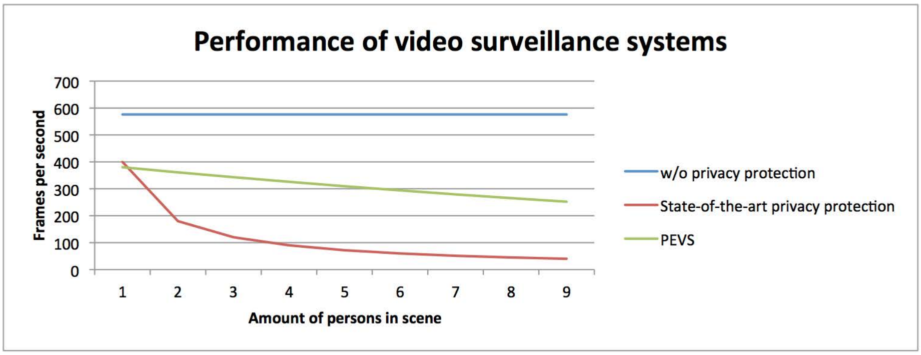


Fig. 46. Results of measurements of different systems.

5.1.7 PEVS for forensic privacy protection

In order to evaluate the usage of PEVS in forensic privacy protection, as discussed in Chapter 6, live measurements of masking video data were taken.

Fig. 47 and Table 22 show measurements of masking people in video frames manually and using PEVS. For each category, the minimum, maximum and average time to mask frames is given. As can be seen, using PEVS the task of masking people in video frames could be sped up by a factor of more than 15 compared to manual masking. To better visualize the results, the measurements with PEVS in Fig. 47 and Table 22 were scaled up with a factor of ten.

These results show that using PEVS for forensically masking video frames, a significant amount of time could be saved by the operator. For masking just one hour of video the time to mask this video could be reduced from 2.1 hours to just 7.8 minutes. This does not only save critical time but is a significant cost saving factor for criminal investigations.

Table 22. Comparison of time to mask frames (PEVS up-scaled by factor 10).

Hours	Max w/o PEVS	Max PEVS	Min w/o PEVS	Min PEVS	AVG w/o PEVS	AVG PEVS
1	4	2.5	0.2	0.14	2.1	1.32
10	40	25	2	1.25	21	13.12
20	80	50	4	2.5	42	26.25
30	120	75	6	3.75	63	39.37

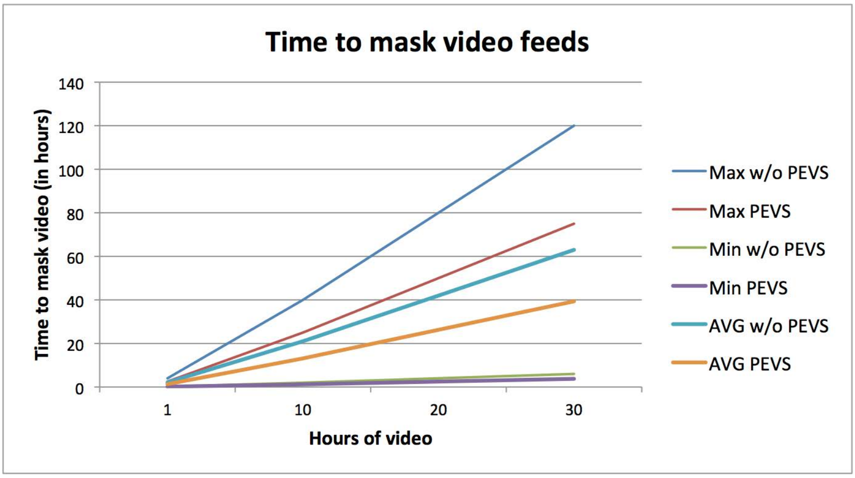


Fig. 47. Comparison of time to mask frames (PEVS up-scaled by factor 10).

5.2 Survey-based evaluation

In order to establish the requirements for the system, 37 people from three different groups (security chiefs, users, employees) were interviewed at the beginning of the research (see Section 3.3.1). After the development of the construct, this user group was interviewed again to investigate how the final result corresponded to expectations and requirements of the users. Of these 37, 31 responded to the survey. While 31 responses do not allow the performance of statistical analysis for quantitative research, it provides a good indication if the goals of the research could be reached from the perspective of the environment, which provided the requirements. The goal of the survey was to assess the subjective feeling towards the system in regard to security and privacy, to find out if the requirements were reached and if improvements to sub-components were necessary. Sub-section 5.2.1 presents questions asked during the survey while in 5.2.2 the results are discussed.

5.2.1 Survey questions

The users were asked to answer the following questions on a Likert-type scale (1: Strongly disagree, 2: Disagree, 3: Neither agree nor disagree, 4: Agree, 5: Strongly agree):

1. *Have your requirements towards a privacy preserving video surveillance system changed since the last survey?*
2. *Do you feel that after the system was installed at your workplace your privacy is better protected than before?*
3. *Can you still provide the same level of security than without the system?*
4. *Has the system increased your workload compared to usage without the system?*
5. *Do you think the relative responsiveness of the overall system compared to other tasks or what is required in critical situations is satisfactory (time to unscramble people, etc.)?*
6. *How do you assess the security of the system: do you think private information is secure in the system?*
7. *How much do you agree with the following statements?*
   7.1. *The system does not unlock people fast enough.*
   7.2. *The system helps me investigate incidents.*
   7.3. *Data in the system can easily be accessed without authorization.*
8. *Do you find all of the functionality needed in the user interface in less than 30 seconds?*

The first question was designed to assess whether the user has changed his requirements towards such a system. If that were the case, the answers would not be very significant since the system would not fit the user's expectations and possible outliers could be detected. The second question was designed to assess the overall satisfaction of the system and whether the goal of protecting the privacy of people was reached in the subjective feeling of the user. The third and fourth questions were designed to assess if the system made the situation for security chiefs and users worse or not. The fifth question dealt with performance, since a real-time requirement was part of the work. The sixth question aimed at getting responses whether subjective security of the stored data could be reached.

The seventh question and sub-questions were designed to check answers to other questions, again to find outliers. The final question dealt with the ease-of-use of the user interface.

5.2.2 Results

Of all the people asked to answer the survey, 31 responded to the questions posed, meaning that six people did not provide any answers. Of these, nine were employees without security responsibility of companies employing video surveillance (29%), 17 were users that would potentially use the system (55%) and 5 were security chiefs (16%). The average answers to the survey for each group are shown in Table 26. While due to the small sample size these results might not represent the wider population, several insights into the opinion of the group surveyed could be concluded. None of the respondents stated that their requirements towards a privacy preserving system have changed (no 4 and 5 responses for question 1). In fact, in talks after the survey, some revealed that the situation at their work place had become worse. Employees referred in this context to more video surveillance cameras that were installed, all of them with no measures to protect privacy. Security chiefs, who mentioned that the situation had become worse, stated this in the sense that they have had difficulties installing new video surveillance cameras due to privacy concerns of employees and thus could not provide the security they thought was required. Hence, the need for a system developed during the research even grew in recent years.

Table 23. Results of the interviews (average answers for each group).

Question / Group	1.	2.	3.	4.	5.	6.	7.1.	7.2.	7.3.	8.
Security chief	1.6	4.2	4.0	1.2	4.2	4.0	2.0	3.2	2.8	4.8
User	1.9	3.7	4.1	1.5	4.4	4.2	1.8	3.5	2.1	4.5
Employee	1.8	4.8	-	-	-	-	-	-	-	-

Further, the subjective feeling that the privacy of individuals is better protected with the system than before (question 2) was higher for employees than for users and security chiefs, while security chiefs scored higher than users. This could be due to the fact that users had direct access to unlock specific individuals in the video but were not as aware regarding the legal restrictions in place as their superiors. Both security chiefs as well as users thought that they could provide the same security as before (question 3), again with a better score for security chiefs.

This effect might be due to added “hassle” for users to deal with unlocking perpetrators instead of identifying them right away while security chiefs, who establish security protocols and security strategy, do not view this as an added risk. In fact, unlocking perpetrators was not considered as added workload (question 4), since this task rarely had to be performed and took up a very small part of the overall workload. Responsiveness of the system was considered to be satisfactory (question 5) for both groups as well. This sentiment further corresponded with responses to question 7.1, which was used as a control question to detect outliers. This was true for all respondents except one, who provided contradicting answers, indicating that answers from this person would not be reliable. However, this person’s answers did not alter the overall average scores.

The overall security of the system and the storage of private information were considered to be satisfactory (question 6) and corresponded with the answers to question 7.3 as well. Answers to question 7.2, asking if the system helped with investigation of incidents, were neither negative nor positive, as it was no goal of the system to further support investigation but to provide privacy protection. This was a positive result, since it was important that the system did not hinder the work of security professionals. This corresponded with answers to questions 3 and 4 as well.

Finally, ease of finding required functionality (question 8) scored very well, suggesting that the user interface provided all necessary functionality in an easy-to-understand way.

Based on the results to the survey presented above for the three different groups of involved parties, one can conclude that the overall goal was reached in the opinion of those surveyed. The system developed indeed satisfied the needs and the requirements of the environment, which were established before development of the system. A further discussion of the overall results of the research is presented in Chapter 6.

6 Discussion and implications

This chapter discusses results that were achieved during building of the construct described in Chapter 4. Furthermore, implications of evaluation results presented in Chapter 5 are elaborated upon. This chapter is structured as follows: first, the purpose of the study and accompanying research results are repeated. Next, theoretical, managerial as well as ethical implications are given. Then, methodological considerations are presented. Finally, an outlook to future works is given.

6.1 Purpose of the study

Video surveillance in public and at office buildings is omnipresent today. While technology in this area has seen constant advancement, for example in the form of automatic analysis of the behavior of individuals under surveillance, solutions for the issue of the protection of personal privacy are lagging behind. However, this issue is important, since it touches the very foundations of our freedom and should not be taken lightly. Consequently, we should not be content with letting our personal freedom and privacy be taken away by technological advancements without searching for alternatives. First attempts were made to formulate regulations on the use of data and data protection. However, few technological solutions, which aim at protecting the privacy of people were developed. This study was based on the belief that technological solutions on the issue of privacy protection could be constructed. The purpose of this thesis was to provide a contribution to this issue and to solve the main research question, which was formulated as follows:

> *"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"*

The requirements of the construct, which led to the main research question, were gathered during a series of interviews with security chiefs, employees of companies employing video surveillance as well as users of video surveillance systems. This study revealed that there was a need for privacy protection in video surveillance but any system employed should not diminish security as well as reaction speed of security personnel at a given site.

6.2 Research results

This section summarizes the results obtained with each sub-component of the construct, which are presented in detail in Chapter 5.

In order to evaluate the security and privacy metrics, proposed in the form of video surveillance coverage quality (VSCQ) and privacy quality levels (PQL) (see Section 4.2) a case study was set up at an international airport (see Section 5.1.1). There, according to requirements of the end customer, the current set-up of cameras was used to set coverage and privacy quality levels for each important area. After this step, the system integrator on site was interviewed on the benefits of these metrics for his work. The results of this interview showed that the metrics did indeed help him in his work. He pointed out three main benefits: reduced time for security planning, reduced cost and better protection for the end customer.

Background subtraction and shadow detection, which were proposed in Section 4.4, were evaluated empirically by performing the algorithm developed on a video sequence used to evaluate other, established shadow detection algorithms and an own video sequence, which showed the benefits of the algorithm (see Section 5.1.2). For each sequence values composed of true positive and false negative detections were used. Results showed that while postprocessing the results could improve the results only slightly, the overall algorithm performed better than previous methods.

The two image segmentation methods proposed, blob-merging (Section 4.5.2) and mean shift (Section 4.5.3), were each evaluated with the same image sequence but using different methods. Blob-merging was evaluated by establishing a ground truth for all correct pixels and the results were measured (see Section 5.1.3). The results showed a true positive rate of 89.2%. Mean shift segmentation was evaluated by testing results of the segmentation using traditional (unweighted) kernels and the proposed weighted kernels (see Section 5.1.4). Segmentation results were represented by rectangular bounding boxes. The results showed that the weighted kernel performed better than unweighted kernels and segmented otherwise merged people correctly.

Storage techniques proposed (Section 4.6) were evaluated in a case study and empirically by measuring their performance (see Section 5.1.5). In the case study, storage technique 3 was used at a customer site and it was presented how the customer used the system. Furthermore, storage requirements and time to decode were measured for each technique. It was shown that a trade-off between time to decode and storage requirements existed and that a decision according to requirements on site had to be made. In most real-world scenarios storage technique 3 would be the most feasible.

The overall performance of the system was evaluated by comparing time to compute to another state-of-the-art privacy protection system as well as to a traditional video surveillance system (see Section 5.1.6). The results showed that performance in the traditional video surveillance system was linear, independent of how many people were present in the scene, since no additional calculations had to be performed. However, compared to the state-of-the-art privacy protection system, PEVS performed significantly better the more people were present in the scene. This benefit was already significant starting with two people in the scene. Further, in order to evaluate managerial implications, the time to mask video frames manually compared to the proposed system was measured (see Section 5.1.7). It was shown that using PEVS, the time to mask one hour of video could be reduced from 2.1 hours to just 7.8 minutes, thus providing significant cost implications.

The system developed could be applied in various use cases and organizations. For example, the system could provide significant benefit in a shopping center. Shopping centers are practically a public area but also the working place of employees. Shoplifting and pickpockets are major concern for shops. Hence, shopping centers might employ video surveillance in order to see if such an incident takes place. However, since employee unions are defending the right to privacy at the working place for their employees, their privacy needs to be protected as well. Using PEVS, both could be achieved. On the one hand the privacy of the employees would be protected and only be lifted for a specific person who triggered an incident. On the other hand, security personnel could still see if an incident took place and could react accordingly.

6.3 Research contributions

The overall contribution of this work to the existing knowledge base was the construction of a novel video surveillance system featuring selective privacy protection, namely the Privacy Enhancing Video Surveillance system PEVS. Using PEVS, the privacy of individuals was protected and the time of security personnel was optimized, with everything performing in real-time. This way, privacy of individuals could be controlled or regulated on demand, depending on proper authorization. This was achieved by developing each sub-component of the system with a focus on robustness, performance and speed. This created a significant contribution to the knowledge base because up until this time, no overall system for privacy protection was constructed or proposed (see Section 6.4). As shown in Table 24, five main contributions could be attributed to this work: a system, which provided privacy protection to selected individuals, the overall system with all its components, the PEVS system architecture, improvements to algorithms of sub-components and novel security and privacy metrics.

Table 24. Summary of main contributions of the presented system.

Contribution	Description
Selective privacy protection	A system was presented that allowed to control privacy protection for individuals. Authorized users could either choose to scramble or unscramble an individual. This had significant managerial implications, since such a system was not employed before.
Complete system	A complete system for privacy protection was presented, which was ready to be used in real-world scenarios. No other technology-based approach provided a complete system with real-world applicability. This was an important contribution with managerial applications as well.
Privacy architecture	A novel architecture for privacy preserving video surveillance was proposed. This architecture could be used as a basis for future research in the area of privacy preserving video surveillance and had theoretical implications as a contribution to the knowledge base of how such systems could be constructed.
Algorithm improvements	For each sub-component of the system, new algorithms or improvements to existing algorithms were proposed which represented theoretical implications and contributed to the existing knowledge base. Improvements were either an increase in performance or in accuracy, such as with shadow detection, segmentation and storage techniques.
Security & privacy metrics	Novel security and privacy metrics were proposed, which allowed system integrators to measure security and privacy at a given video surveillance installation. This had significant implications on speed, cost and ease-of-use of video surveillance deployments as well as on the overall cost and security of video surveillance installations with privacy protection.

Similar to March's approach in March and Storey (2008) the results of this study are outlined in Table 25, including the problem, the artifact that provides the solution and the evaluation method.

Table 25. Results of this study formalized as in March and Storey (2008).

Item	Description
Problem	Video surveillance systems are widely employed but privacy of people under surveillance is hardly considered. Current systems do not adequately protect the privacy of people.
IT Artifact (Solution)	Construction of a video surveillance system, which allowed to protect the privacy of uninvolved persons while providing security.
Evaluation Method	Empirically measuring the performance of sub-components as well as the overall system. Further, describing case studies to show usage in real-applications as well as employing surveys to verify that the system met requirements.

In the knowledge contribution framework proposed by Gregor and Hevner (2013), the overall system presented in this work would be situated in the lower right corner as an exaptation (see Fig. 48).

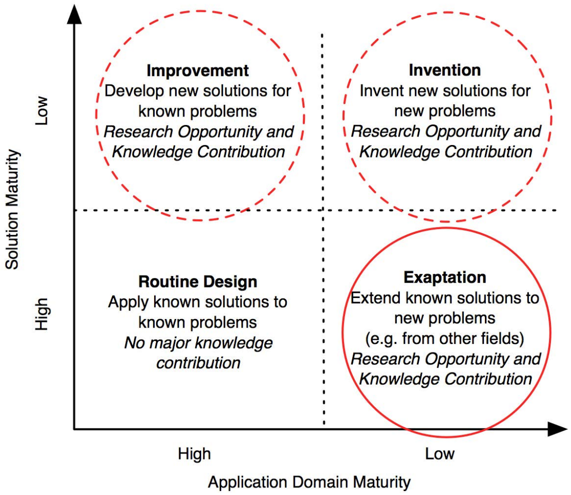

Improvement
Develop new solutions for known problems
Research Opportunity and Knowledge Contribution
Invention
Invent new solutions for new problems
Research Opportunity and Knowledge Contribution
Routine Design
Apply known solutions to known problems
No major knowledge contribution
Exaptation
Extend known solutions to new problems
(e.g. from other fields)
Research Opportunity and Knowledge Contribution
Solution Maturity
Low
High
High
Low
Application Domain Maturity

Fig. 48. Placement of this work in the knowledge contribution framework.

The application domain maturity was quite low yet, with the awareness of privacy issues in video surveillance currently being raised. The solution to the problem consisted of several method types, which were used for other application domains already. However, sub-components of the system might be classified as improvements, since new solutions to known problems were provided. This was, for example, the case for segmentation or shadow detection, where problems in these fields existed for several years, however in this work new solutions were provided, which solved these problems better. Security and privacy metrics could be categorized as invention, since no solutions to the problem were present and the solution itself did not exist in other domains before.

6.4 Theoretical implications

In Chapters 4 and 5 the privacy enhancing video surveillance system PEVS was presented that represented a complete video surveillance system, which protected the privacy of people under surveillance while allowing one or more specific individuals to be unscrambled and thus visible. This kind of system could not be found in the earlier knowledge base and is a significant contribution. Sohn *et al.* (2011) present a privacy preserving video surveillance system, which can scramble faces in the images by using features of JPEG XR. However, contrary to PEVS, Sohn *et al.* (2011) scramble all people in the image; there is no possibility to scramble only individuals, which was the main goal of this study. In PEVS, a user could select, which individuals were to be scrambled and which not. This way, the privacy of uninvolved individuals was protected while suspects could easily be identified. Sohn *et al.* (2011) further use JPEG XR, which is not a common video compression format in video surveillance. Thus, the approach by Sohn *et al.* (2011) cannot be used in real applications at the moment. While PEVS could handle any common video format, the H.264 standard was used. Further, Sohn *et al.* (2011) use face detection for identifying regions of interest. Using face detection is a sub-optimal technique to detect regions of interest from a data protection point of view, since there is no 100% face detection available, leaving the chance of not detecting a face, which then is not scrambled. Further, people can always be identified by the clothes they are wearing, which is not taken into account as well.

Different from earlier approaches, PEVS relied on background subtraction with shadow detection. This was a highly reliable method for detecting regions of interest. Selected individuals were further tracked using object tracking. If uncertain, PEVS scrambled rather too many pixels than too little in order to minimize the possibility of accidentally unscrambling a person. Kitahara *et al.* (2004) and Zhang *et al.* (2010) use face detection as well (detection using multiple cameras or detection of elliptical shapes, respectively), which poses the same issues as discussed above. Korshunov and Ebrahimi (2013) use face detection to detect regions of interest and use warping to transform the image.

Additional to the aforementioned issues, using a scrambling technique, which is reversible, such as warping, is questionable from a data protection perspective since there should be no possibility to reverse videos with the privacy of people being protected, otherwise privacy cannot truly be considered protected. Martínez-Ballesté *et al.* (2013) and Boult (2005) face similar issues regarding reversability and detection reliability. Videos scrambled by PEVS were irreversible. This means that if a video stream was sent or exported, e.g. with one person unscrambled, there was no possibility to unscramble the other people. The original, unscrambled video was stored separately in a highly encrypted manner.

Saini *et al.* (2012) take these issues into account and combine face and blob-based detectors for privacy protection. When the algorithm is unsure if a person is present in the room, the whole image is blurred. However, again, Saini *et al.* (2012) provide no possibility to unscramble individuals, which was achieved in PEVS by segmentation and tracking, as presented in Section 4.5. Baaziz *et al.* (2007) further present an approach, which, such as PEVS, uses background subtraction to detect regions of interest. This approach is much more reliable than face detection. Detected image regions are scrambled in order to protect the privacy of individuals. However, as Saini *et al.* (2012), Baaziz *et al.* (2007) do not offer an approach, which would allow to individually unscramble people.

There are several approaches for storing private information next to original video streams. The approach proposed by Paruchuri *et al.* (2013) aims at hiding private information in the original video stream itself. Similar, Peng *et al.* (2013) aim at storing encrypted private data in a H.264 stream. Yabuta *et al.* (2005) present a method that hides privacy information inside a JPEG stream. However, from a data protection perspective, similar to reversible scrambling approaches, it is questionable whether it is wise to store and send private information in the same video stream as original video data. In PEVS, different storage techniques could be chosen, depending on available storage space and processing performance (see Section 4.6.3). Depending on available storage space and processing performance, the optimal storage technique could be chosen in PEVS. However, independent of the storage technique chosen, the location of private data should always be different to the storage location of original video feeds.

Upmanyu *et al.* (2009) present a system framework to protect the privacy of individuals when using untrusted computers. By splitting the image into several parts, the whole image cannot be intercepted and viewed but rather only the correct receiver is able to recompose the complete image. While this ensures that the video cannot be intercepted, all individuals in the image can still be viewed at the receiver's end. No scrambling as in PEVS is employed. In PEVS, a number of scrambling methods were available, depending on the level of privacy desired (see Section 4.8). Further, in PEVS, secure communication was assumed using SSL and private networks for transmission.

Brassil (2005) presents a system that lets mobile phone users decide, which level of privacy they prefer. Similarly, Barhm *et al.* (2011) present a system where users can set their individual level of desired privacy. This enables each person to set their specific privacy preferences. However, this method requires users to use a specific program on their mobile phones. Thus, while an interesting concept, this system is not applicable in real-world scenarios. PEVS allowed to selectively unscramble individuals, thus to set their "privacy settings". No interaction by the individuals themselves and no special device were required. With PEVS, the decision if a person was scrambled or not depended on authorized individuals. Usually, this would be a security chief, however in a setup, where four-eye authentication was required the decision might be made by a workers union as well.

Chen *et al.* (2007) present a system for hospitals that could mask individuals and unmask others based on a training set. However, with this approach manual training of the algorithms is necessary. This makes the system only usable in restricted cases where the system can be trained by humans, for example in a hospital scenario where staff does not change regularly. Wickramasuriya *et al.* (2004) propose a similar system based on standard sensor technology such as RFIDs. As with the system proposed by Chen *et al.* (2007), it is not applicable in general purpose real-world scenarios. PEVS on the other hand, could be used in any scenario, without requiring a special setup or training of specific individuals. PEVS used standard cameras, which could be set up in virtually any environment and which had to be configured once. It was a general-purpose system, which protected the privacy of any person that did not classify as a suspect.

De-identification techniques as proposed by Newton *et al.* (2005) can furthermore not be used for preserving the privacy of individuals since, additional to issues to face detection as elaborated before, faces are still recognizable by humans. In PEVS, all private information of a person was scrambled, not only facial information. Each scrambling technique hid private information so it could not be recognized by humans.

In order to protect privacy already on smart cameras, several approaches have been published, such as Winkler and Rinner (2010a), Winkler and Rinner (2012) and Chattopadhyay and Boult (2007). Fleck and Straßer (2010) present a distributed system that automatically detects predefined critical events on the smart camera in a geo-referenced world-model and protects the privacy of people by masking them directly in the smart camera. PEVS focused on server-based implementation since hardware limitations had to be considered, which made only simple privacy protection on cameras possible. From a privacy and security viewpoint, encryption of transmitted data between cameras and central servers should be employed in any case.

In conclusion, no approach in the existing knowledge base could be found, which provided a complete system for selective privacy protection (see Table 24).

6.5 Managerial implications

The construction of PEVS did not only have theoretical implications, it did and will have significant managerial implications. The protection of personal space and privacy of employees is an important topic in many organizations. Workers' councils demand that the privacy of employees at their place of employment is protected and are a major hindrance when management would like to employ video surveillance systems. Especially in organizations where employees work in semi-public places, such as shopping malls, shops, jewelers or banks, this can become a major issue. On the one hand, video surveillance is necessary for security, on the other hand it reduces the privacy of employees to a minimum. With PEVS, security could be provided while still protecting the privacy of employees. For example, at a bank everything that was happening was still visible but all individuals were scrambled. If a robbery took place, the robber could individually be unscrambled. Thus, a video feed could be provided to the public by the police to search for the robber, while not intruding into the privacy of employees or other customers. Furthermore, the video surveillance metrics proposed for PEVS were providing better information on the current security situation at an organization and a better tool to plan security more efficiently than before. As shown in Section 5.1.7, PEVS had significant managerial implications when employed to forensically mask individuals compared to manual masking. The time to mask a defined video sequence could be reduced from 2.1 hours to just 7.8 minutes, resulting in significant timesaving, which in turn meant reduced cost. Thus, PEVS had mainly three managerial implications:

- More security: by providing an alternative system with PEVS, which valued the needs of individuals responsible for security as well as the needs of employees, video surveillance could be employed, where it would otherwise not be possible due to resistance of workers' councils or data protection officers. Thus, more video surveillance cameras could cover larger areas and provide more security than without PEVS. Further, video surveillance metrics further increased security at the organization. More security in turn meant reduced loss or shrinkage and larger profits.

- Reduced cost: by providing a tool in the form of video surveillance metrics, security systems could be planned in an optimal way, thus reducing unnecessary components while providing the same level of security. This reduced cost significantly. Further, by enabling system integrators to plan faster and more efficiently, installation and maintenance cost, which were one of the main cost drivers of video surveillance systems, could be reduced significantly, as well. Further, by speeding up the process of masking video frames, also time as well as cost for masking could be reduced.
- Satisfied employees: by providing a means to help protect the privacy of employees, they felt that their needs in the company were valued. This created more satisfaction among employees, which in turn led to increased performance and output in the organization. This again improved profit and the overall success of the company.

6.6 Ethical implications

Privacy and security are topics that inevitably touch upon personal freedom and ethics. How much privacy are we willing to sacrifice for our security? What are the benefits when everything we do is seen and analyzed permanently? Should a government be allowed to permanently watch citizens for the sake of security? These questions were some of the drivers for this work. They triggered the desire of the author to find technological solutions, which could provide a solution to the otherwise diametrically opposite needs of security and privacy.

The case study at the international airport presented in Section 5.1.1 exemplifies that for each area of a site, ethical and privacy issues should be considered. Since the airport in general was considered a high security area with special permission by the data protection body, no specific consent by each visitor needed to be given. However, it was aimed to achieve maximum privacy protection while still providing necessary security. The more secure and restricted an area was, the less privacy was employed. Public areas, which everyone could access, were provided with maximum privacy protection.

With PEVS, a technological solution to the needs of security and privacy could be provided. Using PEVS, the privacy of all innocent persons under surveillance could be protected while still providing the same security as with a common video surveillance system. For the first time, a technology-based solution was available, which gave a little bit of freedom back that was lost to technological advancements. However, PEVS just provided the tool for privacy protection. In order to make it useful it had to be used accordingly and the following questions needed to be answered for each organization: Who is authorized to decide, which individuals should be unscrambled? Are four-eyes principles employed? Who has access to the system in general? If the video surveillance network spans across different countries, how does one deal with the different legal instances of identity definitions? In most countries, this is a very sensitive topic as it touches basic human rights.

In this dissertation ethical considerations were given to the usage of used image material as well. All persons on the images either gave their consent (e.g. if the image material was created for research or marketing purposes) or could not be identified anymore in the image.

6.7 Methodological considerations

In this thesis the research guidelines by Hevner *et al.* (2004) as well as the design science research cycles, proposed by Hevner (2007), were followed. By using both concepts, the quality of the research benefited. Following the research guidelines ensured that the quality of the research was kept at a high level and no important steps in the research were forgotten. Table 26 shows how each guideline was applied in this work.

Guideline 1 (design as an artifact) was followed by building a complete construct (artifact). Guideline 2 (problem relevance) was followed by researching user requirements from the environment before starting the build process (see Section 3.3.1). Guideline 3 (design evaluation) was followed by evaluating each sub-artifact as well as the main construct either by measurement or case study (see Section 5). Guideline 4 (research contributions) was followed by clearly stating research contributions to the existing knowledge base (see Sections 6.3 and 6.4). Guideline 5 (research rigor) was followed by using Hevner’s research guidelines and design science research cycles. Guideline 6 (design as a search process) was followed by using all available information on the topic, including conference proceedings, scientific publications, books and user knowledge.

Finally, guideline 7 (communication of research) was followed by presenting the research at scientific conferences as well as creating presentations for a managerial oriented audience in the form of marketing presentations.

Table 26. How design science research guidelines were applied in this work.

Guideline	How it was applied in this work
1: Design as an artifact	The result of this work was a complete system (artifact) with several sub-artifacts.
2: Problem relevance	Before starting to build the construct, extensive research of the environment was performed in the form of interviews.
3: Design evaluation	Each sub-component as well as the complete system was evaluated and checked against existing methods.
4: Research contributions	Based on prior research and state-of-the-art research, research contributions were clearly stated and discussed.
5: Research rigor	Hevner's design science research cycles as well at the research guidelines were used in order to research with the utmost rigor.
6: Design as a search process	All available material in the present area was used, including conference proceedings, scientific publications, books and user knowledge.
7: Communication of research	Results were presented in scientific publications and at international conferences. Furthermore, they were introduced to a non-scientific audience at business working groups, customer presentations and one radio interview.

Using the research cycles ensured that requirements from the environment as well as foundations from the knowledge base were first researched before starting to build the construct (see Fig. 49). Moreover, in the design process, the design cycle ensured that sub-artifacts were permanently evaluated and as a result of this further improved. In the relevance cycle, developed sub-artifacts were presented to the environment, e.g. in the form of surveys and case studies. The results of these triggered loops in the build process to improve the artifact. Finally, in the rigor cycle, each part was checked with the knowledge base and contributions to the knowledge base were made in the form of scientific publications. The following works were published at scientific conferences during the research: Matusek *et al.* (2008), Matusek and Reda (2008), Sutor *et al.* (2008a), Matusek (2010), Matusek *et al.* (2010), Matusek (2011), Matusek (2012). Each publication was presented at the corresponding conference and feedback was exchanged with the peer group, who gave valuable input to the on-going research. Fig. 49 shows the design science research cycles including what has been done in each cycle during the research as well as the resulting artifacts.

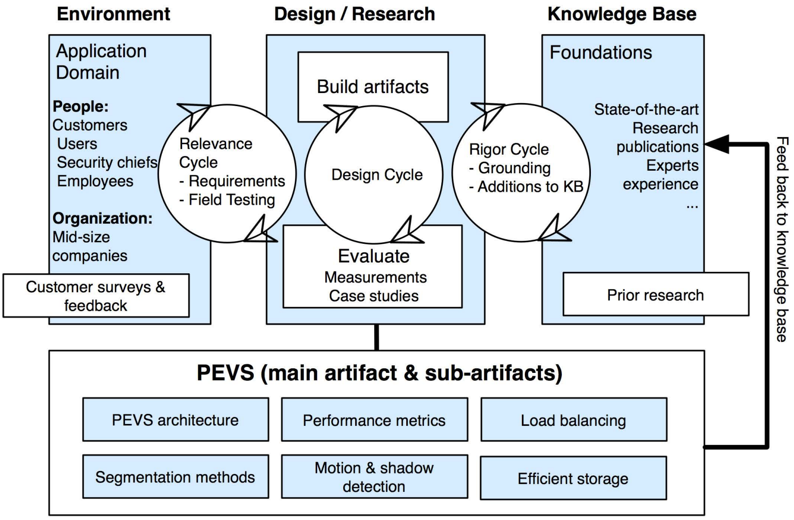


Fig. 49. Design science research cycles applied to research in this work.

Investigation of the prior research showed that no earlier knowledge existed for a system constructed during this research. The results of the research were added to the knowledge base and contributed to the on-going scientific discussion on privacy protection in video surveillance. Other researchers could benefit from this contribution and could use it as a basis for future privacy preserving video surveillance constructs.

6.8 Future outlook

The PEVS system provided a tool to protect the privacy of innocent persons under surveillance and presented a basis for future generations of privacy preserving video surveillance systems. However, a lot has still to be done in this area by future researchers. First, individual algorithms used in the system could be improved. Even though in this work, state-of-the-art methods were improved, tasks such as person tracking are an on-going research field, where significant contributions will be made in the future. Furthermore, PEVS still relied on human input to decide which person was unscrambled and which was not. In future systems, this could be done automatically, either by identifying known individuals and scrambling these, e.g. by RFID chips or face recognition, or by automatically detecting suspicious individuals and unscrambling these.

A topic of future research should be if such systems as mentioned above would still be in accordance with the goal of protecting the privacy of individuals. One would intuitively expect identification systems to decrease the level of privacy in a system. Additionally, legal issues regarding the storage of sensitive information in systems, which span multiple countries, should be investigated. This would be in accordance with issues storing sensitive data in the cloud where it is not always known where the servers are located. In most cases, data protection laws of countries apply where the data is physically stored. Finally, certifications and standards should be defined for privacy preserving video surveillance. With EuroPriSe (Bock 2008) there is already a privacy certificate present in the EU, however, it is a general certificate for IT products and not focused specifically on video surveillance.

7 Conclusions

Video surveillance has become ubiquitous and a part of our everyday lives. While technology, such as the automatic analysis of an individual's behavior, constantly moved forward, little has been done to prevent a further erosion of our personal privacy due to these developments. For a long time, privacy and security were viewed as mutually exclusive factors. One could not have security and privacy at the same time. This study, however, was based on the belief that this sentiment is fundamentally false. The purpose of this study was to find technological solutions to the issue of privacy protection in video surveillance by building a video surveillance system that allows the protection of privacy of innocent individuals while still enabling the identification of criminals and provision of security. After a number of interviews with people at companies that employ video surveillance systems, the main research question that had to be answered in this study was formulated:

> *"How could a system be constructed to protect the privacy of selected individuals while maintaining security in video surveillance applications?"*

For this study, Hevner *et al.* (2004) design science research guidelines as well as Hevner (2007) design science research cycles were followed. Both helped to create meaningful research results that built on the existing knowledge base as well as solved real-world problems posed by the environment. As suggested by Hevner (2007), already during the build process parts of the system were evaluated, so that these results influenced again the build process. Similarly, results were checked against the environment and the knowledge base in several cycles until a satisfying result was found.

After defining the main research question, together with sub-questions, prior research in the area of interest was investigated. First, security and privacy in ICT was researched since it played an important role in ICT systems in general (Buscher *et al.* 2013). This is especially important in the health care sector where sensitive data needs to be handled (Appari & Johnson 2010, Alemán *et al.* 2013, Caine & Hanania 2013). Furthermore, recently security and privacy in the cloud has become a topic of interest (Hamouda 2012).

In order to build a privacy preserving video surveillance system, regions of interest in the image, where identity information should be hidden, have to be identified by background subtraction (Bharti 2013), segmentation (Li *et al.* 2012, Caleiro *et al.* 2013) and object association (Kim *et al.* 2013). Additionally, face detection (Badii & Einig 2012) and recognition (Ahonen *et al.* 2004) could be employed.

Privacy preserving methods in video surveillance are an emerging field (Cavallaro 2007) with different methods for hiding privacy information proposed (Saini *et al.* 2012, Paruchuri *et al.* 2013). Systems, which enable the selective unscrambling of individuals, usually require some sort of identification, such as face recognition or RFID tags (Wickramasuriya *et al.* 2004, Chen *et al.* 2007). Another trend is employing privacy preserving methods directly on smart cameras, such as proposed by Winkler and Rinner (2013).

The study of prior research showed that no complete video surveillance systems have been proposed, which allowed to selectively unscramble selected individuals and which would be build on an architecture taking privacy from the ground up into consideration and enabling real-world applications.

After the study of prior research, the construct was built and evaluated by using case studies, surveys and empirical measurements. The construct was divided into several sub-constructs, which were necessary to build the overall system to be used by customers in a meaningful way. First, security and privacy metrics where proposed, which enabled to measure the level of security and privacy in a given area and allowed easy and efficient security planning for a privacy preserving system. In a case study at a customer's site and a survey with a local system integrator it could be shown that using these metrics, video suveillance installations could be deployed faster at lower cost and to provide better overall security and privacy protection. Next, an overall architecture for the system was designed with a focus on security and privacy by design from the beginning.

Following that, algorithms to identify regions of interest were sought. The focus of these algorithms was accuracy and speed since in order to employ all algorithms at once in one system, calcuation speed had to be several times faster than real-time. The first step included background subtraction and shadow detection. There, accuracy as well as performance improvements compared to previous methods could be achieved. Next, two segmentation methods were proposed. During evaluation, it was shown that both methods performed more accurately and faster than previous methods. Moreover, different techniques for storing privacy sensitive data were proposed. It was shown that a tradeoff between calculation speed and storage consumption had to be made, depening on the concrete use case. In addition, different locations in the system for storage where proposed. For employing the system on smart cameras, where resources are limited, load balancing approaches were presented.

The evaluation of the overall system was done empirically as well as in a survey with the same representatives of the environment, who provided the initial requirements of the system. This evaluation showed, first, that the performance of the overall system was significantly better than previous privacy preserving video surveillance systems. Second, it showed that the time to mask individuals in a video stream could be significantly improved by a factor of over 15 compared to previous, manual methods. The survey with the environment showed that during the course of the research the demand for a such as system even grew and that the system built met the requirements by all three groups surveyed (employees without security responsibility, users of security systems, as well as security chiefs).

This study provided significant contributions to the existing knowledge base in several ways. First, a system was presented that allowed to selectively unscramble specific individuals while all others in the video stream remained scrambled, thus protecting their privacy. This had meaningful managerial implications since overall satisfaction of employees could be increased by protecting their privacy. Second, a complete technology-based privacy protection system was presented, which did not previously exist. Third, an architecture for a privacy preserving video surveillance system was presented, which could be used as a basis for future researchers building such a system. Fourth, new algorithms and improvements to existing ones resulted in higher accuracy and better performance leading to theoretical implications and contributions to the knowledge base. Finally, novel security and privacy metrics were presented that contributed to faster and cost-effective deployment of video surveillance systems with privacy protection.

However, this work left room for improvement. Most importantly, while the accuracy of shadow detection and segmentation could be improved, overall object detection and tracking was still not perfect. In scenes with a multitude of objects to track simultaneously, occlusions and people in close proximity posed challenges. This limitation resulted in a semi-automatic approach, where the user had to tag the relevant individuals if they were lost by the algorithm. Object tracking is an active research field where valuable contributions will still be made. Possibilities to improve results would be by using multiple cameras or additional sensors such as Bluetooth to improve tracking accuracy. This would allow the tracking of individuals over multiple cameras, reducing the amount of user interactions necessary to a minimum.

Another limitation of the system was that it could only be deployed in installations where only static cameras were employed. Background subtraction required a static image. As soon as pan-tilt-zoom cameras would be employed, person tracking would not be possible anymore, thus not allowing the unscrambling of individuals anymore. In order to achieve tracking on pan-tilt-zoom cameras, different tracking methods would have to be employed, based on feature point detection, which would make real-time applications difficult.

Further, legal requirements in different countries were not taken into consideration during implementation of the system. These could become important not only for deploying a privacy preserving system, different legal considerations could be implemented in the system directly, as well. For example, a requirement to use four-eyes authentication for certain countries could be employed. This could be implemented as a “certified privacy mode”, which would take all legal requirements for a specific country into account.

All mentioned limitations leave room for future research. Object tracking is an active research field where improvements will be made in the future. Furthermore, storage and computing of private information is an open research field that gains importance in light of a trend to store more information in datacenters, i.e. the cloud. Special focus should be given to which part of an overall video surveillance system could be employed in the cloud itself and which part should remain on local, protected servers. In this context, legal issues will be important for future investigations and rules and standards should be defined on how private information should be stored in a secure system.

References

Ahonen T, Hadid A & Pietikäinen M (2004) Face recognition with local binary patterns. Lecture Notes in Computer Science of the 8th European Conference on Computer Vision. Prague, Czech Republic**:** 469–481.

Aken JE (2004) Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. Journal of Management Studies 41(2): 219–246.

Alemán JLF, Señor IC, Lozoya PÁO & Toval A (2013) Security and privacy in electronic health records: A systematic literature review. Journal of Biomedical Informatics 46(3): 541–562.

Appari A & Johnson ME (2010) Information security and privacy in healthcare: current state of research. International Journal of Internet and Enterprise Management 6(4): 279–314.

Baaziz N, Lolo N, Padilla O & Petngang F (2007) Security and privacy protection for automated video surveillance. IEEE International Symposium on Signal Processing and Information Technology**:** 17–22.

Badii A & Einig M (2012) MediaEval 2012 Visual Privacy Task: Privacy and Intelligibility through Pixellation and Edge Detection. Proceedings of the MediaEval 2012 Workshop. Pisa, Italy. 927**:** 2p.

Bankok-Post (September 6th 2012) 1 million more security cameras in Bangkok. http://www.bangkokpost.com/lite/topstories/311173/one-million-more-cctvs-for-bangkok (last accessed Jan 29th 2014).

Barhm MS, Qwasmi N, Qureshi FZ & el-Khatib K (2011) Negotiating privacy preferences in video surveillance systems. Proceedings of the 24th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems. Syracuse, NY, USA. 6704**:** 511–521.

Beer W, Kurschl W, Matusek F, Moser B, Mitsch S & Sutor S (2009) Application development and management of smart camera networks. In Belbachir AN (ed) Smart Cameras. Springer: 259–266.

Beleznai C, Fruhstuck B & Bischof H (2004) Human detection in groups using a fast mean shift procedure. Proceedings of the International Conference on Image Processing, 1, Singapore**:** 349–352.

Benbasat I & Zmud RW (1999) Empirical research in information systems: The practice of relevance. MIS Quarterly 23(1): 3–16.

Beucher S (1992) The watershed transformation applied to image segmentation. Scanning Microscopy International 6: 299–299.

Beyan Ç & Temizel A (2012) Adaptive mean-shift for automated multi object tracking. Institution of Engineering and Technology - Computer Vision 6(1): 1–12.

Bharti TT (2013) Background subtraction techniques review. International Journal of Innovative Technology and Exploring Engineering 2(3): 166–168.

Bock K (2008) EuroPriSe trust certification. Datenschutz und Datensicherheit 32(9): 610–614.

Boult TE (2005) Pico: Privacy through invertible cryptographic obscuration. Computer Vision for Interactive and Intelligent Environment**:** 27–38.

Boult TE & Woodworth R (2008) Privacy and security enhancements in biometrics. In Ratha NK & Govindaraju V (eds) Advances in Biometrics. London, Springer London, p 423–445.

Bouma H, Baan J, Landsmeer S, Kruszynski C, van Antwerpen G & Dijk J (2013) Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall. SPIE Defense, Security, and Sensing, International Society for Optics and Photonics**:** 87560A-87560A-13.

Bowyer KW (2004) Face recognition technology: security versus privacy. IEEE Technology and Society Magazine 23(1): 9–19.

Bradski GR (1998) Computer vision face tracking for use in a perceptual user interface. Proceedings IEEE Workshop on Applications of Computer Vision. Princeton, NJ**:** 214–219.

Brassil JT (2005) Using mobile communications to assert privacy from video surveillance. Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium. Princeton, NJ. 18**:** 290–298.

Breaux TD & Anton AI (2008) Analyzing regulatory rules for privacy and security requirements. IEEE Transactions on Software Engineering 34(1): 5–20.

Buscher M, Wood L & Perng S-Y (2013) Privacy, security, liberty: Informing the design of EMIS. Proceedings of the 10th International ISCRAM Conference, Baden-Baden, Germany**:** 11p.

Cai Y & Pietikäinen M (2011) Person re-identification based on global color context. Computer Vision–ACCV 2010 Workshops: 205–215.

Caine K & Hanania R (2013) Patients want granular privacy control over health information in electronic medical records. Journal of the American Medical Informatics Association 20(1): 7–15.

Caleiro PM, Neves AJ & Pinho AJ (2013) Color-spaces and color segmentation for real-time object recognition in robotic applications. Electronics and Telecommunications 4(8): 940–945.

Cavallaro A (2004) Adding privacy constraints to video-based applications. European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology. London**:** 8p.

Cavallaro A (2007) Privacy in video surveillance [in the spotlight]. IEEE Signal Processing Magazine 24(2): 166–168.

Cavoukian A (2013) Privacy and security by design: An enterprise architecture approach. Ontario, Oracle Corporation.

CCMB (2012) Common Criteria for Information Technology Security Evaluation. http://www.commoncriteriaportal.org/cc/ (last accessed Jan 29th 2014), Common Criteria.

Chang R-I, Wang T-C, Wang C-H, Liu J-C & Ho J-M (2012) Effective distributed service architecture for ubiquitous video surveillance. Information Systems Frontiers 14(3): 499–515.

Chattopadhyay A & Boult TE (2007) Privacycam: a privacy preserving camera using uclinux on the blackfin dsp. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**:** 1–8.

Chen D, Chang Y, Yan R & Yang J (2007) Tools for protecting the privacy of specific individuals in video. EURASIP Journal on Applied Signal Processing 2007(1): 107–116.

Choi S, Han J-W & Cho H (2011) Privacy-preserving H. 264 video encryption scheme. ETRI Journal 33(6): 935–944.

Comaniciu D, Ramesh V & Meer P (2003) Kernel-based object tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25(5): 564–577.

Compagna L, Khoury PE, Krausova Aze, Massacci F & Zannone N (2009) How to integrate legal requirements into a requirements engineering methodology for the development of security and privacy patterns. Artificial Intelligence and Law 17(1): 1–30.

Cristania M, Raghavendraa R, Buea AD & Murinoa V (2013) Human behavior analysis in video surveillance: A Social Signal Processing perspective. Neurocomputing 100: 86–97.

Cumming DJ & Johan S (2013) Cameras tracking shoppers: The economics of retail video surveillance, http://dx.doi.org/10.2139/ssrn.2298019 (last accessed Jan 29th 2014).

Dalai N, Triggs B, Rhone-Alps I & Montbonnot F (2005) Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego. 1**:** 886–893.

Demiröz BE, Arı İ, Eroğlu O, Salah AA & Akarun L (2012) Feature-based tracking on a multi-omnidirectional camera dataset. Proceedings of the 5th International Symposium on Communications Control and Signal Processing. Rome**:** 1–5.

Denning PJ (1996) A new social contract for research. Communications of the ACM 40(2): 132–134.

Descamps A, Carincotte C & Gosselin B (2012) Person detection for indoor videosurveillance using spatio-temporal integral features. Communications in Computer and Information Science 277: 110–118.

Dufaux F (2010) A framework for the validation of privacy protection solutions in video surveillance. Proceedings of the 2010 IEEE International Conference on Multimedia and Expo. Suntec City**:** 66–71.

Dufaux F & Ebrahimi T (2006) Scrambling for video surveillance with privacy. Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop**:** 160–167.

Earp J, Anton A & Jarvinen O (2002) A social, technical and legal framework for privacy management and policies. Americas Conference on Information Systems: 605–612.

European Parliament (1995) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal. 281: 31-50.

Ferryman J & Crowley JL (2006) Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance PETS. In Conjunction with IEEE Conference on Computer Vision and Pattern Recognition.

Fleck S & Straßer W (2010) Privacy sensitive surveillance for assisted living - A smart camera approach. In Nakashima H, Aghajan H & Augusto JC (eds) Handbook of Ambient Intelligence and Smart Environments. Springer: 985–1014.

Foucault M (1975) Discipline and Punish. Gallimard, Paris.

Fukunaga K & Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory 21(1): 32–40.

Gao T, Li G, Lian S & Zhang J (2012) Tracking video objects with feature points based particle filtering. Multimedia Tools and Applications 58(1): 1–21.

Greenwald G & MacAskill E (2013) NSA Prism program taps in to user data of Apple, Google and others. The Guardian, http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data (last accessed Jan 29th 2014). 2013.

Gregor S & Hevner AR (2013) Positioning and presenting design science research for maximum impact. MIS Quarterly 37(2): 337–355.

Halonen R & Paavilainen J (2005) Hierarchical Model of Problems in Implementing Information Systems. Proceedings of the Pacific Asia Conference on Information Systems: 350–362.

Hamouda S (2012) Security and privacy in cloud computing. Proceedings of the International Conference on Cloud Computing Technologies, Applications and Management, Dubai, IEEE: 241–245.

Heikkilä M, Pietikäinen M & Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recognition 42(3): 425–436.

Heikkinen K, Eerola J, Jäppinen P & Porras J (2004) Personalized View of personal information. WSEAS Transactions on Information Science and Applications 2(4).

Herrmann DS (2007) Complete guide to security and privacy metrics: Measuring regulatory compliance, operational resilience, and ROI. Auerbach Publications.

Hevner AR (2007) A three cycle view of design science research. Scandinavian Journal of Information Systems 19(2): 87–92.

Hevner AR, March ST, Park J & Ram S (2004) Design science in information systems research. MIS Quarterly 28(1): 75–105.

Horprasert T, Harwood D & Larry SD (1999) A statistical approach for real-time robust background subtraction and shadow detection. Proceedings of IEEE International Conference on Computer Vision Frame-Rate Workshop. Kerkyra, Greece: 1–19.

Huang J, Li Z & Tang N (2013) A background subtraction method with spatial-temporal information analysis for false detection suppression. Journal of Computational Information Systems 9(2): 565–574.

Izadinia H, Ramakrishna V, Kitani KM & Huber D (2013) Multi-pose multi-target tracking for activity understanding. Proceedings of the IEEE Workshop on Applications of Computer Vision, Tampa: 385–390.

Järvinen P (2004a) On a variety of research output types. Monographs and series D-2004-6.

Järvinen P (2004b) Research questions guiding selection of an appropriate research method. Monographs and series D-2004-5.

Johnson M, Karat J, Karat C & Grueneberg K (2010a) Usable Policy Template Authoring for Iterative Policy Refinement. Proceedings of the IEEE International Symposium on Policies for Distributed Systems and Networks. Fairfax**:** 18–21.

Johnson M, Karat J & Karat C-M (2010b) Optimizing a policy authoring framework for security and privacy policies. Proceedings of the 6th Symposium on Usable Privacy and Security. Redmond.

Jurie F & Dhome M (2001) A simple and efficient template matching algorithm. Proceedings of the 8th IEEE International Conference on Computer Vision 2**:** 544–549.

Kalman RE (1960) A new approach to linear filtering and prediction problems. Transactions of the ASME Journal of Basic Engineering 82: 35–45.

Kavasidis I, Palazzo S, Salvo RD, Giordano D & Spampinato C (2012) A semi-automatic tool for detection and tracking ground truth generation in videos. Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications 6: 1–5.

Kim S, Guy SJ, Liu W, Lau RW, Lin MC & Manocha D (2013) Predicting pedestrian trajectories using velocity-space reasoning. Proceedings of the Tenth Workshop on the Algorithmic Foundations of Robotics, 86, Springer**:** 609–623.

Kitahara I, Kogure K & Hagita N (2004) Stealth vision for protecting privacy. Proceedings of the International Conference on Pattern Recognition, 4**:** 404–407.

Korshunov P, Araimo C, De Simone F, Velardo C, Dugelay J-L-L & Ebrahimi T (2012) Evaluation of visual privacy filters impact on video surveillance intelligibility. Proceedings of the Fourth International Workshop on Quality of Multimedia Experience. Yarra Valley**:** 150–151.

Korshunov P & Ebrahimi T (2013) Using warping for privacy protection in video surveillance. Proceedings of the 18th International Conference on Digital Signal Processing.

Kratz L & Nishino K (2012) Tracking Pedestrians Using Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5): 987–1002.

Kraus K (2010) Security Management Process in Distributed, Large Scale High Performance Systems. Proceedings of the World Congress on Power and Energy Engineering. Alexandria, Egypt**:** 228–247.

Kraus K, Martikainen O & Reda R (2009) An advanced data fusion architecture for high performance surveillance systems. Proceedings of the Fourth International Conference on Intelligent Computing and Information Systems, Cairo.

Kraus K, Uiberacker M, Martikainen O & Reda R (2008) Hot-Spot Blob Merging for Real-Time Image Segmentation. Proceedings of the International Conference on Pattern Recognition and Computer Vision, Bangkok, World Academy of Science, Engineering and Technology**:** 430 - 435.

Kroener I (2013) 'Caught on Camera': The media representation of video surveillance in relation to the 2005 London Underground bombings. Surveillance & Society 11(1/2): 121–133.

Kruger H, Drevin L & Steyn T (2006) A framework for evaluating ICT security awareness. Proceedings of the ISSA 2006 from Insight to Foresight Conference, Sandton.

Langheinrich M (2001) Privacy by design — principles of privacy-aware ubiquitous systems. Proceedings of the International Conference on Ubiquitous Computing, 2201, Atlanta**:** 273–291.

Leung T & Malik J (2001) Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1): 29–44.

Li BN, Chui CK, Chang S & Ong SH (2012) A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. Expert Systems with Applications 39(10): 9661–9668.

Li M, Yu S, Zheng Y, Ren K & Lou W (2013) Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption.

Li SZ & Jain AK (2011) Handbook of Face Recognition. Springer London, London.

Määttä J, Hadid A & Pietikäinen M (2011) Face spoofing detection from single images using micro-texture analysis. Proceedings of the 2011 International Joint Conference on Biometrics, Washington**:** 1–7.

March ST & Smith GF (1995) Design and natural science research on information technology. Decision Support Systems 15(4): 251–266.

March ST & Storey VC (2008) Design science in the information systems discipline: an introduction to the special issue on design science research. MIS Quarterly 32(4): 725–730.

Markus ML, Majchrzak A & Gasser L (2002) A design theory for systems that support emergent knowledge processes. MIS Quarterly 26(3): 179–212.

Martikainen O & Halonen R (2011) Model for the benefit analysis of ICT. Proceedings of the Americas Conference on Information Systems, Detroit.

Martínez-Ballesté A, Rashwan HA, Castellà-Roca J & Puig D (2013) A trustworthy database for privacy-preserving video surveillance. Proceedings of the Joint EDBT/ICDT 2013 Workshops, ACM**:** 179–183.

Matusek F (2010) Hot-spot blob merging for real-time image segmentation for privacy protection. Proceedings of the 2nd International Conference on Emerging Network Intelligence, Florence, Italy**:** 44–49.

Matusek F (2011) Mean shift clustering for privacy protection in the PEVS system. Proceedings of the International Conference on Intelligent Network and Computing. Bangkok**:** 697–705.

Matusek F (2012) Real-time selective privacy protection: Case study and validation of the PEVS system. Proceedings of the 1st Congress on Business Management and Corporate Social Responsibility. Baku.

Matusek F, Kraus K, Martikainen O & Reda R (2010) Selective ID protection in advanced high performance intelligent video surveillance systems: the SIDP smartcam. Proceedings of the 7th International Conference on Informatics and Systems. Cairo**:** 1–6.

Matusek F, Martikainen O & Reda R (2009) Performance and load balancing in advanced video surveillance. Proceedings of the International Conference on Intelligent Computing and Information Systems. Cairo: 5p.

Matusek F, Pujolle G & Reda R (2008) Shadow detection for increased accuracy of privacy enhancing methods in video surveillance edge devices. Proceedings of the International Conference on Pattern Recognition and Computer Vision. Bangkok**:** 217–222.

Matusek F & Reda R (2008) Efficient secure storage of privacy enhanced video surveillance data in intelligent video surveillance systems. Proceedings of the 23rd International Symposium on Computer and Information Sciences, Istanbul**:** 1–5.

Mazzon R & Cavallaro A (2013) Multi-camera tracking using a multi-goal social force model. Neurocomputing, Special issue on Behaviours in video 100: 41–50.

McPhail B, Ferenbok J, Dehghan R & Clement A (2013) "I'll Be Watching You": what do Canadians know about video surveillance and privacy? iConference 2013 notes.

Miller AR & Tucker CE (2009) Privacy protection and technology diffusion: the case of electronic medical records. Management Science 55(7): 1077–1093.

Nadimi S & Bhanu B (2004) Physical models for moving shadow and object detection in video. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(8): 1079–1087.

Newton EM, Sweeney L & Malin B (2005) Preserving privacy by de-identifying face images. IEEE Transactions on Knowledge and Data Engineering 17(2): 232–243.

Noh S & Jeon M (2013) A new framework for background subtraction using multiple cues. Proceedings of the 11th Asian Conference on Computer Vision, 7726, Daejeon, Springer Berlin Heidelberg**:** 493–506.

Noordin MF (2013) Application of privacy, security and ethics in islamic concerned ICT. Middle-East Journal of Scientific Research 14(11): 1548–1554.

Ondrisek B (2008) Sicherheit elektronischer Wahlen: Eine Methode zur Bewertung und Optimierung der Sicherheit von E-Voting-Systemen. Vdm Verlag Dr. Müller.

Orwell G (1949) Nineteen Eighty-Four. Secker and Warburg, London.

Paruchuri IK, Luo Y & Cheung S-CS (2013) Preserving and managing privacy information in video surveillance systems. In Flammini F, Setola R & Franceschetti G (eds) Effective Surveillance for Homeland Security: Balancing Technology and Social Issues. Boca Raton, CRC Press: 1087.

Peng F, Zhu X-w & Long M (2013) A ROI Privacy Protection Scheme for H. 264 Video Based on FMO and Chaos.

Pietikäinen M, Hadid A, Zhao G & Ahonen T (2011) Local binary patterns for still images. Computer Vision Using Local Binary Patterns, Springer London: 13–47.

Piva S, Calbi A, Angiati D & Regazzoni CS (2005) A multi-feature object association framework for overlapped field of view multi-camera video surveillance systems. Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Como, IEEE**:** 505–510.

Porikli F, Brémond F, Dockstader SL, Ferryman J, Hoogs A, Lovell BC, Pankanti S, Rinner B, Tu P & Venetianer PL (2013) Video surveillance: past, present, and now the Future. IEEE Signal Processing Magazine 30(3): 190–198.

Porras J, Hiirsalmi P & Valtaoja A (2004) Peer-to-peer communication approach for a mobile environment. Proceedings of the 37th Annual Hawaii International Conference on System Sciences, IEEE**:** 7p.

Portokalidis G, Homburg P, Anagnostakis K & Bos H (2010) Paranoid Android: versatile protection for smartphones. Proceedings of the 26th Annual Computer Security Applications Conference, ACM**:** 347–356.

Prati A, Mikic I, Cucchiara R & Trivedi MM (2001) Comparative evaluation of moving shadow detection algorithms. Proceedings of the IEEE CVPR workshop on Empirical Evaluation Methods in Computer Vision. Kauai, Hawaii.

Räty T (2008) Architectural improvements for mobile ubiquitous surveillance systems. PhD thesis. University of Oulu.

Reeder RW, Karat C-M, Karat J & Brodie C (2010) Usability challenges in security and privacy policy-authoring interfaces. Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction 4663: 141–155.

Rinner B & Wolf W (2008) An Introduction to Distributed Smart Cameras. Proceedings of the IEEE 96(10): 1565–1575.

Rosales R & Sclaroff S (1998) Improved tracking of multiple humans with trajectory prediction and occlusion modeling. Proceedings of the IEEE CVPR Workshop on the Interpretation of Visual Motion: 117–123.

Saini M, Atrey PK, Mehrotra S & Kankanhalli M (2012) Adaptive transformation for robust privacy protection in video surveillance. Advances in Multimedia 2012.

Schulz D, Burgard W, Fox D & Cremers AB (2001) Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. Proceedings of the IEEE International Conference on Robotics and Automation, 2, Karlsruhe, IEEE**:** 1665–1670.

Schwartz A (2013) Chicago's video surveillance cameras: a pervasive and poorly regulated threat to our privacy. Northwestern Journal of Technology and Intellectual Property 11: 47–61.

Sein MK, Henfridsson O, Purao S, Rossi M & Lindgren R (2011) Action design research. MIS Quarterly 35(1): 37–56.

Senior A, Pankati S, Hampapur A, Brown L, Tian YL & Ekin A (2003) Blinkering surveillance: enabling video surveillance privacy through computer vision. Yorktown Heights, IBM Research.

Shi J & Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8): 888–905.

Shi J & Tomasi C (1994) Good features to track. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle: 593–600.

Shitrit HB, Berclaz J, Fleuret F & Fua P (2013) Multi-commodity network flow for tracking multiple people. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Shu G, Dehghan A, Oreifej O, Hand E & Shah M (2012) Part-based multiple-person tracking with partial occlusion handling. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island: 1815–1821.

Siebel N (2003) Design and implementation of people tracking algorithms for visual surveillance applications. PhD thesis. The University of Reading.

Simon HA (1996) The sciences of the artificial. MIT Press, Cambridge.

Sohn H, Neve WD & Ro YM (2011) Privacy protection in video surveillance systems: analysis of subband-adaptive scrambling in JPEG XR. IEEE Transactions on Circuits and Systems for Video Technology 21(2): 170–177.

Solove DJ (2011) Nothing to hide: the false tradeoff between privacy and security. Yale University Press, Yale.

Soy SK (1997) The Case Study as a Research Method. Unpublished paper. University of Texas at Austin: https://http://www.ischool.utexas.edu/~ssoy/usesusers/l391d1b.htm (last accessed Jan 29th 2014).

Stahl BC (2007) Privacy and security as ideology. IEEE Technology and Society Magazine 26(1): 35 - 45.

Stauffer C & Grimson W (1999) Adaptive background mixture models for real-time tracking. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Ft. Collins.

Stutzman F, Gross R & Acquisti A (2013) Silent listeners: the evolution of privacy and disclosure on facebook. Journal of Privacy and Confidentiality 4(2).

Such JM, Espinosa A & Garcia-Fornes A (2013) A survey of privacy in multi-agent systems. Knowledge Engineering Review: 1–31.

Sutor S, Matusek F, Kruse F, Kraus K & Reda R (2008a) Large-scale video surveillance systems: new performance parameters and metrics. Proceedings of the Third International Conference on Internet Monitoring and Protection, Bucharest: 23–30.

Sutor S, Matusek F & Reda R (2008b) WSSU: high performance wireless self-contained, surveillance unit; an ad hoc video surveillance system. Proceedings of the Fourth Advanced International Conference on Telecommunications, Athens: 157–161.

Sutor S, Röhr R, Pujolle G & Reda R (2008c) Efficient Mean Shift Clustering Using Exponential Integral Kernels. Proceedings of the International Conference on Pattern Recognition and Computer Vision, 26, World Academic of Science, Engineering and Technology: 376 - 380.

Tang Y, Ma B & Yan H (2013) Intelligent video surveillance system for elderly people living alone based on ODVS. Advances in Internet of Things 3: 44–52.

Taylor H (2003) Most people are 'privacy pragmatists' who, while concerned about privacy, will sometimes trade it off for other benefits. The Harris Poll 17(19): 44.

Ullah S, Xuefeng Z, Feng Z & Haichun Z (2013) Tcloud: Challenges And Best Practices For Cloud Computing. International Journal of Engineering Research and Technology 1(9).

Upmanyu M, Namboodiri AM, Srinathan K & Jawahar (2009) Efficient privacy preserving video surveillance. Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto**:** 1639–1646.

van der Velden M & El Emam K (2013) "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. Journal of the American Medical Informatics Association 20(1): 16–24.

Viola P & Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1.

Forbes. last accessed January 20th 2014 (2012) The Next Privacy Battle: Cameras That Judge Your Every Move.

Walls JG, Widmeyer GR & Sawy OAE (1992) Building an information system design theory for vigilant EIS. Information Systems Research 3(1): 36–59.

Warren SD & Brandeis LD (1890) The right to privacy. Harvard Law Review 193.

Wickramasuriya J, Datt M, Mehrotra S & Venkatasubramanian N (2004) Privacy protecting data collection in media spaces. Proceedings of the 12th annual ACM international conference on Multimedia**:** 48–55.

Winkler T & Rinner B (2010a) A systematic approach towards user-centric privacy and security for smart camera networks. Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras. Atlanta**:** 133–141.

Winkler T & Rinner B (2010b) TrustCAM: security and privacy-protection for an embedded smart camera based on trusted computing. Proceedings of the Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance. Boston**:** 593–600.

Winkler T & Rinner B (2012) User Centric Privacy Awareness in Video Surveillance. Multimedia Systems Journal 18(2): 99–121.

Winkler T & Rinner B (2013) Sensor-level security and privacy protection by embedding video content analysis. Proceedings of the International Conference on Digital Signal Processing.

Xiao Z & Xiao Y (2013) Security and privacy in cloud computing. Communications Surveys & Tutorials 15(2): 843–859.

Xing J, Ai H & Lao S (2009) Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, IEEE**:** 1200–1207.

Yabuta K, Kitazawa H & Tanaka T (2005) A new concept of security camera monitoring with privacy protection by masking moving objects. Advances in Mulitmedia Information Processing: 831–842.

Yang HS (2013) Robust people tracking using an adaptive sensor fusion between a laser scanner and video camera. International Journal of Distributed Sensor Networks 2013.

Yilmaz A, Shafique K & Shah M (2003) Target tracking in airborne forward looking infrared imagery. Image and Vision Computing 21(7): 623–635.

Yizong & Cheng (1995) Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(8): 790–799.

Yu P, Sendor J, Serme G & de Oliveira AS (2013) Automating privacy enforcement in cloud platforms. Data Privacy Management and Autonomous Spontaneous Security, Springer 7731: 160–173.

Zervos M (2013) Multi-camera face detection and recognition applied to people tracking. Master's thesis. Ecole Polytechnique Fédérale de Lausanne.

Zhang P, Thomas T, Emmanuel S & Kankanhalli MS (2010) Privacy preserving video surveillance using pedestrian tracking mechanism. Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence. Florence**:** 31–36.

Zhang T, Liu S, Xu C & Lu H (2013) Mining semantic context information for intelligent video surveillance of traffic scenes. IEEE Transactions on Industrial Informatics 9(1): 149–160.

Zhang W, Cheung S-cS & Chen M (2005) Hiding privacy information in video surveillance system. Proceedings of the IEEE International Conference on Image Processing, 2005. ICIP 2005, 3, Genova**:** 868–871.

Zhang X, Chen L, Pan L & Xiong L (2012) Study on the image segmentation based on ICA and watershed algorithm. Proceedings of the Fifth International Conference on Intelligent Computation Technology and Automation, Zhangjiajie, IEEE**:** 505–508.

Zhou H, Yuanb Y & Shic C (2009) Object tracking using SIFT features and mean shift. Computer Vision and Image Understanding 113(3): 345–352.

606. Hokkanen, Juho (2013) Liquid chromatography/mass spectrometry of bioactive secondary metabolites – *in vivo* and *in vitro* studies

607. Kuokkanen, Matti (2013) Development of an eco- and material-efficient pellet production chain—a chemical study

608. Jansson, Eeva (2013) Past and present genetic diversity and structure of the Finnish wolf population

609. Myllykoski, Matti (2013) Structure and function of the myelin enzyme 2′,3′-cyclic nucleotide 3′-phosphodiesterase

610. Lehto, Tuomas (2013) The importance of persuasive systems design in enhancing consumers' perceptions and adoption of health behavior change support systems

611. Hernoux-Villière, Audrey (2013) Catalytic depolymerisation of starch-based industrial waste : use of non-conventional activation methods and novel reaction media

612. Lawrence, Carl (2013) Innovating with information technology in a globalized world : being proactive about culture

613. Ardanov, Pavlo (2013) Priming capacities of endophytic *Methylobacterium* sp. on potato (*Solanum tuberosum* L.)

614. Koskela, Anni (2013) Wolverine habitat selection, diet and conservation genetics

615. Holm, Jana (2013) Catalytic pretreatment and hydrolysis of fibre sludge into reducing sugars

616. Kemi, Ulla (2013) Adaptation to growing season length in the perennial *Arabidopsis lyrata*

617. Aalto, Esa (2013) Genetic analysis of demography and selection in Lyrate rockcress (*Arabidopsis lyrata*) populations

618. Rodríguez, Pilar (2013) Combining lean thinking and agile software development : how do software-intensive companies use them in practice?

619. Vatka, Emma (2014) Boreal populations facing climatic and habitat changes

620. Isomursu, Marja (2014) Host–parasite interactions of boreal forest grouse and their intestinal helminth parasites

621. Ponnikas, Suvi (2014) Establishing conservation management for avian threatened species

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-952-62-0414-7 (Paperback)
ISBN 978-952-62-0415-4 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO