Stephan R. Sutor

LARGE-SCALE HIGH-PERFORMANCE VIDEO SURVEILLANCE

Stephan R. Sutor

UNIVERSITY OF OULU GRADUATE SCHOOL,
UNIVERSITY OF OULU,
FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING,
DEPARTMENT OF INFORMATION PROCESSING SCIENCE
STEPHAN R. SUTOR

LARGE-SCALE HIGH-PERFORMANCE 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 the Wetteri auditorium (IT115), Linnanmaa, on 17 October 2014, at 12 noon

UNIVERSITY OF OULU, OULU 2014
Abstract

The last decade was marked by a set of harmful events ranging from economical crises to organized crime, acts of terror and natural catastrophes. This has led to a paradigm transformation concerning security. Millions of surveillance cameras have been deployed, which led to new challenges, as the systems and operations behind those cameras could not cope with the rapid growth in number of video cameras and systems. Looking at today’s control rooms, often hundreds or even thousands of cameras are displayed, overloading security officers with irrelevant information.

The purpose of this research was the creation of a novel video surveillance system with automated analysis mechanisms which enable security authorities and their operators to cope with this information flood. By automating the process, video surveillance was transformed into a proactive information system. The progress in technology as well as the ever increasing demand in security have proven to be an enormous driver for security technology research, such as this study. This work shall contribute to the protection of our personal freedom, our lives, our property and our society by aiding the prevention of crime and terrorist attacks that diminish our personal freedom.

In this study, design science research methodology was utilized in order to ensure scientific rigor while constructing and evaluating artifacts. The requirements for this research were sought in close cooperation with high-level security authorities and prior research was studied in detail. The created construct, the “Intelligent Video Surveillance System”, is a distributed, highly-scalable software framework, that can function as a basis for any kind of high-performance video surveillance system, from installations focusing on high-availability to flexible cloud-based installation that scale across multiple locations and tens of thousands of cameras. First, in order to provide a strong foundation, a modular, distributed system architecture was created, which was then augmented by a multi-sensor analysis process. Thus, the analysis of data from multiple sources, combining video and other sensors in order to automatically detect critical events, was enabled. Further, an intelligent mobile client, the video surveillance local control, which addressed remote access applications, was created. Finally, a wireless self-contained surveillance system was introduced, a novel smart camera concept that enabled ad hoc and mobile surveillance.

The value of the created artifacts was proven by evaluation at two real-world sites: An international airport, which has a large-scale installation with high-security requirements, and a security service provider, offering a multitude of video-based services by operating a video control center with thousands of cameras connected.

Keywords: ad-hoc surveillance, intelligent video surveillance, mobile video surveillance, multi-sensor analysis, security, security management process, sensor fusion, sensor processing, smart cameras, system architecture, video analysis, video surveillance, video surveillance networks, wireless video surveillance

Tämän tutkimuksen tarkoitus oli luoda uusi videovalvontajärjestelmä, jossa on automaattiset analyysimekanismit, jotka mahdollistavat turva-alan toimijoiden ja niiden operaattoreiden suoritus voimaan informaatiotulvasta. Automatisoidun videovalvontaprosessin avulla videovalvonta muokattiin proaktiiviseksi tietojärjestelmäksi. Teknologian kehitys ja kasvanut turvallisuusvaatimus osoittautuivat olevan merkittävä aju turvallisuustekniikan tutkimukselle, kuten tämä tutkimus oli. Tämä tutkimus hyödyttää yksittäisen ihmisen henkilökohtaista vapautta, elämää ja omaisuutta sekä yhteisöä estämällä rikoksia ja terroristihyökkäyksiä.

Tässä tutkimuksessa suunnitteluudetettiin sovellettii varmistamaan tieteellinen kurinalaisuus, kun artefakteja luotiin ja arvioitiin. Tutkimuksen vaatimukset perustuvat laheiseen yhteistyöhön korkeatasoista turva-alan viranomaisten kanssa, ja lisäksi aiempi tutkimus analysoitiin yksityiskohtaisesti.

Luotujen artefaktien arvo voidaan todentaa arvioimaan laajamittaisessa toteutuksessa kansainvälinen lentokenttä, jonka laajamittaisessa toteutuksessa on korkeat turvavastaan ja turvallisuuspalveluntuottajia, jotka tarjoaa moninaisia videopohjaisia palveluja videovalvontakeskuksen avulla käyttäen tuhansien kameroiden.

Asiastanan: ad-hoc valvonta, järjestelmäarkkiitehtuuri, langaton videovalvonta, mobiili videovalvonta,videovalvontakeskuksen avulla käyttäen tuhansien kameroiden.
Acknowledgements

As research may sometimes be practiced in reclusion, locked away in an ivory tower, it was absolutely not the case regarding this dissertation. Hence, I would like to thank a few people who were vital in this journey.

First, I would greatly like to thank Adjunct Professor Raija Halonen for having taken interest in my research topic and having supervised this thesis. You have fought through mountains of bureaucracy and opposition due to the atypical setting of this research and you never hesitated once in your commitment. Further, I would like to thank Dr. Reda Reda, not only for being my secondary supervisor, but for having supervised my research from the early beginning, long before we came to Oulu and were in search of the perfect university for this endeavor. I would also like to thank Professor Olli Martikainen for heading my follow-up group. Just as well, I would like to thank Dr. Jouni Markkula and Dr. Alice Sendera who also joined my follow-up group. I would further like to thank Professor Jarmo Harju and Associate Professor Jouni Ikonen for taking interest in my work and having accepted to review my thesis, just as I would like to thank Professor Doug Vogel for having accepted to be my opponent in my last exam, my doctoral defense. Further, I would like to thank Dean Petri Pulli, who made it possible for us to study form abroad in the first place and Dr. Kaisu Juntunen for having come to Vienna to teach us.

I would like to thank Klemens Kraus, you were like a brother to me from the very beginning of our lives, and Dr. Florian Matusek for your friendship, your great work and the exciting times we spent together founding and growing our start-up company KiwiSecurity, tackling any challenge on our way to success. Further, I would like to thank Ahmed Asfour joining us and becoming an essential part of KiwiSecurity, but also for being a great friend and mentor. Without your guidance our start-up company would never have become the international enterprise it is today. I would also like to thank the entire KiwiSecurity team, all of you who have joined and supported us in our undertaking. Most certainly I want to thank Joanna Babicka, not only for the long nights of proof reading, but also for being extra critical and supportive in all kinds of ways. Nobody has read this thesis as often as you have. Finally, I would like to thank my family, especially my Mom, Dr. Martina Sendera, my sisters Sandra Sutor and Sonja Sutor as well as my Dad, Dr. Ronald Sutor, for providing the loving and caring environment which allowed me to proceed in all my endeavors. Thank you for believing in me. Thank you for supporting me all this time. Without you I would not be here today!
Abbreviations & definitions

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G/3G/4G</td>
<td>2nd/3rd/4th Generation: Mobile Communication Standard</td>
</tr>
<tr>
<td>CCTV</td>
<td>Closed Circuit Television</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal-Oxide-Semiconductor</td>
</tr>
<tr>
<td>DARPA</td>
<td>Defense Advanced Research Projects Agency (USA)</td>
</tr>
<tr>
<td>DMZ</td>
<td>Demilitarized Zone</td>
</tr>
<tr>
<td>DSL</td>
<td>Digital Subscriber Loop</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processor</td>
</tr>
<tr>
<td>DSR</td>
<td>Design Science Research</td>
</tr>
<tr>
<td>FoV</td>
<td>Field of View</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field-Programmable Gate Array</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>HCI</td>
<td>Human–Computer Interaction</td>
</tr>
<tr>
<td>HSPA</td>
<td>High Speed Packet Access</td>
</tr>
<tr>
<td>ICT</td>
<td>Information Communication Technology</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IVSS</td>
<td>Intelligent Video Surveillance System</td>
</tr>
<tr>
<td>KB</td>
<td>Knowledge Base</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MSAP</td>
<td>Multi-Sensor Analysis Process</td>
</tr>
<tr>
<td>MoG</td>
<td>Mixture of Gaussians</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio-Frequency Identification</td>
</tr>
<tr>
<td>SDK</td>
<td>Software Development Kit</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>SYSARCH</td>
<td>System Architecture</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>VS</td>
<td>Video Surveillance</td>
</tr>
<tr>
<td>VSSaaS</td>
<td>Video Surveillance as a Service</td>
</tr>
<tr>
<td>VSHQ</td>
<td>Video Surveillance Head Quarters</td>
</tr>
<tr>
<td>VSLC</td>
<td>Video Surveillance Local Control</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wire Local Area Network</td>
</tr>
<tr>
<td>WSSU</td>
<td>Wireless Self-Contained Surveillance Unit</td>
</tr>
</tbody>
</table>
Anomaly
Rare event which deviates from what is considered normal in a scene

Artifact
A constructed IT artifact according to design science research

Cloud computing
Distributed computing where processing and storage mainly reside in a data center, and data is transmitted over the Internet

Exaptation
A category of research contributions where existing solutions (e.g. from other fields) are extended to new problems

Incident
Security critical event which demands human response

Intelligent video surveillance
Video surveillance with automated video analysis

PTZ camera
A motorized camera that can change its point of view by physically panning, tilting and zooming

Security
The state of being protected against harm

Smart camera
A camera with on-board processing capabilities such as video analysis
Content

Abstract
Tiivistelmä 7
Acknowledgements 7
Abbreviations & definitions 11
Content 15
1 Introduction
1.1 Purpose .................................................................................................... 16
1.2 Motivation for the research ................................................................. 17
1.3 Prior research ........................................................................................ 19
1.4 Research methods ............................................................................... 21
1.5 Research environment ........................................................................ 22
1.6 Research questions ............................................................................... 23
1.7 Contributions ....................................................................................... 24
1.8 Structure of the thesis ........................................................................... 25
2 Prior research
2.1 System architecture ............................................................................... 28
2.2 Communication tier ............................................................................. 31
2.3 Sensor tier ............................................................................................ 34
2.4 Operator / HCI tier ............................................................................... 35
2.5 Storage and retrieval tier ..................................................................... 37
2.6 Processing tier ..................................................................................... 38
   2.6.1 Detection and tracking................................................................. 39
   2.6.2 Activity recognition .................................................................... 41
   2.6.3 Anomaly recognition ................................................................ 43
   2.6.4 Multi-camera processing ............................................................ 45
   2.6.5 Other processing ......................................................................... 47
2.7 Conclusion of prior research analysis .................................................. 49
3 Research approach
3.1 Design science research ....................................................................... 51
3.2 The research process ........................................................................... 56
   3.2.1 Research phases .......................................................................... 56
   3.2.2 Design science research applied ................................................ 58
   3.2.3 The research questions ............................................................... 61
   3.2.4 The artifacts ............................................................................... 63
   3.2.5 Evaluation ................................................................................... 63


6 Discussion

6.1 Purpose of the study ................................................................. 131
6.2 Research results ....................................................................... 132
6.3 Research contributions ............................................................. 134
6.4 Theoretical implications ............................................................. 136
6.5 Managerial implications ............................................................. 138
6.6 Methodological considerations ................................................ 139
6.7 Future outlook ......................................................................... 141

7 Conclusion ................................................................................. 143

References .................................................................................... 147
1 Introduction

The last decade was marked by a set of harmful events ranging from economical crises to organized crime, acts of terror and natural catastrophes. This has led to a paradigm transformation concerning security, resulting in the new trend security economy (Keenan 2005). Taking a look at the evolution of surveillance, first there was the human observer who could see whatever was in front of him in close proximity. The “eyes” and the “brains” in form of that person had to be on-site. The first revolution arose by the invention of CCTV (Kruegle 2007). Here, multiple “eyes”, the cameras, viewing remote locations were brought together to be processed by a centralized “brain”, the human operator. Siemens’ installation at Test Stand VII in 1942 in Peenemünde, Germany, is said to be the first CCTV system (Dornberger 1952), however, in the early 1970s, when analog tape-based recording became widely available, CCTV was also on the rise becoming ubiquitous in the 1980s (Kruegle 2007), especially in the UK. Fig. 1 shows such an early CCTV control room used for traffic surveillance in 1973 in Munich, Germany. As can be seen, two officers observe only few cameras streams on few monitors.

Fig. 1. An early CCTV control room in 1973 (permission by U.S. National Archives, photo no. 551905).
The progress in technology as well as the ever increasing demand in security have proven to be an enormous driver for the entire security industry and security technology research which it thrives on. The largest contribution to this increase was achieved in the physical security sector, where the global video surveillance market is expected to reach $40 billion by 2015 (Honovich 2011). Thus, millions of surveillance cameras have been deployed (Porikli et al. 2013), which leads to new challenges, as the systems and operations behind those cameras cannot cope with the rapid growth in number of video cameras and systems.

Looking at today’s control rooms, often hundreds or even thousands of cameras are displayed, hence, the likelihood of an officer detecting an incident happening on a random display is very little due to the massive information flood, assuming that the image information was transported to the control room in the first place (Räty 2010). Fig. 2 shows a mock-up of a typical video surveillance control room, where the security officer has to maintain situational awareness across 91 cameras and then trigger certain responses according to a security management processes in case of an incident. He is overloaded with information and will thus very likely miss critical events.

![Fig. 2. A mock-up of a typical control room situation](image)

**1.1 Purpose**

New system architectures as well as automated analysis mechanisms are necessary for security authorities and their operators in order to cope with the amount of data generated by the amount of cameras and sensors deployed. Thus, video surveillance systems need to be transformed into information systems, with
all the corresponding considerations and challenges (Martikainen & Halonen 2011). Further, the transition from analog CCTV to IP-based systems, which has been happening over the last decade, is a critical factor. These developments in video surveillance pose a whole new set of challenges, both organizational and technological (Räty 2010).

Addressing these newly arisen demands in coping with this vast increase in video surveillance cameras and systems, the purpose of this study was to scientifically assess user requirements and create a reference model of a system fulfilling these requirements. The resulting prototype shall be the basis of a commercial product, which is to be developed based on the results and findings of this study. Hence, the applicability and relevance of this research was of great importance and constantly ensured by close collaboration with security authorities such as international airport operators, police forces or private security service providers as they are the future users and customers of the resulting product.

This study shall contribute to the knowledge base by providing scientifically assessed user requirements as well a blueprint for next generation implementations of video surveillance systems.

1.2 Motivation for the research

Today, many of us have the privilege of living in a civilized democratic society and are able to enjoy many of its liberties. The price for this has been argued for centuries, dating back to the seventeenth century when John Locke wrote the essay the Second Treatise of Government (Locke 1689), which had significant influence on the development of modern concepts of democracy and whose principles are valid until today. He stated that all individuals have natural rights to freedom, independence and political equality and explained how the powers of government may be used to protect these.

According to Locke, the 'state of nature', the condition preceding the development of society, is a state of insecurity, where individuals are exposed to infringement of their natural rights by other individuals. Thus, the purpose of establishing a civil government is to protect the freedom and well-being of all members of society. Locke further states that only a government which has been established by common consent has the right to power and that people of society have a right to defend themselves against those individuals who use violence to try to take away their freedom and security.
These issues and principles have not changed since Locke’s work. Many such attacks on our freedom and security today, in the form of crime and terrorism, have led to a strong demand for new solutions. Thus, the growth in the security market is a driving force for a vast number of innovations and research in the field of video surveillance, as video surveillance is an effective tool in fighting these attacks. However, surveying and analyzing prior research results revealed the fact that much work remains to be done regarding system and component performance of video surveillance systems. A lot of research has been done in rather isolated sub-topics of video surveillance, but only very little on large-scale system architecture design.

Analyzing the big picture in the security sector (Sutor et al. 2010), including the converging of ICT security and physical security and the security potential and its role in the progress of the information communication technology market, the following areas of research could be identified, and were all addressed in this study:

– Video Surveillance System Architecture
– Video Surveillance Performance Metrics
– Multi-Sensor Surveillance
– Mobile Video Surveillance Access
– Smart Cameras

Another important topic in video surveillance research is the protection of privacy and the avoidance of misuse of such systems (Bartosinski et al. 2012, Kim & Han 2012), especially as surveillance cameras have become ubiquitous and surveillance systems interconnected (Porikli et al. 2013). However, this topic is beyond the scope of this work and is thoroughly dealt with within the same research group in the start-up company by Florian Matusek (2014). The earlier mentioned new organizational challenges, such as the Security Management Process are only touched upon in this work and also addressed in the same research group by Klemens Kraus (2010).

Concluding the motivational factors for this work, next to the promising market situation acting as a driver for security research and the tackling of technological challenges, thus contributing to the knowledge base by presenting a reference model for future video surveillance systems based on real-world user requirements, this work shall contribute to protect our personal freedom, protecting our lives, our property and our society by aiding the prevention of
crime and terrorist attacks diminishing our personal freedom, ultimately, making this world a better place to live in.

1.3 Prior research

In recent years, a vast amount of research has been conducted concerning high-performance video surveillance, transforming a once reactive tool, which was used after the fact, into a pro-active intelligent information system, initially strongly driven by defense and homeland security (Corrall 1991, Kanade et al. 1999), and public safety (Räty 2010). Systems for city surveillance are developed (Fernández et al. 2013) fusing data from multiple sites, centralizing processing and are thus moving to cloud-based system architectures (Zhu et al. 2011, Dey et al. 2012, Lin et al. 2012, Wu et al. 2012, Chattopadhyayr et al. 2013).

In order to structure the analysis of prior research, next to video surveillance system architectures in general, their sub-components are analyzed individually. This results in a five-tier model, which splits video surveillance systems into the following tiers:

- Communication tier
- Sensor tier
- Operator / HCI tier
- Storage & retrieval tier
- Processing tier

The communication tier addresses all connections in a digital video surveillance system, ranging from different wired (Duelk & Zirngibl 2006, Timmers et al. 2013, Beygi et al. 2014) to wireless connectivity (Haider et al. 2011, Abdullah & Yonis 2012, Kelly 2014), comparing different technologies and analyzing their applicability for video surveillance applications. This results in systems utilizing commercially available networks and devices (Sandu et al. 2010, Fasui et al. 2013) or designing entirely experimental systems with wearable cameras (Dao et al. 2012). Regarding video surveillance, data streaming and utilization of bandwidth are the most crucial topics (Räty et al. 2008, Chen et al. 2009).

The sensor tier deals with the input devices of a video surveillance system: the cameras. Nowadays, these are a lot more than just image acquisition devices, and feature extensive on-board processing capabilities and multi-modal sensing capabilities (Belbachir 2009, Chunjing et al. 2012, Sánchez et al. 2012, Ye et al. 2012, Baoxia et al. 2013).
The human-computer interfaces for interaction with the video surveillance system are addressed in the operator / HCI tier by providing different methods of efficiently locating events in recorded video (Aydemir et al. 2012, Sun et al. 2013), creating methods that try to learn the operator’s behavior (Meessen et al. 2009, Filonenko et al. 2013, Martinel et al. 2013) or by creating user-centric description languages (Feng et al. 2013) to simplify querying for non-technical users. Utilizing mobile devices for video surveillance is also of major importance (Imai et al. 2008, Rashmi & Latha 2013).

Storage of large-scale video surveillance data as well as the querying of such databases is covered by the storage and retrieval tier. This ranges from general queries (Presti et al. 2012, Xue et al. 2012) to more specific tasks, such as license plate analysis (Shu et al. 2012), to annotation frameworks (Vezzani & Cucchiara 2010b) in order to improve such querying.

The processing tier addresses the processing of any kind of sensor information. This ranges from entire solutions or processing pipelines (Lewis 2012, Liu et al. 2013) to the analysis of applications of such systems (Thornton et al. 2009, Argioli & Bisogni 2011) or can further be broken down into specific tasks in such processing pipelines. The first step in such processing pipelines is detection and tracking of persons (DiCaterina & Soraghan 2011, Kushwaha et al. 2012, Zhang et al. 2012b) or objects in general (Haque & Murshed 2012, Feris et al. 2013). Special attention driven by public safety is also given to the detection of unattended objects (Fan & Pankanti 2012, Lai et al. 2012).

In the next step, the processing tier deals with the detection of predefined activities (Wolf & Baskurt 2012, Xu et al. 2012, Taha et al. 2013), such as detecting drawn fire arms (Grega et al. 2013) or detecting fights (Esen et al. 2013), people falling on the ground (Rougier et al. 2011) or previously undefined anomalies (Wang & Snoussi 2012, Zhang et al. 2013, Zhu et al. 2013, Li et al. 2014). The processing of information across multiple cameras can substantially add to the capabilities of video surveillance analysis (Behera et al. 2012, Hong et al. 2012, Martelli et al. 2012, Snidaro et al. 2012, Zhang et al. 2012a) and lead to automatic control of PTZ cameras (Starzyk & Qureshi 2011, Abdelkader et al. 2012).

Finally, the processing tier covers specialized image or video recognition algorithms such as face recognition (Hadid & Pietikäinen 2009, Yang et al. 2014), license plate recognition (Anagnostopoulos 2014), or algorithms that reach into other modalities, such as audio analysis (Kim & Ko 2011, Salvati et al. 2011) or the combination of multiple modalities (Niemenen et al. 2009, Viani et al. 2012).


1.4 Research methods

This study was based on research in the design science research theory conducted by Hevner et al. (2004), which represents a human approach focusing on the applicability and the incorporation of the end user, and was thus ideal for this study. Their theory is based on three pillars: the environment, the actual design science research, which is again split into the artifact design and its evaluation, and the knowledge base.

In this work, the research environment consisted of the research group and their lab, university collaboration, and especially the security authorities and their real-world surveillance sites. The knowledge base for this study consisted of prior research in the respective fields as well as the customers’ experience and processes, which finally led to the actual design and evaluation of the created artifacts.

In addition to the three pillars, Hevner et al. (2004), then connected these with three cycles: First the relevance cycle, which connects the real-world requirements and sites from the customers and the artifact design, and especially evaluation to ensure the outcome of the research to be relevant, hence useful for the customer. Second, the design cycle ensures that created artifacts are constantly evaluated and redesigned if necessary. Finally, the rigor cycle connects the knowledge base to the artifact creation and evaluation, which ensures not only scientific grounding, but also dissemination of findings and thus contributing to the knowledge base.

Gregor & Hevner (2013) further propose a so-called knowledge contribution framework to enable the classification of research. This positioning is conducted in two dimensions: Solution maturity and application domain maturity, resulting in four quadrants.

- The Invention Quadrant for fields with little knowledge of the domain or existing artifacts
- The Improvement Quadrant for new solutions to known problems
- The Exaptation Quadrant for the application of known solutions to new problems
- The Routine Design Quadrant for applying known solutions to known problems

As neither video surveillance nor security in general were novel fields of research and the problems in the area were well known and researched as shown in the
prior research section, this design science project could clearly be categorized into the Improvement Quadrant, although, while designing the artifacts, it was also possible to utilize some solutions from others fields of research for certain issues, hence also touching upon the Exaptation Quadrant.

1.5 Research environment

The research was conducted in the environment of a high-tech start-up company. This allowed close interaction with end-users and security authorities using video surveillance systems with vast expert knowledge, which directly influenced the gathering of requirements and ultimately formulating research questions. This further ensured that the research conducted was highly relevant. On the other hand, very close collaboration and supervision by universities (Sutor et al. 2008b, Sutor et al. 2009, Sutor et al. 2010) ensured scientific rigor.

The author recognizes that working in such an environment, many discussions and input from colleagues had great influence. Simultaneously, presented concepts have partially been implemented and built upon in a commercial product by the company’s software development team. The main contribution of the author, however, is the conception and design of the overall system as presented in this thesis.

This design science research project addresses challenges in creating a novel system architecture and extending artifacts for high-performance large-scale video surveillance. The research in this thesis was partially funded and conducted within a number of research projects. Table 1 shows an overview of these research projects.
Table 1. Research projects contributing to this work.

<table>
<thead>
<tr>
<th>Project</th>
<th>Year</th>
<th>Funding Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Multi-Camera Tracking System I</td>
<td>2007</td>
<td>Austria Wirtschaftsservice(^1)</td>
</tr>
<tr>
<td>Smart Camera Concept (Best projects award in the ZIT ICT competition)</td>
<td>2007</td>
<td>Zentrum für Innovation und Technologie(^2)</td>
</tr>
<tr>
<td>A Multi-Camera Tracking System II</td>
<td>2008</td>
<td>Austria Wirtschaftsservice</td>
</tr>
<tr>
<td>An Automated PTZ Camera System</td>
<td>2008</td>
<td>Österreichische Forschungsförderungsgesellschaft(^3)</td>
</tr>
<tr>
<td>A Multi-Sensor Surveillance System I</td>
<td>2009</td>
<td>Österreichische Forschungsförderungsgesellschaft</td>
</tr>
<tr>
<td>A Multi-Sensor Surveillance System II</td>
<td>2010</td>
<td>Österreichische Forschungsförderungsgesellschaft</td>
</tr>
<tr>
<td>A Cloud-Based Video Surveillance System</td>
<td>2012</td>
<td>Österreichische Forschungsförderungsgesellschaft</td>
</tr>
</tbody>
</table>

\(^1\)The Austrian federal development and financing bank
\(^2\)The Technology Agency of the City of Vienna
\(^3\)The Austrian Research Promotion Agency

1.6 Research questions

The research questions were carefully deducted from the meeting minutes from more than 40 end-user interviews with different security authorities, revealing their expert insight on high-performance large-scale video surveillance as well as the problems with current systems, and their wishes in general. Accordingly, this led to the following problem formulation:

*What kind of video surveillance system is needed to maximize situational awareness, thus minimizing reaction time in critical situations, while being fully scalable to any number of cameras?*

Starting from this problem formulation, the following sub-questions arose.

1. What does a video surveillance system architecture need to look like in order to be fully scalable, distributable and highly secure?

This could be further explored by the following sub questions: What are the specific demands regarding video surveillance? What are the weak spots of current video surveillance systems? What measures need to be taken in order to create a system architecture for large-scale high-performance video surveillance, sufficiently robust to withstand any type of attack or fault? How can the human observer be assisted in processing an information flood too large to be manually observed?
Addressing the fact that CCTV cameras can only observe a limited part of the frequency spectrum, the idea of utilizing multiple sensor modalities arose. Hence, the following research question was created:

2. How can multi-sensor analysis improve the performance of a surveillance system?

This led to further questions, such as: What increase in system performance in means of detection accuracy and false alert elimination can be achieved? Does the type of sensor affect the scenarios of improvement?

As we are in a time where mobile devices are predominant and every security officer is equipped with a range of mobile devices, the next research question arose:

3. How can mobile field-devices improve a video surveillance system?

Further asking: What are the main issues to consider? How much can the reaction time in critical situations be decreased?

Finally, as there is a trend in increasingly more intelligent cameras to even self-contained camera systems, so it was clear this also had to be considered in this work, which lead to the next research question:

4. How can an intelligent camera improve video surveillance performance?

Further asking, how can intelligent ad hoc cameras decrease the time of reaction, or even the time of deployment in critical situations? How much time can be saved? What are the main issues to consider?

The results of this design science research project shall contribute to the knowledge base by providing not only real-world requirements derived directly from end-users, but also a reference design for large-scale high-performance video surveillance systems, allowing a standardized research and better understanding of such systems, as well as a broad insight into work of security authorities, the challenges they face, and how to assist them by designing proven and evaluated artifacts.

1.7 Contributions

Aligned to the design science research methodology by Hevner et al. (2004), the foundation of this work was built on an analysis of prior research as well as numerous interviews with security authorities, gathering their expert knowledge
in order to create specific user requirements. Based on these requirements, which were then formulated into research questions, artifacts were designed in order to create a scientifically grounded solution that fulfills these requirements. In order to evaluate the created artifacts different use cases were defined and measured in order to quantify the benefit of the created solution.

The solution included a novel system architecture, which incorporated a multi-sensor analysis process, a novel smart camera concept, as well as a novel mobile client device. These artifacts shall ultimately form a blueprint for future high-performance large-scale video surveillance systems, thus contributing to the knowledge based by providing a reference design as well as a deep insight into real-world user requirements and challenges. Table 2 summarizes the results of this dissertation as suggested by March & Storey (2008).

Table 2. Research results.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
<td>Due to the ubiquity of video surveillance the amount of information gathered cannot be processed manually any longer. Operators are overloaded with too much information. Critical incidents are often missed.</td>
</tr>
<tr>
<td>Artifact (Solution)</td>
<td>Design and construction of a novel multi-sensor video surveillance system which is virtually unlimited in scalability, automates the viewing process, and extends to mobile devices.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Several use cases were created in collaboration with security authorities in order to objectively measure the performance increase when using the created system.</td>
</tr>
</tbody>
</table>

1.8 Structure of the thesis

The thesis continues as follows: Chapter 2 gives an extensive analysis of prior research. Chapter 3 introduces the research methodology applied in this dissertation. Chapter 4 introduces the created artifact and sub-artifacts, while Chapter 5 presents their respective evaluation. Chapter 6 gives a thorough discussion relating the created artifacts, their evaluation and the prior research, analyzing resulting implications, while Chapter 7 summarizes and concludes the thesis.
2 Prior research

There have been numerous developments in the area of large-scale high-performance video surveillance. Sensors are becoming almost ubiquitous and Moore’s law (Moore 1965) doubling the available processing power every eighteen months constantly allows more sophisticated algorithms to be processed in real-time. In order to structure the analysis of prior research, the reviewed literature has been split into works regarding the overall system architecture and works regarding sub-components, which accordingly have been split into five tiers as illustrated in Fig. 3:

- The communication tier, incorporating all data transmission, wired and wireless
- The operator / HCI tier, which consist of all interfaces towards the human user operating the video surveillance system
- The sensor tier, including a variety of sensors with optional on-board processing
- The storage & retrieval tier, incorporating different data-storage strategies (e.g. cloud storage) and off-line data processing for retrieval
- The processing tier, which is further split into five sub categories due to the enormous amount of research in each area:
  - Activity recognition: Automatic recognition and classification of peoples’ activities
  - Anomaly recognition: Automatic detection of unusual states in a scene without prior classification
  - Detection & tracking: Detection of objects in a scene which are then tracked over consecutive frames
  - Multi-camera processing: Algorithms utilizing the information of multiple cameras
  - Other processing: Selected work utilizing non-video sensors, such as audio analysis or RFID sensors
2.1 System architecture

It is almost impossible to pinpoint the start of intelligent video surveillance architectures. Early works in which a systematic approach to this topic can be found date back to the early 1990s, where e.g. Corrall (1991) aims to obtain a description of airport ground traffic and road traffic utilizing computer vision algorithms. However, the first large-scale endeavor in research concerning this topic can be traced back to the VSAM (video surveillance and monitoring) project by Kanade et al. (1999) at the Carnegie Mellon University, backed by DARPA and the US Information Systems Office. For the first time, a test bed was created mixing color cameras, thermal cameras and night vision cameras, applying a vast number of computer vision algorithms.

Taking a look at more recent research on video surveillance architecture, Li et al. (2013) proposed a video surveillance management platform which was designed for public safety and is in use on a subway line. It is based on a hierarchical system architecture to integrate various hardware devices and utilized web browsers for video monitoring and interaction. An intelligent surveillance
platform based on the usage of large numbers of inexpensive sensors is introduced in the European Eureka Celtic project HuSIMS in which Fernández et al. (2013) apply motion detection, tracking algorithms and a versatile communication network in order to facilitate data collection from the visual sensors, targeting mainly city surveillance for smart cities and possibly large facilities. Chen et al. (2013), on the other hand, present a system fusing automatic license plate recognition engines operating in remote data centers, thus resulting in a cloud solution capable of scalable detection and tracking of a target vehicle in a city with a given license plate number.

A cloud-based architecture which allows edge and back-end processing is presented by Chattopadhyay et al. (2013). They create a sensor markup language which facilitates communication between sensors and an interface that allows access to event information remotely, so third party applications can provide services using the surveillance data. An architecture based on peer-to-peer streaming and cloud-based storage to provide an economic, scalable, reliable approach to store video data is presented by Wu et al. (2012) and Zhu et al. (2011) for supporting thousands of users in an unstructured network. A more centralized cloud-based approach is presented by Lin et al. (2012), whereas Dey et al. (2012) present a multimedia surveillance back-end system architecture utilizing cloud storage for large-scale city surveillance.

Yu et al. (2012) present a video surveillance architecture which works in a ubiquitous sensor network (USN) environment, utilizing the open geospatial consortium’s sensor web enablement standard, creating a multi-hop mesh networking topology. In order to achieve this, they present an intermediate processing and communication module in between the sensors and the management servers. Another video surveillance system aiming at lower-end customers, based on commercially available cloud computing is presented by Rodriguez-Silva et al. (2012) focusing on the optimization of the transmission rates according to available network bandwidth.

Segor et al. (2012) present a generalized video surveillance system architecture with the focus on achieving maximum reusability. This makes it simple and cost-effective to integrate new requirements such as hardware or software components at a later stage. A prototype for an intelligent video surveillance system is introduced by Ning et al. (2012). It features a client-server system architecture and has the ability of selective video archiving to store only relevant data by motion analysis.
Vezzani and Cucchiara (2010a) analyze the main issues regarding the development of distributed surveillance systems and propose an integrated framework particularly suitable for research purposes, featuring a three layer tracking system, which copes with the integration of both overlapping and non-overlapping cameras. The system is based on an event-driven communication infrastructure, which assures scalability and flexibility, and features a static service-oriented architecture for classification tasks. Saini et al. (2009) present a flexible surveillance system architecture which can easily be ported to different environments, is dynamic without compromising system performance, and can be extended to integrate newer technological developments in the future. This is achieved by applying an abstraction layer, which is referred to as an environmental model in order to adapt quickly to a new scenario or to incorporate new types of sensors. Table 3 summarizes publications on video surveillance system architectures presented in this section.

Table 3. Summary of publications on video surveillance system architectures.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video surveillance platform for public safety</td>
<td>Li et al. (2013)</td>
</tr>
<tr>
<td>An intelligent surveillance platform for inexpensive sensors</td>
<td>Fernández et al. (2013)</td>
</tr>
<tr>
<td>Cloud-based architecture for license plate recognition</td>
<td>Chen et al. (2013)</td>
</tr>
<tr>
<td>Cloud-based video surveillance architecture with 3rd party application</td>
<td>Chattopadhyay et al. (2013)</td>
</tr>
<tr>
<td>Video surveillance architecture as ubiquitous sensor network</td>
<td>Yu et al. (2012)</td>
</tr>
<tr>
<td>Peer-to-peer-based cloud storage architecture</td>
<td>Wu et al. (2012)</td>
</tr>
<tr>
<td>Video surveillance architecture focusing on reusability</td>
<td>Segor et al. (2012)</td>
</tr>
<tr>
<td>Optimized video transmission in cloud-based video surveillance</td>
<td>Rodríguez-Silva et al. (2012)</td>
</tr>
<tr>
<td>System for selective data storage</td>
<td>Ning et al. (2012)</td>
</tr>
<tr>
<td>Cloud-based video recording system</td>
<td>Lin et al. (2012)</td>
</tr>
<tr>
<td>Multimedia surveillance back-end system architecture</td>
<td>Dey et al. (2012)</td>
</tr>
<tr>
<td>Peer-to-peer-based video surveillance architecture</td>
<td>Zhu et al. (2011)</td>
</tr>
<tr>
<td>Architecture flexible for environmental adaption</td>
<td>Saini et al. (2009)</td>
</tr>
<tr>
<td>Framework for research purposes in surveillance systems</td>
<td>Vezzani and Cucchiara (2010a)</td>
</tr>
<tr>
<td>Multi-sensor surveillance system</td>
<td>Kanade et al. (1999)</td>
</tr>
<tr>
<td>A system for airport ground monitoring</td>
<td>Corrall (1991)</td>
</tr>
</tbody>
</table>
2.2 Communication tier

Before analyzing prior research, the most common data transmission standards used in digital video surveillance are reviewed. First, most commonly applied wired standards are introduced, followed by wireless standards:

- **Fiber Optics**: The study of coded modulation techniques for fiber optic networks by Beygi et al. (2014) revealed that bitrates ranging from 100Gbit/s up to 300Gbit/s are possible over fiber optical media.

- **Ethernet LAN**: This wired technology started in the early 70s with the IEEE 802.3, providing up to 2.9Mbit/Sec, until the 1000bT providing up to 1Gbit/Sec was introduced. Then the 10GBase followed, providing up to 10 Gbit/Sec, and finally 100Gbit/Sec (Duelk & Zirngibl 2006). Thus, Ethernet LAN technology qualifies for all communication connections and is the most widely used in practice.

- **DSL Technique**: This wire line technology utilizing telephone connections offers up to 1 Gbit/Sec (Timmers et al. 2013). Thus, DSL technology qualifies for all communication connections as a substitute if neither fiber optics nor Ethernet LAN are available. It is the most commonly used technique in consumer-level, cloud-based installations.

- **Wi-Fi**: Wireless Fidelity, which is a class of WLAN based on the IEEE 802.11 standards. Wi-Fi offers a wide spectrum of bit-rates, started at 2Mbit/Sec in the original specification and increased over 6 Gbit/Sec in the current IEEE 802.11ac standard (Kelly 2014). Thus, Wi-Fi qualifies for wireless connections within a video surveillance system.

- **Wireless Mobile Technologies**: Taking the possibility of using the video surveillance systems also within regions with low technology penetration into account, 3G technologies can be built upon:
  - **UMTS**: Provides a bit-rate up to 384 kbit/Sec in the original release. This is only adequate for status reports and low quality transmission of single cameras in video surveillance, whereas UMTS HSPA+ providing up to 42 Mbit/Sec transmission rates is sufficient for transmission of multiple medium to high quality channels video (Haider et al. 2011).
  - **Finally, the two high-throughput technologies**: Wireless Metropolitan Area Networks (commercially WiMax), based on the IEEE 802.16 standard, provides over 360 Mbit/Sec up- and downstream and by combining multiple channels allows up to 1 Gbit/Sec (Forum 2010), and
LTE/LTE Advanced (4G), which also allows peak rates of up to 1 Gbit/Sec downlink and 500 Mbit/Sec uplink (Abdullah & Yonis 2012). Both are sufficient for high quality video transmission.

Table 4 summarizes the presented communication technologies.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Range</th>
<th>Max. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiber Optics</td>
<td>Direct connection</td>
<td>300 Gbit/Sec</td>
</tr>
<tr>
<td>Ethernet</td>
<td>Direct connection (local)</td>
<td>100 Gbit/Sec</td>
</tr>
<tr>
<td>DSL</td>
<td>Direct connection (wide area network)</td>
<td>1 Gbit/Sec</td>
</tr>
<tr>
<td>Wi-Fi</td>
<td>Short-range wireless</td>
<td>6 Gbit/Sec</td>
</tr>
<tr>
<td>UMTS</td>
<td>Cellular Wireless</td>
<td>42 Mbit/Sec</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Cellular Wireless</td>
<td>1 Gbit/Sec</td>
</tr>
<tr>
<td>LTE</td>
<td>Cellular Wireless</td>
<td>1 Gbit/Sec</td>
</tr>
</tbody>
</table>

Utilizing mobile wireless technologies, Fasui et al. (2013) present a system consisting of mobile devices, organized as a fault tolerant network, capable of automatic monitoring and detection of intrusions based on a client-server system architecture. The client application is created for Android devices implementing a challenge-response protocol for authentication. A case study of real-time data transmission in mobile 2G and 3G networks is given by Sandu et al. (2010). The importance of Quality of Service for data streaming, which is essential for video surveillance, is shown. Further, in Shakeel and Ramani (2007) the requirements for mobile ad-hoc networks are thoroughly investigated.

Valera et al. (2011) present a framework for the design of visual surveillance systems based on components derived from the principles of real-time networks/data-oriented requirements implementation scheme. They further propose the implementation of these components using the common object request broker architecture (CORBA) middleware technology. A network model for video surveillance featuring a false data sensitive protocol is presented by Jeong and Yang (2011). The protocol tests every packet and every image sent from wireless cameras authentically and semantically, thus malicious messages can only travel one hop. In addition to reducing the possibility of key disclosure, a dynamic key scheme that lets wireless sensors employ a different key in each communication is suggested.

Chang et al. (2010) present a distributed and secure architecture for cloud-based video surveillance services, especially protecting service privacy without
compromising scalability. Here, a symmetric encryption mechanism, multicast overlay network and forward error correction are suggested. An approach towards ad-hoc video surveillance with the focus on situations like natural disasters is proposed by Asahizawa et al. (2009). A configuration of airborne cameras using helium balloons connected via wireless LANs is shown, providing an overview of a large area. A mix of omnidirectional and PTZ cameras is used.

Kim and Han (2012) analyze different network security-based threats for IP-based video surveillance systems and present a security model to overcome these, while Chen et al. (2009) present a system architecture supporting Quality of Service, re-compensating the instability in the available mobile bandwidth. A token generation algorithm achieves this in a self-adaptive manner. The system has been prototyped and showed improvement in instable network configurations.

A framework of hybrid mobile-fixed video surveillance systems is introduced by Dao et al. (2012), utilizing wireless wearable cameras in addition to other wireless cameras, while Räty et al. (2008) introduce an intelligent video transmission client-server system, ideal for mobile clients. Depending on the user’s area of interest, all transmitted video data is scaled only to the necessary resolution, considerably reducing the required bandwidth without compromising situational awareness and saving time in critical situations. Table 5 summarizes publications on the communication tier presented in this section.

Table 5. Summary of publications on the communication tier.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fault tolerant network for mobile devices</td>
<td>Fasui et al. (2013)</td>
</tr>
<tr>
<td>Network security-based threat analysis</td>
<td>Kim and Han (2012)</td>
</tr>
<tr>
<td>Framework of hybrid mobile-fixed video surveillance systems</td>
<td>Dao et al. (2012)</td>
</tr>
<tr>
<td>Corba-based surveillance middleware</td>
<td>Valera et al. (2011)</td>
</tr>
<tr>
<td>A false data sensitive protocol</td>
<td>Jeong and Yang (2011)</td>
</tr>
<tr>
<td>A case study of real-time data transmission in mobile networks</td>
<td>Sandu et al. (2010)</td>
</tr>
<tr>
<td>Encryption for privacy in cloud-based video surveillance</td>
<td>Chang et al. (2010)</td>
</tr>
<tr>
<td>Quality of service compensating instability in mobile bandwidth</td>
<td>Chen et al. (2009)</td>
</tr>
<tr>
<td>Airborne cameras using helium balloons</td>
<td>Asahizawa et al. (2009)</td>
</tr>
<tr>
<td>Intelligent video transmission for mobile clients</td>
<td>Räty et al. (2008)</td>
</tr>
<tr>
<td>Requirements for mobile ad hoc networks</td>
<td>Shakeel and Ramani (2007)</td>
</tr>
</tbody>
</table>
2.3 Sensor tier

In this section research concerning the sensor tier is presented, including different smart camera approaches and innovative sensors such as presented by Tang et al. (2014). They introduce an omni-directional vision sensor that provides a 360° panoramic image for obtaining situation awareness for elderly people at home in real-time to detect various behavioral abnormality, such as tripping and falling or other medical emergencies. Another omni-directional camera design, targeted at surveillance purposes is presented by Kon et al. (2012).

A multi-modal wireless smart camera is introduced by Magno et al. (2013). It is equipped with a pyroelectric infrared sensor and a solar energy harvester in order to create a low-power, low-cost wireless video sensor node with on-board video processing, whereas Brzoza-Woch et al. (2013) built a reconfigurable hardware platform for general-purpose video processing. It is targeted towards portable, web-enabled devices and is based on programmable logic, microcontrollers and dedicated communication modules.

Hamida et al. (2013) present a spatiotemporal video filtering method to filter captured data without altering information of interest within a smart camera. Their approach claims to improve the H.264 encoding phase by reducing prediction error values during the motion estimation step, which leads to a faster and better compression. An approach for compressing the data within each pixel prior to storage in order to reduce the size of the required memory using CMOS image sensors, hence reducing required silicon area, is presented by Milin and Bermak (2010).

Cameras with on-board processing and wireless communication are introduced by Baoxia et al. (2013) and Ye et al. (2012), while Sánchez et al. (2012) present a hardware and software architecture based on a camera design with multiple DSPs and an FPGA with a flexible modular processing pipeline. Chunjing et al. (2012), on the other hand, present a smart camera specifically designed for automated traffic analysis, which includes the calculation of average vehicle speed, traffic density or dynamic flow information. A verification platform for evaluation and comparison of smart camera designs is presented by Lee and Jeong (2012), whereas Qureshi (2011) develops smart camera nodes with dynamic resolutions and frame rates in order to minimize energy consumption. Another low power smart camera implementation is presented by Bartosinski et al. (2012).
A comprehensive collection of work on smart cameras is given by Belbachir (2009). Smart cameras as the front-end of an intelligent sensor network are investigated and different sensor types are evaluated, including entire processing pipelines optimized to run as a “system on a chip”. Table 6 summarizes publications on the sensor tier presented in this section.

Table 6. Summary of publications on sensor tier.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omni-directional 360° panoramic camera</td>
<td>Tang et al. (2014)</td>
</tr>
<tr>
<td>Multimodal wireless smart camera</td>
<td>Magno et al. (2013)</td>
</tr>
<tr>
<td>Spatiotemporal video filtering for improved H.264 compression</td>
<td>Hamida et al. (2013)</td>
</tr>
<tr>
<td>Reconfigurable hardware platform for video processing</td>
<td>Brzoza-Woch et al. (2013)</td>
</tr>
<tr>
<td>Camera with on-board processing and wireless communication</td>
<td>Baoxia et al. (2013)</td>
</tr>
<tr>
<td>Camera with on-board processing and wireless communication</td>
<td>Ye et al. (2012)</td>
</tr>
<tr>
<td>Camera with flexible hardware and software architecture</td>
<td>Sánchez et al. (2012)</td>
</tr>
<tr>
<td>Verification platform for evaluation of smart cameras</td>
<td>Lee and Jeong (2012)</td>
</tr>
<tr>
<td>Omni-directional camera design</td>
<td>Kon et al. (2012)</td>
</tr>
<tr>
<td>Camera for automated traffic analysis</td>
<td>Chunjing et al. (2012)</td>
</tr>
<tr>
<td>Low power smart camera</td>
<td>Bartosinski et al. (2012)</td>
</tr>
<tr>
<td>Smart camera nodes with dynamic resolutions and frame rates</td>
<td>Qureshi (2011)</td>
</tr>
<tr>
<td>Compressing data within each pixel prior to storage</td>
<td>Millin and Bermak (2010)</td>
</tr>
<tr>
<td>Comprehensive collection of work on smart cameras</td>
<td>Belbachir (2009)</td>
</tr>
</tbody>
</table>

2.4 Operator / HCI tier

Robust filtering of relevant information is only one side of a successful video surveillance system and the way in which this information is presented to the operator is often just as important. Hence, Feng et al. (2013) propose a framework targeted at bridging between user applications and video analysis based sensor networks by semantic descriptions. This shall allow operators with less technical knowledge to leverage more complex underlying processing algorithms.

The problem of manual labeling work in interactive retrieval systems is addressed by Meessen et al. (2009). In this approach, an active learning technique is proposed for the selection of results, which is constantly learning from the security authority’s interactions utilizing support vector machines. Further, different user interface layouts are compared in the evaluation.

The often inefficient task of seeking through recorded video is addressed by Aydemir et al. (2012). They introduce an approach towards video summarization
based on human movement understanding. Sun et al. (2013) present another hierarchical video summarization method for surveillance video in which synopsis videos are generated for a quicker overview. Users can efficiently browse through video content and dynamically zoom into collage images and skim through synopsis video or original video for details.

Web-based video surveillance systems are presented by Rashmi and Latha (2013) as well as Imai et al. (2008), implementing IP-cameras as sensors and utilizing different mobile phones as front-end for the operators. Another mobile phone application that enables several mobile devices to form a sensor network dedicated to video surveillance and automatic intrusion detection is presented by Fasui et al. (2013). Wang et al. (2012c) present a system that integrates different video analysis algorithms for video surveillance, followed by a multi-tier visualization scheme in order to optimize the information queues presented to the operators in wide-area city surveillance.

A system featuring a predictive camera selection for assisting video surveillance operators is introduced by Martinel et al. (2013). By analyzing video streams the system preemptively displays cameras that will most likely be of interest to the operator, whereas Filonenko et al. (2013) present an autonomous monitoring system for supervising environmental parameters, including temperature, humidity, poisonous gas concentration, with the ultimate goal to develop an augmented reality system for environmental conditions surveillance on mobile devices.

Vural and Akgul (2012) present a surveillance system that continuously learns the properties of objects that are interesting for a human operator. The system automatically learns the operator’s preferences by tracking eye gaze positions while the operator monitors the surveillance video, while Kuhn et al. (2012b) address image retrieval in large-scale archives of video data employing a human-machine hybrid process. They propose a variety of visualizations for a better interpretive power by the operator as well as customizable iterative querying based on tagged key words.

A system based on video analysis which is targeted at the retail sector is presented by Senior et al. (2007). They combine video analysis and transaction logging, resulting in a comprehensive set of reporting and loss prevention applications. Table 7 summarizes publications on the security authority tier presented in this section.
Table 7. Summary of publications on the operator / HCI tier.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical video summarization</td>
<td>Sun et al. (2013)</td>
</tr>
<tr>
<td>Web-based video surveillance system using mobile phones</td>
<td>Rashmi and Latha (2013)</td>
</tr>
<tr>
<td>Predictive camera selection</td>
<td>Martinel et al. (2013)</td>
</tr>
<tr>
<td>Monitoring system for supervising environmental parameters</td>
<td>Filonenko et al. (2013)</td>
</tr>
<tr>
<td>Bridging between user applications and video analysis</td>
<td>Feng et al. (2013)</td>
</tr>
<tr>
<td>Several mobile devices as a sensor network</td>
<td>Fasui et al. (2013)</td>
</tr>
<tr>
<td>Multi-tier visualization scheme</td>
<td>Wang et al. (2012c)</td>
</tr>
<tr>
<td>Tracking eye gaze positions for self learning system</td>
<td>Vural and Akgul (2012)</td>
</tr>
<tr>
<td>Image-retrieval in large-scale archives</td>
<td>Kuhn et al. (2012b)</td>
</tr>
<tr>
<td>Video summarization based on human movement understanding</td>
<td>Aydemir et al. (2012)</td>
</tr>
<tr>
<td>Active learning technique for labeling in retrieval systems</td>
<td>Meessen et al. (2009)</td>
</tr>
<tr>
<td>Web-based video surveillance system using mobile phones</td>
<td>Imai et al. (2008)</td>
</tr>
<tr>
<td>Video analysis and transaction-logging for retail</td>
<td>Senior et al. (2007)</td>
</tr>
</tbody>
</table>

2.5 Storage and retrieval tier

Addressing the task of efficient retrieval in video surveillance, Xue et al. (2012) present an ontology-based content archive and retrieval framework for surveillance videos where video data is analyzed to form description files in Web Ontology Language (OWL). Presti et al. (2012), on the other hand, propose a framework for querying a distributed video surveillance database in order to retrieve a set of likely paths of a person moving in the area under surveillance. Operators can further pose queries about where a certain person appeared while moving on the site during a specified temporal window. This is presented with the retrieved set of likely paths in ranked order.

A system for license plate number retrieval in urban video surveillance employing ubiquitous cameras with automated license plate recognition is presented by Shu et al. (2012). They employ an optimization problem formulation using only the shortest driving distance between pairs of cameras to calculate a global optimum. Thus, traffic violations such as speeding can be detected. Pankanti et al. (2012) present a more generalized computer video based system for a variety of different data analysis tasks, including finding people and cars in unconstrained images and analysis of retail compliance work.

Vezzani and Cucchiara (2010b) present an open platform framework for the collection, annotation and retrieval of surveillance videos. The framework allows video browsing by querying and compressed video previewing, and is targeted
towards research projects, whereas Feris et al. (2012) present a framework for learning specific multi-view object detectors from large datasets for video indexing and retrieval purposes, demonstrated for tracking vehicles. They use an algorithm based on a cascade of complementary classifiers, learning from datasets larger than a hundred thousand positive and negative examples. In order to improve data storage, a compression technique to improve the coding efficiency of H.264 for surveillance videos is presented by Cheok and Gagvani (2010). There, video content is analyzed and video semantics are extracted using video analysis algorithms, in order to store only relevant object information.

The large-scale European research project CANDELA, which has been described by Pietarila (2005), focuses on the integration of video analysis and network/storage technologies in a service-oriented manner, similar to the work presented by Moßgraber et al. (2010), who introduce an innovative video storage and retrieval architecture, focusing on adaptive resource allocation applying human-like reasoning by the application of ontologies. The system provides a single coherent scene model, which for instance allows tracking results from video analysis of individual cameras to be projected. Table 8 summarized the presented research on the storage and retrieval tier.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology-based content archive and retrieval framework</td>
<td>Xue et al. (2012)</td>
</tr>
<tr>
<td>A system for license plate number retrieval</td>
<td>Shu et al. (2012)</td>
</tr>
<tr>
<td>A framework for querying a distributed video surveillance database</td>
<td>Presti et al. (2012)</td>
</tr>
<tr>
<td>Data analysis system</td>
<td>Pankanti et al. (2012)</td>
</tr>
<tr>
<td>Multi-view object detectors for indexing and retrieval</td>
<td>Feris et al. (2012)</td>
</tr>
<tr>
<td>Open platform framework retrieval of surveillance videos</td>
<td>Vezzani and Cucchiara (2010b)</td>
</tr>
<tr>
<td>Event-driven service-oriented data analysis</td>
<td>Moßgraber et al. (2010)</td>
</tr>
<tr>
<td>Object based video compression</td>
<td>Cheok and Gagvani (2010)</td>
</tr>
<tr>
<td>Storage and retrieval based on adaptive resource allocation</td>
<td>Pietarila (2005)</td>
</tr>
</tbody>
</table>

2.6 Processing tier

The vast field of research regarding sensor processing is presented in the following. It ranges from entire processing pipelines to single-purpose algorithms and applications from a user-centric view, such as Liu et al. (2013) present. They give a comprehensive survey on the theory and applications of intelligent video systems and video analytics, highlighting the processing architectures and
analysis algorithms. It demonstrates the value of intelligent video systems in a variety of domains, such as transportation and surveillance. Lewis (2012) presents a survey of video analysis solutions from the point of view of a government advisor, concluding that even though research in this field has been active for many years, it is expected that the technology has gained enough confidence and will be deployed on a larger scale. Argioli and Bisogni (2011) present a state-of-the-art review for the European Commission, evaluating the performance and reliability of different video analysis tasks, ranging from people counting to face recognition. They suggest the implementation of a standardized system of evaluation, functioning as driver to obtain users’ trust toward video analysis systems.

A survey of video analysis technology for critical infrastructure protection, particularly the transportation sector, is given by Thornton et al. (2009). Based on discussions with security personnel at multiple facilities, a list of desired video analysis functionality was assessed, while a survey of behavior analysis is presented by Ko (2008), ranging from low-level computer vision to high-level artificial intelligence algorithms. Table 9 gives an overview of the presented surveys regarding the processing tier.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory and applications of intelligent video systems</td>
<td>Liu et al. (2013)</td>
</tr>
<tr>
<td>Video analysis from the point of view of a government advisor</td>
<td>Lewis (2012)</td>
</tr>
<tr>
<td>State-of-the-art review for the European Commission</td>
<td>Argioli and Bisogni (2011)</td>
</tr>
<tr>
<td>Video analysis technology for critical infrastructure protection</td>
<td>Thornton et al. (2009)</td>
</tr>
<tr>
<td>Survey on behavior analysis</td>
<td>Ko (2008)</td>
</tr>
</tbody>
</table>

### 2.6.1 Detection and tracking

As sensor processing is a broad field, it has been structured further into the most important and researched sub-tasks, starting with low-level processing, which in most computer vision systems starts with object detection based on background modeling and tracking. Feris et al. (2013) present such an approach towards automatically creating object detectors trained specifically for individual video surveillance cameras, utilizing samples acquired with the help of computationally more expensive, general-domain detectors. The approach is demonstrated for the task of vehicle detection in crowded surveillance videos. Lijun and Kaiqi (2013),
on the other hand, present a method for crowd density estimation and prediction for situations where it becomes impossible to detect individual persons due to massive occlusions. Their approach is based on optical flow and temporal statistics and has been employed in practical scenarios such as a park, a plaza or a public transport station.

Zhang et al. (2012b) present a method for detecting individual pedestrians in surveillance videos by the combination of motion information, human skin color information, human shape information and variation of ambient lighting, whereas Kushwaha et al. (2012) propose another algorithm for detection and tracking of people in video by combining trained classifiers based on haar-like features with a particle filter tracker. A mean shift based algorithm for the tracking of people in surveillance video is presented by DiCaterina and Soraghan (2011). Their approach is capable of handling partial and complete occlusions, and is thus also applicable in more crowded scenes. An evaluation framework that attempts to provide qualitative information about the tracker performance is presented by Sankaranarayanan et al. (2012), while Brutzer et al. (2011) compare the performance of nine different background subtraction methods, evaluating the strengths and weaknesses of each algorithm utilizing a ground-truth annotated data set.

Rajpurohit et al. (2012) present a specialized method for motion detection avoiding the use of a background model, utilizing inter-frame differencing on pixel level to obtain spatiotemporal histograms which are analyzed to determine the significance of motion, while Haque and Murshed (2012) propose a new background subtraction technique, for the detection of objects in videos. They use a Mixture-of-Gaussians per pixel to model a multiple-hypothesis classifier, which results in a dynamic statistical background image that can then be differenced from each incoming frame. A detection and tracking framework for video surveillance is proposed by del-Blanco et al. (2012), focusing on simple installation and configuration, and unsupervised working conditions, utilizing a Bayesian tracking model that manages multimodal distributions. Results of an outdoor intrusion detection evaluation based on video analysis at Sandia National Laboratories are presented by Norman (2012). The probability of detection as well as false alerts over a six-month test period are analyzed.

Lai et al. (2012) present a new framework for detecting missing and unattended objects in complex environments for video surveillance systems, based on detecting the center of mass of static foreground objects. They claim that the method is able to handle illumination change, repetitive background and
occlusions in complex crowded environments. Fan and Pankanti (2012), on the other hand, present a system for large-scale abandoned object detection, focusing their work on robust foreground segmentation, regardless of lighting changes, low texture contrast and cluttered background. Takala et al. (2010) propose a view-independent object-tracking algorithm. A wide collection of features is extracted from the object of interest (such as shape, pose, color, texture) of which a model is created. These features are boosted, resulting in an individual classifier for each specific object. Huang and Tan (2010) introduce a system with a sophisticated computer vision-processing pipeline. Their system consists of object detection, object classification and trajectory tracking. Three-dimensional human body models are applied, which allow features such as gait analysis. This is followed by a behavior analysis stage, which applies ontologies for modeling different situations, trying to detect objects’ behaviors. Table 10 summarizes prior research on detection and tracking.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training of camera specific object detectors</td>
<td>Feris et al. (2013)</td>
</tr>
<tr>
<td>Method for detecting pedestrians</td>
<td>Zhang et al. (2012b)</td>
</tr>
<tr>
<td>Evaluation framework for tracking algorithms</td>
<td>Sankaranarayanan et al. (2012)</td>
</tr>
<tr>
<td>Motion detection without a background model</td>
<td>Rajpurohit et al. (2012)</td>
</tr>
<tr>
<td>Exterior intrusion detection evaluation</td>
<td>Norman (2012)</td>
</tr>
<tr>
<td>Framework for detecting missing and unattended objects</td>
<td>Lai et al. (2012)</td>
</tr>
<tr>
<td>Algorithm for detection and tracking of people</td>
<td>Kushwaha et al. (2012)</td>
</tr>
<tr>
<td>MoG based background subtraction technique</td>
<td>Haque and Murshed (2012)</td>
</tr>
<tr>
<td>System for large-scale abandoned object detection</td>
<td>Fan and Pankanti (2012)</td>
</tr>
<tr>
<td>Bayesian tracking model</td>
<td>del-Blanco et al. (2012)</td>
</tr>
<tr>
<td>Mean shift-based person tracker</td>
<td>DiCaterina and Soraghan (2011)</td>
</tr>
<tr>
<td>Performance evaluation of background models</td>
<td>Brutzer et al. (2011)</td>
</tr>
<tr>
<td>View-independent object tracking algorithm</td>
<td>Takala et al. (2010)</td>
</tr>
<tr>
<td>Computer vision processing pipeline</td>
<td>Huang and Tan (2010)</td>
</tr>
</tbody>
</table>

### 2.6.2 Activity recognition

Opposed to anomaly recognition, activity recognition tries to detect predefined states or actions in a video. An introductory survey on this topic is given by Taha et al. (2013), where the performances of the different algorithms on a number of
datasets are compared, while Wolf and Baskurt (2012) present a survey of dominant models and methods, and discuss recent developments in this domain.

A specific algorithm for detecting a person carrying an uncovered firearm based on a trained neural network classifier operating a shape appearance is proposed by Grega et al. (2013). Basically, it tries to detect the outlines of a person holding a gun in front of them. Esen et al. (2013) propose an algorithm for fight detection utilizing a motion block matching algorithm, which is then used for spatiotemporal clustering in order to detect significant outliers which imply persons fighting in the video, whereas Wang et al. (2012a) present an approach for detecting violence in general by creating a mixed feature classifier. They operate on single images instead of a video stream.

Cristani et al. (2013) reach beyond computer vision for analyzing human behavior in video by exploring social signal processing, which incorporates principles from the social, affective, and psychological literature, which leads to the inclusion of face expressions and gaze, body posture and gestures, vocal characteristics and relative distances in space, while Choudhary et al. (2012) address the fact that activities strongly rely on their temporal significance in an area under surveillance. They propose a model that can discover and learn the activities as well as their hidden dependencies on particular periods of time. Xu et al. (2012) present an approach towards recognizing different activities in crowded scenes by analyzing particle trajectories, which estimates long range motion opposed to optical flow methods, whereas Tu et al. (2012) present an approach for creating action models for the detection of different human activities in videos. Patino et al. (2012), on the other hand, introduce a method for activity recognition in video by the modeling of context. They propose an algorithm that automatically learns and then analyzes the main activity zones of an observed scene.

Wang et al. (2011) introduce a view-invariant measure for human action recognition. They apply video synchronization by imposing both the similarity ratio and the consistency in a trifocal tensor over entire video sequences, while Rougier et al. (2011) introduce a specialized method for detecting humans falling on the ground by analyzing human shape deformation during a video sequence. They apply a shape matching technique to track the person’s silhouette along the video sequence and analyze the deformation of these silhouettes.

Weiyao et al. (2008) present an innovative approach to behavior recognition in video surveillance by describing activities like walking, running or fighting with specific trained models. Their focus lies on reducing the necessary training
data for classifiers, as well as creating a model which allows new activities to be added at a later stage without re-starting of the entire training phase. Another approach to recognizing human actions is presented in Kellokumpu et al. (2009) by utilizing a spatiotemporal texture classifier using local binary patterns. In this two-phase process, first, movement dynamics are captured, followed by a characterization of these observed movements. Table 11 summarizes prior research on activity recognition.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey on behavior analysis</td>
<td>Taha et al. (2013)</td>
</tr>
<tr>
<td>Detection of fire arms</td>
<td>Grega et al. (2013)</td>
</tr>
<tr>
<td>Fight detection using motion blocks</td>
<td>Esen et al. (2013)</td>
</tr>
<tr>
<td>Social, affective, and psychological principles for behavior analysis</td>
<td>Cristani et al. (2013)</td>
</tr>
<tr>
<td>Particle trajectory analysis for activity recognition</td>
<td>Xu et al. (2012)</td>
</tr>
<tr>
<td>Survey on action recognition</td>
<td>Wolf and Baskurt (2012)</td>
</tr>
<tr>
<td>Violence detection on still images</td>
<td>Wang et al. (2012a)</td>
</tr>
<tr>
<td>Action models for detecting different human activities</td>
<td>Tu et al. (2012)</td>
</tr>
<tr>
<td>Activity zones for activity recognition</td>
<td>Patino et al. (2012)</td>
</tr>
<tr>
<td>Time-based activity recognition</td>
<td>Choudhary et al. (2012)</td>
</tr>
<tr>
<td>A view-invariant measure for human action recognition</td>
<td>Wang et al. (2011)</td>
</tr>
<tr>
<td>Silhouette-based fall detection</td>
<td>Rougier et al. (2011)</td>
</tr>
<tr>
<td>Spatiotemporal texture classifier using local binary patterns</td>
<td>Kellokumpu et al. (2009)</td>
</tr>
<tr>
<td>Reducing the necessary training data for activity classifiers</td>
<td>Weyao et al. (2008)</td>
</tr>
</tbody>
</table>

### 2.6.3 Anomaly recognition

Anomaly recognition is the task of detecting a state or action in an observed scene which deviated from what has been learned as normal. Addressing this issue, Li et al. (2014) propose a detector for anomalous behavior detection in crowded scenes by analyzing temporal and spatial anomalies, based on a video representation that accounts for both appearance and dynamics, using a set of mixture of dynamic textures models. Zhu et al. (2013), on the other hand, propose a framework to jointly model-related activities with both motion and context information for activity recognition and anomaly detection. They utilize a structural model to learn the motion and context patterns within and across activity classes from training sets.
A framework mining semantic context information in intelligent video surveillance of traffic scenes is introduced by Zhang et al. (2013). For each kind of object, corresponding semantic scene-specific context information, such as motion pattern, width distribution, paths, and entry/exit points, are learned in order to subsequently detect outliers. Wang and Snoussi (2012) also propose an algorithm for detecting abnormal events in video streams. Their algorithm is based on histograms of the orientation of optical flow descriptors fed to a support vector machine classifier.

Sadeghi-Tehran and Angelov (2012) present an algorithm for detecting and tracking objects in a video and detecting anomalies by clustering a multi-feature object trajectory and detecting outliers, while Nguyen et al. (2012) propose a framework for non-parametric data segmentation and multi-modal abnormality detection by multiple detection models focused on different coherent sections of the video stream. Data is segmented over each day into multiple contiguous sections based on the coherence of the total activity levels for which anomaly detectors are constructed. Trinh et al. (2012) go a step further and present an approach detecting non-compliant activities in a retail surveillance system in order to reduce shrinkage. They combine video analysis with transaction logs and train a support vector machine classifier in order to generate alerts.

Lin and Wu (2012) present an algorithm for camera tampering detection in video surveillance by detecting large edge differences and gray scale histogram comparisons between frames. Thus, they are able to detect manipulation of the camera, such as occlusion, defacing or ego-motion, while Wang and Fang (2008) propose a surveillance system that incorporates the interpretation of facial expressions in people’s faces, arguing that such a system could be capable of saving time in order to prevent crime. Mueller et al. (2011), on the other hand, focus in a combined approach on the detection of fire in videos using prior knowledge and online-learning.

A general framework for anomalous event detection in uncrowded scenes is presented by Li et al. (2012). They apply statistical scene modeling, which combines trajectory-based and region-based information for fully unsupervised anomaly detection. Kuhn et al. (2012a) present a framework that combines optical flow and Lagrangian analysis of time-dependent vector fields in order to capture complex dynamic motion for automated detection of abnormal occurrences.

Hommes et al. (2011) introduce an approach towards unsupervised detection of abnormal events in video surveillance data by initially allowing a training
phase in which spatial distribution of objects is learned and subsequently outliers are detected in a chart-based tracking process, similar to Jeong et al. (2011), who propose an algorithm for abnormal behavior detection by trajectory analysis. They cluster spatial information into semantic regions in order to detect outliers, avoiding the comparison of every single trajectory. Table 12 summarizes prior research on anomaly recognition.

Table 12. Summary of publications on anomaly recognition.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture of dynamic textures models</td>
<td>Li et al. (2014)</td>
</tr>
<tr>
<td>Modeling of related activities with motion and context information</td>
<td>Zhu et al. (2013)</td>
</tr>
<tr>
<td>Mining semantic context information</td>
<td>Zhang et al. (2013)</td>
</tr>
<tr>
<td>Histograms of the orientation and SVM</td>
<td>Wang and Snoussi (2012)</td>
</tr>
<tr>
<td>Detection of non-compliant activities in retail surveillance</td>
<td>Trinh et al. (2012)</td>
</tr>
<tr>
<td>Framework for nonparametric data segmentation and multi-modal anomaly detection</td>
<td>Nguyen et al. (2012)</td>
</tr>
<tr>
<td>Camera tampering detection</td>
<td>Lin and Wu (2012)</td>
</tr>
<tr>
<td>Unsupervised anomaly detection in uncrowded scenes</td>
<td>Li et al. (2012)</td>
</tr>
<tr>
<td>Optical flow and Lagrangian analysis of time-dependent vector fields for anomaly detection</td>
<td>Kuhn et al. (2012a)</td>
</tr>
<tr>
<td>Fire detection</td>
<td>Mueller et al. (2011)</td>
</tr>
<tr>
<td>Abnormal behavior detection by trajectory analysis</td>
<td>Jeong et al. (2011)</td>
</tr>
<tr>
<td>Unsupervised detection of abnormal events</td>
<td>Hommes et al. (2011)</td>
</tr>
<tr>
<td>Interpretation of facial expressions</td>
<td>Wang and Fang (2008)</td>
</tr>
</tbody>
</table>

### 2.6.4 Multi-camera processing

Approaches towards extracting information from multiple cameras are presented in this section. Zhang et al. (2012a) present such a distributed video surveillance framework which allows tracking of multiple objects across multiple cameras by calibrating camera views and brightness. Wang et al. (2012b), on the other hand, present a computer vision processing architecture that estimates objects’ locations based on the fusion of the information gained from multiple cameras to generate a 3D tracking process which incorporates multiple views of the same objects.

Palmieri et al. (2013) present a method for camera calibration in harbor surveillance utilizing AIS (Automatic Identification System) data for matching pixel locations in cameras with overlapping views. Due to mechanics, PTZ cameras are prone to mechanical and random errors after longer operation periods
and thus initial calibration settings would no longer be accurate. Wu and Radke (2012) introduce a dynamic correction method for keeping such PTZ cameras calibrated in wide-area surveillance networks.

Snidaro et al. (2012) introduce a fusion framework for combining data from multiple, possibly heterogeneous sensors observing a surveillance area. They focus on feature level data fusion utilizing an array of features such as Haar-like features, local binary patterns and color histograms, while Martelli et al. (2012) propose a set of features suitable for pedestrian detection in stereo settings, i.e., when range information is also available. They pool visual features, such as human body parts and shape, in order to overcome occlusions. Hong et al. (2012) present a method for recognizing and tracking vehicles utilizing homography, thus projecting the views of multiple cameras onto one coherent image, similar to Behera et al. (2012), who present a multi-camera surveillance system which features overlapping camera views to overcome occlusions. They apply a handshake algorithm to hand off objects from one camera to the next.

Abdelkader et al. (2012) present an approach to maximize visual coverage by combining static wide-angle cameras which then remotely control PTZ-cameras to focus on targets of interest and track these in a detailed view, whereas Starzyk and Qureshi (2011) present a video surveillance system comprising of PTZ cameras that function either as passive, static overview cameras or active focus cameras. The cameras automatically tune sensing parameters in response to the scene activity and switch roles according to the activity in the observed scene. An approach utilizing multiple miniaturized cameras is presented by Pham and Makhoul (2010). They strive to achieve visual coverage by randomly distributing many of these sensors. Table 13 summarizes the presented prior research on multi-camera processing.
Table 13. Summary of publications on multi-camera processing.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera calibration utilizing AIS data</td>
<td>Palmieri et al. (2013)</td>
</tr>
<tr>
<td>Multi-camera multi-object tracking</td>
<td>Zhang et al. (2012a)</td>
</tr>
<tr>
<td>Dynamic correction of calibration for PTZ cameras</td>
<td>Wu and Radke (2012)</td>
</tr>
<tr>
<td>Multi-camera 3D tracking process</td>
<td>Wang et al. (2012b)</td>
</tr>
<tr>
<td>Multi-sensor fusion framework on feature level</td>
<td>Snidaro et al. (2012)</td>
</tr>
<tr>
<td>Pedestrian detection based on depth information</td>
<td>Martelli et al. (2012)</td>
</tr>
<tr>
<td>Multi-camera vehicle tracking with overlapping FoV</td>
<td>Hong et al. (2012)</td>
</tr>
<tr>
<td>Multi-camera person tracking with overlapping FoV</td>
<td>Behera et al. (2012)</td>
</tr>
<tr>
<td>Automatic PTZ camera control</td>
<td>Abdelkader et al. (2012)</td>
</tr>
<tr>
<td>PTZ cameras with multiple tasks</td>
<td>Starzyk and Qureshi (2011)</td>
</tr>
<tr>
<td>Multiple miniaturized cameras for visual coverage</td>
<td>Pham and Makhoul (2010)</td>
</tr>
</tbody>
</table>

2.6.5 Other processing

This section presents prior research which does not fall into any of the other categories, such as specialized recognitions or multi-sensor analysis. Such specialized recognition algorithms are introduced by Yang et al. (2014), who propose a method for efficient face recognition by coding a residual map learning scheme from training samples. This allows the recognition of partially occluded or masked faces. Further work on specialized recognition is proposed by Anagnostopoulos (2014), who presents a survey on license plate recognition, comparing different approaches and discussing unresolved issues. Filonenko et al. (2013) describe an autonomous monitoring system to supervise environmental parameters such as temperature, humidity or poisonous gas concentration as a set of sensor nodes connected via a wireless network.

Viani et al. (2012) introduce a wireless sensor network for surveillance in museums, thus enabling heterogeneous functionalities like the detection of theft attempts, the monitoring of environmental parameters for better artwork conservation and analysis of visitor behavior, while Suutala (2012) presents statistical machine learning and pattern recognition methods used for multi-sensor human context recognition. Different kinds of sensors such as fixed environmental sensors and wearable sensors are utilized.

Salvati et al. (2011) present an approach towards locating multiple acoustic sources in far-field environments, separation of the sources by means of beamforming and by comparison of the incident signal power, whereas Roda and Micheloni (2011) present a method for tracking moving sound events by
exploiting a priori information derived from medium and long-term observations of the monitored area. Kim and Ko (2011) introduce a hierarchical method to detect and classify abnormal acoustic events occurring in an elevator environment. First, detected acoustic events are separated into vocal and non-vocal, and are subsequently classified; e.g. conversation, scream or announcement in case of a vocal event, crash, door opening/closing or footsteps in case of non-vocal events.

Nieminen et al. (2009) present a multi-sensor data fusion architecture, which aims at consolidating data for simple decision making. A distributed configuration, including a logical decision making server and multiple single location surveillance points, is shown. Yu and Ganz (2010) on the other hand, propose an identity-aware video analysis system that generates identity annotated video records for forensics and training purposes by augmenting the video surveillance system with RFID sensors in order to recognize the identity of detected persons in the video streams. An approach to gender and face recognition in videos is presented in Hadid and Pietikäinen (2009). The proposed algorithm is based on the same spatiotemporal texture classifier as utilized by Kellokumpu et al. (2009), analyzing not only static features, such as shape, but also dynamic features, for instance motion (for example the way persons talk or move their heads). These local binary patterns can also be applied for the purpose of face recognition as shown in Ahonen et al. (2006). Table 14 summarizes the presented prior research.

Table 14. Summary of publications on miscellaneous processing.

<table>
<thead>
<tr>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face recognition of partially occluded faces</td>
<td>Yang et al. (2014)</td>
</tr>
<tr>
<td>Survey on license plate recognition</td>
<td>Anagnostopoulos (2014)</td>
</tr>
<tr>
<td>System for supervising environmental parameters</td>
<td>Filonenko et al. (2013)</td>
</tr>
<tr>
<td>Wireless sensor network surveillance in museums</td>
<td>Viani et al. (2012)</td>
</tr>
<tr>
<td>Locating multiple acoustic sources in far-field environments</td>
<td>Salvati et al. (2011)</td>
</tr>
<tr>
<td>Tracking moving sound events</td>
<td>Roda and Micheloni (2011)</td>
</tr>
<tr>
<td>Detecting and classifying abnormal acoustic events in an elevator</td>
<td>Kim and Ko (2011)</td>
</tr>
<tr>
<td>Identity aware video analysis system with RFID sensors</td>
<td>Yu and Ganz (2010)</td>
</tr>
<tr>
<td>Multi-sensor data fusion architecture</td>
<td>Nieminen et al. (2009)</td>
</tr>
<tr>
<td>Gender and face recognition</td>
<td>Hadid and Pietikäinen (2009)</td>
</tr>
<tr>
<td>Local binary patterns for face recognition</td>
<td>Ahonen et al. (2006)</td>
</tr>
</tbody>
</table>
2.7 Conclusion of prior research analysis

Much research in the area of intelligent video surveillance has been done and it remains a very active field of research until today. However, in most cases, the focus is rather narrow, e.g. being either processing algorithms, or just sensor fusion. The larger context is often ignored, which leads to ingenious work in certain aspects, but lacks the creation of an extensive system for solving real-world problems addressing the needs of security authorities.
3 Research approach

This research was strongly driven by the demand of security authorities. It was necessary to create novel information systems, as well as a research methodology which was suitable for such situations. Requirements gathered from user experience, as done in this research by deduction from expert interviews with security authorities, are difficult to structure narrowly without losing information and are often contradictory, thus design science research as research methodology was ideally suited to serve as scientific framework.

3.1 Design science research

Design science research is defined by Hevner and Chatterjee (2010), pp. 5, as follows:

“Design science research is a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artifacts, thereby contributing new knowledge to the body of scientific evidence. The designed artifacts are both useful and fundamental in understanding that problem.

Thus, design science research is ideal for addressing challenges arising in real-world applications by creating artifacts in order to find scientifically grounded solutions. Hevner et al. (2004) present seven guidelines as to what can be considered a design science research project as shown in Table 15.
Table 15. Design science research guidelines according to Hevner et al. (2004), pp. 83.

<table>
<thead>
<tr>
<th>Guidelines</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline 1: Design as an Artifact</td>
<td>Design science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.</td>
</tr>
<tr>
<td>Guideline 2: Problem Relevance</td>
<td>The objective of design science research is to develop technology-based solutions to important and relevant business problems.</td>
</tr>
<tr>
<td>Guideline 3: Design Evaluation</td>
<td>The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.</td>
</tr>
<tr>
<td>Guideline 4: Research Contributions</td>
<td>Effective design science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.</td>
</tr>
<tr>
<td>Guideline 5: Research Rigor</td>
<td>Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact</td>
</tr>
<tr>
<td>Guideline 6: Design as a Search Process</td>
<td>The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.</td>
</tr>
<tr>
<td>Guideline 7: Communication of Research</td>
<td>Design science research must be presented effectively both to technology-oriented as well as management-oriented audiences.</td>
</tr>
</tbody>
</table>

Regarding design science research, Hevner (2007) further introduces three design science research cycles, providing clear and consistent definitions, ontologies, boundaries, guidelines and deliverables for the design and execution of high-quality design science research. The model is based on the previous work of Hevner et al. (2004), which introduces an information science research framework. It is based on three main pillars, the environment, the actual design science research and the knowledge base superimposing three inherent research cycles, which are illustrated in Fig. 4:

- The relevance cycle, which connects the environment to the conducted research.
- The design cycle, which shows the iterative process of creation and validation of artifacts.
- The rigor cycle, which connects the conducted research to the knowledge base, ensuring the scientific grounding by taking prior research into account and, on the other hand, ensuring the dissemination of findings, thus contributing to the knowledge base.
Gregor and Hevner (2013) further propose a knowledge contribution framework in order to be able to classify the type of research done. This positioning is conducted in two dimensions: Solution maturity and application domain maturity, resulting in four quadrants as shown in Fig. 5:

- The invention quadrant is for radical breakthroughs, for fields with little knowledge of the domain or existing artifacts. This type of research is difficult to publish because of lacking grounded theory or insufficient understanding of the field to make a proper evaluation.
- The improvement quadrant, which, as the name implies, improves and builds upon existing knowledge. Thus, grounding is easier as the domain is more mature and evaluation of new artifacts can be based upon the improved knowledge.
- The exaptation quadrant, which applies knowledge from other domains to a new task.
- The routine design quadrant, which is necessary to distinguish form general professional design or system building by scientific rigor.
Sein et al. (2011) propose a method for design science projects which they refer to as “Action Design Research”, in which they explicitly try to include the organizational context and the evaluation into the research as inseparable and inherently interwoven activities. They argue that previous methods are lacking this separation and the inclusion of organizational context, often causing designed artifact to lack relevance. Their method is structured into four stages:

1. Problem Formulation
2. Building, Intervention and Evaluation
3. Reflection and Learning
4. Formalization of Learning

These stages interact as illustrated in Fig. 6, showing an iterative process similar to Hevner (2007).
Another technique in the context of design science research is presented by Stewart et al. (2007). They introduce focus groups as a suitable technique for the improvement of the design of an artifact, inherently providing proof of an artifact’s utility in the application field. Thus, focus groups can be used as confirmatory method to test hypotheses. They gather expert knowledge in a structured framework, subsequently collecting and interpreting data learned from experience in a scientific way, usually proceeding in the following way: First, hypotheses or research questions are formulated, then a sample set is identified determining the number of participants, their backgrounds and the number of groups. In order to objectify responses a moderator is necessary. Before participants are recruited the questioning route is set. After these preparatory steps the focus groups are conducted, the results are analyzed and presented. Fig. 7 illustrates these steps.

Fig. 7. Basic steps of conducting focus groups as described by Stewart et al. (2007).
Markus et al. (2002) present a complex design theory framework which allows contradicting requirements and emergent processes. This framework has been developed considering large-scale projects and has been brought to a commercialized product named “TOP Modeler”, while Walls et al. (1992) present an approach to building and testing design theories. Another approach is introduced in March and Smith (1995); They present a framework on how to classify different types of research, while March and Storey (2008) portray different research problems, the artifacts for solving these problems and the evaluation methods which can be applied.

3.2 The research process

Video surveillance is an essential tool in physical security, which is expected to play an increasingly significant role in our daily life. It affects not only our social environment, but also the micro, macro and even the global economy. Thus, the goal of this work was to reveal the security gaps in current systems and subsequently to develop a reference model for future high-performance, large-scale video surveillance systems, in research as well as in real-world usage. Building and evaluating artifacts, according to design science research principles, created such an information system.

3.2.1 Research phases

The research process in this work was split into the following three phases, while being carried out in alignment to Hevner’s design science research methodology (Hevner 2007). While Phase 1 was considered pre-research to gain a better understanding of the environment on different levels, from the technical point of view to management and decision makers’ point of view, Phase 2 analyzed the knowledge base ensuring scientific grounding, and in Phase 3 the artifacts were finally constructed and evaluated:

- Phase 1: The Pre-Research Phase. Studying information communication technology in general, evolution and future trends, the intertwining between ICT and physical security, technology penetration and the global economy.
- Phase 2: Security - The Big Picture: Analyzing prior research in physical security as well as ICT, and all aspects that need to be considered in large-scale high-performance video surveillance.
Phase 3: Video Surveillance. According to the research question deducted from interviews with security authorities and experts, designing and creating artifacts in order to create a novel video surveillance system.

While Phase 1 and 2 ensured overall understanding, which was essential when collaborating with a security authority, Phase 3 created the scientific grounding on which this thesis was built. This was, on the one hand, based on an extensive study of prior research according to the rigor cycle (Hevner 2007), as well as the expertise of the users, their experiences and processes and an analysis of the systems they use in their daily business. In this process a number of research questions were formulated which is presented in the following chapter. This research was conducted in a high-tech start-up company and allowed very close cooperation with users as well as universities. This ensured very careful attention to real-world relevance since it is the vision of the start-up company to create products beyond-state-of-the-art based on research results, thus addressing the relevance cycle. On the other hand, the close cooperation and supervision with universities ensured that close attention was paid to scientific rigor and methodology, addressing the rigor cycle.

Fig. 8 illustrates the research phases. As can be seen, as grounding broader topics such as e-business, online services and the ICT market were studied, as well as the effects on the global economy and ICT technology were analyzed. In the next phase, the studies became more specific towards the topic of this work, by first addressing security in general, including physical security as well as IT security and before finally studying video surveillance. This was also the basis for the artifacts created in this research.
Fig. 8. The research strategy of this thesis in three phases.

Based on the research questions, artifacts were created accompanied by a constant evaluation, which corresponds to the design cycle. Finally, field trials and final discussions with the customers were conducted, closing the loop in the relevance cycle. The research results of this thesis shall contribute to the knowledge base by dissemination of relevant findings and designs.

3.2.2 Design science research applied

Mapping the research conducted in this thesis on Hevner’s Design Cycles (Hevner 2007) leads to Fig. 9. As can be seen, the three-cycle model fits the
situation of this research, due to Hevner’s human approach, focusing on the applicability and the incorporation of the end user.

Fig. 9. The research conducted in this thesis mapped on Hevner’s Design Cycles (Hevner 2007).

The environment consists of the research group and their lab within a start-up company, university collaboration, and also the customers and their real-world sites. Prior scientific research as well as expert knowledge form security authorities covers the knowledge base, so the artifacts can be designed and evaluated ensuring relevance and rigor. According to Järvinen (2000), this research can clearly be categorized as artifact building approach.

The knowledge base consists of the prior research, the customers’ experiences and processes, finally leading to the actual design of the artifacts. Regarding the seven guidelines for design science research projects by Hevner et al. (2004), which range from artifact design to the evaluation and communication of the findings, this research can be mapped onto these guidelines. As Table 16 shows, all prerequisites for a design science research project are met.
Table 16. The proposed DSR project mapped to the seven DSR guidelines.

<table>
<thead>
<tr>
<th>Guidelines</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline 1: Design as an Artifact</td>
<td>An artifact is being created in this research.</td>
</tr>
<tr>
<td>Guideline 2: Problem Relevance</td>
<td>Due to the close collaboration with the customers, problem relevance is ensured from beginning to the end.</td>
</tr>
<tr>
<td>Guideline 3: Design Evaluation</td>
<td>The designed artifact will be evaluated by the customers in real-world use cases.</td>
</tr>
<tr>
<td>Guideline 4: Research Contributions</td>
<td>The proposed design science research project shall contribute to the knowledge base by creating and evaluating a reference design for large-scale video surveillance systems.</td>
</tr>
<tr>
<td>Guideline 5: Research Rigor</td>
<td>Research rigor is given through a thorough grounding with the state-of-the-art and research methodology.</td>
</tr>
<tr>
<td>Guideline 6: Design as a Search Process</td>
<td>All available means will be utilized to achieve the best possible results for the customers.</td>
</tr>
<tr>
<td>Guideline 7: Communication of Research</td>
<td>Due to the close collaboration with customers and funding agencies, it is inherent that this research is communicated to technical-oriented as well as management-oriented audiences.</td>
</tr>
</tbody>
</table>

Making use of Gregor and Hevner’s knowledge contribution framework (Gregor & Hevner 2013) the research in this thesis was analyzed and positioned. It can be broadly positioned in the area of information systems, targeted towards security applications, specifically video surveillance.

As neither video surveillance nor security in general are new fields of research and the problems in the area are well known and researched areas as shown in the analysis of prior research, the proposed design science research project cannot be categorized in the invention quadrant. Considering the amount of unsolved problems in the domain, the proposed research also cannot be categorized according to the routine design quadrant. This work can be positioned in the improvement quadrant in the knowledge contribution framework as shown in Fig. 10, although, while designing the artifact, it might also be possible to utilize solutions from other field for some issues, so it may be touching upon the exaptation quadrant.
3.2.3 The research questions

The research questions were carefully deducted from the meeting minutes from more than 40 end-user interviews with different security authorities, revealing their expert insight on high-performance large-scale video surveillance as well as the problems with current systems, and their wishes in general. These confidential interviews were structured in order to create focus groups similar to the methodology presented by Stewart et al. (2007), providing a thorough grounding for this design science research project. Summarizing these interviews, the main issues were all about too much information having to be analyzed by human operators, thus not having situational awareness. Hence, often reaction times are too slow to prevent damage or loss. A further issue was that systems in use were mostly non-standard, non-scalable, and could not perform as required. Thus, the following problem formulation could be deducted:

*What kind of video surveillance system is needed to maximize situational awareness, thus minimizing reaction time in critical situations, while being fully scalable to any number of cameras?*
Starting from this problem formulation, the following research questions arose:

1. How must a video surveillance system architecture be designed in order to be fully scalable, distributable and highly secure?

This could be further explored by the following sub questions: What are the specific demands regarding video surveillance? What are the weak spots of current video surveillance systems? What measures need to be taken in order to create a system architecture for large-scale high-performance video surveillance, sufficiently robust to withstand any type of attack or fault? How can the human observer be assisted in processing an information flood too large to be manually observed?

Addressing the fact that humans can monitor only a very limited number of cameras simultaneously, the idea of utilizing automated analysis on these cameras and additional multiple sensors arose. Hence, the following research question was formulated:

2. How can automated multi-sensor analysis improve the performance of a surveillance system?

This lead to further questions such as: Which increase in system performance in means of detection accuracy and false alert elimination can be achieved? How much additional information can operators process? Does adding different sensors affect the scenarios that can be improved?

As there is a trend in increasingly more intelligent cameras and even self-contained camera systems, it was clear that this also had to be considered in this work, which led to the research question:

3. How can an intelligent camera improve video surveillance performance?

Going into more detail, more sub-questions could be formulated: How can intelligent ad-hoc cameras decrease the time of reaction or even the time of deployment in critical situations? How much time can be saved? What are the main issues to consider?

Finally, as we are in a time where mobile devices are ubiquitous and every security officer is equipped with a range of mobile devices, the next research question arose:

4. How can mobile field devices improve a video surveillance system?
This question was followed by more detailed sub-questions: What are the main issues to consider? To which extent can the reaction time in critical situations be optimized? Can mobile field devices bring other advantages to security officers?

### 3.2.4 The artifacts

After formulating the research questions, which were deducted from the expert interviews, first, an extensive research of the knowledge base, which can be found in Chapter 2, was conducted in order to provide a strong scientific foundation for the creation of the artifacts. These were then created according to the presented three-cycle design science research methodology in order to answer the research questions. The creation of artifacts was structured into an overall main artifact:

- **IVSS**: The intelligent video surveillance system

This artifact has further sub-artifacts, which cover specific parts of the systems:

- **SYSARCH**: The IVSS system architecture
- **MSAP**: The multi-sensor analysis process
- **WSSU**: The wireless self-contained surveillance unit
- **VSLC**: The video surveillance local control

### 3.2.5 Evaluation

After constructing the artifacts in multiple iterations, which included lab testing and field-testing, the construct was finally evaluated and compared to existing systems in order to verify the contribution of the conducted research. Use cases were created with the security experts who were interviewed in order to obtain the requirements on which the research questions were based. Finally, the benefits of the created artifacts as well as ideas for future research were discussed. The research questions and chapters answering these correspondingly are summarized in Fig. 11.
<table>
<thead>
<tr>
<th>Main research question</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>What kind of video surveillance system is needed to maximize situational awareness,</td>
<td>4</td>
</tr>
<tr>
<td>thus minimizing reaction time in critical situations, while being fully scalable to</td>
<td></td>
</tr>
<tr>
<td>any number of cameras?</td>
<td></td>
</tr>
<tr>
<td>Sub-question 1</td>
<td></td>
</tr>
<tr>
<td>How must a video surveillance system architecture be designed in order to be fully</td>
<td>4.2</td>
</tr>
<tr>
<td>scalable, distributable and highly secure?</td>
<td></td>
</tr>
<tr>
<td>Sub-question 2</td>
<td></td>
</tr>
<tr>
<td>How can multi-sensor analysis improve the performance of a surveillance system?</td>
<td>4.3</td>
</tr>
<tr>
<td>Sub-question 3</td>
<td></td>
</tr>
<tr>
<td>How can an intelligent camera improve video surveillance performance?</td>
<td>4.4</td>
</tr>
<tr>
<td>Sub-question 4</td>
<td></td>
</tr>
<tr>
<td>How can mobile field-devices improve a video surveillance system?</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Fig. 11. The research questions and the corresponding chapters in this thesis.
4 The intelligent video surveillance system

This chapter introduces the construct which was created in order to answer the research questions, which represent demands and challenges of different security authorities, based on prior research and the state-of-the-art in video surveillance systems. First, the overall system is introduced, followed by a detailed description of the sub-components. The presented research resulted in a prototype on which a commercial product has later been created.

4.1 Introduction

With the goal of designing a construct that fulfills all the requirements posed by end-users, the security authorities, an artifact named the “Intelligent Video Surveillance System” (IVSS) was created. It aimed to serve as a reference model for large-scale high-performance video surveillance systems, by answering to the posed research questions and ultimately presenting a next generation system design.

This chapter is structured as follows: First, the connection to the research questions is elaborated and the boundary conditions of creating the construct are discussed, eliciting challenges that needed to be tackled and standards to be built upon. Then the artifact is introduced according to the five-tier model and illustrated in Fig. 12, first starting with system architecture and its respective implementation. This is followed by the next sub-artifact, the multi-sensor analysis process which clearly covers the sensor processing tier, finally leading to mobile devices on each end of the system: The smart camera concept of the wireless self-contained surveillance which covers the communication and the sensor tier unit, and the video surveillance local control adding the operator/HCI tier and touching upon the storage and retrieval tier.
4.1.1 Addressing the research questions

The sections in this chapter have been structured in order to clearly answer the posed research question. While the entire construct, namely the IVSS, answers the main research question “What kind of video surveillance system is needed to maximize situational awareness, thus minimizing reaction time in critical situations, while being fully scalable to any number of cameras?”, the sub-questions are each answered by a corresponding sub-artifact. The IVSS system architecture as presented in Section 4.2 is a distributed, fully scalable system architecture implementing the highest possible security standards. The multi-sensor analysis process is introduced in 4.3 and is an artifact designed to to minimize the data flood that operators in common video surveillance control rooms are confronted with, and greatly automates large parts of surveillance. The smart-camera artifact, the WSSU, introduced in Section 4.4, shows how such a mobile device can improve the performance of a video surveillance system, while the final sub-artifact, the VSLC in Section 4.5 further shows how the system can be improved by a mobile field device. Fig. 13 gives an overview of the research questions and the corresponding artifacts.
4.1.2 Boundary conditions

Creating the IVSS, one of the aims was to design an architecture that achieves maximum availability, optimal reliability and best performance in a system. According to the security authorities’ requirements, the IVSS had to support the following boundary conditions:

- Fully automated sensor analysis in order to maximize the operator’s situational awareness.
- Full redundancy or hot standby components: There had to be a substitute for every system critical component. This means that any physical single point of failure had to be avoided.
– Redundant networks, including wired data connection networks, power supplies and network components. Every network element was power supplied through at least two separate circuits.
– Tampering protection: Reliable software mechanisms in order to prevent tampering of cameras, network components and other hardware components.
– Scalability (both in terms of sensor analysis and network): The system should be designed to be scalable up to 10,000 sensors in a single site. For larger installations or multi-site installations the system had to be configurable into sub-sites in order to allow virtually unlimited scalability of cameras and sensors.

In addition to the hot standbys or full redundancy of the main functional components, data transfer was carried out through two different units ending in two separate data networks. The site engineering was done according to the following criteria:
– The total number of cameras should assure maximum coverage of the area (ideally in multiple angles).
– Cameras should be selected and mounted according to the video analysis tasks that are performed.
– Additional PTZ cameras should be utilized for close-up identification tasks and can take over as “context cameras” in case any of the other cameras malfunctioned or were tampered with.

4.1.3 Engineering challenges

In order to design and create a large-scale high-performance video surveillance system, in addition to the input from the security authorities, security management process standards needed to be considered carefully in order to identify potential threats against the created system on a theoretical level. The International Organization for Standardization’s code of practice for information security management (ISO 2013) provides a general blueprint for such a security management process. Fig. 14 shows selected topics, which were considered for the creation of the IVSS. The building blocks of a video surveillance security management process are the following:
Fig. 14. Security Management Process Components, relevant for the IVSS.

- Access Control, addressing all situations where assets need to be accessed and can possibly be manipulated. Identification methods and technologies for authentication and authorization needed to be applied; Policies for the implementation had to be developed. The scope ranged from user management aspects for the users accessing the database to the privileges under which an application process could access the operating system’s kernel processes. Advanced security models dealt with the formalized description of how information flows within a system in a secure way. This included the states in which a secure system should be.
- Physical security, regarding the system itself is in general concerned with any kind of physical influence. For the proposed architecture this was a major concern, as it was physically distributed and the locations of the cameras, sensors and servers mainly a given fact. So the planning of measures against natural disasters as well as intentional tampering needed to be addressed.
- Network security is concerned with the design of secure data transmission systems. This encompassed the creation of a network architecture, which
allows the separation of different types of networks (data, operations, security management, etc.), that provides secure areas for the databases and demilitarized zones for data access of various applications. It also dealt with the lock down of various network elements so that they did not provide any more functionality than absolutely necessary and also the access to respective functionalities was limited to authorized users (human or processes). Cryptography was utilized in order to protect security relevant information and maintain privacy aspects of transported or stored information. This was especially important as not only security critical data, but also personal data was transmitted over different networks (possibly passing insecure channels). The data must not have been intercepted or have been tampered with. Methods and cryptographic algorithms needed to be chosen carefully especially for situations where video surveillance systems operate geographically distributed.

- Application Security deals with the questions pertaining to the secure development, deployment, installation and maintenance of hardware and software. Already in the planning phase detailed policies needed to be designed.

- Operational Security is concerned with all details about setup, execution and decommissioning of the system. The management process here needed to ensure that all processes are well described, operational security measures were broken down to procedures, and operational descriptions for all parts of the system were present. Business continuity needed to be addressed with high availability in mind. When designing a surveillance system, there is a number of policies for the implementation of software (e.g. backup & restore system), of redundant hardware (e.g. a complete offsite backup facility) and procedures to be followed (e.g. definition of first response team and its actions) in case of disruptive events. Thus, business continuity played an important role in network security and application security of the design, where secure and continuous operation of networks and applications were planned.

System threats

In order to define the IVSS and its protection mechanisms, all types of threats against video surveillance systems needed to be investigated. These threats were divided into two main categories: The first category was the threat of content
manipulation, i.e., forging or removing of data, or adding useless or misleading data. The second category was based on system disruption, which ranges from simple attacks like destruction of cameras and cables to more sophisticated ones like network or software-based attacks.

Content manipulation was defined as tampering and manipulation of data in the system, ranging from cameras to the storage of the video streams. In practice, it could include the forging of access logs, modification of stored video, or even the injection of a false live video stream. As tampering with data is categorized as part of network security when discussing digital, IP-based systems, content manipulation on a physical level was not possible. When analog legacy systems were to be integrated, the possibility of injection of a manipulated signal had to be considered. The following threats were taken into account:

- Social engineering: Using interactions to obtain confidential information was considered as the greatest operational security threat for a video surveillance system.
- Network security: An attack on data transferred between one of the cameras and the final stage of processing, including attacks on analog signals and digital TCP/IP streams. When the attacker got in a position to observe and intercept data streaming, this was called man-in-the-middle attack. This would allow the forging of video data by playing pre-recorded video streams or by editing the actual video transmitted by the camera.
- Application security: The goal of application security attacks is to gain control over the video surveillance system processing the pre-processed information or the machine storing data via standard attacks like Trojans, worms, buffer overflows or exploiting backdoors, which could cause consequences ranging from a denial of service in the system to sensitive data stolen.

4.2 The IVSS system architecture

The IVSS was created in order to fulfill the highest security demands of high-level security authorities in order to be able to protect assets such as critical infrastructure on a large scale. It was designed according to the presented design science research principles: Creating artifacts, evaluating and improving them (Hevner 2007). The architectural overview is illustrated in Fig. 15 and was structured as follows:
Fig. 15. High-Level view of the IVSS architecture.

- **Peripherals**: This included all types of video surveillance cameras, IP cameras, smart cameras, WSSUs, additional sensors and the Video Surveillance Local Control. Cameras and sensors were double connected to the Video Surveillance Headquarters (VSHQ) to assure availability even in cases where one of the connections was unavailable. The redundant
connection concept was applied to the power supply and the wired types of data connections. It was possible to incorporate existing analog cameras, however, it was not recommended for achieving the highest levels of security. In addition to video processing, smart cameras were able to encrypt and sign video streams to protect them against tapping and manipulation, and were thus favored. Miscellaneous sensors, such as audio or thermal sensors, helped to identify events and tampering of cameras. Using multiple kinds of sensors, in addition to video, has considerably increased the security of the system.

- **Geo Redundancy**: The collected and processed information was additionally stored at a remote location, which was only accessible for the highest-ranking authorities in emergency situations. Even in disaster situations, e.g. if the VSHQ was shut down due to an attack, all information was secured.

- **Video Surveillance Headquarters VSHQ**: This included the data processing center, the central operation, system administration, maintenance and control. It further included the processing units, either fully redundant or with hot-spare. Video information, including meta-information from sensor processing, was entering the "core zone" of the security system, the server network, through multiple firewalls with advanced stateful packet inspection and specialized network intrusion detection systems, blocking anything but encrypted and correctly signed packets on network level. Both processing and storage on the respective server hardware took place simultaneously in consistency-checked self-healing clusters. Each of these three zones was subject to high security conditions. Additionally, the overall security concept and the security management process of the respective security authorities were implemented to ensure maximum system performance.

In regards of redundancy in the system at the camera level, power and network connections were redundant. This could simply be achieved in digital systems based on IP-based protocols, utilizing different network technologies, whereas realizing the same redundancy in an analog system would not have been feasible. This was one more reason for avoiding analog cameras in critical areas. In order to ensure power redundancy, power over Ethernet (PoE) was utilized aside local power connections and buffer batteries. Parallel networks were based on different media (e.g. copper wires, fiber optics or radio communication). Especially for a camera the co-usage of radio networks and wired connections was used as much as possible since they could only be attacked in completely different ways. The wide availability of wireless communication networks was another advantage of
digital systems; nevertheless, protection against tapping was crucial on any wireless camera network. Multi-tier encryption (application and network level) and signing had to be implemented. The same applied to any communication channel, connecting node in the same secure server space DMZ or in a physical redundant network within this space.

The security management process

After system design and engineering, implementation and configuration were executed. The security authority’s security management process was implemented, in order to assure a maximum level of security. Accordingly, maximum availability and optimal performance of the system were ensured. User interface software and relevant information about the IVSS was presented to the security authority on various interfaces. This included control rooms, authorized workstations and mobile devices such as the VSLC, which were used by security officers on patrol to receive live alerts and video access so they could instantaneously react to security threats.

In addition to the interfaces for the end user, the security authority could define application rules and conditions to be detected by the system in advance. An important issue was the design of these user interfaces. They had to be as intuitive and natural to the operators as possible, so that no time was lost in critical situations, especially considering the frequent lack of technical know-how of the operating users.

4.2.1 The IVSS software implementation

Based on the requirements of high-end security authorities and the theoretical threat analysis, the IVSS was implemented as a modular software platform, which allowed deployment in any small to large-scale scenario with varying degrees of redundancy, single or multi-site structure supporting an unlimited amount of sensors and cameras. Depending on the security authority’s requirements, the IVSS was fully scalable and offered redundancy at system level by replicating every single building block as often as required. Fig. 16 shows a technical view of the software building blocks of the IVSS system architecture. Each building block represents a separate software server process with different functionalities and is presented in the following.
Fig. 16. The IVSS software implementation.
The module controller

The module controller represented the bottom layer of the IVSS implementation on top of the operating system. It basically functioned as a watchdog, performing system-monitoring tasks, ensuring a self-healing system in case any process in the system behaved unexpectedly. Further, it was responsible for automated load balancing by constantly monitoring all system processes and by distributing the processing load across individual processing nodes. An instance of the module controller operated on each physical machine in the IVSS.

In a commercially deployed system, the module controller would further incorporate license control and management within the system. It had full overview over every video stream, every application or service that was running on the respective machine. All instances of the video management system, including all server components as well as video analysis applications in the multi-sensor analysis process, could be managed centrally utilizing a remote configuration tool which is shown in Fig. 17.

Fig. 17. Remote configuration of the IVSS.
The video server

The video server was responsible for all camera or sub-system connections, as well as incorporating the storage layer in dedicated machines. When passed the firewalls, entering the DMZ, it was the first connection point for all incoming video streams. Thus, it was also responsible for camera management and control, which means it would be necessary to implement all important open standards such as ONVIF\(^1\) as well as camera manufacturers SDKs in order to support virtually any incoming video stream.

The video server further controlled the database in which all video could be stored in flexible ring buffers. By default, the video server stored recorded video data without decoding for maximum performance, but it was optionally possible to re-encode incoming video to harmonize video compression formats, optimizing it for further processing or transmission over narrow bandwidth connections.

The IVSS architecture was designed to support multiple video server instances, either to split a large system into multiple sub-systems (e.g. multiple sites), but also for redundancy by simply replicating the same video server instances on different physical machines or geographic locations.

The data fusion server

The data fusion server implemented the multi-sensor analysis process as introduced in Section 4.3. It analyzed all available sensors and metadata generated by analysis applications and fused them into a rich multi-modal scene model, which was then further analyzed. The data fusion server further incorporated a number of specialized analysis algorithms especially for anomaly detection.

The alert server

The alert server managed events and created alerts in the system, allowing complex rule modeling. It further incorporated calendar functionality allowing scheduling of analysis and event processing and further integrated I/O triggers (IP-based or analog relay-based). An example for such a rule could be that an alert is triggered only if at least two cameras detect a person close to the building.

\(^1\) http://www.onvif.org
they are protecting, during nighttime. If an alert is then triggered, all lights in the area are turned on in order to scare the intruder away while simultaneously the alert is sent to a security service provider.

**The gateway server**

The gateway server is the last point in the DMZ, before connecting to any client. It featured detailed user rights management, as shown in Fig. 18, which could be defined for every single resource in a system. These rights included e.g. live and replay access, exporting or access to audio streams. Further, automated bandwidth management was implemented, ensuring the best possible connection over the available bandwidth implementing the quality levels introduced in Section 4.5.3.

![Fig. 18. User rights management in the IVSS.](image-url)
Video analysis applications

Video analysis applications analyzed video streams to extract specific information utilizing the multi-sensor analysis process presented in Section 4.3. For different applications specialized algorithms have been developed and implemented by the company’s software development team, based on the IVSS. These included:

- Perimeter Protection
- Direction Monitoring
- Face Collection
- Object Detection
- People Counting
- Vehicle Counting
- Queue Detection
- License Plate Recognition
- Parking Space Analysis
- Privacy Masking

On the left, Fig. 19 shows a virtual intrusion zone, where an alarm was triggered if a person got close to the building, while the image on the right shows a vehicle that was detected while parking. Fig. 20 shows people being counted on a top-down camera on the left image, while the right image shows a directed motion analysis, which triggers an alert if an object crosses the zone against the defined direction. Finally, Fig. 21 shows the automated collection of faces from a video stream.

Fig. 19. Perimeter protection (left) and parking space analysis (right).
The web server

The web server was implemented to be a bandwidth-optimized server for accessing an IVSS via Internet, either using a VSLC, browser, tablet or smartphone connection. It could be accessed via https or by a proprietary highly secure protocol. Fig. 22 shows the front-end on a tablet computer, displaying a live camera stream for mobile surveillance.
Fig. 22. The web server accessed by a tablet computer.

The client

The client was designed to be a full-featured workstation client application, allowing multi-view, multi-monitor viewing of videos and events, timeline with activity index and event markers, ideal for video surveillance command centers. It was fully scalable from one-person operations to large-scale operations with multiple operators, hundreds of screens and video walls. Fig. 23 shows an example of a set-up with one screen displaying 16 cameras simultaneously.
4.2.2 The IVSS in the cloud

At the time of the study, the use of high-performance video surveillance systems was limited mostly to large-scale security authorities and high-end users. Due to the necessary investments in video surveillance infrastructure, small to medium enterprises as well as private users were mostly excluded. This issue could be addressed by transforming the IVSS into a mandate-capable cloud service. Here, the same measure of sensor analytics was required, however, not in the scale of the high-level authorities, which has led to the idea of creating a virtual VSHQ in the cloud, serving many customers in a video-surveillance-as-a-service (VSaaS) based manner. This significantly reduced the up-front investment in infrastructure as well as maintenance on the customer’s side, providing the same high-performance IVSS analysis capabilities. Fig. 24 shows such a cloud scenario in a high-level view. Multiple sites (respective users) could be connected to a virtualized processing and storage center on one end, while security authorities or the customers themselves could access on the other.
Every component of the proposed system was designed to be fully distributable and is thus suited to be utilized in a cloud scenario. Every component could be virtualized in order to transparently utilize the hardware infrastructure of data centers. Fig. 25 shows an example configuration. In addition to the introduced IVSS components, the cloud scenarios required extra servers for session/user management and billing.
4.3 The multi-sensor analysis process

Utilizing different types of sensors and addressing the sensor processing tier, brought new short-term challenges concerning the choice of sensor type in order
to increase the security level or long-term challenges (i.e. how to deal with different data gained from heterogeneous devices and technologies). First, available sensors were analyzed regarding how they could provide additional value in a surveillance setting, which then led to the design of the multi-sensor analysis process. This analysis process was executed in the following steps:

1. Sensors and pre-processing
2. Sensor processing and generation of metadata
3. Multi-sensor fusion into a coherent 3D scene model
4. Efficient event and video management
5. Effective storage and retrieval processes
6. User interfaces and visualization

The overall goal of the multi-sensor analysis process was the generation of a dynamic virtual 3D scene model containing as much information about the observed scene as possible, which was then analyzed for security critical situations. Alerts, following a security management process, were created assisting the security operators in reacting optimally to every situation. Hence, video surveillance could be turned from a reactive into a proactive tool, greatly enhancing security as well as cutting costs as security personnel could be appointed more efficiently.

Fig. 26 gives an overview of the presented process: First, data is gathered by the sensors, which is then consequentially analyzed and the analysis results added as metadata. Next in the data fusion step, the processed sensor information is fused into the scene model, which is then distributed and made accessible to the visualization and user interface by the event and video management. Simultaneously, the storage and retrieval engine stores the original sensor data, the corresponding metadata as well as the scene model, while giving access to this stored data through the event and video management.
4.3.1 Sensors

The bulk of information was extracted from the image sensors since they delivered the most information in a surveillance scenario. The idea of augmenting the system by deploying additional sensors with different modalities in order to make a scene model more robust in many situations, improving the analysis process and the inference of critical situations, arose. Further, many sensors, so-called “smart sensors”, already featured on-board processing capabilities that could be utilized to pre-process the multi-sensor analysis, as Section 4.4 shows.

4.3.2 Sensor processing

In the multi-sensor analysis process, every sensor was calibrated individually. This included noise filtering, mapping the sensor output and an optimization process to achieve a comprehensive format for the system (e.g. conversion of logarithmic sensor response to linear scale). Further, the location and orientation of each sensor was accurately measured. For imaging sensors automatic and semi-automatic approaches for camera calibration were employed in order to optimize the set-up procedures.
When attached to the system, every single sensor was processed in a data abstraction layer; hence, maximum flexibility and maintainability were ensured. When a new sensor (e.g. a new camera manufacturer or a new type of sensor) was to be connected to the system, just a new plug-in for the data abstraction layer needed to be implemented. Thus, no recompiling of code was necessary. Analog to telecommunication networks, all smart sensors, processors and active system peripheries were synchronized, thus avoiding offset, thus assuring a synchronous real-time event detection among all components.

In case of binary sensors, such as break-beam sensors, door contacts, or pressure mats, additional processing, except for debouncing, was not required; nevertheless, this type of sensor could provide much information about a scene. In case of audio sensors, first the magnitude of the signal was analyzed in its peak mean values. This was used parallel to video sensors in order to achieve better behavior classifications. The video-processing pipeline was one of the most important aspects of the entire analysis process, hence it is described in the next section in detail.

4.3.3 The video processing pipeline

The IVSS video processing pipeline has been designed to overcome typical problems of prior systems, which were prone to generating too many false alerts in order to being used in demanding situations such as shaking cameras, bad weather (rain, snow, fog…) or animals in the observed scenes. The following processing pipeline as shown in Fig. 27 was constructed: Stabilization needed to be performed in order to eliminate the camera’s ego-motion, while the optimization tried to improve the video for analysis purposes. Environmental filters should eliminate snow, rain or other environmental influences that could harm the image analysis. The 3D projection made use of depth information prior to the detection and tracking of objects, while the spatiotemporal logic filter functioned as high-level reasoning to compensate detection errors. Finally, effects influencing the entire image region were analyzed.
Fig. 27. The IVSS video processing pipeline.

**Stabilization and optimization**

In order to compensate poor image quality and shaking cameras, in the IVSS video processing pipeline the image stream was stabilized and optimized utilizing feature point matching algorithms on every single frame of the incoming video. Fig. 28 shows a screenshot of an image sequence which was stabilized. The top left shows the input frame, the top right the displaced output. The bottom left shows the difference between the current and the last frame, and the bottom right, finally, shows the displacement vectors which were calculated by matching the detected feature point in the consecutive images. Further, histogram stretching and color correction could be performed if necessary.
Environmental filters

Rain, fog and snow were among the most frequent sources of false alerts in video analysis. In the IVSS video processing pipeline, filter methods to remove rain and snow were applied. Fig. 29 shows an example of a surveillance image before (left) and after filtering (right). As can be seen, the oblique streaks across the image are completely removed, creating a clear image ideally suited for object detection algorithms.

Fig. 29. A surveillance image before and after an environmental filter was applied.
3D Projection

Every monocular video, such as a surveillance video in this study, is a projection of a three-dimensional space onto a two-dimensional image plane. As the objects of interest usually could be assumed to be moving along the ground plane (humans or vehicles), it was possible to project local coordinates onto a virtual ground plane and thus utilize depth information. Fig. 30 shows an example of two persons in a typical surveillance scenario, which greatly differ in size in the image due to perspective, but not in reality.

![Fig. 30. Calculating the actual size of a person from a surveillance video.](image)

In order to calculate the perspective cameras needed to be calibrated. This could be achieved by manually defining reference points and the ground plane as shown in Fig. 31.
Fig. 31. Calibrating the camera.

The x & y dimensions, which represented the ground plane, as well as the z dimension, which was eliminated with the projection, are indicated by superimposed lines. The boxes on the far ends show objects of the approximate same size in three locations in the image, while the grid with the box in the middle indicates the ground plane.

Detection and tracking of objects

At this stage of the processing pipeline, objects were detected and tracked throughout an image sequence. Fig. 32 indicates detected persons with a bounding box and displays their trajectories showing their movement history over the duration of the sequence.
Fig. 32. Tracking of persons and vehicles.

Spatiotemporal logic filter

Even the best object detection algorithm delivers false detection at times (e.g. due to image noise or compression artifacts). Thus, it was necessary to further filter the tracking results. This was achieved by utilizing a state-based spatiotemporal occupancy filter. Thus, it was possible to detect outliers, such as an object detection which indicated that a person “suddenly appeared” in the middle of a scene, which could not be correct and thus could be discarded by the filter. Fig. 33 shows a mock-up of this processing step.

Fig. 33. Outlier detection by spatiotemporal filtering.
Analysis of global features

In addition to the image content, the overall image needed to be analyzed in order to detect disturbances such as artifacts covering the camera, sudden changes of light or sabotage. This also includes direct manipulations such as the detection if a camera is moved, out of focus or completely dismounted. Fig. 34 shows an example of such a situation where a bird close to the cameras partially occludes the camera’s field of view.

Fig. 34. A bird partially occluding the camera’s field of view.

Data fusion

Before the processed sensor data could be fused into a scene model, all meta-information was brought into a standardized XML-like meta-language, which is combined with binary data. In the following, an example for a tracked human is given, including a timestamp, the location of the object in the camera image by the four bounding box corners, a classification tag, and additional information such as a color histogram and feature points:
Each sensor output could be considered a piece of the puzzle in reconstructing the observed scene. The data fusion step was where this puzzle was put together. Here, all sensor information available was brought together in a spatiotemporal scene model. Sensor coordinates were transformed into scene coordinates, and sensor times were synchronized to scene time.

This scene model was then analyzed in two steps: First, a rule engine was applied, using a user-defined set of rules for detection of specific situations. The second step was an automated anomaly detection algorithm. This was implemented as a machine-learning module, which was trained over time to learn normal situations and behaviors (according to e.g. time, location or object types). After an initial learning period, it was possible to detected outliers and subsequently detect abnormal situations.

**Event & video management**

The event and video management was the backend of the user interface as well as the entire configuration management. This included sensor configuration,
parameter settings of algorithm or user rights and access management. Further in this step of the analysis process, an alert management according to the security management process for each situation was implemented, assisting the user to react accordingly to every critical situation. The alert management is responsible for:

- Access of alerts indexed by type, time and location.
- Alert priority to ensure an optimized reaction time to the most critical situations.
- A notification engine alerting the security personnel according to the alert type (e.g. via on-screen alert, e-mail, SMS, automated telephone calls).
- Configuration and user assistance for the reaction to each alert type according to the security management process.
- Verification of the alert workflow, ensuring that necessary measures were taken and not overlooked by the security personnel. If this did happen, an alert could for instance automatically be redirected to a different operator.

The forensic engine was also implemented in this step. It allowed the user to browse and query through recorded sensor data as well as re-analyze recorded data in case it was desired to re-analyze the scene with different algorithm settings or rules than those defined at run-time. For example, in a situation where an object was stolen, which was not explicitly protected at run-time, an “object detection/removal” could be applied offline, which requires a re-analysis of the stored video information, if querying existing object data does not lead to the desired result. Since re-analysis was only performed on a specific desired area, it is calculated in multiple real-time.

Since metadata was collected at the analysis stages, it is possible to search and query the database. A representative example of such a query for selecting green cars that are speeding in a desired area is:


SELECT ...
object type: vehicle,
PROPERTIES
color: green,
speed: >50
WHERE
location: area 5
WHEN
start: timel
end: time2

This query delivered a list of all green vehicles that were speeding in the given time span. A simple click on each icon instantly displayed the corresponding video clip and location in the 3D visualization, including tracking information. This type of querying is done instantly, since only existing metadata is analyzed. Hence, operators had a real-time experience in modifying and adapting their searches for rapid decision-making.

Storage & retrieval

The storage and retrieval step consisted of high-performance databases, ensuring not only the streaming of thousands of sensors in the video and data server, but also a rapid random access for forensic searches. The metadata storage represented the backend of the forensic engine for query searches as well as the automated anomaly detection for storage of meta-information and behavior patterns. Fig. 35 shows an example of a query searching in the client software for all license plates that were detected on a specific camera during the last 24 hours.
Visualization & user interaction

The final step in every high-performance video surveillance system such as the IVSS is the presentation of data analysis results, such as alerts and relevant videos, to the user (see Fig. 16). The overall goal was to give the security operators maximum situational awareness and assist them in making the best decisions in security critical situations. This step is split into two parts:

- The live user interface: After applying the multi-sensor analysis process only relevant information was displayed. Hence, only critical scenes aggregated as much as possible had to be presented. Distraction by irrelevant events had to be eliminated. Further, data had to be presented in an intuitive way to the operator. This led to the development of a 3D visualization as shown in Fig. 36. Since humans are used to seeing a three dimensional world, a visualization of a scene mapped into such a model could be understood quicker, since it corresponds to natural perception, unlike two-dimensional maps or single camera views. The operators could also browse through available sensors by simply clicking the respective icons in the 3D-model, without the need of cumbersome browsing through long unintuitive lists.
Additionally, the more traditional GIS maps were implemented in order to create overlays, that indicated cameras and their respective video streams as well as possible alerts in the correct locations, as shown in Fig. 37.

The forensic user interface: Here, browsing and queries of recorded material were possible. Further, automated, interactive intelligence reporting could be visualized in real-time.
A virtually unlimited number of sensors was instrumented intelligently in order to achieve a higher level of situational awareness. Fig. 38 summarizes the multi-sensor analysis process in detail.

Again, data is gathered by different sensors, which are calibrated and noise filtered. The data is further synchronized and brought into a common format to allow data fusion in the next step. There, the scene model is created by integrating all sensor and analysis information in order to then analyze the entire data in a holistic approach. In parallel, the storage and retrieval engine stored the original sensor data, the corresponding metadata as well as the scene model, while giving access to this stored data through the event and video management which can then be accessed via the live or forensic user interfaces.
4.4 The wireless self-contained surveillance unit

The next sub-artifact, addressing the research question “How can an intelligent camera improve video surveillance performance?” was the Wireless Self-Contained Surveillance Unit (WSSU). It was a smart camera concept, functioning
fully autonomously, allowing ad-hoc integration in any surveillance system and featuring rich on-board processing functionality for detection of critical events.

As end-user have stated that rapid ad-hoc placement of cameras was a severe problem with currently available systems, an innovative approach for utilizing advances of wireless mobility and computer vision for the application of video surveillance was sought. The Wireless Self-Contained Surveillance Unit could be deployed in emergency cases, when no time for preparation and installation is available. Further, utilizing the performance of the WSSU in combination with a mobile Video Surveillance Local Control (VSLC), an ad hoc video surveillance system was created. It should be noted that the term ad-hoc was used in the context of “Plug and Play”, where no pre-installation or configuration process is required.

4.4.1 The WSSU concept

From a hardware point of view, the WSSU was designed to fit into the regular housing of the video surveillance camera, with minor modifications. The functionality of the WSSU was designed so that it is able to act independently of any power supply up to one day. Moreover, the WSSU was able to act independently whether a wired connection to the video surveillance head quarters was present or not. Further, the WSSU was designed not to be dependent on a single type of wireless technology. The general structure of the WSSU is depicted in Fig. 39.
The WSSU consisted of four main functional units:

- Intelligent video surveillance unit (or smart camera)
- Power supply and power management
- Processing unit
- Smart antenna (networking interface)

One of the greatest challenges during the design phase of the WSSU device as well as the VSLC was to define and validate transmission concepts. The main goal was to be able to provide a broadband access component for the mobile devices through current and future mobile cellular / wireless systems. The following boundary conditions were considered for the concept design:

- Automated switching from one network technology to the other, thus attaining different bandwidth, i.e. higher or lower data rate according to the standards and the devices within the system.
- Increasing the movement speed in mobile usage within the same technology resulted in the reduction of the available bandwidth and smaller transmission throughput.

However, moving from one technology to another would mean an increase of the possible moving velocity for the same given throughput. The smart antenna of the WSSU automatically selected the best available wireless network; it considered
not only the signal strength, but also its limitations regarding throughput and mobility. Fig. 40 illustrates a mock-up of the camera unit of the system.

**Fig. 40. A mock-up of the camera unit of the WSSU.**

The design goal was to meet the need for higher quality video, higher resolution, and more flexibility and features. Higher frame rates and resolution require efficient compression techniques, such as H.264 and more specialized derivatives, which in turn require greater processing power in the camera and in the processing units. The need to extract maximum image content in a wide range of lighting conditions (high and low light, high contrast) has led to digital cameras adopting a new class of WDR sensors which require dynamic range compression within the camera. Another required feature of the WSSU was the possibility of performing video analysis in-situ within the camera.

Currently, up to 18 megapixel sensors have been deployed in the WSSU, however, the data flood generated by such a sensor was intelligently filtered by applying dynamic resolution and compression ratio reduction, controlled by the available transmission bandwidth.

### 4.4.2 Mobile video quality metrics

To guarantee a certain level of quality in mobile video surveillance it was necessary to define quantitative measures of required bandwidth for a certain set of features. This was especially important for operators with non-technical background. For calculating the required bit rate of the video, the quality was measured in a two-dimensional space with the compression ratio as one dimension and the frames per second on the other; this is illustrated in Fig. 41.
Quality levels were measured in two dimensions: Image quality (compression ratio) and speed (frames per second). The calculations were based on an Axis P1357\(^2\) camera.

![Fig. 41. Wireless video surveillance quality levels.](image)

Low quality (LQ) video transmission was suitable for maintaining overview in a scene, optimized to require only a minimum of bandwidth. Medium quality (MQ) was designed in order to be prioritized to the particular task a camera needs to fulfill. Recognition tasks, such as identifying a face or license plate, no matter if automatic or by human, desired a better quality image (resolution and less compression), while the actual frame rate was less important. Tracking fast moving objects, such as vehicles, was the opposite: Since it was not necessary to re-identify the object in every frame, less image quality was sufficient, but a higher frame rate was necessary to maintain a smooth trajectory. High quality (HQ) was ideal for any surveillance task, utilizing the camera’s maximum quality. Table 17 summarizes the defined video quality levels.

\(^2\)http://www.axis.com
Table 17. Video quality levels.

<table>
<thead>
<tr>
<th>Quality Level</th>
<th>Video Parameters</th>
<th>~ Bit rate (Mbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQ</td>
<td>25 fps@5MP, H.264</td>
<td>20</td>
</tr>
<tr>
<td>MQ (Tracking)</td>
<td>25 fps@2MP, H.264</td>
<td>5</td>
</tr>
<tr>
<td>MQ (Recognition)</td>
<td>8 fps@2MP, H.264</td>
<td>5</td>
</tr>
<tr>
<td>LQ</td>
<td>6 fps@640x480, H.264</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### 4.5 The video surveillance local control

This artifact addresses the research concerning the HCI / operator tier. The video surveillance local control (VSLC) was an approach towards designing a mobile client interface for high-performance video surveillance systems, as security authorities reported the need for remote access and control of their video surveillance systems.

In order to demonstrate the need for VSLC the following scenario was taken as an example and analyzed accordingly: A security authority staff member was patrolling around a surveillance site, such as an airport, where a critical situation was discovered, e.g. a suspected vehicle was seen in a restricted area. Here the security officer had following options:

- Send an alert to the security authorities with the details via text message, which might take too long to edit and send if detailed information about the scene was given.
- Call the emergency or military police, and give a verbal description of the scene. However, voice transmission may not be clear and is prone to miscommunication and inaccuracies.

In the meantime the observed vehicle might have enough time to escape. Scenarios like this have been reported by security authorities to happen frequently. Accordingly, a mobile control platform, that enables full instant access to the video surveillance headquarters was designed and prototyped, the Video Surveillance Local Control.

#### 4.5.1 The VSLC concept

The VSLC was designed to be a specialized device for mobile security officers, based on a handheld or a smart phone. This device should allow the security team in the field to communicate directly with the Video Surveillance Headquarters and
even to execute a set of commands to the system (depending on his authorization profile and access rights) and receive filtered or all system messages. The VSLC was designed in order to be fully integrated in the IVSS, providing security authorities with a set of real-time functionalities. The VSLC needed to provide the following features:

- Real-time mobile administration and control in critical situations.
- Online coordination with the video surveillance headquarters for patrolling staff.
- Real-time observation and tracking across all surveillance videos for security personnel/decision makers.
- Bi-directional video communication.
- Secure controlling and command broadcasting in real-time.
- Direct secure access to the system database.
- Access or modification of the system’s central security management process.
- Direct access to each camera in real-time and to recorded data.
- According to the communication with and access to the central video surveillance database, the VSLC’s functionality needed to be extendable to various new applications.
- The VSLC should provide two main functional modes, which could instantly be accessed by user:
  - The “in situ” mode, where viewing of pre-selected scenes as well as a video back channel in order to stream video from the device to the VSHQ was provided. Accordingly, an individual command or a broadcast command could be released as depicted in Fig. 42.

Fig. 42. A mock-up of the VSLC device in the “in-situ mode”.

106
The security management process “SMP Mode”, where an overview of the different tasks of the system security management process was displayed. Thus, it was possible to select a certain function, e.g. network security or access control, then release the required system command in real-time, see Fig. 43.

Fig. 43. A mock-up of the VSLC in “SMP mode”.

VSLC design challenges

In order to obtain maximum performance and flexibility, the VSLC was implemented as a mobile software application in order to function on a range of devices. The following considerations had to be made:

- Heterogeneous Network Challenges: The device had to be successfully deployable in every possible and available wireless network, technology or standard, such as LTE/UMTS/HSDPA, GSM/EDGE or Wi-Fi. Thus, on the fly network switching was imperative.
- Engineering Challenges: Various challenges on the engineering level arose when building a wireless video surveillance system. Availability, reliability, survivability were to be maximized. Through the power management system, power supply had to be kept flexible to adapt to a given task, while keeping power consumption at a minimum level.
- Operational Challenges: Mechanically, thermally and physically extreme situations had to be taken into account in order to enable reliable performance under all circumstances.
– Software Challenges: An embedded high performance chip as a processing unit and high-speed memory needed to be provided in order to cope with performance bottlenecks.

4.5.2 VSLC in the IVSS

While designing the VSLC, another significant challenge was to find the right balance between security requirements and system performance. This balancing act required a holistic, integrated approach for a complete security concept with security and event management, identity and access management, as well as system and change management. All disciplines had to interact automatically and seamlessly in order to ensure an effective level of service that enables optimal component and system performance.

Fig. 44 displays a general overview of an IVSS with mobile access: The VSLC connected to the video surveillance headquarters, containing the video surveillance processing and database, through a secure connection. On the VSLC, only the scenes triggering alerts were automatically displayed. If no alert was triggered, individual scenes or cameras could be accessed. Every access to the system was handled through a gateway server or access control server, providing all access functionalities such as firewalls, authentication and authorization. The video surveillance database included four types of data:

– Permanent video surveillance data, which was system specific
– Semi-permanent video surveillance data, e.g. information about current suspicious persons or objects
– Quasi-transient video surveillance data, which included the recorded video surveillance data and videos
Utilizing the presented prototype of the VSLS in a video surveillance system, the possibility of connecting to a variety of security relevant databases was explored. Hence, the creation of a number of novel applications for high security requirements was possible. Databases that could be connected to the VSLS include:

- Law Enforcement Criminal Justice Database
- Vehicle Central Database
- National Security Database
- Military Database
- Medical Database
- Hazardous Material Information Database
- Local Video Surveillance Database

For example, connecting the VSLS to a vehicle database, the following tasks could be performed:

- Immediate identification of license plates
- Identification of vehicle owners
- Capturing possible offenders in identified vehicles
Thus, an officer on the streets could quickly access the database, retrieve missing vehicles and identify suspected offenders. This feature should give the officer on the street the ability to react faster to a critical situation and make the right decisions, supported by the deployed security management processes.

Accessing a national security database, on the other hand, could grant the following functionalities:

- Automatic identification of known criminals or terrorists
- Adding a suspect to the database
- Identification of accomplices of captured criminals or terrorists

This gives the security authority the possibility to access the national security databases and receive immediate warnings and updates of attacks in the country, and is thus especially interesting for homeland security applications.

### 4.5.3 Performance quality levels for mobile clients

Based on the previously defined video quality levels (see Section 4.4.2), the following performance levels for mobile video surveillance clients were defined. These will also be referred to as MVS-levels (Mobile Video Surveillance) in this thesis. These quality levels gave the operator a measure of what kind of access to the video surveillance system is to be expected while being mobile and using a VSLC in “in situ” mode. The advantage of such quality levels was that operators with little technical knowledge could obtain a quick understanding of what system performance to expect in a critical mission.

- **MVS-level 0** was the most basic operation for mobile video surveillance; it allowed messaging in real-time and a single low quality video transmission at low frame rates on demand. It was sufficient e.g. for a patrolling officer to detect if something was happening in a restricted area.

- **MVS-level 1** allowed the transmission of a single video stream that corresponds to the given quality when requested. The video stream had to feature MQ and can hence be prioritized either to higher frame rate or to a higher image quality. This level was sufficient for specific surveillance tasks that did not require a broader overview of the site.

- **MVS-level 2** allows streaming of at least five MQ video streams or one HQ video stream and a back channel for controlling (i.e. controlling of a
motorized pan-tilt-zoom-camera). This level was sufficient for critical surveillance missions.

- MVS-level 3 could be seen as the extended mobile video surveillance headquarters. The operator had the possibility of utilizing a mobile client with multiple HQ video streams and full control of a video surveillance system, including rapid channel switching, and full access to virtually any control available in the command center (given the access rights).

Table 18 summarizes the mobile video quality levels and the corresponding bandwidth requirements.

**Table 18. User-centric mobile video quality levels.**

<table>
<thead>
<tr>
<th>MVS Level</th>
<th>Bandwidth requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>on demand (&gt; 0.25 Mbit/s)</td>
</tr>
<tr>
<td>Level 1</td>
<td>&gt; 5 Mbit/s</td>
</tr>
<tr>
<td>Level 2</td>
<td>&gt; 25 Mbit/s</td>
</tr>
<tr>
<td>Level 3</td>
<td>&gt; 100 Mbit/s</td>
</tr>
</tbody>
</table>
5 Evaluation

In order to ensure scientific rigor as well as real-world applicability, the created artifacts have been thoroughly evaluated. The results of this evaluation are presented in this chapter in two different user sites.

According to the third research guideline by Hevner et al. (2004), evaluation is a crucial component of the research process in which the utility, quality, and efficacy of a design artifact must be rigorously demonstrated. As the design of the artifacts created in this study were closely aligned to Hevner’s design cycles (Hevner 2007), the creation of the artifacts was a constant loop between the creation and implementation of ideas and constant evaluation, analog to the design cycle. The methods of evaluation were the following:

- Theoretical evaluation
- Lab tests
- Field tests (case studies at the security authority’s sites)

The use cases that were carried out as final field tests at the security authorities’ sites are presented in this chapter, in order to evaluate the created artifact and its sub-artifacts especially in regards of applicability. This addressed the relevance cycle properly, by ensuring that the construct really did solve real-world problems, and thus fulfilled the requirements of the security authorities on which the research questions were based on.

5.1 User sites for evaluating the IVSS

The goal of the evaluation was to show the usefulness, the applicability and quality of the created artifacts as well as some objective measurements of the created system’s performance. The measurement criteria were defined by carefully selecting representative use-cases with the respective security authorities, in order to evaluate the IVSS in general as well as the respective sub-artifacts, the multi-sensor analysis process, the wireless self-contained surveillance system, and the video surveillance local control. Two different user sites were selected in order to conduct a scientific evaluation of the created artifacts and will be presented in the following.

As real-world authorities operate the security critical evaluation sites, it was necessary to pay close attention to confidentiality restrictions in order to prevent identification.
5.1.1 Site 1: International airport

The first site in which the IVSS was evaluated was an international airport. At the time of the study it had more than 2000 cameras installed as well as multiple control rooms belonging to different security authorities accessing the systems. Such an airport has many different areas with different security requirements, ranging from parking spaces to the terminal to the actual runway.

As such an airport is considered a critical infrastructure, the highest levels of security, reliability and performance across an entire surveillance system are required. Redundancy regarding all aspects of the system as well as intelligent automation was demanded, which resulted in the following two maxims when applying the IVSS:

– Maximization of reliability
– Minimizing reaction times in critical situations

5.1.2 Site 2: Large-scale security service provider

The second site for evaluation of the IVSS was an international large-scale security provider that offers a multitude of video surveillance and control room services to a wide variety of customers. This resulted in the demand of a large-scale central video surveillance platform, which had to be fully geo-redundant across at least two different locations, integrating sub systems in three different ways:

– Remote IVSS connects to the IVSS in the control center
– Third party video surveillance system connects to the IVSS in the control center
– Cloud IVSS: Direct camera connection to the IVSS

Remote IVSS

In the case of a remote IVSS there was an individual IVSS installation at each site and the security service provider was given access to perform a number of services. Thus, the remote site is integrated hierarchically into the security provider’s system, but remains independently fully functional.

In this case, the video was stored and analyzed at the remote location. Only metadata and alerts were transmitted via Internet, and processed and stored by the
security service provider. In case of an incident, the security provider had to have access to each remote IVSS in order to trigger necessary reactions according to the implemented security management process.

**Third party remote video surveillance**

In this case there is a third party video surveillance system operating at a remote site. The desired use-case is basically the same as with the remote IVSS, however, a number of additional challenges arise, as not all third party vendors apply the same standards for video transmission or even rely on proprietary formats or protocols.

Another challenge are the differing qualities of video analysis in different third party systems, so the security provider cannot rely on them and is often overwhelmed with false alerts.

**IVSS as a cloud service**

In this case, the remote site has no video surveillance system at all, but just cameras or smart cameras with some preprocessing that connect directly via Internet to the IVSS, which is hosted by the security service provider. All storage as well as analysis is performed in the security provider’s data center.

This configuration is most often applied with small-scale remote systems. The security service provider expects massive growth in this area analogous to the availability of broadband Internet access. Thus, new markets are created by supplying small-scale customers high-performance video surveillance.

**5.2 Evaluation of the IVSS**

In order to evaluate the IVSS, first the overall system was addressed on a theoretical level, verifying its usefulness and applicability by theoretically analyzing potential weak spots where the system could be attacked, either in single-site operation or a multi-site setup. Next, the system architecture is applied, first for the airport use-case focusing on redundancy and high-security features (see 5.3.1), then for the security service provider, where integration, distribution and flexibility is paramount (see 5.3.2). The multi-sensor analysis process is then validated in two use-cases, first showing how reaction time can be decreased at
the airport (see 5.4.1), then showing how more information can be processed by a security officer in the control room of the security service provider (see 5.4.2). The smart camera concept WSSU is evaluated in use-cases demonstrating ad-hoc surveillance (see 5.5.1) and temporary site monitoring (see 5.5.2). Finally, the VSLC is evaluated with use-cases showing again how reaction time can be decreased at the airport (see 5.6.1) and how security personnel’s efficiency can be improved (see 5.6.2). Fig. 45 summarizes the evaluation presented in this chapter.

<table>
<thead>
<tr>
<th>Evaluation of the construct</th>
<th>User site: International Airport</th>
<th>User site: Security Service Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artifact IVSS</td>
<td>single site - analytical threat analysis</td>
<td>multi site - analytical threat analysis</td>
</tr>
<tr>
<td>Sub-Artifact System Architecture</td>
<td>highly secure redundant architecture</td>
<td>hybrid cloud architecture</td>
</tr>
<tr>
<td>Sub-Artifact Multi-sensor analysis process</td>
<td>multi-sensor vs. conventional analysis</td>
<td>(re-) analysis in the control room</td>
</tr>
<tr>
<td>Sub-Artifact Wireless self-contained surveillance unit</td>
<td>short-term ad-hoc surveillance</td>
<td>temporary remote site monitoring</td>
</tr>
<tr>
<td>Sub-Artifact Video surveillance local control</td>
<td>measure reaction times</td>
<td>hybrid control room / field officer</td>
</tr>
</tbody>
</table>

**Fig. 45. Structure of the construct evaluation.**

After analyzing the IVSS for possible attacks on an analytical level and after passing the stress test in the lab, it was implemented at an international airport, in parallel to the already existing video surveillance system. The IVSS was hardened
against vandalism, sabotage and natural catastrophes on a conceptual level. A system level failure could be defined in two ways: the status of having a single non-observed area or the lack of an alert when a critical event occurs (e.g. a suspicious object, person or behavior). Different attack situations, e.g. destruction or vandalism against single components or the entire system were taken into account.

In the periphery zone all the equipment outside the secured processing center could be found, including cameras, sensors or local video processing servers connected to cameras, sub-networks and devices. In Fig. 46 the infrastructure of a single-site IVSS, where all cameras and equipment are connected to a centralized video surveillance head quarter, is depicted. Further, possible attack points are indicated by numbered lightning bolts (1-5) and discussed in the following.

Fig. 46. Potential attacks on a single-site IVSS.

Analogous to Fig. 46, Fig. 47 shows the identified attack points (lightning bolts 1-5) for a multiple-site IVSS, such as operated by the large-scale security provider. Every sub-site has a local system and is connected to the central video surveillance operation and control room.
The five critical attack points (indicated as lightning bolts in Fig. 46 and in Fig. 47) in single-site and multi-site operation had to be taken into account:

- In Attack-Point 1 one of the cameras was destroyed or damaged, hence a redundancy in visual coverage needed to be available according to the required security level.
- In Attack-Point 2 an interruption of the network connection occurred: An alarm needed to be triggered, containing a description of the event, and the switching to wireless transmission. The entire system further was equipped with watchdog processes that constantly monitor components of the system in order to realize situations where even redundant connections are separated or components were destroyed or malfunctioned.
- In Attack-Point 3 the wireless network was jammed. The connection then was maintained via wired connection. Further, an alarm was triggered. If the cable-network was damaged first, an alarm was triggered through the wireless connection.
- In Attack-Point 4 one of the data processing servers or a storage server was attacked or malfunctioned. The system would not be interrupted, as all components need to be redundant so all tasks can be distributed among other units. In a worst-case scenario, where the entire system was damaged, for
instance by a natural catastrophe, a remote system, recognizing the loss of connection, would act as failover and take over the operation immediately.

- In Attack-Point 5 the control room or one of the user interfaces is attacked. This was dealt with by redundancy as well, by having either fully redundant control rooms or on demand control rooms, which could be initiated instantly.

5.3 Evaluation of the IVSS system architecture

In the following the IVSS system architecture was evaluated according to the two use cases: First, the international airport in which reaction time in critical situations is imperative. Then, the large-scale security provider that offers remote video monitoring to a multitude of customers and needs to maximize the number of cameras/screen to be monitored without missing any critical event.

5.3.1 The IVSS system architecture applied to an airport

Applying the IVSS to the entire airport, the decision was made to structure the system in six hierarchical subsystems as different areas of operation had different requirements. Each zone had its own processing, network and communication channels, which were connected to the control rooms. The control rooms were inside the airport facilities, but also replicable at a remote location as fallback for critical situations.

At the time the airport had more than 2000 cameras deployed which were processed and recorded redundantly in two separate buildings on the airport premises, but it was imperative that the system architecture could arbitrarily be extended to suit even a larger number of sensors and processing, providing the foundation for a virtually unlimitedly scalable system. Fig. 48 illustrates a modified illustration of this airport.
The structure of the subsystem is color-coded. The ramp and runway are grouped into one zone. Thus, they are the first subsystem in this use case. The entire indoor passenger area is also grouped into one zone, starting at the check-in areas and extending all the way to the pier and everything in between. In this critical area complete video surveillance coverage from different angles and directions was necessary.

The logistics and cargo areas are also divided into separate subsystems, followed by the airport surrounding, such as business centers, hotels, parking or public transportation stations, such as bus or train. The final structure of the system is shown in Fig. 49.
5.3.2 The IVSS system architecture as a hybrid cloud system

Utilizing the IVSS as a hybrid cloud platform for the security service provider, the three earlier mentioned use cases had to be covered. First, direct connection of an independent IVSS, which is installed locally on a different site. This was straightforward as the interfaces connected seamlessly due to the same system design, just as in the airport use-case where multiple subsystems were connected.

In the second use case, remote third party video surveillance systems had to be integrated. In order to do this, multiple specialized drivers, mostly based on
SDKs, provided by the vendors of the respective video surveillance systems needed to be implemented. This could be done effectively by extending the video server of the IVSS, however, the achievable quality of the video streaming depended on the respective vendor. Nevertheless, in all cases it was sufficient for the security service provider to deliver their services. In practice, one of the main challenges was overcoming the heterogeneous IT environments, which were often controlled by different companies, in order to pass all firewalls between the remote locations and the video monitoring center.

In the third use-case, the security service provider utilized the IVSS in order to market a cloud-based surveillance system to its customers. In this case, at the remote locations IP-based video surveillance cameras were directly connected to IVSS in the security provider’s data center, delivering a constant stream. Thus, no local storage or analysis of video data was necessary, as it is done most efficiently in the data center. The customers were then granted access to their own videos via the web server.

The system needed to support 2500 cameras in the initial stage, but needed to be designed in order to support more than 100000 cameras. This further opens the possibility for multiple data centers and control rooms to function as fallback for each other. Thus, availability can further be increased.

The IVSS, due to its completely distributed and scalable design, perfectly covered all three use-cases. In order ensure availability, the entire control room was replicated on two sites in order to maintain geo-redundancy. For critical installations, at least to fully redundant Internet connections needed to be ensured. Fig. 50 shows the IVSS system architecture applied as a hybrid cloud-based system.
5.4 Evaluation of the multi-sensor analysis process

The multi-sensor analysis process is what made the IVSS “intelligent”; it turned the IVSS into an automated, proactive video surveillance system. In order to evaluate the MSAP two use cases were created to demonstrate its effectiveness, first, saving valuable time in critical situations at the airport, second, by reducing the information load on the security officers in the control room of the security service provider.
5.4.1 Critical situations at the airport

As the security authority stated, a performance increase in means of detection accuracy and false alert elimination is highly desired from current systems. Hence, measurements in a case study at a critical infrastructure installation, an international airport, were made in order to validate the multi-sensor analysis process.

A set of 116 test scenes with different situations has been recorded using multiple sensors. The locations included a control room, a corridor and the perimeter surrounding the building. Subsequently, analysis, first solely based on video, has been conducted, and then incrementally sensors were added to show how the system performance increases. Table 19 shows the situations and how often they occurred in the test set. All situations were specifically acted to be difficult to detect for automated analysis, trying to trick the system whenever possible.

Table 19. Validation measurements for the MSAP.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Location</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>person passes through</td>
<td>corridor</td>
<td>23</td>
</tr>
<tr>
<td>person shouts/screams</td>
<td>corridor</td>
<td>10</td>
</tr>
<tr>
<td>gunshot</td>
<td>control room</td>
<td>3</td>
</tr>
<tr>
<td>person places object</td>
<td>corridor/perimeter</td>
<td>16</td>
</tr>
<tr>
<td>person breaches perimeter</td>
<td>outside</td>
<td>35</td>
</tr>
<tr>
<td>vehicle breaches perimeter</td>
<td>outside</td>
<td>8</td>
</tr>
<tr>
<td>smoke emerges</td>
<td>control room</td>
<td>5</td>
</tr>
<tr>
<td>two persons fighting</td>
<td>corridor</td>
<td>7</td>
</tr>
<tr>
<td>person falls</td>
<td>corridor</td>
<td>9</td>
</tr>
</tbody>
</table>

As evident, the sensors chosen to augment a video surveillance system can greatly improve the performance, but have to be chosen very well to suit the specific applications of the system. For instance, audio was only of use in indoor scenarios while thermal vision only outdoors. Fig. 51 shows the aggregated results of this evaluation in regard of detection rate, measured in true positives (TP, how many incidents were correctly detected as such), false positives (FP, an incident falsely alerted), and false negatives (FN, an incident missed).
As can be seen, the number of detections when utilizing multiple sensors could be increased from 87 to 114, while eliminating all false detections. Further, the missed detections in the test scenes could be decreased from 29 to 2.

5.4.2 Sensor analysis for a hybrid cloud system

Another use-case utilizing mainly the video-processing pipeline of the MSAP was created. In this case the amount of customer sites with the respective number of cameras which could be monitored by a control room officer should be measured.

Originally, every video camera was constantly displayed on a screen. In order to save screen space, layouts with 50 cameras on a standard monitor were created, which meant that each camera was represented no larger than 3x2 cm at regular 72dpi resolution. However, at this size, it has been reported that critical events could be missed, so there was a strong desire to better utilize the available screen space to relevant information.

In this case two operators were used to constantly monitor 500 cameras, where each operator was responsible for six screens. Applying the MSAP in the IVSS, the security service provider reported that it was possible to monitor all 500 cameras with just one operator utilizing just two screens: One main alert screen,
where alerts generated by the MSAP automatically, show up in real-time on a considerably larger grid and one general system screen.

5.5 Evaluation of the WSSU

The need for mobile video surveillance ranges across different industries and applications, however, two use-cases were created in order to evaluate the created smart camera concept, namely the WSSU.

5.5.1 Ad-hoc surveillance at the airport

The WSSU was validated in three scenarios at the international airport. The critical parameter, which was measured, was the installation time in critical situations. These included:

- An unexpected state visit requiring additional cameras
- High-level security warning, requiring additional high-resolution monitoring for face recognition
- Network fault at an entire section

In the case of the state visit, 14 additional cameras were to be installed at critical locations. In the experience of the authority, until additional 14 cameras are operational, at least 70 minutes per camera are calculated for installation. Due to the ad-hoc capabilities of the WSSU, it was possible to prove that the time can be cut down to below 40 minutes or even lower when utilizing the battery powered version. In this particular set-up, a crew of four persons had set up the additional 14 WSSUs in 2 hours and 38 minutes.

Similar results could be achieved when adding the WSSUs for face recognition. In this case, five cameras were placed, however, due to the more critical positioning 53 minutes were measured on average. The largest of the case studies, the network fault, was tackled the fastest, since the 20 battery powered WSSUs deployed were prepared in the VSHQ in less than 15 minutes, and deployed in another 10 minutes on average. Table 20 gives an overview on the measurement results, comparing the results to the same situations utilizing the existing processes and technology at the premises.
Table 20. WSSU validation results.

<table>
<thead>
<tr>
<th>Case study</th>
<th>Number of cameras</th>
<th>Avg. deployment time, current system [min]</th>
<th>Avg. installation time with WSSU [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>State visit</td>
<td>14</td>
<td>980</td>
<td>540</td>
</tr>
<tr>
<td>Security threat</td>
<td>5</td>
<td>350</td>
<td>265</td>
</tr>
<tr>
<td>Network fault</td>
<td>20</td>
<td>1400</td>
<td>486</td>
</tr>
</tbody>
</table>

The WSSU was further evaluated on a theoretical level, in order to show how different attacks can be dealt with. Fig. 52 illustrates these attack points (numbers 1-4) in a fictional mobile surveillance setup.

- **Attack 1**: Manipulation of the WSSU. In this scenario the attacker either covered it or modified its point of view by manipulating its position. This could be quickly detected, since the WSSU has embedded image-processing capabilities. As soon as any changes beyond a threshold in the background model occur, a health check routine is initialized to determine if the camera’s view still corresponds to the view it was set up to observe.

- **Attack 2**: Interception of a transmitted signal. All video streams are encrypted and an offline brute force attack to regain the keys affords considerable hardware and time. Even if the attacker steals a WSSU in an attempt to gain access to the keys, the WSSU wipes its memory after a set period of time if the connection is lost unexpectedly.

- **Attack 3**: Injection of a modified signal. Similar to the previous attack scenario, an attacker requires the WSSU’s keys to be able to sign transmitted video streams.

- **Attack 4**: Jamming at an aerial mast. Here, the attacker jams the network connection. In this case the WSSUs will try to transmit using another network technology (e.g. 4G instead of WLAN). If no connection can be established, the WSSU will temporarily record on its internal memory and try to transmit the collected image data as soon as a connection is established again.
5.5.2 Temporary remote site monitoring

In this use-case the WSSU was utilized for remote monitoring, for customers of the security service provider that only required temporary surveillance, such as construction sites. Here, there was no IT infrastructure available at the site that would support a traditional surveillance system and it was not feasible to build one for the limited time period.

The WSSU was utilized for this scenario, placed on mobile masts. As the site was illuminated all night, the WSSU could be powered from the same source and did not need to deal with low lighting issues. The WSSU was configured to automatically detect any person entering the construction site after the set curfew and triggered an alert in the control center.

5.6 Evaluation of the VSCLC

In order to validate the video surveillance local control two case studies were created. First, an experimental case study at the international airport was conducted measuring the reaction time to specific events with and without the VSCLC. Second, the VSCLC was utilized for a mobile field officer.
5.6.1 Use Case: Critical situations at an airport

At the airport up to five security officers were equipped with a VSLC, directly interacting with the VSHQ and the system. Typical critical situations that occur at such airports were played out, including:

- Abandoned luggage / objects
- Illegal border crossing
- Irregular crowding
- Appearance / tracking of a sought criminal

These situations were acted out over ten hours, measuring all interactions between the patrolling officers and the VSHQ. Fig. 53 illustrates the aggregated results.

As can be seen, the reaction time utilizing the presented system could be improved in all cases by at least 84%. In the case of criminal tracking, the reaction speed could be increased by 422% due to the fact that the field officers had all necessary information available at all times without delay.

Fig. 53. The VSLC evaluation results.
5.6.2 Use Case: Mobile security officer

As a second use case to evaluate the VSLC, security officers at a security provider were equipped with VSLCs. It should demonstrate the effectiveness of the device, allowing a more flexible usage of human resources for the security service provider. It has been reported that at certain times fewer persons in the control center could easily handle the expected number of alerts and operators are idle some of the time, just as backup.

In a field trial, a spare officer was equipped with a VSLC in order to have a remote miniature control room at all times, allowing other tasks to be carried out (e.g. patrolling), but was still able receive and react to alerts. During a test period of three days, the security service provider reported a significant improvement in utilizing their officers when utilizing the VSLC, because idle times could be used for other tasks, allowing the operator to leave the control room without being disconnected.
6 Discussion

This chapter discusses the results and outcome of the research presented in this thesis. The created artifacts and their corresponding evaluation are connected with prior research and further discussion of their implications on theoretical, managerial and methodological levels. This chapter is structured as follows: First, the purpose of the study is recalled, followed by research results. Next, theoretical and managerial considerations are elaborated, followed by methodological considerations. Finally, an outlook on future works is presented.

6.1 Purpose of the study

As video surveillance has become ubiquitous (Porikli et al. 2013), new system architectures as well as automated analysis mechanisms have become necessary in order to cope with the amount of data generated by the amount of cameras and sensors deployed (Räty 2010). Video surveillance systems need to be intelligent information systems that optimize and automate a large part of the security operators’ workflows, as they are overloaded with irrelevant data. This often makes video surveillance inefficient or even ineffective causing an illusion of security without providing it.

The purpose of this research was to scientifically assess user requirements first and then create a reference model of a system fulfilling these requirements. The research questions were created in close collaboration with different security authorities that gave deep insight into their operations and processes in order to find better solutions to their problems and challenges. From all different wishes and requirements, one was predominant and valid for all security authorities. Thus, it was the basis of the main research question that drove this research:

What kind of video surveillance system is needed to maximize situational awareness, thus minimizing reaction time in critical situations, while being fully scalable to any number of cameras?

This study showed how a distributed, software-based video surveillance platform could be created in order to be unlimited in scalability, thus supporting a virtually unlimited amount of cameras or sensors in a single system. It further showed how this system could be augmented with a multi-sensor analysis processing pipeline, and finally extended by mobile devices, first on the sensor side, by designing a
smart camera, then on the operator side, by creating a smart mobile client interface.

The prototype resulting from this study will finally be transformed into a commercial product. Hence, the applicability and relevance of the research was crucial. The prototype was field-tested in collaboration with the security authorities at an international airport and at the premises of a private security service provider, as they are the future users and customers of the resulting product.

6.2 Research results

This section summarizes the results obtained when evaluating the construct and its artifacts, which are presented in detail in Chapter 5. The methods of evaluation were theoretical analysis, constant lab tests during the development and finally field tests at two real-world sites. Thus, the usefulness of this research was ensured from the formulation of requirements up to the creation and evaluation of the artifacts.

When designing the construct in order to fulfill the requirements posed by the security authorities, the “Intelligent Video Surveillance System” emerged (see Chapter 4). It is a multi-sensor surveillance system that was designed in order to turn video surveillance, which was mainly used reactively after the fact, into a proactive real-time tool. After general considerations and boundary conditions, the IVSS system architecture is presented in Section 4.2. The respective software implementation of the IVSS is then elaborated in Section 4.2.1, presenting the individual components of the implementation. The multi-sensor analysis process is presented in Section 4.3, showing how the surveillance process can be automated. Section 4.4 then introduces the wireless self-contained surveillance system, thus extending the IVSS by a smart camera device that is capable of functioning either completely autonomous or as an extended sensor of an IVSS. Finally, the video surveillance local control is presented in Section 4.5, extending the IVSS’s user interface to a mobile device.

The IVSS was first evaluated on a theoretical level as presented in Section 5.2. The system was analyzed for potential attack points, first in a single-site setup, then in a multi-site configuration. Measures in order to overcome all of the identified potential threats were presented. Subsequently, the IVSS system architecture was evaluated by application in two large-scale scenarios with very different requirements. First, the international airport (see Section 5.3.1), which
was a critical infrastructure and has the highest demands in regards of availability and performance. In critical situations instant reaction was paramount. It could be shown that the IVSS system architecture fulfilled these requirements. Due to its distributed and modular design, any level of redundancy could easily be achieved and the multi-sensor analysis process was able to improve the reaction times of the security officers in critical situations.

The second scenario was a large-scale security service provider operating a video control center where hundreds of customers with thousands of cameras were connected in order to provide real-time video monitoring as well as other services (see Section 5.3.2). It could be shown that the IVSS system architecture fulfilled these requirements, ultimately fusing the systems of the security provider’s customers together with the IVSS creating one large overall system, which was further replicated geo-redundantly in order to maximize availability. Further, the IVSS enabled the security provider to offer a cloud-based solution to its customers from within the same system, showing that the system architecture, due to its flexibility, fully met their requirements and even enabled new services.

The multi-sensor analysis process was evaluated by measuring reacting times in different use-cases at the airport, comparing the utilization of different sensors. Further, it could be shown how the security provider can handle a significantly greater number of video streams per security officer, not only improving security for the customer, but also optimizing cost by a more efficient appointment of personnel, while minimizing the risk of liability claims in case a critical event was missed.

The wireless self-contained surveillance system unit, WSSU, was first evaluated (see Section 5.5.1) at the international airport by measuring deployment time in ad-hoc surveillance situations. The WSSU’s benefit was clearly evident, making the deployment process more than twice as fast on average. Further, the WSSU was utilized for the security service provider in order to enable temporary mobile remote video monitoring, opening up a new field of business.

The video surveillance local control, the VSLC, extended the IVSS for usage on mobile devices by providing a mobile client interface for field officers. In order to evaluate the VSLC, different use-cases were conducted at the airport measuring the improvement in reaction time when utilizing the VSLC. Depending on the use-case, improvements of 80% and more could be measured. On a wide-range area, a mobile field officer of the security provider was equipped with a VSLC and thus had a miniature version of his control room work place with him, allowing him to carry out other duties in times of low activity without
compromising reaction times in critical situations. In addition, mobile video quality levels were introduced to give security operators with little technical background a better idea of which quality and number of video streams they can access when being mobile.

### 6.3 Research contributions

The main contribution of this work was the creation of an intelligent video surveillance system design, which fulfills the requirements of real-world security authorities, turning a formerly reactive tool into an automated information system. Thus, it is possible to partially automate the video surveillance process, optimize the appointment of security personnel and, most importantly, increase the level of security while utilizing existing infrastructure.

In further detail, this work contributes to the knowledge base by introducing a highly robust and scalable system architecture, which was evaluated in two different use-cases: an international airport and a large-scale security service provider. The architecture was then extended by the multi-sensor analysis process (MSAP), which turned the video surveillance system into an automated multi-sensor surveillance system. Further, the system was extended by two mobile device concepts: the smart camera WSSU and the mobile client device VSLC. Table 21 summarizes the contributions of this research.
Table 21. The created artifacts and their contributions to the knowledge-base.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The intelligent video surveillance system</td>
<td>A system design capable of fulfilling the requirements of real-world users: An international airport with the demand for high-performance analysis at maximum availability and a large-scale security service provider, creating a large-scale cloud-based surveillance system.</td>
</tr>
<tr>
<td>The IVSS system architecture</td>
<td>The distributed system architecture of the IVSS for high-performance large-scale installations. It is fully modular and virtually unlimited regarding scalability. Any component can be replicated or virtualized for redundancy.</td>
</tr>
<tr>
<td>The multi sensor analysis process</td>
<td>A processing pipeline that is capable of transforming a video surveillance system into an automated multi-sensor surveillance system. It creates a virtual model of an observed scene that fuses information available from any sensor in an observed scene in order to automatically detect critical events.</td>
</tr>
<tr>
<td>The video surveillance local control</td>
<td>A mobile client implementation extending the IVSS to remote surveillance and control functionality. It allows security officers in the field to access video streams and communication, and to control parts of the IVSS. User-centric quality levels have been created in order to assist operators with non-technical background</td>
</tr>
<tr>
<td>The wireless self-contained surveillance unit</td>
<td>A smart camera concept, allowing ad-hoc deployment and temporary remote video surveillance. It features on-board processing capabilities and flexible wireless communication for video transmission.</td>
</tr>
</tbody>
</table>

As March & Storey (2008) suggest, the results of this study are shown in Table 22, clearly repeating the problem, the artifact that shall solve the problem and the evaluation of the created artifact.

Table 22. Results of this study.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
<td>The amount of information generated by cameras and other sensors cannot feasibly be processed manually. Operators are overloaded with too much information, increasing the risk of missing critical events or using video surveillance just as a reactive tool, after the fact.</td>
</tr>
<tr>
<td>Artifact (Solution)</td>
<td>Design and construction of an automated large-scale high-performance video surveillance system.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Theoretical analysis as well as empirical measurement of the performance of the system in different use cases at real-world surveillance sites.</td>
</tr>
</tbody>
</table>
6.4 Theoretical implications

The IVSS, as presented in Chapter 4 and evaluated in Chapter 0, is a design for an entire video surveillance system that ranges from the sensor to the operator, as described by the five-tier model in Chapter 2. It is an automated multi-sensor system that is virtually unlimited in scalability and focuses on large-scale high-performance installations.

The idea of automating video surveillance has been worked on for more than two decades, where Corrall (1991) analyzed airport ground traffic with image analysis algorithms. Later, Kanade et al. (1999) presented the first multi-sensor surveillance system that integrated color cameras, thermal cameras and night vision cameras. However, opposed to the IVSS, these early approaches analyzed every sensor individually and were not built on a system architecture that was designed to scale up to thousands of sensors across heterogeneous networks and infrastructure.

More recent research on video surveillance, such as presented by Li et al. (2013), who present a distributed video surveillance architecture, however, does not feature any automated analysis, which was one of the most important parts in order for the IVSS to answer the main research question. Fernández et al. (2013), on the other hand, introduce a multi-sensor surveillance system based on utilizing many low-cost sensors, which are then compared to commercially available cameras; Opposed to the IVSS, it is not intended to utilize existing infrastructure or target real-world high-security applications.

One of the requirements, especially from the large-scale security service provider, was to be able to operate their video in a cloud-based manner, as it is done with the IVSS. Looking at other cloud-based video surveillance systems, such as proposed by Chen et al. (2013) that focuses on traffic analysis by distributed license plate recognition, is not intended for general purpose surveillance tasks such as the security service provider offers. Another cloud-based architecture which allows edge and back-end processing is presented by Chattopadhyay et al. (2013). It facilitates communication between sensors and an interface that allows access to event information remotely, to enable third party applications to provide services using the surveillance data. The system, however, was not designed for high-security purposes, a fundamental issue when designing the IVSS. Similar to the IVSS, the system introduced by Zhu et al. (2011) is intended for large-scale installations, however, the focus lies on peer-to-peer streaming and again, automated analysis has not been considered. Unlike the
IVSS, more video surveillance systems targeting lower-end applications are presented by Rodriguez-Silva et al. (2012), by Lin et al. (2012) and Dey et al. (2012), utilizing consumer grade cloud storage. They focus mostly on the optimization of transmission data rates according to available network bandwidth.

Segor et al. (2012) present a generalized video surveillance system architecture with the focus on achieving maximum reusability opposed to maximum performance as in the IVSS, while Saini et al. (2009) present a modular, multi-modal surveillance system architecture with the focus on being easily portable to different environments, however, high-availability and redundancy issues, which were paramount in the design of the IVSS system architecture, are not addressed. A distributed system architecture for intelligent video surveillance research is presented by Vezzani and Cucchiara (2010a), introducing a three-layer tracking system for overlapping and non-overlapping cameras, followed by a static service-oriented architecture, allowing the addition of different applications. Opposed to the IVSS, however, it features no multi-sensor capabilities, and no scalability or reliability issues are accounted for. Chang et al. (2010) present an architecture for distributed video surveillance including fixed and wireless networks. The focus of their research are the different topologies and streaming strategies that are building a hierarchical system. This work, however, does not incorporate the demand for highly available, mission-critical systems.

Huang and Tan (2010) present a system focusing mainly on the analysis of video, not accounting for scalability and reliability issues needed for high-security installations that the IVSS was designed for. Another approach focusing on analysis, thus lacking scalability and reliability, is employed by Weiyao et al. (2008), who present an approach towards behavior recognition in video surveillance by describing activities like walking, running or fighting with specifically trained models.

An approach for recognizing human actions, a spatiotemporal texture classifier using local binary patterns, is presented by Kellokumpu et al. (2009). In this two-phase process, first, movement dynamics are captured, followed by a characterization of these observed movements. Suutala (2012) presents statistical machine-learning and pattern recognition methods used for multi-sensor human context recognition.

Räty et al. (2008) describe an intelligent video transmission client-server system, similar as implemented in the WSSU. Depending on the user’s area of interest, all transmitted video data is scaled only to the necessary resolution,
considerably reducing the required bandwidth without compromising situational awareness and saving time in critical situations.


### 6.5 Managerial implications

As the IVSS was built upon user requirements in order to solve existing problems in close cooperation with security authorities, strong managerial implications can be expected. Moreover, the prototype created in this study shall be the basis for a commercial product, thus emphasizing the focus on applicability in real-word scenarios, which has been reported multiple times during the evaluation at the security authorities’ sites.
Video surveillance has become ubiquitous in our daily lives (Porikli et al. 2013), ranging from every corner on the street to shops, public transportation, banks, and office buildings, up to high security premises such as an airport or prison. Most of the deployed video surveillance systems nowadays are used only as a reactive tool after the fact, when it is too late to prevent damage or loss (Räty 2010). Worse, these systems often give a false sense of security, since the cameras are not always being watched. From a managerial point of view this is a disastrous situation, as much has been invested into these systems that cannot deliver the expected level of security. By creating the IVSS, it became possible to incorporate and build on these systems, integrating them into an automated multi-sensor processing system, thus leveraging existing investments by allowing a transition to the new technology without the need for replacing previous investments. Just as well, the IVSS is well suited for designing and implementing new systems from the scratch, being future-proof due to open standards and virtually unlimited scalability. In a nutshell, the most important managerial implications can be summarized in three steps, as Table 23 shows: Maximizing Security, Optimizing Cost & Utilizing Statistics.

### Table 23. The most important managerial implications summarized.

<table>
<thead>
<tr>
<th>Implication</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximizing Security</td>
<td>Video surveillance is turned into a pro-active automated information system, ideally preventing damage or loss by real-time analysis and alerting.</td>
</tr>
<tr>
<td>Optimizing Cost</td>
<td>Security personnel can be appointed more efficiently, increasing the amount of data that can be processed. Further, existing infrastructure can be leveraged in order to secure previous investments.</td>
</tr>
<tr>
<td>Utilizing Statistics</td>
<td>The automated analysis can further be used to gain statistical insight into observed scenes and provide valuable data for either strategic security planning or marketing (e.g. customer analysis in retail)</td>
</tr>
</tbody>
</table>

### 6.6 Methodological considerations

This research was strictly carried out according to design science methodology (Hevner et al. 2004) in order to ensure scientific rigor, while developing a solution for real-world problems by creating and evaluating artifacts. Each of the proposed research guidelines was carefully checked in order to ensure that no important step has been missed while conducting the research.
Guideline 1 (design as an artifact) was clearly met, as an artifact, the IVSS, and several sub-artifacts were created. Guideline 2 (problem relevance) was followed, since the entire research was conducted in close cooperation with security authorities, from assessing the requirements to the final evaluation, which leads to Guideline 3 (design evaluation): The created artifact has been theoretically evaluated, constantly been evaluated in the laboratory while being developed, and finally evaluated by case studies at an international airport and in the control room of a large-scale security service provider. Guideline 4 (research contributions) was also accounted for, as this research resulted in a reference design for high-performance multi-sensor video surveillance systems. Guideline 5 (research rigor) was followed by utilizing design science research methodology for the entire research process. Guideline 6 (design as a search process) was also accounted for, as not only the scientific knowledge base, including conference proceedings, journals and other scientific publications were studied, but also expert knowledge by the security authorities was utilized for finding the best solution. Finally, Guideline 7 (communication of research) was also clearly met, as results and findings were presented at international conferences, scientific publications, business work groups and other mainstream media. Table 24 summarizes how the guidelines were followed in this research.

Table 24. The presented DSR project mapped to the seven DSR guidelines.

<table>
<thead>
<tr>
<th>Guidelines</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline 1: Design as Artifact</td>
<td>An artifact, the IVSS was created, which consists of several sub-artifacts</td>
</tr>
<tr>
<td>Guideline 2: Problem Relevance</td>
<td>Close collaboration with security authorities, from gathering the requirements to the evaluation</td>
</tr>
<tr>
<td>Guideline 3: Design Evaluation</td>
<td>Theoretical evaluation and evaluation by case studies at the real-world surveillance sites</td>
</tr>
<tr>
<td>Guideline 4: Research Contributions</td>
<td>Reference design for large-scale high-performance video surveillance systems</td>
</tr>
<tr>
<td>Guideline 5: Research Rigor</td>
<td>Thorough research rigor due to the application of design science research methodology</td>
</tr>
<tr>
<td>Guideline 6: Design as a Search Process</td>
<td>The scientific knowledge base, as well as the expert knowledge from security authorities were the basis of the research</td>
</tr>
<tr>
<td>Guideline 7: Communication of Research</td>
<td>This research is communicated to technical-oriented audience, through scientific publications and conferences as well as management-oriented audience through talks at work group meetings and other presentations</td>
</tr>
</tbody>
</table>
According to Hevner’s design cycles (Hevner 2007), as shown in Fig. 9, this research was conducted in a constant loop between the creation of artifacts and their respective evaluation. From this process a number of scientific publications emerged during the search process. The WSSU was presented for the first time in 2008 (Sutor et al. 2008a), which initiated the research on ad-hoc surveillance (Sutor et al. 2008b). The idea of utilizing sensors from different modalities and combining them by data fusion was presented (Sutor & Reda 2008), soon followed by the first version of the multi-sensor analysis process (Sutor et al. 2009). Early versions of the research on system architecture were also published (Sutor 2010). Finally, a contribution to a scientific book chapter (Beer et al. 2009) on management and application development aspects of distributed smart camera systems was given.

6.7 Future outlook

Driven by the successful research activities as well as the scientific research during this study, the idea to create a product line based on these results arose, which was already started simultaneously. Further research will continue to be driven by user requirements from security authorities, leveraging the artifacts created in order to further enhance them and create new artifacts.

A number of future topics have been identified which pose challenges for research, and show promising business opportunities beyond academic research. As video surveillance is gaining more and more applications beyond security, this study is an ideal basis for exploring applications that can utilize and build upon the IVSS. The following application fields have been identified which go beyond security, and will be taken into account for future work:

- Compliance monitoring in production facilities
- Construction progress monitoring
- Performance and safety monitoring
- Customer analysis
- Traffic monitoring and traffic regulation enforcement
- Statistics for marketing and business intelligence
- Environment monitoring
- Smart city applications
- Traffic management
7 Conclusion

The ubiquity of surveillance cameras and systems led to new challenges, as the systems and operations behind those cameras could not cope with the rapid growth any longer. In today’s control rooms often hundreds or even thousands of cameras are displayed and security officers are overwhelmed by irrelevant information, distracting them from detecting and reacting to critical events (Porikli et al. 2013).

The purpose of this study was to scientifically assess user requirements first and then create a reference model of a system fulfilling these requirements. The research questions were created in close collaboration with experts from security authorities. Thus, deep insight into operations and processes of these authorities was given in order to support the development of better solutions for the problems and challenges they face in their daily business. Aggregating the security authorities’ demands, one main research question driving this research could be formulated:

What kind of video surveillance system is needed to maximize situational awareness, thus minimizing reaction time in critical situations, while being fully scalable to any number of cameras?

This research was based on design science research methodology, as introduced by Hevner et al. (2004), ensuring scientific rigor while constructing and evaluating design artifacts. This methodology is ideally suited for research aiming at finding solutions to real-world problems, incorporating expert and user knowledge into the process. Further, the research was conducted in the environment of a high-tech start-up company that plans to commercialize results and findings from this research. Hence, the applicability and usefulness of the conducted research in real-world scenarios was of great importance.

The scientific knowledge base was studied in a structured way by introducing a five-tier model. Work that addresses video surveillance and system architectures in general, such as presented by Kanade et al. (1999) or Fernández et al. (2013) was analyzed, then zooming in on a more detailed view, work that focuses on subcomponents was analyzed. This includes work in the area of communication such as presented by Fasui et al. (2013) and by Sandu et al. (2010), in the area of HCI by Feng et al. (2013) or Aydemir et al. (2012), in the area of sensors by Belbachir (2009), storage & retrieval by Vezzani and Cucchiara (2010b) or sensor processing introduced by Liu et al. (2013). However, no approach in the
knowledge base could be found which provided a complete system, designed specifically for automated multi-sensor surveillance in large-scale high-performance applications.

In this research a construct was created, the “Intelligent Video Surveillance System” (IVSS). It is a distributed, automated, highly scalable video surveillance software framework that can function as basis for any kind of high-performance video surveillance system. It is well suited to be utilized in a variety of scenarios, ranging from high-security installations such as at airports, prisons, or even military facilities, to applications such as large-scale, distributed video systems, enabling telecoms or security service providers to offer cloud-based video surveillance. The IVSS’s holistic approach contributes a video surveillance reference design that has a strong focus on applicability and usefulness in real-world scenarios.

First, in order to provide a strong foundation, a modular, large-scale system architecture was created. It was important to have an entirely distributed design in order to ensure virtually unlimited scalability and reliability by being able to replicate any component in the system. Attempting to tackle the information flood produced by these large-scale video surveillance systems, the multi-sensor analysis process was created. Thus, the analysis of data from multiple sources, combining video and other sensors in order to automatically detect critical events was possible. Hence, it became feasible to utilize video surveillance as a proactive tool, ideally preventing damage or loss instead of just browsing through video after an incident.

Further, an intelligent mobile client, the video surveillance local control, which addressed remote access applications, was created. This device should allow the security team to communicate directly with the control room and even to execute a set of commands to the system, or simply observe transmitted video data. Finally, the wireless self-contained surveillance unit was introduced, a novel smart camera concept that is capable of ad hoc and mobile surveillance. It is a camera with on-board memory and processing, and can be equipped with batteries in order to be fully self-contained. For communication a smart antenna utilizing different wireless networking technologies was utilized.

The applicability and usefulness of the IVSS could be demonstrated theoretically and practically by several use-cases at two real-world sites: an international airport, which has a large-scale video surveillance installation with high-security requirements, and a private security service provider, offering a multitude of video-based services by operating a video control center with
thousands of cameras connected in a cloud-based manner. Significant improvement when utilizing the IVSS could be demonstrated in multiple situations and configurations, thus contributing to the knowledge base by proving a reference model for large-scale high-performance video surveillance systems.

Finally, this work shall contribute to the protection of our personal freedom, the protection of our lives, our property and our society by supporting the prevention of crime and terrorism diminishing our personal freedom; ultimately, making this world a better place to live in.
References


Dornberger W (1952) V2 - Der Schuss Ins Weltall. Esslingen, Bechtle.


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


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


622. Matusek, Florian (2014) Selective privacy protection for video surveillance

623. Virtanen, Elina (2014) Effects of haulm killing and gibberellic acid on seed potato (Solanum tuberosum L.) and techniques for micro- and minituber production in northern latitudes

624. Kopatz, Alexander (2014) Genetic structure of the brown bears (Ursus arctos) in Northern Europe

625. Loukola, Olli (2014) Information networks among species: adaptations and counter-adaptations in acquiring and hiding information

626. Langrial, Sitwat (2014) Exploring the influence of persuasive reminders and virtual rehearsal on the efficacy of health behavior change support system


630. Stibe, Agnis (2014) Socially influencing systems: persuading people to engage with publicly displayed Twitter-based systems

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
LARGE-SCALE HIGH-PERFORMANCE VIDEO SURVEILLANCE

Stephan R. Sutor