ESPOO 2008 VTT PUBLICATIONS 691

Tomi Räty

Architectural Improvements for Mobile Ubiquitous Surveillance Systems

VTT PUBLICATIONS 691

Architectural Improvements for Mobile Ubiquitous Surveillance Systems

Tomi Räty

Academic Dissertation to be presented with the assent
of the Faculty of Science, University of Oulu,
for public discussion in Auditorium IT 116, Linnanmaa,
on the 28th of November, 2008, at 12 p.m.

ISBN 978-951-38-7113-0 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7114-7 (URL: http://www.vtt.fi/inf/pdf/)
ISSN 1455-0849 (URL: http://www.vtt.fi/inf/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 7001

VTT, Bergsmansvägen 5, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 7001

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O.Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax +358 20 722 7001

VTT, Kaitoväylä 1, PL 1100, 90571 OULU
puh. vaihde 020 722 111, faksi 020 722 2320

VTT, Kaitoväylä 1, PB 1100, 90571 ULEÅBORG
tel. växel 020 722 111, fax 020 722 2320

VTT Technical Research Centre of Finland, Kaitoväylä 1, P.O. Box 1100, FI-90571 OULU, Finland
phone internat. +358 20 722 111, fax +358 20 722 2320

Technical editing Leena Ukskoski

Edita Prima Oy, Helsinki 2008

Räty, Tomi. Architectural Improvements for Mobile Ubiquitous Surveillance Systems. Espoo 2008. VTT Publications 691. 105 p. + app. 55 p.

Keywords surveillance systems, mobile ubiquitous surveillance systems, Single Location Surveillance Point, logical decision making servers, multi-sensor systems, intelligent surveillance systems, video surveillance, video analysis, Quality of Service, audio surveillance, system architecture

Abstract

Surveillance systems have begun to be integrated into the common lives of humans and improved surveillance systems will spread even further. Systems manufacturers will continue to provide powerful surveillance systems with different aspects from single sensors to an abundance of different intelligent sensors. These different devices will have the ability to deliver a large variety of information to either remote or local surveillance personnel for immediate utilization or for extracting information about occurred events. The objective of this dissertation is to analyse the reduction of excessive information delivered to security personnel and the immediate delivery of essential alarms to security personnel by refining a design of a distributed multi-sensor intelligent surveillance system. The surveillance system created is reflected against the mobile and ubiquitous requirements of the end users of the surveillance system. The mobile requirement contains the reduction of excessive information distributed to the end user, a.k.a., the service personnel. The ubiquitous requirement consists of sensor data fusion and situation deduction. This dissertation uses a constructive research method, in which the results are validated by technical implementation and experimentation against mobile and ubiquitous requirements.

The major results of this dissertation are the prototype implementations of the Single Location Surveillance Point (SLSP) system. It consists of a selective amount of sensors that collect readings from a single location, which is the surveillance point. Each sensor transmits its crude sensor data to a session server, which handles the connections between the components. The session server routes the crude sensor information to the logical decision making service. The logical decision making server automatically deducts the situation at the surveillance point based on the received sensor information. The logical decision making server informs the security manager server of the situation at the surveillance point. The security manager server's user interface displays essential information about the surveillance point to a human security administrator. The security manager server can transmit information to the nomadic security personnel's smart phones over wireless networks.

Preface

I wish to thank my supervisor, Professor Petri Pulli, from the Department of Information Processing Science, University of Oulu, Finland, for his constructive counselling concerning the dissertation.

I would also like to thank the reviewers of this dissertation, Professor Jorma Jormakka from the National Defence University of Finland and Professor Yoshitsugu Manabe, Nara Institute of Science and Technology of Japan for their insightful suggestions and comments.

List of original publications

1. Räty, T. 2007. High-Level Architecture for a Single Location Surveillance Point. Proceedings of the Third International Conference on Wireless and Mobile Communications, ICWMC'07. Guadeloupe, French Caribbean, 4–9 March, 2007.

2. Räty, T., Oikarinen, J. & Sihvonen, M. 2007. A Scalable Quality of Service Middleware System with Passive Monitoring Agents over Wireless Video Transmission. Proceedings of the Sixth International Conference on the Quality of Information and Communications Technology, QUATIC 2007. Lisbon New University, Lisbon, Portugal, 12–14 September, 2007.

3. Räty, T., Lehikoinen, L. & Bremond, F. 2008. Scalable Video Transmission for a Surveillance System. Proceedings of the Fifth International Conference on Information Technology: New Generations, ITNG 2008. Las Vegas, Nevada, USA, 7–9 April, 2008.

4. Räty, T., Oikarinen, J., Nieminen, M. & Lindholm, M. 2008. Sensor Data Collection of the Single Location Surveillance Point System. Proceedings of the Seventh International Conference on Computer and Information Science, ICIS 2008. Portland, Oregon, USA, 14–16 May, 2008. To appear also in the Journal of Information Technologies and Control.

5. Räty, T., Lindholm, M., Nieminen, M. & Oikarinen, J. 2008. Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder. Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM 2008. Valencia, Spain, September 29 – October 4, 2008.

6. Räty, T., Luo, M., Oikarinen, J. & Nieminen, M. 2008. Testing and Validation of a Multi-sensor Distributed Surveillance System. Proceedings of the Seventh International Caribbean Conference on Devices, Circuits and Systems, ICCDCS 2008. Cancun, Mexico, 28–30 April, 2008.

Contents

Abbreviations

1GSS	1st Generation Surveillance Systems
2GSS	2nd Generation Surveillance Systems
2D	2-dimensional
3D	3-dimensional
3GSS	3rd Generation Surveillance Systems
4CIF	Four times Common Interchange Format
16CIF	Sixteen times Common Interchange Format
ADVISOR	Annotated Digital Video for Intelligent Surveillance and Optimized Retrieval
AES	Advanced Exterior Sensor
AI	Artificial Intelligence
AoI	Area of Interest
API	Application Programming Interface
CA	Cross-correlation Algorithm
CCD	Charge Couple Device
CCTV	Closed Circuit TV System
CDA	Cross-correlation Derivative Algorithm
CDS	Correlated Double Sampling
CEAD	Confidence Encapsulated Atomic Data
CIF	Common Interchange Format

COA	Course Of Action
CORBA	Common Object Request Broker Architecture
CPA	Closest Point of Approach
CPU	Central Processing Unit
CRLB	Cramer-Rao Lower Bound
CRVSS	Cluster Remote Video Surveillance System
DCT	Discrete Cosine Transition
DFIG	Data Fusion Information Group
DIVA	Distributed Interactive Video Array
DM	Decision Making
DOA	Difference Of Arrival
DSP	Digital Signal Processing
DTMF	Dual Tone Multi-Frequency
EAST	Environment for Automatic Systems Testing
EU	European Union
FAA	Federal Aviation Administration
FPS	Frames Per Second
GCC	Generalized Cross-Correlation
GIF	Graphics Interchange Format
GIS	Geographic Information System
GPS	Global Positioning System

GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
GUI	Graphical User Interface
HMM	Hidden Markov Model
HTTP	HyperText Transfer Protocol
IAA	Intelligent Alarm Analysis
ICT	Information Communications Technology
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IF	Information Fusion
IFS	Information Fusion System
IIR	Infinite Impulse Response
IMS	Internet protocol Multimedia Subsystem
ISO/IEC	International Organization for Standardization / International Engineering Consortium
IP	Internet Protocol
IS	Information System
ITU	International Telecommunications Union
ITU-T	International Telecommunications Union – Telecommunication
IVS	Intelligent Video Surveillance

J2EE	Java 2 Platform Enterprise Edition
JDL	Joint Directories of Laboratories
JMPD	Joint Multi-target Probability Density
JPEG	Joint Photographic Experts Group
JPEG2000	Joint Photographic Experts Group 2000
LAN	Local Area Network
LDMS	Logical Decision Making Server
LMS	Least Means-Square
KBIF	Knowledge-Based Information Fusion
M2M	Machine-to-Machine / Man-to-Machine / Mobile-to-Machine
MAS	Mobile Agents System
MEMS	Micro Electronic-Mechanical System
MIPSA	Modular Integrated Passenger Surveillance Architecture
MJPEG	Motion Joint Photographic Experts Group
ML	Maximum Likelihood
MPEG	Moving Pictures Expert Group
MPEG-1	Moving Pictures Expert Group 1
MPEG-2	Moving Pictures Expert Group 2
MPEG-4	Moving Pictures Expert Group 4
MSIS	Multiple Sensor Indoor Surveillance
NHPP	Non-Homogeneous Poisson Process

OODA	Object-Oriented-Decide-Art
OSO	Overall Spatial Observance
PDA	Personal Digital Assistant
PEI	Performance Effectiveness Index
PRISMATICA	PRo-active Integrated Systems for Security Management and Communication Assistance
PTRS	Program Tracking and Reporting Subsystems
PTZ	Pan, Tilt, Zoom
QCIF	Quarter Common Interchange Format
QoS	Quality of Service
QOSO	Q-observance measure Overall Spatial Observance
R-D	Rate - Distortion
R&D	Research & Development
RAM	Security Risk Assessment Method
RAM-C	Security Risk Assessment Methodology for Communities
RFID	Radio Frequency Identification
RR	Receiver Report
RT	Real-Time
RTP	Real-time Transport Protocol
RTS	Real-Time System
RTSP	Real-Time Streaming Protocol

SA	Situation Awareness
SA	Stereausis Approach
SASO	Sustainable and Security Operations
SGA	Spatial Gradients Approach
SIF	Standard Interchange Format
SLF	Spatial Likelihood Function
SOA	Service-Oriented Architecture
SOF	Spatial Observability Function
SLSP	Single Location Surveillance Point
SMS	Short Message Service
SMSU	Security Manager Server and UI
SQoS	Scalable Quality of Service
SRB	System “Knowledge” Base
SRP-PHAT	Signal-to-Reverberation PHAse Transform
SS-LMS	Sign-Sign Least-Mean-Squares
STI	Surveillance Test Intervals
TCP	Transmission Control Protocol
TDOA	Time Difference Of Arrival
TCR	Test Case Runner
TDOA	Time Difference Of Arrival
TV	Television

UDP	User Datagram Protocol
U.S.	United States
UI	User Interface
VCR	Videocassette Recorder
VHS	Video Home System
VIS	Vital Information Subsystem
VOL	Video Object Layer
VOS	Video Object Segmentation
VS	Video Surveillance
VSAM	Visual Surveillance and Monitoring
VSIP	Video Surveillance Interpretation Platform
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network
WRVISS	WLAN-based Remote Video Intelligent Surveillance System
WWW	World Wide Web
XML	eXtensible Markup Language

1. Introduction

Public safety and surveillance systems have become part of everyday life today. An increasing need for distributed multi-sensor intelligent surveillance systems has presented a substantial challenge to the software industry. The software industry has begun to examine approaches to use multiple sensors in a distributed environment to automatically raise alarms for surveillance personnel. Also, progress has transformed society into a more mobile and ubiquitous civilisation.

The inherent limitation in the effectiveness of CCTV surveillance systems is the cost of offering adequate human monitoring to cover for what is a considerably tiresome task. Consequently, CCTV tends to be used as a reactive tool and the perspective that a public transport operator is in charge of its space is lost if no response is acquired when a problem happens. The proactive approach is desirable, in which the likelihood of events can be recognized automatically to guide the attention and action of the human operators in charge of conducting a transport network. It is vital to perform this in a manner that sees surveillance systems as decision-support tools that human operators can use to address problems within complex and vast environments. [1]

There are immediate needs for automated surveillance systems in commercial, military, and law enforcement applications. Mounting video cameras is inexpensive, but locating available human resources to survey the output is expensive. What is required is an incessant 24 hour monitoring of surveillance video to alert security officers, while there still is time to prevent the criminal felony. [2]

Homeland security is an inherent concern for governments worldwide, which must protect their people and the critical infrastructures that uphold them. Information technology plays a significant role in such initiatives. It can assist in mitigating risk and enable effective responses to disasters of natural or human origin. [3]

Video monitoring usually deploys multiple video cameras, channelling video signals to a central monitoring room, where multiplexing is utilized to render a subset of the images to security personnel. Event detection and recognition use the perceptual capabilities of a human operator to discern objects moving within the field-of-view (FOV) of the cameras and to conclude their actions. However vigilant the operators, manual monitoring inevitably suffers from information overload, which results in periods of operator inattention due to fatigue, distractions, and interruptions. Automating all or part of this process would obviously offer dramatic benefits, ranging from the capability to alert an operator of a potential event of interest, through to a completely automatic detection and analysis system. [4]

CCTV devices have played a crucial role in the management of public places pertaining to safety and security. The explosion in the amount of cameras that must be monitored, the accruing costs of offering monitoring personnel and the limitations of human operators to uphold sustained levels of concentration severely circumscribe the efficaciousness of these systems. Alternatively, subsequent advances in information and communication technologies can potentially offer considerable improvements. The deployment of technology to maintain surveillance is used in modern urban environments. [5]

The key to security is situation awareness. Awareness requires information, which spans multiple scales of time and space. To offer comprehensive, non-intrusive situation awareness, it is vital to ply the challenge of multi-scale, spatiotemporal tracking. From the perspective of real-time threat detection, it is a known fact that human visual attention decreases below acceptable levels even when trained personnel are assigned to visual monitoring. [6]

Intelligent remote monitoring systems allow users to survey sites from significant distances. These systems exert rapid and efficacious remedial actions to be executed immediately once a suspicious activity is detected. An alert system can be employed to warn security personnel of impending vicissitudes and numerous sites can be concurrently monitored. This substantially abates the load of the security personnel. With the decreasing cost of computational power and advancement in Internet technologies, the implementation of a web-based security surveillance system becomes a considerable option to the traditional manually operated systems. Streaming technology enables video servers to transmit content in a subsequent stream, which can be decoded and played back shortly after it has been received by the client contraption. This is the preferred mode of operation. [7]

The security of places and public events are attempted to be improved by developing an innovative software approach in which distributed data coming from distinct devices are automatically correlated and analysed to provide security personnel with the right information at the right time.

Numerous tragic recent events have illustrated that despite the vast amount of technology deployed, adequate security of places and public events is not achieved. Citizens are demanding a higher level of protection. The analysis of the technologies available illustrates that the bottleneck of security of public places does not reside in surveillance hardware, but in the real-time analysis and correlation of the data provided by different sensors. They emphasize the lack of global management of threats for the people and their environments. Therefore, more “intelligence” is applied to surveillance systems.

Recent progress in computing, communication, and sensor technology are accelerating the development of multiple new applications. This trend is apparent in pervasive computing, sensor networks, and embedded systems. During the past two decades, surveillance systems have been an area of heavy research. Recently, considerable research efforts have been concentrated on video-based surveillance systems, especially for public safety and transportation systems. [8]

The increasing demand for safety and security has resulted in more research in constructing more efficient and intelligent automated surveillance systems. A future challenge is to develop a wide-area distributed multi-sensor surveillance system which has robust, real-time computer algorithms able to execute with minimal manual reconfiguration on variable applications. These systems should be adaptable enough to automatically adapt and withstand with the changes in the environment, such as lighting, scene geometry or scene activity. The system should be expandable; hence it should be based on standard hardware and use plug-and-play technology. [9]

Two substantial dilemmas of the security personnel are 1) the abundant amount of information that is distributed to them, and 2) discovering alarming events from this information. These, in consolidation, are distinguished into two requirements that, in unison, form a resolution to aforementioned dilemmas. The mobile requirement contains the requirement of reducing the superfluous information which is distributed to the end user. The ubiquitous requirement contains the requirement of fusing sensor data and deducing situations.

The resolution of these two dilemmas imposes stringent requirements to the architecture of surveillance systems. This dissertation attempts to resolve these two dilemmas through the proposed architectural improvements, and validate the resolution through an implemented prototype.

The definitions for the most important terms used throughout the dissertation summary are presented in the *introduction to the topic*. Then, the *motivation* for the distributed multi-sensor intelligent surveillance system is presented to express the desire for a distributed multi-sensor intelligent surveillance system, to locate the research gap and to reason why a distributed multi-sensor intelligent surveillance system is a desirable goal. Next, the *research questions, objectives and scope* of the study are laid out. The *research approach* is presented, containing a description on how the research was done and how the research results were validated. The *structure to the dissertation* provisions an overview of the dissertation summary.

1.1 Introduction to the topic

Middleware provides implementation guidelines and frameworks to ease the development of heterogeneous distributed systems [10]. They are typically computer software components that provide generic services which can be used by more than one application or an end user service. Middleware is typically used to support complex and distributed systems and applications. Middleware components are located at web servers, mobile devices, multimedia devices, application servers, content management systems and Personal Digital Assistant (PDA) equipment.

Surveillance personnel, or *security surveillance personnel*, are the individuals who survey an area which is under surveillance. The surveillance personnel, or security surveillance personnel, are either surveillance administration personnel, who reside in the control room, or nomadic surveillance personnel, who are the ambulating guards of the premises.

Nomadic surveillance personnel are people that perambulate from one location to another while using a wireless service [11]. The word nomadic is the indication of a person who moves constantly from one place to another.

Surveillance administration personnel are people that are located in the control room and survey the area under surveillance from a remote location.

End user refers to a person that uses a product that is a computer application. Therefore an end user is a human individual who utilizes a computer application. He may belong to either the surveillance administration personnel or nomadic surveillance personnel.

Software architecture is an essential part of software intensive products. Software architecture is the structure or structures of the system containing components, their relationships to each other and to the environment [12]. Software architecture also incorporates the principles guiding its design and evolution [13].

The biometrical sensor of the SLSP system is a fingerprint sensor which registers fingerprints at a door and transmits the access information derived from the access rights based on the access rights of the fingerprint.

The audio sensor of the SLSP system monitors the environment for threatening sound events. A threatening event is distinguished as an audio event that exceeds a pre-defined threshold of volume. The audio sensor indicates the bearing of the sound location. Sound recognition is omitted from this dissertation.

The video recorder of the SLSP system records and distributes video information.

The network activity monitoring sensor surveys all the IP-level network traffic inside the SLSP. The network activity monitor observes the data, both the amount and type, passing in the network and the devices.

The end devices are perceived as the devices that are handled by nomadic security guards.

The Area of Interest (AoI) is a distributed scalable video transmission subsystem, for a surveillance system, which concentrates on decrementing the amount of video information transmitted to the end-user equipped with a mobile device.

Quality of Service (QoS) indicates the nature of the packet delivery service provided, as represented by parameters such as achieved bandwidth, packet delay, and packet loss rates [14].

The Scalable Quality of Service (SQoS) is a middleware system which improves the control of the video transmission over a mobile system.

The Situation Deduction comprises of the employment of logical deductions to formulate an authentic comprehension of the event(s) happening in the surveyed area.

1.2 Research questions, approach, objectives and scope

Two substantial dilemmas of the security personnel are 1) the abundant amount of information that is distributed to them, and 2) discovering alarming events from this information. These, in consolidation, are divided into two requirements that, in unison, form a resolution to the aforementioned dilemmas. The mobile requirement contains the requirement of reducing the superfluous information which is distributed to the end user. The ubiquitous requirement contains the requirement of fusing sensor data and deducing situations.

The resolution of these two dilemmas imposes stringent requirements on the architecture of surveillance systems. This dissertation attempts to resolve these two dilemmas through the proposed architectural improvements, and validate the resolution through an implemented prototype.

The question of this dissertation is to examine how to collect, correlate and analyse automatically distributed data resulting from distinct devices, and instantaneously provide the security personnel essential, accurate information in distributed multi-sensor intelligent surveillance systems for public locations. Two specific requirements are taken into account, the mobile requirement and the ubiquitous requirement. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction. The answer to this question is divided into the following aspects. Resulting from the lack of a comprehensive review of existing distributed multi-sensor intelligent surveillance systems collecting data from a public location, the first question aims at reviewing the existing approaches of distributed multi-sensor intelligent surveillance systems collecting data from a public location. There is not an extensive body of literature on distributed multi-sensor intelligent surveillance systems, therefore the first question is categorized into five segments, video surveillance, audio surveillance, data fusion, architecture and communication, and testing surveillance systems. This analysis may prove useful for practitioners who require an immediate, careful review of the current state. The analysis informs researchers of the areas and fields that have been studied. The answer to the second question provides information on how to collect, correlate and analyse automatically distributed data resulting from distinct devices, and provide the security personnel essential, accurate information instantaneously in distributed multi-sensor intelligent surveillance systems for public locations.

The dissertation renders a design for a distributed multi-sensor intelligent surveillance system, which is called the Single Location Surveillance Point (SLSP). The SLSP system collects sensor data from multiple and distributed sensors. Logical deductions are established based on the sensor data, and alarms are automatically indicated to the security surveillance personnel. The aim of developing the SLSP system is to reduce the amount of excessive information shown to the modern surveillance personnel and improving the capabilities of acquiring authentic alarms instantaneously. This results in a greater collection of authentic alarms and a decrease in false alarms. These goals attempt to achieve the mobile and ubiquitous requirements applied to the dissertation. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction.

The two specific requirements of mobility and ubiquitousness are answered in the resolution of the second question.

The Mobile Requirement contains the requirement of reducing the superfluous information which is distributed to the end user.

The Ubiquitous Requirement contains the requirement of fusing sensor data and deducing situations.

Research question: Define architectural improvements to the 3GSS that

1) allow the utilization of mobility for security personnel (comprising the *mobile requirement*), and
2) allow ubiquitous utilization for wireless security personnel (comprising the *ubiquitous requirement*).

The approach to the antecedent research question can be divided further into the two subsequent research sub-questions:

Research sub-question 1: To what extent are distributed multi-sensor intelligent surveillance systems collecting data from a public location and transmitting intelligent information to surveillance administrators examined and answered by modern science? (Categorized into five segments, video surveillance, audio surveillance, data fusion, architecture and communication, and testing surveillance systems.)

Research sub-question 2: How to collect, correlate and analyse automatically distributed data resulting from distinct devices in indoor public locations, and instantaneously provide the security personnel essential, accurate information by the means of a distributed multi-sensor intelligent surveillance systems that abide to the mobility and ubiquitousness requirements?

Research intention 1:

The intention of the first research step is to peruse the different approaches proposed by the existent research to address distribute multi-sensor intelligent surveillance systems. The first objective is to identify distributed multi-sensor intelligent system's approaches that attempt to correlate to the essential information of distributed multi-sensor intelligent systems. This approach is segmented into subsections to address the subject in an addressable and complete manner. The review can assist practitioners in selecting approaches that correspond to the intention of an approach.

The second intention of the first research step is to contribute to the comprehension of the theoretical background of the existing distributed multi-sensor intelligent surveillance systems approaches. This is advantageous for both scholars and practitioners, as understanding the theoretical background expands their knowledge of why a particular distributed multi-sensor intelligent surveillance system approach has (or is expected to have) the desired impact on the surveillance system.

Research intention 2:

As earlier indicated, distinguishing the theoretical background of a distributed multi-sensor intelligent surveillance system's approaches benefits both scholars and practitioners. In addition, a distributed multi-sensor intelligent surveillance system's practitioners would benefit from concrete research on implemented approaches for their own systems. An indoor location was selected to prevent changing weather and lighting conditions, which reside out of the focus area of this dissertation. A public location was selected to bring out the assembly of alarms that could be raised through a vast area that is accessible to miscellaneous individuals, who may possess either evil intentions or good ones. By describing the implemented automatic collection, correlation, and analysis of distributed multi-sensor data from a public location to establish instantaneous, essential and accurate logical deductions for the surveillance personnel forms a validation of achieving the problem of the second research problem. It also forms a basis to proceed with the research in this segment of surveillance systems. This also formulates a basis for the commercialization of distributed multi-sensor intelligent surveillance systems. Consecutively, the first target of the second research step is to examine how the aspect of the research problems can be addressed. Finally, the second goal of the second research step is to design an approach that handles the problem and to test the approach pragmatically. The SLSP surveillance system is reflected against the mobile and ubiquitous requirements. The mobile requirement contains the reduction of superfluous information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction.

1.3 Research strategy

The dissertation utilizes a constructive research approach ([15], [16]; see Table 1). Conceptual analysis is applied to the first research question. The second research question employs conceptual analysis to distinguish the research concepts of existing distributed multi-sensor intelligent surveillance systems. Constructive research is performed to prove a novel, distributed multi-sensor intelligent surveillance systems approach and to test the distributed multi-sensor intelligent surveillance systems in practice.

Table 1. Research strategy: research approaches for resolving the research questions.

Research step	Chapters	Research approaches
To what extent are distributed multi-sensor intelligent surveillance systems examined and resolved by modern science?	2, 3, 4	Conceptual analysis
How to collect, correlate and analyse automatically distributed data resulting from distinct devices in public locations, and provide security personnel essential, accurate information instantaneously with a distributed multi-sensor intelligent surveillance system?	4, 5	Conceptual analysis and constructive research

1.4 Structure of the dissertation

The remainder of this dissertation is constructed as follows. The second chapter presents an overview of existing distributed multi-sensor intelligent surveillance systems approaches. The third chapter reviews these approaches and a critical analysis of the existing distributed multi-sensor intelligent surveillance systems approaches is presented.

The fourth chapter presents the introduction to the original publications, which forms a new model for distributed multi-sensor intelligent surveillance systems, which is a grouping of (1) multiple sensors collecting information from their environment, (2) logical intelligence to derive essential information based on the output of the sensors, (3) a scalable video transmission component which enables video reception over a wireless network in low coverage conditions, (4) a device for the reception of all the information transmitted by the sensors and logical intelligence, and (5) an intercommunication network for transmitting and receiving information related to the previously mentioned applications and devices. This is joined with logical intelligence that provides essential and accurate information of an area under surveillance. The fifth chapter presents the validation of the distributed multi-sensor intelligent surveillance system and emphasizes its connections to the publications. The sixth chapter presents the conclusions (including a summary), the limitations and the future research related to the achieved results.

2. A review of existing distributed multi-sensor intelligent surveillance systems

2.1 Distinguishing the source material for the literature review

A thorough review of the literature should contain all the essential literature on the topic without being restricted to one research methodology, one collection of journals or one geographic region ([17] pp. xv–xvi). The literature review presented in the second and third chapters of this dissertation attempts to cover all existing surveillance systems automatically collecting, correlating and analyzing distributed data from distinct devices, and providing security personnel with accurate information instantaneously. To achieve this attempt, the following process was used to distinguish the source material for the review. Surveillance systems were examined through the aide of digital databases (e.g., ACM Digital Library, IEEE/IEE Electronic Library, CiteSeer.IST). Additionally, conference proceedings were examined directly and by user the previously mentioned electronic databases.

2.2 A review of existing distributed multi-sensor intelligent surveillance systems' approaches

The basic rule of surveillance systems is to collect information from an area and distribute the information to security personnel. There are four main branches in surveillance systems: 1) according to the type of sensor acquiring the information, 2) transmission of information, 3) data fusion and event detection (if any), and 4) information rendering to the security personnel. The first branch can be sub-divided related to the sensor type utilized. The majority of surveillance systems focus on video and audio surveillance.

First, a high-level summary is presented on the topic of surveillance systems in general, containing a brief history of surveillance systems and their generations. Then video surveillance will be presented, accompanied with a sub-chapter regarding video analysis. This is followed by audio surveillance, with the focus point being on audio detection, not interpretation. The utilization of wireless networks in the area of surveillance systems is reviewed next. The subsequent chapter will consider data fusion and situation derivation. The review will be concluded with a short summary of testing surveillance systems.

2.3 Introduction to surveillance systems related to public areas

Valera and Velastin indicate that intelligent visual surveillance systems address the real-time monitoring of static and dynamic objects within a specific environment [9]. The primary goals of these systems are to offer an automatic interpretation of scenes, to understand and predict the actions and interactions of the observed objects based on the information gathered by sensors [9]. The recent interest in surveillance regarding public, military, and commercial scenarios is raising the need to establish and deploy intelligent or automated visual surveillance systems [9]. Currently, there tends to be a lack of contribution from the field of system engineering to the research [9]. Table 2 reviews the technological evolution of intelligent surveillance systems, viz. 1st, 2nd, and 3rd generations, sketching the main problems and modern research in each of them [9]. Regazzoni et al. inform that the three generations are according to the evolution of communications, processing, and storage and they have evolved in recent years with the same increasing rate as these technologies [18]. Tabar et al. denote that third-generation surveillance systems is the term occasionally utilized in literature to refer to systems created to handle with a large number of cameras, a geographical spread of resources, many monitoring points, and to mirror the hierarchical and distributed nature of the human process of surveillance [19].

Valera and Velastin acknowledge that society's increasing demand for security results in a growing need for surveillance activities in many environments. Recently, the demand for remote monitoring relative to safety and security reasons has received significant attention, particularly in the following areas: 1) transport applications, such as airports, to survey traffic; 2) public places, such as department stores; 3) remote surveillance of human activities, such as attendance at soccer matches; and 4) surveillance to procure a certain quality of control in many industrial processes, surveillance in forensic applications and remote surveillance in military applications. The basic goals that are expected of a 3rd generation vision surveillance application, based on end-user requirements, are to offer good scene understanding, oriented to attract the attention of the human operator in real-time, possibly in a multi-sensor environment, surveillance information and utilizing low cost standard components. [9]

Table 2. Review of technical evolution of intelligent surveillance systems [9].

1st Generation	
Techniques	Analogue CCTV systems
Advantages	– They give good performance in some situations – Mature technology
Problems	Use analogue techniques for image distribution and storage
Current research	– digital versus analogue – digital video recording – CCTV video compression
2nd Generation	
Techniques	Automated video surveillance by combining computer vision technology with CCTV systems
Advantages	Increase the surveillance efficiency of CCTV systems
Problems	Robust detection and tracking algorithms required for behavioural analysis
Current research	– Real-time robust computer vision algorithms – Automatic learning of scene variability and patterns of behaviours – Bridging the gap between the statistical analysis of a scene and producing natural language interpretations
3rd Generation	
Techniques	Automated wide-area surveillance system
Advantages	– More accurate information as a result combining of different kinds of sensors – Distribution
Problems	– Distribution of information (integration and communication) – Design methodology – Moving platforms, multi-sensor platforms
Current research	– Distributed versus centralized intelligence – Data fusion – Probabilistic reasoning framework – Multi-camera surveillance techniques

In the previous work of Li et al., they had developed a multimedia application, an Internet-based surveillance service, which allows users to perceive real-time snapshots on the spot, from anywhere and at anytime. This could be done to gather suspicious historical scenarios, to access heterogeneous media, and to transmit emails or paging signals. Such a service can be only applied to wired-connection desktops and could not accommodate to increasing pervasive requirements from mobile users, who carry handheld devices, such as personal digital assistants (PDAs). This prevents the delivery of Internet-based mobile information services. To accommodate to the environment of low bandwidth, the scene snapshots need to be transmitted efficaciously and briefly. [20]

Ho et al. recognize that recent progressions in the third-generation mobile communication systems allow the transmission of video information using mobile channels. One of the possible applications in this aspect is real-time road traffic monitoring utilizing the mobile videophones or similar handheld video communication devices. Even though the third-generation mobile system has a wide communication bandwidth at its disposal, an efficient coding technique is still in demand for transmitting real-time road traffic video. [21]

According to Hampapur et al., the key to security is situation awareness. Awareness requires information, which spans multiple scales of time and space. Assuring high levels of security at public access facilities, such as airports and seaports, is a complex challenge. Modern video surveillance systems perform as large-scale video recorders, either analogue or digital. These systems serve two purposes: offering a human operator images to detect and react to potential threats and recording evidence for investigation reasons. While these are the initial steps in employing video surveillance to improve security, they are inadequate for supporting both real-time threat detection and forensic investigation. [6]

Hampapur et al. denote that from the perspective of real-time threat detection, it is a known fact that human visual attention decreases below acceptable levels even when trained personnel are assigned to visual monitoring. Automatic video analysis technologies can be applied to develop smart surveillance systems, which can assist the human operator in both real-time threat detection and forensic investigatory tasks. The most critical challenge in video-based surveillance from the perspective of a human intelligence analyst, is interpreting the automatic analysis data to detect events of interest and identify trends. Modern systems have just begun to review automatic event detection. The region of context-based interpretation of the events in a monitored space needs to be examined. [6]

Reiter and Rohatgi propose that homeland security is an essential concern for governments worldwide, which must protect their people and the critical infrastructures that uphold them. Information technology plays a significant role in such initiatives [3]. It can assist in reducing risk and enable effective responses to disasters of natural of human origin [3]. Pham & Xie acknowledge that with the development of modern technologies and the evolution of the industrial society, sundry engineering systems have been developed and are becoming more detailed and complex [22]. Typically, a system comprises multiple subsystems and every subsystem is frequently inspected to keep them functional [22].

Fong & Hui aver that security surveillance systems are becoming important in situations in which personal safety could be compromised resulting from criminal activity. As security personnel typically monitor multiple locations simultaneously, this manual task is labour intensive and inefficient. Significant stress may be placed on the security personnel involved. Intelligent remote monitoring systems allow users to survey sites from significant distances. This is especially useful when numerous sites require security surveillance concurrently. These systems practice rapid and efficient corrective actions to be performed immediately once a suspicious activity is detected. An alert system can be employed to warn security personnel of foreboding problems and numerous sites can be concurrently monitored. This substantially reduces the load of the security personnel. [7]

Fong & Hui denote that by utilizing the Internet as the communications medium for real-time transmission of video signals in such a security-sensitive operation, many technological issues need to be resolved. First, the system needs to tolerate potentially significant bandwidth restrictions. At any time the available Internet bandwidth is restricted and needs to be shared among all the Internet users. A great amount of data flow can cause network congestion. The system must provide real-time transmission of video signals even though there might be only a small amount of bandwidth available. Robust and efficient error control mechanisms and video compression techniques need to be utilized to subvert the problems related to limited bandwidth. The second point is that the streaming technology enables video servers to transmit content in a following stream, which can be decoded and played back shortly after it has been received by the client device. [7]

Regazzoni et al. recognize that a surveillance system can be defined as a technological tool that assists humans by offering an extended visualization and reasoning capability about situations of interest that occur in the monitored environments [18]. Human perception and reasoning are constrained by the capabilities and the limits of human senses and mind to simultaneously collect, process and store a limited amount of data [18]. Collins et al. denote that an individual human operator cannot efficiently monitor a vast area by viewing dozens of monitors displaying raw video output [2]. That quantity of sensor overload virtually asserts that information will be ignored and the information requires an unacceptable amount of transmission bandwidth [2].

Regazzoni et al. acknowledge that a surveillance system should be complete and it should enable user oriented data accessibility both for the raising of direct alarms and for off-line inspection. There is a requirement for 3GSS system researchers and designers to understand realistic use case scenarios of these systems and to interpret end user requirements to design practical and effective systems. In table 3, real-world applications are categorized by Regazzoni et al. It includes their functional requirements and cost/performance requirements. [18]

Regazzoni et al. recognize that there is a rapid growth of metropolitan localities that need to offer improved safety and security to the public. If the automated system creates too many false alarms, the human operator would tend to ignore the automated system and the intelligent function will be switched off. The problem is compounded when multiple event types are automatically created. A false alarm in this case is the detection of a change in the scene as a person. Another system requirement is the reaction time for these systems, i.e., the time required for the system to create an alarm. Normal reaction times may vary depending on the event, but it is reasonable to expect reaction times in a few seconds. Another substantial pitfall in embedding these intelligence functions in real-world systems is the lack of robustness, the inability to test and validate these systems under a variety of use cases, and the lack of quantification of these systems' performance. Additionally, the system should gracefully degrade in performance as the complexity of data enlarges. This is a very open research issue that is vital to the deployment of these systems. The main problems currently considered are related to either real-time distributed or centralized processing, and robustness issues in multi-sensor surveillance networks. [18]

Table 3. The real-world applications according to Regazzoni et al. [18].

Application Domain	**Primary Benefits**	**Intelligent Functionality desired**	**Cost and Performance requirements**
Public area monitoring, large area monitoring	Safety, security	Person/vehicle detection, tracking and event analysis	Low system cost, false alarm/detection requirements rather stringent
Building exterior and interior monitoring, parking garage monitoring	Security, safety, access control, building automation	Person/vehicle detection, parking space monitoring, license plate recognition, face recognition	High-end market. High reliability desired in access control. Illumination is controlled/ unconstrained
Subway, highway, tunnel monitoring, transportation applications	Safety, security, resource management and improved quality of service	People detection and tracking, vehicle, truck detection/tracking, classification of object type, recognition of events	Few high-end systems exist on the market. Very low false alarms rates. All weather and illumination conditions
Indoor monitoring (malls, lobbies, banks, shopping complexes)	Security and safety	Person detection, tracking, event analysis	Low cost systems, minimal false alarms

Detmold et al. take a thorough survey of the entire field of automated video surveillance [23]. The approach of Detmold et al. does not consider video surveillance to be principally a real-time application, and neither their architecture nor middleware implementing it are oriented towards real-time requirements [23]. Pavlidis et al. inform that the modern security infrastructure could be summarized as follows: 1) security systems act locally and they do not cooperate in an efficient manner; 2) Extremely high value assets are inadequately protected by old-fashioned technology systems; 3) Dependence on intensive human concentration to detect and assess threats [24].

Ott et al. denote that a generic surveillance and security is built of three essential parts: data collection, information analysis, and on-field operation. Any surveillance system requires the means to monitor the environment and gather data in the form of video, still images, audio, etc. Such data is to be processed and analysed by a human, a computer or a collection of both at a command centre. An administrator can decide on performing an on-field operation to put the environment back into a situation considered as normal. On-field control operations are practised by on-field agents who require efficient communication channels to keep a close interaction with the command centre. Security personnel review their wireless video systems for critical incident information. The need for providing detailed real-time information to the surveillance agents has been identified and is being addressed by the research community. [25]

Petrushin et al. acknowledge that the growth of a wide variety of sensors in public areas has created opportunities for development of security and business applications. A scalable system built for this class of tasks should be able to integrate these sensor data with contextual information and domain information offered by both the humans and the physical environment to maintain a coherent picture of the world over time. The performance of the majority of the systems is far from what is required for real-world applications. [26]

Valencia-Jimenez & Fernandez-Caballero recognize that the incorporation of distributed artificial intelligence has brought forward the development of new technologies in detection (sensors and captors), robotics (actuators), and data communication. Communication among the system's elements is essential, because alarms need to be spread along the subsystems and help or collaborate with other platforms to define the situation and act accordingly. If the multi-sensor platform is dynamic, the communication link should be wireless. If the position is static, the link can either be wireless and wired. Wireless communication is preferred, because the distance to other nodes may be considerable. [27]

Atrey et al. indicate that resulting from the increase of public security threats, the majority of the cities around the world are being equipped with thousands of sensors, including video cameras and audio sensors, with a primary goal of monitoring and recording interesting events as they occur in the area under surveillance. In the modern generation of surveillance systems, in which multiple asynchronous and different sensors are used, the combination of the information gathered from them to derive the events from the environment is an important and challenging research problem. Information combination refers to the process of combining the sensor and non-sensor information using the context and past experience. The issue of information combination is vital, because the information gathered from multiple sources when combined offers more precise inferences of the environment than individual sources. [28]

Cucchiara acknowledges that society requires the results of research activities addressing new solutions in video surveillance and sensor networks. The demand for security and safety calls for new generations of multimedia surveillance systems, in which computers will act not only as supporting platforms, but they will work as the substantial core of real-time data understanding process. The adjective “multimedia” typically refers to systems and services created for human end-users for accessing and utilizing multimedia data, multimedia streams, multimedia content, and multimedia interfaces in many different applications. According to this abstraction, a multimedia surveillance system should easily be a surveillance system capable of achieving the output of the task in a multimedia format. The concept of multimedia surveillance systems is a system that is capable of furnishing multimedia data, as well as gathering, processing in real-time, correlating and addressing multimedia data resulting from different sources. [29]

Velastin et al. present the EU-funded project PRO-active Integrated systems for Security Management by Technological, Institutional, and Communication Assistance (PRISMATICA) was part of the effort to make public transport systems more appealing to passengers, safer for passengers and staff and operationally cost effective [1]. The major goal of PRISMATICA is to detect certain types of behaviours, which are distinguished from public transport management requirements [30]. Attwood & Watson proclaim that ADVISOR (Annotated Digital Video for Intelligent Surveillance and Optimized Retrieval) was developed in an EU-funded project on innovative architectures for public transport systems [31].

Valera & Velastin recognize that even though both systems are classified as distributed architectures, they have substantial differences in that PRISMATICA utilizes a centralized approach and ADVISOR can be considered as a semi-distributed architecture. PRISMATICA is built with the concept of a main or central computer, which controls and supervises the entire system. ADVISOR can be seen as a network of independent dedicated processor nodes, avoiding a single point-of-failure at first sight. In each node there is a central computer, which controls the entire node. Hence, there is a single point-of-failure within each node. [30]

According to Velastin, there is a growing interest and demand for the development and distribution of surveillance systems in private and public environments. Traditional approaches rely on the installation of wide-area closed-circuit television (CCTV). CCTV requires a relatively small amount of operators to constantly monitor a significant number of cameras and other devices. [32]

2.4 Video surveillance

According to Greiffenhagen et al., in the commercial sector, there is an increasing need for monitoring and video surveillance. There is an increased use of video surveillance system in urban areas. Visual surveillance and monitoring (VSAM) systems are constantly becoming stronger factors in prevention and reduction of criminal offences and in the improvement of efficient management of resources. The accuracy requirements are typically defined in terms of detection and false alarm rates for objects, while the computational requirement is specified commonly by the system response time to an object’s presence, e.g., real-time or delayed. It is still an art to engineer systems that satisfy application-specific requirements. There are two basic steps in the design process: the choice of the system architecture and the modules for achieving the task, and the statistical analysis and validation of the system to check if it fulfils user requirements. In real life, the system design and analysis phases usually follow each other in a cycle until the engineer establishes a design and a suitable analysis that satisfies the user specifications. [33]

Bramberger et al. inform that recent progress in computing, communication, and sensor technology are inciting the development of multiple new applications [8]. Recently, considerable research efforts have been concentrated on video-based surveillance systems, particularly for public safety and transportation systems [8]. Bartolini et al. denote that recent advances in telecommunication and electronic technology, enchained with the development of improving powerful signal and image processing techniques, essentially broaden the scope and quality of automatic video surveillance (VS) systems [34]. Research is currently being performed by several industrial and academic institutions to improve automatic surveillance in terms of continuous and efficient monitoring, cost reduction and reliable control of dangerous and remote sites [34].

Makris & Ellis announce that video surveillance has become a ubiquitous aspect of the modern urban landscape [35]. In some cases surveillance functions as a persuader, preventing unacceptable social behaviour that can no longer be engaged anonymously, recording and logging events for evidential reasons, or offering remote observation of sensitive locations where tight access control is important [35]. Trivedi et al. indicate that recently video surveillance activity has grown significantly [36]. According to Trivedi et al., research interests have moved from invisible static image-based analysis to video-based dynamic monitoring and analysis [36]. Installing multiple sensors proposes new design aspects and challenges [36].

Muller et al. denote that visual surveillance systems are used for observation and protection of private and public regions [37]. For this purpose, a multiview video streaming system has been developed, which can contain a decisive station [37]. Desurmont et al. recognize that video surveillance is a large market as the amount of installed cameras can indicate [38]. There are several requirements for these systems [38]. They must be network connected, entail multiple cameras, modular, the user interface needs to be user-friendly, and the entire system has to be reliable and robust [38].

Foresti et al. acknowledge that safety and security have become critical in numerous public areas, and there is a specific need to enable human operators to remotely monitor activity across large environments, such as: 1) transport systems, 2) shopping malls, 3) industrial environments, and 4) government establishments. Modern video-based surveillance systems use real-time image analysis techniques for efficient image transmission, colour image analysis, event-based attention focusing, and model-based sequence comprehension [39]. According to Velastin et al., the surveillance of public places is assembled with numerous key factors, such as: 1) the widespread geographical extent of what must be addressed; 2) a vast region of behaviours that require the attention of human operators; 3) the variety of type of information that must be handled to estimate a situation, e.g. vision and sound; and 4) the requirement of transmitting information within a hierarchical system of control [1].

Detmold et al. inform that that at the hardware level, it is possible to construct thousands of camera networks at a reasonable cost by utilizing IP networking devices and IP video cameras. Monitoring such networks through human inspection is inefficient in its usage of human resources. Trained operators lose their concentration to the extent of missing a significant percentage of considerable events after only ten minutes of viewing camera images. Detmold et al. proclaim that many of the challenges are general to the video surveillance domain, rather than designated to certain surveillance algorithms. Middleware can assist with these general aspects of video surveillance network construction, containing support for both computation and communication. [23]

Detmold et al. inform that the field of automated video surveillance is quite novel, and the majority of modern approaches are engineered in an ad hoc manner. Recently, researchers have begun to consider architectures for video surveillance. Middleware that provides general support for video surveillance architectures is the logical next step. It should be noted that while video surveillance networks are a class of sensor networks, the engineering challenges are quite different. Especially the requirement for extreme savings in use of power and network bandwidth, which is a dominating factor in most sensor networks, is left out from most surveillance networks. A video surveillance network is a detailed distributed application, and requires sophisticated support from middleware. The middleware’s role is primarily to support communication between modules. [23]

Cucchiara acknowledges that multimedia surveillance systems can improve visual data with audio streams and information resulting from other sensors. In vast distributed environments, the exploitation of networks of small cooperative sensors should substantially improve the surveillance capability of a few higher levels sensors, such as cameras. [29]

2.4.1 Video analysis

Micheloni et al. inform that object tracking is an important task for many applications in the region of computer vision and particularly in those relevant to video surveillance. Recently, the research community has concentrated its interests on developing smart applications to improve event detection capability in video surveillance systems. Every detected object is tracked and their trajectories are analysed to deduct their movement in the scene. [40]

Bowden & KaewTraKulPong acknowledge that intelligent visual surveillance is a vital application area for computer vision [41]. Situations in which networks of hundreds of cameras are used to cover a wide area, the obvious restriction is the user's ability to manage vast amounts of information [41]. Kreucher et al. state that the difficulty of tracking an individual manoeuvring target in a cluttered environment is a well-examined region [42].

Hu et al. recognize that as an active research topic in computer vision, visual surveillance in dynamic scenes attempts to detect, recognize and track certain objects from image sequences, and more typically to understand and describe object behaviours [43]. According to Kumar et al., a thorough video-based surveillance system executes the following functions: 1) detection of mobile objects; 2) tracking of mobile objects through the image sequence; 3) classification of tracked targets; and analysis of the tracked targets' behaviour [44].

Bremond et al. indicate that one of the most demanding problems in the domain of computer vision and artificial intelligence is video understanding [45]. The research in this area mainly focuses of the development of methods for analysis of visual data to extract and process information about the behaviour of physical objects in a scene [45]. Carincotte et al. acknowledge that advancements in sensor, communications and storage capacities render it easier to gather a large amount of multimedia material [46].

2.4.2 Video Quality of Service

Maier et al. suggest that in traffic surveillance for example, services like MPEG-video streaming, typically have high demands in QoS. Typical QoS parameters contain frame rate, transfer delay, image resolution, and video compression rate. Further power savings are attained by graceful degradation of QoS. There has been research executed between the trade-off of image quality and power consumption. [47]

Korshunov & Ooi propose that a large-scale distributed video surveillance system usually contains many video sources distributed over a large area, transmitting live video streams to a central location for monitoring and processing. Modern advances in video sensors and the increasing availability of networked digital video cameras have allowed the distribution of large-scale surveillance systems over existing IP network infrastructure. Numerous commercial enterprises offer IP-based surveillance solutions. Implementing an intelligent, scalable and significantly distributed video surveillance system remains a research problem. Researchers have not paid too much attention on the scalability of video surveillance systems. [48]

May et al. acknowledge that in a large surveillance system, the digital network that enables remote monitoring, storage, control and analysis is not in a single LAN [49]. It typically indicates a collection of interconnected LANs, wired or wireless, with different bandwidths and QoS [49]. Different types of clients connect to these networks and access one or multiple video sources, decode them at the temporal and spatial resolution they require, and provide different functions [49]. Frescura et al. announce that in wireless standards there is the need for robust multimedia transmission [50]. The applications require the best trade-offs between QoS, e.g., image quality at the end receiver, bandwidth, e.g., transmission rate, and delay [50].

Bramberger et al. inform that in video-based surveillance, normal QoS parameters contain frame rate, transfer delay, image resolution, and video-compression rate. The surveillance tasks might also provide multiple QoS levels. Mobile agents are employed to support the development of their distributed surveillance system. Mobile agents are the most appropriate for this distributed application, because each surveillance task can be encapsulated within a mobile agent, which can then move between cameras. This approach is also highly endurable and scalable. [8]

2.5 Audio surveillance

Stanacevic & Cauwenberghs announce that accurate and robust localization and tracking of acoustic sources is of interest to a variety of applications in surveillance, multimedia, and hearing enhancement [51]. Julian et al. claim that sound localization employing compact sensor nodes deployed in networks has applications is surveillance, security, and law enforcement [52]. Coherent methods are based on the arrival time differences of the acoustic signal to the sensors. In standard systems, microphones are separated to maximize precision, hence the nodes must attain synchronization to introduce a valid estimate [52].

Smeaton & McHugh aver that audio surveillance is typically performed using one or multiple microphones that are wired up to a central unit. The audio information for that location is captured and the analysis of the captured information is executed by a separate processor, either continuously as the audio is streamed in, or afterwards on stored audio. With audio, one can have event detection on a graded scale, from minor events to abnormal sounds. The purpose of an audio sensor network would be to assist the end user in reviewing through data and to return the points of interest. It is potentially inexpensive to distribute, thus it is a good complement to CCTV. [53]

Aarabi acknowledges that the sound localization methods, such as the ones that are presented in this study, usually presume that the location and orientation of the microphone array is known [54]. In practical situations, such information may not be available [54]. Julian et al. recognize that the determination of sampling frequency with cross-correlation algorithms is essential to the accuracy of bearing detection [52]. The lower the sampling frequency is, the greater the distance between the microphones must be [52].

2.6 Sensor and data fusion

Wald proposes the following definition: Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources [55]. Hall suggests that one brief way to define sensor and data fusion is the following: sensor Fusion is “Data Fusion from Multiple Sensors (same or different sensor types)” and data Fusion is “Combining information to estimate or predict the state of some aspect of the world” [56].

Steinberg et al. acknowledges that data fusion involves combining information [57]. In the broadest sense, data fusion is used to estimate or predict the state of some aspect of the universe [57]. These may be represented in terms of attributive and relational states [57]. Steinberg informs that fusion involves the use of multiple data, which typically result from multiple sources, to estimate or predict the state of some aspect of reality, which are treated as if they were independent of the states of other entities. [58].

Jaeger indicates that RAM-C (Security Risk Assessment Methodology for Communities) is a systematic, risk-based process to assist communities in evaluating threats, prioritizing targets, identifying consequences, and reviewing completeness and effectiveness of physical security and response systems. RAM-C assists communities and facility managers in determining how well potential targets are protected. [59].

Blasch & Plano aver that increasingly complex scenarios require more intelligent and efficient processing strategies for multi-sensor information fusion and target tracking. This is integral to any information processing is decision making (DM). A smart sensor is either a single sensor or a subsystem of different sensor components coordinating to provide data and intelligent algorithmic output to aid or conduct decision making in the larger system. The combination of sensor data has to be delivered to a computer for processing and displayed to a user as information for decision making. [60]

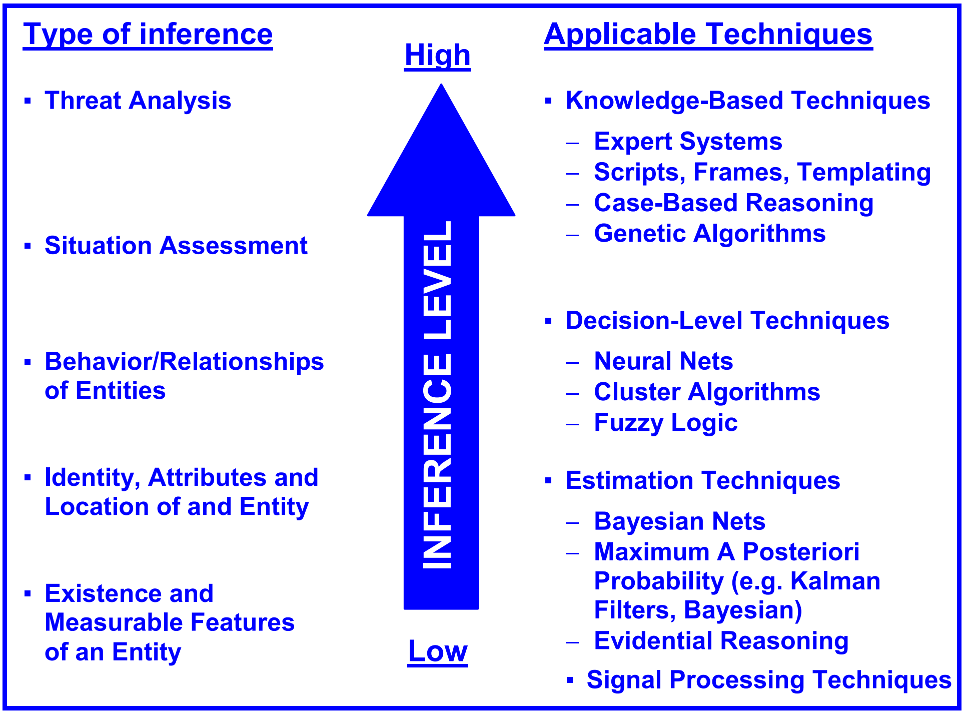
Blasch & Plano announce that a chief evaluation goal related to any system is the ease of adequate situation awareness (SA). SA is not automatically guaranteed for the operator dependant on novel fused hybrid sensor systems. Although these appear to promise much desired increases in capacity, the human awareness process capacity is a bottleneck in overall process operation. The metrics chosen contain timeliness, precision, throughput, confidence, and cost. These metrics resemble the standard QoS metrics in communication theory and human factors literature, as illustrated in Table 4. [60]

Table 4. Metrics for various disciplines according to Blasch & Plano [60].

COMM	Human factors	Info fusion	ATR/ID	TRACK
Delay	Reaction time	Timeliness	Acquisition / Run-time	Update rate
Probability of error	Confidence	Confidence	Prob. (Hit), prob. (FA)	Prob. of detection
Delay variation	Attention	Accuracy	Positional accuracy	Covariance
Throughput	Workload	Throughput	No. images	No. targets
Cost	Cost	Cost	Collection platforms	No. assets
Stallings, 2002	Wickens, 1992	Blasch, 2003	Blasch, 1999	Hoffman, 2000

Blasch & Plano recognize that applications for multi-sensor information fusion (IF) require analysis of how these systems will be distributed and used. Increasingly complex scenarios arise, demanding more intelligent and efficient reasoning strategies. Substantial to information reasoning is decision making (DM) which requires pragmatic knowledge representation for user interaction. IF (information fusion) manages data, sensors, and people. The ability to develop SA (situation awareness) based on the real world environment would have user reasoning about the data to deduct information. The current control requirements are user, sensor, and mission administration. For instance, if sensors are on platforms, then the highest-ranking official distinguishes who gets control of the assets. [61]

Hall acknowledges that numerous multi-sensor systems have been constructed to collect, process, and spread image and non-image data. For many applications, new mobile platforms are being developed for surveillance and monitoring. Modern sensor suites may include image sensors, non-image sensors (acoustic emissions, etc.), and the ability to include reports from human surveyors. Wideband communication links increasingly enable the use of data across distributed systems. The goal of multi-sensor fusion is to achieve inferences about the observed environment or situation that cannot be achieved by a single sensor or source of information. Information about the observed situation is combined to achieve high-level inferences. This is presented in Figure 1, multiple techniques may be used to achieve these high-level inferences [62].



Type of inference

- Threat Analysis
- Situation Assessment
- Behavior/Relationships of Entities
- Identity, Attributes and Location of and Entity
- Existence and Measurable Features of an Entity

High

INFERENCE LEVEL

Low

Applicable Techniques

- Knowledge-Based Techniques
  - Expert Systems
  - Scripts, Frames, Templating
  - Case-Based Reasoning
  - Genetic Algorithms
- Decision-Level Techniques
  - Neural Nets
  - Cluster Algorithms
  - Fuzzy Logic
- Estimation Techniques
  - Bayesian Nets
  - Maximum A Posteriori Probability (e.g. Kalman Filters, Bayesian)
  - Evidential Reasoning
- Signal Processing Techniques

Figure 1. The hierarchy of inference techniques [62].

Hall suggests that if multiple sensors or sources are used in the inference process, they could be fused at one of three levels in the hierarchy: (1) data level fusion, (2) feature-level fusion; or (3) state-vector or decision-level fusion. Data level fusion involves fusion of raw data, for instance multiple images into one fused image, from which inferences are made. Techniques for data level fusion contain model-based methods, statistical estimation methods, and techniques such as least squares methods. In feature-level fusion, representative aspects are extracted from data sources [62].

According to Pavlidis et al., a thorough urban video surveillance system depends primarily on two different technologies: computer vision and threat assessment. The computer vision part contains the optical and system design, the moving object segmentation and tracking and the multi-camera fusion stages. The threat assessment part consists of the feature assembly, the off-line training, and the threat classification stages [63].

Nelson & Fitzgerald introduce a multi-sensor two-stage data fusion system that was created for intelligent alarm analysis. A typical central alarm station for a security environment is an integrated system of people, procedures, and equipment. An alarm communication and display system receives alarm signals from intrusion detection sensors and displays the information to a security operator for performing action. Information displays typically contain only a limited amount of information. Therefore, the operator must manually evaluate each alarm occurrence. The idea behind intelligent alarm analysis (IAA) is to pre-process data from the security sensors and present alarm information in a compact and meaningful way, increase confidence in true alarm events and filter out false alarms. [64]

Nelson & Fitzgerald recognize that machine intelligence requires techniques that can transform incomplete, inconsistent, or imprecise data provided by one sensor into more useful information by fusing it with data provided by other sensors [64]. Multi-sensor data fusion can offer solutions to problems that are characterized by intensive and different sensor information [64]. It can be defined as the process of integrating raw and processed data into a form of meaningful inference that can be used intelligently to improve the performance of the system beyond the level that any one of the components of the system separately could attain [64]. The choice of architecture is central in constructing a data fusion system: where to fuse the data in the processing flow of two or more sensors [64]. Newman acknowledges that automating the data fusion process reduces the burden on today's intelligence analysts who have too much data available [65]. Automation can also reduce the time it takes to distribute intelligence products to the users [65].

Petrushin et al. introduce the Multiple Sensor Indoor Surveillance (MSIS) project. They utilize a Bayesian framework that enables a robust reasoning based on data collected from a network of sensors. In most pragmatic situations, sensors produce streams of multiple, but noisy data. The probabilistic framework gives the ability to reason from this data by containing the local semantics of the sensors and any domain knowledge provided by people involved in these tasks. This framework is believed to be applicable in creating robust and scalable systems that can reason and make inferences from different kinds of sensors that are present in the modern today. [5]

2.7 Architecture and communications

Valera and Velastin denote that spatially distributed multi-sensor environments present interesting possibilities and challenges for surveillance. Recently, there has been some investigation of data fusion techniques to tolerate with information sharing relevant knowledge resulting from different types of sensors. The communication aspects within separate parts of the system play a crucial role, with particular challenges either resulting from bandwidth restrictions or the asymmetric nature of the communication. The distinction between surveillance for indoor and outdoor applications exists because there are differences in the design at the architectural and algorithmic implementation levels. The topology of the indoor environments is different from the outdoor environments. [9]

Valera and Velastin announce that a 3rd generation surveillance system for public transport applications would offer a high-level of automation in the management of information and of alarms and emergencies. The design of a surveillance system with no server deletes the need for centralization, making all the independent subsystems entirely self-contained. Then all the nodes are set up to communicate with each other without having a mutually shared communication point. This approach avoids the disadvantages of the centralized server, and moves all the processes directly to the camera making the system a collection of smart cameras connected over the network. [9]

Attwood & Watson acknowledge that defining a single general-purpose optimal architecture for intelligent surveillance is impossible [31]. There are too many variables and constraints, which vary according to the particular installation and user requirements [31]. Valera and Velastin suggest that a distributed multi-agent approach may provide numerous benefits [9]. First, intelligent co-operation between agents may enable the use of less expensive sensors and therefore a large number of sensors may be distributed over a larger area [9]. Second, robustness is enhanced, because even if some agents fail, others remain to perform the mission [9]. Third, performance is more endurable, there is a distribution of tasks at miscellaneous locations between groups of agents [9].

Valera & Velastin describe that a surveillance system with no server to prevent centralization, making all the independent subsystems entirely self-contained, and then setting up all these nodes to communicate with each other without having a mutually shared communication point. The approach prevents the disadvantages of the centralized server, and moves all the processes directly to the camera making the system a group of smart cameras connected over the network. There are four important objectives that design methods for real-time systems should achieve: to be able to structure the system in concurrent tasks, the capability of developing reusable software by information hiding, to be able to define the behavioural characteristics of the system and be able to analyse the performance of the design by defining its performance and the fulfilment of requirements. [30]

May et al. denote that the development and distribution of digital surveillance systems in public and private areas reveal new challenges in the manner video information is encoded, distributed and utilized [49]. The integration of heterogeneous digital networks in the same surveillance architecture needs a video encoding and distribution technology capable of adjusting to the currently available bandwidth, which is applicable to change in time for the same communication channel, and be robust to transmission errors [49]. Valencia-Jimenez & Fernandez-Caballero indicate that there are benefits in a distributed architecture that may assist in complex surveillance systems [27]. Distributed systems allow the system nodes to possess a certain degree of autonomy to arbitrate their own decisions locally and to act independently of central nodes [27]. This helps in removing possible bottlenecks and improving the system's efficiency [27]. This type of architecture improves the performance of the system through the coordination of the distributed system's components [27].

Ming et al. recognizes that wireless local network (WLAN) typically has two fundamental structures. One is ad-hoc, which does not entail a root, and the status of every wireless network station is peer-to-peer with point-to-point communication. The other is hub-based, which has a wireless station as the centre station, and all stations are operated by the centre station to visit the network. In the hub-based structure, the distribution of stations is less restricted by the environment, and the centre station offers a logic access point to join the base cable network, such as the Internet or LAN. Wireless network based intelligent surveillance systems typically use the wireless hub-based structure, which makes video surveillance information transfer and control appropriate. [66]

Detmold et al. acknowledge that multi-agent systems may have an advantage in relation to scalability and availability, and such an approach would be worth investigating. The middleware offers support for both computational and communication aspects of automated video surveillance networks. Communication on the surveillance network is supported through the instantiation of a service-oriented architecture with publish/subscribe messaging. Scalability, availability, and the ability to integrate distinctly developed surveillance services. The efficiency of the middleware is demonstrated through its application to a vital class of surveillance algorithms. [23]

Fong & Hui announce that video data needs to be appropriately compressed prior to transmission via the Internet to abate bandwidth requirements [7]. Since the video can be seen as a sequence of still images, popular video and image compression techniques are considered [7]. Kreimer indicates that real-time systems (RTS) are utilized for the monitoring and control of physical processes [67]. These systems are imbedded in a significant amount of modern technology structures, for instance production control systems, robotic systems, telecommunication systems, radar systems, self-guided missiles, aircraft, and space stations [67].

Yang et al. acknowledge that with the development of computer technology, real-time video compression and computer networks, digital video surveillance systems have recently been evolving rapidly [68]. The drawbacks in many current video surveillance systems include such items as lower QoS in video transmission and the weaker authentication and extensibility [68]. To design a scalable distributed architecture, Bramberger et al. divide an IVS (intelligent video surveillance) into distributed logical groups of typically located smart cameras, or surveillance clusters [8]. The IVS dynamically and autonomously maps surveillance tasks onto individual cameras depending on the system's current state and the cameras' available resources [8].

The CCTV system presented by Desurmont et al. is based on a digital network architecture. This type of system can be distributed, for instance, in a building. It can also be connected to an existing data network. The system is fundamentally composed of computers connected together through a traditional LAN. The miscellaneous cameras are plugged into the local network hub for IP cameras. A human-computer interface and a storage space are also connected to the system.

The main benefit of this architecture is its endurance. The logical architecture has been designed in a modular manner to enable a fair resource allocation over the cluster. [38]

Micheloni et al. denote that the cooperation among the miscellaneous entities of the network is guaranteed by a communication system, which enables the transmission of useful data. The adopted system is comprised of two components: 1) a Wi-Fi communication among different LANs and 2) a software protocol to specify both the data to be transmitted and the destinations of the messages. [40]

2.8 Testing surveillance systems

Marseguerra et al. announce that the problem of defining the optimal time interval between subsequent surveillance tests is generically handled by constructing a model of the system availability and reliability behaviour. In the majority of cases, the model predictions thereby gathered are affected by uncertainties resulting from both the introduction of simplifying presumptions in the system model itself, and from the lack of complete knowledge regarding the values of the model parameters. The influence of the conditions of the environment in which the components actually function affects the second aspect. When the consequences of failures are considerable, for instance in the safety systems of hazardous nuclear plants, the analyst cannot be fulfilled with 'average' predictions, but must have assurance that the required performance is achieved. [69]

Avritzer et al. recognize that testing of individual modules is called unit testing. The test suite for this phase was unusual since it was developed over a period of five years for earlier versions of the system. Integration testing comprised of rerunning the unit test cases after the system was completely integrated. For feature testing, which is also called system testing, testers developed test cases predicated on the system's requirements. They chose adequate test cases for every expected result to occur. [70]

3. Analysis of existing distributed multi-sensor intelligent surveillance systems approaches

This chapter describes the existing distributed multi-sensor intelligent surveillance systems' components from the following perspective: (1) a comparison of the existing distributed multi-sensor intelligent surveillance approaches to the mobile and ubiquitous requirements stated in this dissertation. The chapter will additionally describe a comparison of the mobile and ubiquitous requirements to the SLSP surveillance system. A critical analysis of the components is presented.

3.1 The purpose of the analysis

Scholars and practitioners have studied and constructed numerous approaches to address distributed multi-sensor intelligent surveillance systems, but the research and pragmatism do not consist of an individual system. Therefore, the essential components were separated into video surveillance, audio surveillance, data fusion, architecture and communication, and testing. In the context of distributed multi-sensor intelligent surveillance systems, the relevant definition had to be performed on each essential component. Video surveillance contained the distribution of digital data according to IP-based protocol transmission. The nature of the content transmitted is irrelevant, since video content analysis is left out from the study. Audio surveillance contained only sound localization, including the distribution of digital data abiding to IP-based protocol transmission. Sound identification was left out. Data fusion contained the reception and transmission of digital data transmitted over IP-based protocol communication. Another requirement for data fusion was the responsiveness of data fusion and information indication. The essentials of architecture and communication had to agree with communication requirements of the previous components, i.e., IP-based protocol communication, reception and transmission of digital data, and responsiveness. An analysis on testing distributed multi-sensor intelligent surveillance systems was performed based on the most relevant distributed systems' testing. The intention of the testing analysis was to establish a basis for validating the realized distributed multi-sensor intelligent surveillance system.

The analysis utilizes conceptual analysis according to [15], [16] to achieve its goals. First, the intention is to clarify how existing surveillance systems can resolve the collection, correlation and analysis of automatically distributed data resulting from distinct devices, and the offering of essential, accurate information instantaneously to the security personnel in distributed multi-sensor intelligent surveillance systems for public locations. This was conducted considering the mobility and ubiquitous requirements. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction.

The analysis may be usable for both researchers and practitioners who desire to review the approach in detail. The analysis offers an essential and thorough understanding of the surveillance system approach corresponding to the collection, correlation and analysis of automatically distributed data resulting from distinct devices, and the providing of essential, accurate information instantaneously to the security personnel in distributed multi-sensor intelligent surveillance systems for public locations.

3.2 Framework for analyzing IS surveillance systems' approaches

This section describes a framework for analysing the IS surveillance systems' approach.

The definition of the comparison of the existing distributed multi-sensor intelligent surveillance approaches to the mobile and ubiquitous requirements.

Various studies, of the state of the art review, entail miscellaneous areas of focus research. By reviewing the focus areas of research, a comparison on how existing modern research of distributed multi-sensor intelligent surveillance systems answers the questions posed in this dissertation regarding the mobile and ubiquitous requirements. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction.

The following focal areas of research have been categorized into seven specific categories according to the aspect emphasized in the study by the author(s). The categories were formed according to the recurring main themes of research performed in the state of the art publications, which were the utilization (or requisite) of **intelligence (ubiquitousness), video, audio, multiple sensors, mobility, and architecture**. Testing was only highlighted in specific publications relevant to testing. These categories are the following:

1) **intelligence (ubiquitousness)**, defined as consisting of automatic and intelligent awareness, such as tracking, raising alarms automatically, data fusion or audio localization for surveillance purposes;
2) **video**, defined as consisting of the usage of video cameras for surveillance purposes;
3) **audio**, defined as consisting of the usage of audio sensors for surveillance purposes;
4) **multi-sensor technology**, defined as consisting of all the other sensors that are used for surveillance purposes;
5) **mobility**, defined as consisting of information distribution to mobile users, or the capability of adapting information distribution according to network and/or communication fluctuations for surveillance purposes;
6) **architecture**, defined as consisting of distinct indications of complex architecture required for transmission or communication of sensor and/or intelligent information distribution in the surveillance system; and
7) **testing**, defined as consisting of specific distinctions applicable for surveillance systems.

The Testing segment is merely to establish the testing issues utilized in the study and validation of the SLSP system. It is removed from detailed examination in the summary and revision table of the focus points of modern research.

In addition, if authors clearly indicate that there is a lack of research in a certain area, it is also accounted.

3.3 Comparison of existing distributed multi-sensor intelligent surveillance systems' modern research against mobile and ubiquitous requirements

This chapter describes an analysis of existing surveillance system approaches. The analysis uses the framework shown in the previous chapter.

3.3.1 Introduction to surveillance systems relevant to public areas

Valera and Velastin [9] concentrate on intelligent visual surveillance systems and address the real-time monitoring. Valera & Velastin [9] presented the state of deployment of intelligent distributed surveillance systems, including a revision of contemporary image processing techniques, which are employed in different modules that constitute part of surveillance systems. The focus of Valera and Velastin [9] is on video and mobility.

Castanedo et al. [19] denote that third-generation surveillance systems handle a large number of cameras and many monitoring points. Castanedo et al. [19] describe a logical framework of autonomous agents working in sensor network environments. The focus of Castanedo et al. [19] is on video.

Li et al. [20] developed a multimedia application, an Internet-based surveillance service, which allows users to view real-time snapshots on the spot in real-time to wired desktops and to mobile users through efficient transmission. The focus of Li et al. [20] is on video and mobility.

Ho et al. [21] denote that the possible applications in this aspect are real-time road traffic monitoring utilizing the mobile videophones or similar handheld video communication devices. Ho et al. [21] introduce a modified H.263 encoder which supports real-time content-based scalable video coding. The introduced technique is applied to real-time video surveillance systems for road traffic monitoring. The focus of Ho et al. [21] is on video and mobility.

Hampapur et al. [6] introduce that modern video surveillance systems function as large-scale video recorders, either analogue or digital. Hampapur et al. [6]

explore the concepts of multiscale spatiotemporal tracking. The focus of Hampapur et al. [6] is on video.

Reiter and Rohatgi [3] inform that information technology plays a significant role in upholding critical infrastructures. The focus of Reiter and Rohatgi [3] is on architecture.

Pham & Xie [22] recognize that, typically, a system consists of multiple subsystems and every subsystem is frequently inspected to keep them functional. Pham & Xie [22] present a generalized surveillance model for predicting the performance of complicated systems comprising of multiple subsystems. The focus of Pham & Xie [22] is on architecture.

Fong & Hui [7] indicate that the adoption of the Internet as the communications medium for real-time transmission of video signals in such a security-sensitive operation, requires many technological issues be resolved. Fong & Hui [7] denote that with the decreasing cost of computational power and advancement in Internet technologies, implementation of a Web-based security surveillance system becomes a considerable option to the traditional manually operated systems. Their article depicts such a system, which offers a low-cost and efficient solution that could be distributed in a variety of situations. The focus of Fong & Hui [7] is on video, architecture, and mobility.

Regazzoni et al. [18] announce that a surveillance system can be defined as a technological tool that assists humans by offering an extended visual and reasoning capability about situations of interest that occur in the monitored environments. They also state that a surveillance system should be complete and it should enable user oriented data accessibility both for the raising of direct alarms and for offline inspection. The main problems currently considered are related to either real-time distributed or centralized processing, and robustness issues in multi-sensor surveillance networks. Regazzoni et al. [18] perform a state-of-the-art review and recapitulate the generations of surveillance systems. The focus of Regazzoni et al. [18] is on video, intelligence, and architecture.

Collins et al. [2] claim that monitoring a vast area by viewing dozens of monitors displaying raw video output requires a restricted amount of transmission bandwidth. Collins et al. [2] present an overview of video understanding algorithms that perform cooperative multi-sensor surveillance. The focus of Collins et al. [2] is on video and mobility.

Detmold et al. [23] take a thorough survey of the entire field of automated video surveillance. Detmold et al. [23] illustrate a middleware supporting computation and communication in automated video surveillance networks. The focus of Detmold et al. [23] is on video and intelligence.

Pavlidis et al. [24] inform that modern security systems act locally and there is a dependence on intensive human concentration to detect and assess threats. Pavlidis et al. [24] depict a monitoring system. The focus of Pavlidis et al. [24] is on the need for architecture and the need for intelligence.

Ott et al. [25] announce that a generic surveillance and security system is composed of three essential parts: data collection, information analysis, and on-field operation. Any surveillance system requires means to monitor the environment and obtain data in the form of video, still images, audio, etc. The need for providing detailed real-time information to the surveillance agents has been identified and is being addressed by the research community. Ott et al. [25] presents a system that utilizes Virtual Reality technologies to establish a surveillance and security system. The focus of Ott et al. [25] is on video, audio, multi-sensor technology, intelligence, and mobility, but there is a need for architecture.

Petrushin et al. [26] acknowledge that a scalable system built for the class of the security and business applications should be able to integrate the wide variety of sensor data with contextual information and domain knowledge provided by both the humans and the physical environment to maintain a coherent picture of the world over time. Petrushin et al. [26] describe a surveillance system that employs a network of different sensors for localizing and tracking people in an office environment. The focus of Petrushin et al. [26] is on multi-sensor technology, and intelligence.

Valencia-Jimenez & Fernandez-Caballero [27] denote that the combination of distributed artificial intelligence has brought forward the development of new technologies in detection (sensors and captors), robotics (actuators), and data communication. Valencia-Jimenez & Fernandez-Caballero [27] present a paradigm of holonic multi-agent systems. The focus of Valencia-Jimenez & Fernandez-Caballero [27] is on multi-sensor technology, intelligence, and architecture.

Atrey et al. [28] announce that the majority of the cities around the world are being equipped with thousands of sensors, including video cameras and audio sensors, with the primary goal of monitoring and recording interesting events as they occur in the area under surveillance. The combination of the information gathered from them to derive the events from the environment is an important and challenging research problem. Atrey et al. [28] propose a hierarchical probabilistic method for information assimilation to detect events of interest in a surveillance and monitoring environment. The focus of Atrey et al. [28] is on video, audio, multi-sensor technology, intelligence, and the need for architecture.

Cucchiara [29] observes that the results of research activities addresses new solutions in video surveillance and sensor networks. Security and safety calls for new generations of multimedia surveillance systems, in which computers act as supporting platforms and as the essential core of real-time data understanding process. Cucchiara [29] presents a corollary of research activities in multimedia surveillance systems. The focus of Cucchiara [29] is on mobility, video, multi-sensor technology, and the need for intelligence, and the need for architecture.

Valera et Velastin [30] & Velastin et al. [1] present the EU-funded project PRO-active Integrated systems for Security Management by Technological, Institutional, and Communication Assistance (PRISMATICA) was to detect certain types of behaviours, which are distinguished from public transport management requirements. The focus of Valera et Velastin [30] & Velastin et al. [1] is on video, intelligence, and architecture.

Attwood & Watson [31] announce that ADVISOR (Annotated Digital Video for Intelligent Surveillance and Optimized Retrieval) was developed in an EU-funded project on innovative architectures for public transport systems. The focus of Attwood & Watson [31] is on video, intelligence, and architecture.

According to Velastin [32], traditional approaches rely on the installation of wide-area CCTV. Velastin [32] recapitulates development and deployment of surveillance system in public and private environments. The focus of Velastin [32] is on video.

3.3.2 Video surveillance

According to Greiffenhagen [33] et al., visual surveillance and monitoring (VSAM) systems have capabilities for detection and false alarms of objects, There are two basic steps in the design process: the choice of the system architecture and the modules for achieving the task, and the statistical analysis and validation of the system to check if it fulfils user requirements. Greiffenhagen et al. [33] review the past studies on a systematic engineering methodology for vision systems performance characterization and depict its adaptation in practice to develop a real-time people detection and zooming system. The focus of Greiffenhagen [33] is on video, intelligence, and architecture.

Bramberger et al. [8] indicate that computing, communication, and sensor technology are accelerating the development of multiple new applications, especially of video-based surveillance systems. To demonstrate their distributed surveillance system's feasibility, Bramberger et al. [8] developed a prototype implementation comprising of multiple smart cameras. The focus of Bramberger et al. [8] is on video, multi-sensor technology, and architecture.

Bartolini et al. [34] announce that recently the scope and quality of automatic video surveillance (VS) systems has grown. Bartolini et al. [34] present a novel algorithm, which is suitable for video surveillance visual data authentication. The focus of Bartolini et al. [34] is on video and intelligence.

Makris & Ellis [35] announce that video surveillance has become a ubiquitous aspect of the modern urban landscape. Makris & Ellis [35] developed an activity-based semantic scene model for an area that is perceived by a video surveillance system. The focus of Makris & Ellis [35] is on video.

Trivedi et al. [36] indicate that video surveillance contains video-based dynamic monitoring and analysis. Trivedi et al. (2005) [36] have developed a multi-camera video surveillance approach, known as DIVA (Distributed Interactive

Video Array). The installation of multiple sensors proposes new design aspects and challenges. The focus of Trivedi et al. [36] is on video, intelligence, and the need for architecture.

Muller et al. [37] indicate that visual surveillance systems are used for observation and protection of private and public regions. Müller et al. [37] have developed a multi-view video streaming system, which can contain an arbitrary station. The focus of Muller et al. [37] is on video.

Desurmont et al. [38] acknowledge that video surveillance systems must be network-connected, modular, entail multiple cameras, have a user interface that is user-friendly, and that the entire system must be reliable and robust. Desurmont et al. [38] propose an approach for a third-generation video surveillance platform and demonstrated performance evaluations for a case study. The focus of Desurmont et al. [38] is on video and architecture.

Foresti et al. [39] indicate that modern video-based surveillance systems use real-time image analysis techniques for efficient image transmission, colour image analysis, event-based attention focusing, and model-based sequence comprehension. Foresti et al. [39] describe the low-level image and video processing techniques required to implement a modern visual-based surveillance system. The focus of Foresti et al. [39] is on video, intelligence, and architecture.

According to Velastin et al. [1], the surveillance of public places is associated with multiple key factors, such as: 1) the widespread geographical extent of what must be addressed; 2) a vast area of behaviours that merit the attention of human operators; 3) the variety of type of information that must be handled to estimate a situation, e.g. vision and sound; and 4) the requirement of transmitting information within a hierarchical system of control. Velastin et al. [1] present an architecture that considers the distributed nature of the detection processes and need for disparate types of devices and actuators. The focus of Velastin et al. [1] is on intelligence, multi-sensor technology, video, audio, and architecture.

Detmold et al. [23] claim that many of the challenges general to the automated video surveillance domain can be assisted with middleware containing support for both computation and communication. Detmold et al. [23] illustrate a middleware supporting computation and communication in automated video surveillance networks. The focus of Detmold et al. [23] is on video and architecture.

Cucchiara [29] claims that multimedia surveillance systems of large distributed environments can improve visual data with audio streams and information resulting from other sensors. Cucchiara [29] presents a corollary of research activities in multimedia surveillance systems. The focus of Cucchiara [29] is on video, audio, multi-sensor technology, architecture and intelligence.

3.3.2.1 Video analysis

Micheloni et al. [40] announce that the research community has concentrated its interests on developing smart applications to improve event detection capability in video surveillance systems. Micheloni et al. [40] present a network of cameras organized in subnets, each dedicated to the surveillance of a designated region. The focus of Micheloni et al. [40] is on video and intelligence.

Bowden & KaewTraKulPong [41] inform that intelligent visual surveillance is an important application area for computer vision. Bowden & KaewTraKulPong [41] present a solution to the problem of tracking intermittent targets that can overcome long-term occlusions and movement between camera views. The focus of Bowden & KaewTraKulPong [41] is on video and intelligence.

Kreucher et al. [42] claim that the difficulty of tracking an individual manoeuvring target in a cluttered environment is a well-examined area. Kreucher al. [42] demonstrate that the implementation of the JMPD (Joint Multi-target Probability Density) technique offers a convenient manner to track of a collection of targets. The focus of Kreucher et al. [42] is on video and intelligence.

Hu et al. [43] announce that as an active research topic in computer vision, visual surveillance in dynamic scenes attempt to detect, recognize and track certain objects from image sequences, and more typically to comprehend and depict object behaviours. Hu et al. [43] present a corollary of recent development in visual surveillance within a general processing framework for visual surveillance systems. The focus of Hu et al. [43] is on video and intelligence.

Bremond et al. [45] claim that one of the most demanding problems in the domain of computer vision and artificial intelligence is video understanding. Bremond et al. [45] introduce a video understanding platform to automatically distinguish human behaviours by detecting visual invariants. The focus of Bremond et al. [45] is on video and intelligence.

Carincotte et al. [46] claim that advancements in sensor, communications and storage capacities facilitate gathering large amounts of multimedia material. Carincotte et al. [46] investigate techniques allowing the automatic extraction of germane semantic metadata from crude multimedia, to examine the value of the extracted information to apposite users, and they demonstrate this in a framework that preserves the privacy of the individual. The focus of Carincotte et al. [46] is on video multi-sensor technology, and architecture.

3.3.2.2 Video Quality of Service

Maier et al. [47] suggest that in traffic surveillance, for example, typically have high demands in QoS. Maier et al. [47] introduce a novel approach that endeavours to maximize the service quality while minimizing the system’s power consumption. The focus of Maier et al. [47] is on video, and mobility.

Korshunov & Ooi [48] denote that a large-scale distributed video surveillance system usually comprises of many video sources distributed over a large area, transmitting live video streams to a central location for monitoring and processing. Implementing an intelligent, scalable and significantly distributed video surveillance system remains a research problem. Researchers have not paid too much attention to the scalability of video surveillance systems. Korshunov & Ooi [48] present an area of video quality that can be employed to abate video bit-rate without significantly affecting the precision of the surveillance tasks. The focus of Korshunov & Ooi [48] is on video, the need for mobility, the need for intelligence, and the need for architecture.

May et al. [49] indicate that in a large surveillance system with multiple video sources, the digital network that enables remote monitoring, storage, control and analysis is in a collection of interconnected LANs, wired or wireless, with different bandwidths and QoS. May et al. [49] present an example of a distributed video surveillance system, for which video requirements are developed. The focus of May et al. [49] is on video and mobility.

Frescura et al. [50] announce that in wireless standards there is the need for robust multimedia transmission. Frescura et al. [50] present a protection scheme for addressing the transmission of JPEG2000 and Motion JPEG2000 codestreams in the 802.11 WLAN environment. The focus of Frescura et al. [50] is on video, mobility, and architecture.

Bramberger et al. [8] inform that in video-based surveillance, there may be multiple QoS levels. Mobile agents are used to support the development of their distributed surveillance system. To demonstrate their distributed surveillance system's feasibility, Bramberger et al. [8] developed a prototype implementation comprising of multiple smart cameras. The focus of Bramberger et al. [8] is on video, mobility, and architecture.

3.3.3 Audio surveillance

Stanacevic & Cauwenberghs [51] acknowledge that accurate and robust localization and tracking of acoustic sources is of interest to a variety of applications in surveillance, multimedia, and hearing improvement. Stanacevic & Cauwenberghs [51] present a micropower mixed-signal system-on-chip for three-dimensional localization of a broadband acoustic source. The focus of Stanacevic & Cauwenberghs [51] is on video, audio, architecture, and intelligence.

Julian et al. [52] recognize that sound localization using compact sensor nodes distributed in networks has applications in surveillance, security, and law enforcement. Julian et al. [52] present that the determination of sampling frequency with cross-correlation algorithms is essential to the accuracy of bearing detection. Julian et al. [52] evaluate four algorithms for sound localization utilizing signals recorded in a natural environment with an array of commercial off-the-shelf microelectronical system microphones. The focus of Julian et al. [52] is on audio and intelligence.

Smeaton & McHugh [53] announce that audio surveillance is typically performed using one or multiple microphones that are wired up to a central unit. The purpose of an audio sensor network would be to assist the end user in reviewing data and to return the points of interest. It is potentially inexpensive to distribute, thus it is a good complement to CCTV. Smeaton & McHugh [53] examined if audio analysis could be employed to assist their existing visual event detection system and to study if there are any improvements. The focus of Smeaton & McHugh [53] is on video, audio, and intelligence.

Aarabi [54] claims that the sound localization methods, such as the ones that are presented in this study, usually presume that the location and orientation of the microphone array is known. Aarabi [54] presented an acoustic method for microphone array localization and orientation estimation. The focus of Aarabi [54] is on audio and intelligence.

3.3.4 Sensor and data fusion

Wald [55] announces that data fusion is a formal framework in which the means and tools are expressed for the alliance of data originating from different sources. Wald [55] propose a new definition of data fusion that is suitable to the remote sensing domain. The focus of Wald [55] is on intelligence, and multi-sensor technology.

Hall [56] indicates that sensor fusion is "Data Fusion from Multiple Sensors (same or different sensor types)" and data fusion is "Combining information to estimate or predict the state of some aspect of the world". The focus of Hall [56] is on multi-sensor technology, and intelligence.

Steinberg et al. [57] claim that data fusion is used to estimate or predict the state of some aspect of the universe. Steinberg et al. [57] report on proposed revisions and expansions of the JDL (Joint Directors of Laboratories) Data Fusion model to remedy some of the deficiencies. The focus of Steinberg et al. [57] is on multi-sensor technology, and intelligence.

Steinberg [58] informs that fusion involves the use of multiple data, which typically result from multiple sources. Performing an estimation or prediction of the state of some aspect of reality, the data are treated as if they were independent of the states of other entities. Steinberg [58] presents new ideas in estimating and predicting threat relationships and situations, given uncertain evidence and uncertain models of such relationships and situations. The focus of Steinberg [58] is on multi-sensor technology, and intelligence.

Jaeger [59] concentrate on RAM-C (Security Risk Assessment Methodology for Communities). Jaeger [59] informs that RAM-C is a systematic, risk-based process to assist communities in evaluating threats, prioritizing targets, identifying consequences, and evaluating the completeness and effectiveness of physical security and response systems. The focus of Jaeger [59] is on intelligence.

Blasch & Plano [60] acknowledge that for multi-sensor information fusion and target tracking, and decision making is integral for information processing. A smart sensor is either a single sensor or a subsystem of different sensor components coordinating to provide data and intelligent algorithmic output. Standard QoS metrics in communication theory and human factors literature are used. Blasch & Plano [60] evaluate a proactive sensor fusion strategy towards successful anticipation of novel threats. The focus of Blasch & Plano [60] is on multi-sensor technology, intelligence, and mobility.

Blasch & Plano [61] recognize that applications for multi-sensor information fusion (IF) require more intelligent and efficient reasoning strategies. Blasch & Plano [61] provide insight into user information needs for knowledge representation and cognitive reasoning. The focus of Blasch & Plano [61] is on architecture, multi-sensor technology, and intelligence.

Hall [62] informs that numerous multi-sensor systems have been constructed to collect, process, and disseminate image and non-image data (acoustic emissions, etc.). Wideband communication links increasingly enable the usage of data across distributed systems. For many applications, new mobile platforms are being developed for surveillance and monitoring. Hall [62] revise the problem of multi-sensor fusion and states that new techniques are emerging that will enable fusion of image and non-image data at multiple levels. The focus of Hall [62] is on intelligence, multi-sensor technology, video, audio, and architecture, and the need for mobility.

According to Pavlidis et al. [63], a complete urban video surveillance system depends primarily on computer vision and threat assessment. The computer also includes system design. Pavlidis et al. [63] describe a state-of-the-art monitoring system. The focus of According to Pavlidis et al. [63] is on architecture, video, and intelligence.

Nelson & Fitzgerald [64] introduce a multi-sensor two-stage data fusion system created for intelligent alarm analysis. The choice of architecture is central in constructing a data fusion system: where to fuse the data in the processing flow of two or more sensors. Nelson & Fitzgerald [64] present the sensor fusion approach taken to execute intelligent alarm analysis for the Advanced Exterior Sensor (AES). The focus of Nelson & Fitzgerald [64] is on multi-sensor technology, intelligence, and architecture.

Newman [65] claims that automating the data fusion process reduces the burden on today's intelligence analysts who have too much data available. It can also reduce the time it takes to distribute intelligence products to the users. Newman [65] introduce the design and prototype of a standard delineation of confidence, pedigree, and security classification information known as Confidence Encapsulated Atomic Data (CEAD). The focus of Newman [65] is on intelligence, and architecture.

Petrushin et al. [5] introduce the Multiple Sensor Indoor Surveillance (MSIS) project, which collects data from a network of sensors. This framework is believed to be applicable in creating robust and scalable systems that can reason and make inferences from different kinds of sensors that are currently present. The focus of Petrushin et al. [5] is on multi-sensor technology, intelligence, and architecture.

3.3.5 Architecture and communications

Valera and Velastin [9] indicate there has been some investigation of data fusion for spatially distributed multi-sensor environments regarding a collection of smart cameras connected over a network. The communication aspects within separate parts of the system play a crucial role, with particular challenges either resulting from bandwidth constraints or the asymmetric nature of the communication. Valera & Velastin [9] presented the state of deployment of intelligent distributed surveillance systems, including a revision of contemporary image processing techniques, which are employed in different modules that constitute part of surveillance systems. The focus of Valera and Velastin [9] is on multi-sensor technology, intelligence, architecture, video, and mobility.

Attwood & Watson [31] announce that defining a single general-purpose optimal architecture for intelligent surveillance is impossible due to many variables and constraints, which vary according to the particular installation and user requirements. Attwood & Watson [31] depicts the evolution of the ADVISOR (Annotated Digital Video for Intelligent Surveillance and Optimized Retrieval) system from design to prototyping. The focus of Attwood & Watson [31] is on the need for architecture.

Valera & Velastin [30] depict that a surveillance system with no server to avoid centralization, and moves all the processes directly to the camera making the system a group of smart cameras connected over the network. Valera & Velastin [30] describe an approach to design an intelligent distributed surveillance system, known as "real-time network approach" or MASCOT. The focus of Valera & Velastin [30] is on multi-sensor technology, architecture, and video.

May et al. [49] acknowledge that the integration of heterogeneous digital networks in the same surveillance architecture needs a video encoding and distribution technology capable of adapting to the currently available bandwidth. The bandwidth is applicable to change in time for the same communication channel, and be robust to transmission errors. May et al. [49] present an example of a distributed video surveillance system, for which video requirements are developed. The focus of May et al. [49] is on architecture, video, and mobility.

Valencia-Jimenez & Fernandez-Caballero [27] indicate that there are benefits in a distributed architecture that may assist in complex surveillance systems to decide their own decisions locally and to act independently of central nodes. Valencia-Jimenez & Fernandez-Caballero [27] present a paradigm of holonic multi-agent systems. The focus of Valencia-Jimenez & Fernandez-Caballero [27] is on intelligence and architecture.

Ming et al. [66] denote that wireless network based intelligent surveillance systems typically adopt the wireless hub-based structure, which makes video surveillance information transfer and control appropriate. Ming et al. [66] introduce a WLAN-based remote video intelligent surveillance system, called WLAN-based Remote Video Intelligent Surveillance System (WRVISS). The focus of Ming et al. [66] is on architecture, mobility, intelligence, and video.

Detmold et al. [23] indicate that multi-agent systems may have an advantage in relation to scalability and availability for automated video surveillance networks. Detmold et al. [23] illustrate a middleware supporting computation and communication in automated video surveillance networks. The focus of Detmold et al. [23] is on video, architecture, intelligence, and mobility.

Fong & Hui [7] inform that video data need to be suitably compressed prior to transmission via the Internet to reduce bandwidth requirements. Fong & Hui [7] depicts a Web-based security surveillance system, which offers a low-cost and efficient solution that could be distributed in a variety of situations. The focus of Fong & Hui [7] is on mobility and video.

Kreimer [67] indicates that real-time systems (RTS) are used for the monitoring and control of physical processes. Kreimer [67] provide several definitions of performance effectiveness index (PEI) for the real-time system (RTS) under consideration. The focus of Kreimer [67] is on architecture.

Yang et al. [68] announce that with the recent development of computer technology, real-time video compression and computer networks, digital video surveillance systems have been evolving rapidly. Yang et al. [68] present an architecture model of the cluster surveillance system based on TCP/IP models. The focus of Yang et al. [68] is on mobility, and video.

Bramberger et al. [8] divide an IVS (intelligent video surveillance), which can dynamically and autonomously maps surveillance tasks, into distributed logical groups of typically collocated smart cameras, or surveillance clusters. To demonstrate their distributed surveillance system's feasibility, Bramberger et al. [8] developed a prototype implementation comprising of multiple smart cameras. The focus of Bramberger et al. [8] is on intelligence, architecture and video.

The CCTV system presented by Desurmont et al. [38] is based on a digital network architecture. Desurmont et al. [38] propose an approach for a third-generation video surveillance platform and demonstrated performance evaluations for a case study. The focus of Desurmont et al. [38] is on architecture, and video.

Micheloni et al. [40] claims that the cooperation among the miscellaneous entities of the network is guaranteed by a communication system, which enables the transmission of useful data, such as a Wi-Fi communication among different LANs. Micheloni et al. [40] present a network of cameras organized in subnets, each dedicated to the surveillance of a designated region. The focus of Micheloni et al. [40] is on architecture, and mobility.

3.3.6 Testing surveillance systems

Marseguerra et al. [69] claim that the problem of defining the optimal time interval between subsequent surveillance tests is generically handled by constructing a model of the system availability and reliability behaviour. Marseguerra et al. [69] introduce a multi-objective optimization approach based on genetic algorithms to determine the optimal Surveillance Test Intervals (STI). The focus of Marseguerra et al. [69] is on testing.

Avritzer et al. [70] inform that testing of individual modules is called unit testing. Integration testing contained of rerunning the unit test cases after the system was completely integrated. In feature testing, which is also called system testing, testers developed test cases based on the system’s requirements. Avritzer et al. [70] have designed a novel strategy that employs historical data for testing a rule-based software system. The focus of Avritzer et al. [70] is on testing.

3.4 Summary of the analysis

Future research should address the insufficiencies of the present analysis of the literature. According to the comparison of the existing modern research on distributed multi-sensor intelligent surveillance systems in reflection to mobile and ubiquitous requirements, the majority of the evaluated publications contained references to video, followed by intelligence and architecture. This was followed by mobility, multi-sensor technology, and finally audio. Every publication in which the authors clearly indicated that the particular topic, i.e., intelligence, video, audio, multi-sensor technology, mobility or architecture, was addressed by the publication, or has been addressed by other research, was denoted as a reference. The comparison did not discover any other individual study having the complete category of intelligence (ubiquitousness), video, audio, multi-sensor technology, mobility, and architecture.

Based on the results of the analysis of the comparison of the existing modern research distributed multi-sensor intelligent surveillance systems against mobile and ubiquitous requirements, the ensuing tables summarize the focus points of the state of the art. The tables are associated to the appropriate subchapter in which it was introduced being: introduction to surveillance systems relevant to public areas, video surveillance, including the subchapters video analysis and video quality of service, audio surveillance, sensor and data fusion, architectures and communication, and testing surveillance systems. The columns are cross-validated with the categories of research, which were **intelligence (ubiquitousness), video, audio, multiple sensors, mobility, and architecture.** These categories were defined in detail at the beginning of chapter 3 “Analysis of existing distributed multi-sensor intelligent surveillance system approaches”. If the research indicated that the category was emphasized in the study, it is indicated with an ‘X’ sign in the table. If the research specifically indicates that there is a need of a certain issue in modern science, it will be indicated with the ‘-‘ sign. Tables 5–11 denote the cross comparison of the categories of research with each publication of the state of the art according to the subchapter it was presented in. Table 12 is a summary of the categories of research against the subchapter groups.

Table 5. Introduction to surveillance systems pertaining to public areas.

	Intelli-gence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Valera and Velastin [9]		X			X	
Castanedo et al. [19]		X				
Li et al. [20]		X			X	
Ho et al. [21]		X			X	
Hampapur et al. [6]		X				
Reiter and Rohatgi [3]						X
Pham & Xie [22]						X
Fong & Hui [7]		X			X	X
Regazzoni et al. [18]	X	X				X
Collins et al. [2]		X			X	
Detmold et al. [23]	X	X				
Pavlidis et al. [24]	-					-
Ott et al. [25]	X	X	X	X	X	-
Petrushin et al. [26]	X			X		
Valencia-Jimenez & Fernandez-Caballero [27]	X			X		X
Atrey et al. [28]	X	X	X	X		-
Cucchiara [29]	-	X		X	X	-
Valera & Velastin [30]; Velasting et al. [1]	X	X				X
Attwood & Watson [31]	X	X				X
Velastin [32]		X				
Total	8 (‘X’s), 2 (‘-‘s)	15 (‘X’s)	2 (‘X’s)	5 (‘X’s)	7 (‘X’s)	6 (‘X’s), 2 (‘-‘s)

Table 6. Video surveillance.

	Intelligence	**Video**	**Audio**	**Multi-sensor technology**	**Mobility**	**Architecture**
Greiffenhagen [33]	X	X				X
Bramberger et al. [8]		X		X		X
Bartolini et al. [34]	X	X				
Makris & Ellis [35]		X				
Trivedi et al. [36]	X	X				-
Muller et al. [37]		X				
Desurmont et al. [38]		X				X
Foresti et al. [39]	X	X				X
Velastin et al. [1]	X	X	X	X		X
Detmold et al. [23]		X				X
Cucchiara [29]	X	X	X	X		X
Total	6 ('X's)	11 ('X's)	2 ('X's)	3 ('X's)	0	7 ('X's), 1 ('-')

Table 7. Video analysis.

	Intelligence	**Video**	**Audio**	**Multi-sensor technology**	**Mobility**	**Architecture**
Micheloni et al. [40]	X	X				
Bowden & KaewTraKulPong [41]	X	X				
Kreucher et al. [42]	X	X				
Hu et al. [43]	X	X				
Bremond et al. [45]	X	X				
Carincotte et al. [46]		X		X		X
Total	5 ('X's)	6 ('X's)	0	1 ('X')	0	1 ('X')

Table 8. Video QoS.

	Intelligence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Maier et al. [47]		X			X	
Korshunov & Ooi [48]	-	X			-	-
May et al. [49]		X			X	
Frescura et al. [50]		X			X	X
Bramberger et al. [8]		X			X	X
Total	1 ('-')	5 ('X's)	0	0	4 ('X's), 1 ('-')	2 ('X's), 1 ('-')

Table 9. Audio surveillance.

	Intelligence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Stanacevic & Cauwenberghs [51]	X	X	X			X
Smeaton & McHugh [53]	X	X	X			
Aarabi [54]	X		X			
Julian et al. [52]	X		X			
Total	5 ('X's)	2 ('X's)	5 ('X's)	0	0	1 ('X')

Table 10. Sensor data and data fusion.

	Intelligence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Wald [55]	X			X		
Hall [56]	X			X		
Steinberg et al. [57]	X			X		
Steinberg [58]	X			X		
Jaeger [59]	X					
Blasch & Plano [60]	X			X	X	
Blasch & Plano [61]	X			X		X
Hall [62]	X	X	X	X	-	X
Pavlidis et al. [63]	X	X				X
Nelson & Fitzgerald [64]	X			X		X
Newman [65]	X					X
Petrushin et al. [5]	X			X		X
Total	12 ('X's)	2 ('X's)	1 ('X')	9 ('X's)	1 ('X'), 1 ('-')	6 ('X's)

Table 11. Architecture and communications.

	Intelligence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Valera and Velastin [9]	X	X		X	X	X
Attwood & Watson [31]						-
Valera & Velastin [30]		X		X		X
May et al. [49]		X			X	X
Valencia-Jimenez & Fernandez-Caballero [27]	X					X
Ming et al. [66]	X	X			X	X
Detmold et al. [23]	X	X			X	X
Fong & Hui [7]		X			X	
Kreimer [67]						X
Yang et al. [68]		X			X	
Bramberger et al. [8]	X	X				X
Desurmont et al. [38]		X				X
Micheloni et al. [40]					X	X
Total	5 (‘X’s)	9 (‘X’s)		2 (‘X’s)	7 (‘X’s)	10 (‘X’s), 1 (‘-’)

Table 12. Summary of all the tables.

	Intelligence	Video	Audio	Multi-sensor technology	Mobility	Architecture
Introduction	8(X), 2 (-)	15 (X)	2 (X)	5 (X)	7 (X)	6 (X), 2 (-)
Video surveillance	6 (X)	11 (X)	2 (X)	3 (X)	0	7 (X)
Video analysis	5 (X)	6 (X)	0	1 (X)	0	1 (X)
Video SQoS	1 (-)	5 (X)	0	0	4 (X), 1 (-)	2 (X), 1 (-)
Audio	5 (X)	2 (X)	5 (X)	0	0	1 (X)
Sensor	12 (X)	2 (X)	1 (X)	9 (X)	1 (X), 1 (-)	6 (X)
Arch	5 (X)	9 (X)	0	2 (X)	7 (X)	10 (X), 1 (-)
Total	41 (‘X’s), 3 (‘-‘s)	50 (‘X’s)	10 (‘X’s)	20 (‘X’s)	21 (‘X’s), 2 (‘-‘s)	33 (‘X’s), 5 (‘-‘s)

The majority of the evaluated publications contained references to video (50 references). The second topic most referred to was intelligence (41 references), followed by architecture (33 references). This was followed by mobility (21 references), multi-sensor technology (20 references), and finally audio (10 references). Every publication in which the authors clearly indicated that the particular category of research (intelligence, video, audio, multi-sensor technology, mobility or architecture) was addressed by the publication, or has been addressed by other research, was indicated as a reference.

The usage of video in surveillance systems appears to be a heavily covered topic. This area will continue to be important and further research will be conducted. Significant research has also been performed on the usage of intelligence with surveillance systems. This includes both automatic alarms and deriving deductions of surveyed information. There are authors who clearly state that a strict requirement for additional research is the usage of intelligence in surveillance systems. Architecture also establishes a foundation in surveillance system. Naturally, multiple systems use an architecture in the system, but some authors clearly stated that there are architectural requirements stemming from the variety and amount of devices and that the question relevant to architecture still remains unresolved. The usage of multi-sensor technology and mobility seems to have a correlation. This may be due to the explosion of using multi-sensor technology resulting with their data being distributed over wireless networks. In the region of mobility, a few authors clearly indicate that this research question relevant to mobility has not been resolved. The research performed considering the usage of audio sensors in the field of surveillance systems brought up the least amount of references in the surveyed publications. The nature of audio sensors is still a new system of research and there are difficulties in successful audio location and detection. The utilization of multiple sensors, including such relevant sensors as video and audio, in consolidation with intelligent components for raising alarms or providing information on ongoing events accompanied with robust, tenable and functional architecture is the next step in surveillance systems. Once current events can be determined at a very high success rate, with a sufficiently low false alarm rate, the technology can be prosperously adopted into the real-world and practical scenarios. A natural progression from this step would be to proceed into a direction in which intelligent components could deduct, for example based on behavioural analysis, threatening situation proactively, i.e., evolving into anticipative intelligent components. These types of components would be deduct prohibited actions based on their precursor activities. There were three fields of the referenced topics in which the researchers claimed a considerable requirement for additional research. These fields were architecture, intelligence (ubiquitousness), and mobility.

These fields of architecture, intelligence (ubiquitousness), and mobility have been addressed in this dissertation. The original publications of the dissertation are based on these subjects, as Table 13 indicates. Publication 1 brings forward the high-level architecture of the SLSP system. Publication 2 addresses the issues of distributing video stream in a scalable procedure to a mobile end user. Publication 3 also addresses the aspects of video distribution to a mobile user by reducing the amount information needed to be distributed through transmitting only the essential images. Publication 4 and Publication 5 both form the main structure of the implemented intelligence, video, audio, multi-sensor technology and architecture topics. These two publications together establish the complete substructure of the SLSP system. Publication 6 is omitted from the table, because it concentrates only on the issues regarding validation and testing.

Through the research and original publications of this dissertation, all the categories of research, i.e., intelligence (ubiquitousness), video, audio, multi-sensor technology, mobility, and architecture, were all covered in the dissertation and the original publications. This combination was not addressed in any individual publication of the state of the art survey.

Table 13. A comparison of the main categories of reflection with the list of publications.

	Intelligence	**Video**	**Audio**	**Multi-sensor technology**	**Mobility**	**Architecture**
High-Level Architecture for a Single Location Surveillance Point (Publication 1)						X
A Scalable Quality of Service Middleware System with Passive Monitoring (Publication 2)		X			X	
Scalable Video Transmission for a Surveillance System (Publication 3)		X			X	
Sensor Data Collection of the Single Location Surveillance Point System (Publication 4)	X	X	X	X		X
Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder (Publication 5)	X	X	X	X		X
Total	2 ('X's)	4 ('X's)	2 ('X's)	2 ('X's)	2 ('X's)	3 ('X's)

4. Introduction to the original publications

This chapter specifies the scientific research performed in the attached publications. The publications bring up novel innovations for distributed multi-sensor surveillance systems. Six publications are included. Their topics are the following:

1. High-Level Architecture for a Single Location Surveillance Point
2. A Scalable Quality of Service Middleware System with Passive Monitoring Agents over Wireless Video Transmission
3. Scalable Video Transmission for a Surveillance System
4. Sensor Data Collection of the Single Location Surveillance Point System
5. Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder
6. Testing and Validation of a Multi-sensor Distributed Surveillance System.

Each of the publications analyses the research problem from its own perspective and brings forward a resolution to the research problem considered in the appropriate publication. The publications can be pieced into parallel, complementary and partially overlapping subjects in which the whole research problem has been analysed from its appropriate focus region. Figure 2 illustrates the workflow and development of innovation understanding the publications. It specifies the main validation method applied over the duration of the research work. The first publication is a concept publication, and it presents a theory of a distributed multi-sensor indoor surveillance system. This publication presents the high-level architectural requirements for a surveillance point from a distributed multi-sensor indoor surveillance perspective on their specific research topic areas. The following four publications (2–5) propose solutions for the distributed multi-sensor indoor surveillance system. The proof-of-concept prototype was constructed to experiment the selected solutions. Publication 6 presents the validation and testing process and methodology executed in the validation and testing of the distributed multi-sensor indoor surveillance system. The novel innovations of these publications are brought up in the following paragraphs. The main revelations of each publication are emphasized. A short description of the new key terms utilized in the publications is illustrated below.

The session server consists of a single component, which contains the main logic of the session server. To communicate with the servers and the sensors in SLSP system, the session server uses network libraries for communication.

The Logical Decision Making Server (LDMS) receives sensor data from the sensors. The received sensor data is the raw data from the sensors. The LDMS makes logical deductions based on sensor data and according to rules distinguished to process sensor data.

The logical deductions are the derivations automatically conducted by the LDMS based on sensor data and according to established rules designed to handle sensor data.

The Security Manager Server & User interface (SMSU) receives all the logical deductions done by the LDMS. The SMSU may also receive all the raw information resulting from the sensors. The human security administrator may order the session server to send refined information from the LDMS and/or raw information from the sensors directly to the end devices.

The Single Location Surveillance Point (SLSP) is a distributed multi-sensor surveillance software system. It contains a decisive amount of sensors that collect readings from a single location, which is the surveillance point. Each sensor transmits its raw sensor data to a session server, which handles the connections between the components. The session server routes the raw sensor information to the logical decision making server. The logical decision making server automatically deducts the situation at the surveillance point based on the received sensor information. The logical decision making server informs the security manager server of the situation at the surveillance point. The user interface of the security manager server displays essential information about the surveillance point to a human security administrator. The security manager server can transmit information to the nomadic security personnel's smart phones over wireless networks.

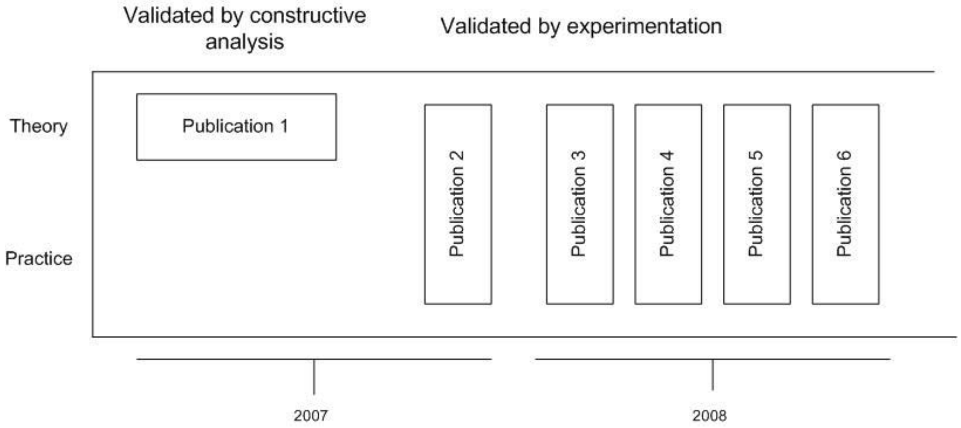


Figure 2. Workflow of publications.

4.1 Publication 1

“High-Level Architecture for a Single Location Surveillance Point” is a concept publication and was published at the Third International Conference on Wireless and Mobile Communications, which was held in Guadeloupe, French Caribbean, 4–9 March 2007. The research problem of the publication is to present an architectural solution of a distributed multi-sensor surveillance system to decrease the amount of unnecessary information that otherwise would be handled by the human security administrator and the security personnel. The distributed multi-sensor surveillance system is the SLSP. The surveillance system consists of a decisive amount of sensors that collect readings from a single location, which is the surveillance point. Each sensor transmits its raw sensor data to a session server, which handles the connections between the components. The session server routes the raw sensor information to the logical decision making server. The logical decision making server automatically deducts the situation at the surveillance point based on the received sensor information. The logical decision making server informs the security manager server of the situation at the surveillance point. The security manager server’s user interface displays essential information about the surveillance point to a human security administrator. The security manager server can transmit information to the nomadic security personnel’s smart phones over wireless networks. The SLSP system informs the human security administrator and the nomadic security personnel about situations that require participation. The SLSP distributes the most important information to the appropriate human users, i.e., the human security administrator and the nomadic security personnel, as quickly as possible. The solution proposed in the publication is an architectural description between the distributed multi-sensor surveillance system's components. The author is the sole writer of the publication.

4.2 Publication 2

"A Scalable Quality of Service Middleware System with Passive Monitoring over Wireless Video Transmission" was published at the 6th International Conference on the Quality of Information and Communications Technology, which was held in Lisbon, Portugal, 12–14 September 2007. The research problem of the publication is to determine a middleware which improves the control of the video transmission over a mobile system. The solution proposed in the publication includes a prototype system that uses a Scalable Quality of Service middleware system, which contains a monitoring user agent client, a monitoring user agent server and a leader agent. A network camera sends video transmission to the smart phone. The video transmission passes through a Scalable Quality of Service server. The monitoring user agent client is in the smart phone. The monitoring user agent server and leader agent is in the Scalable Quality of Service server. Both monitoring user agents monitor the video transmission's bit-rate. The monitoring user agents send their evaluation to the leader agent. Then the leader agent deducts whether to order the network camera to scale the Quality of Service values down or up. The research problem of the paper is to determine a middleware which improves the control of the video transmission over a mobile system. The theories attempt to optimize the video transmission rate to a smart phone over a wireless network. The intention of the theories is to optimize, or improve, the video transmission rate to a smart phone over a wireless network. The operability of the constructed prototype indicates that this attempt was attained. The author is the main writer of the publication and the innovator of all new solutions researched in the publication. Mr. Johannes Oikarinen had the responsibility of implementing the prototype solution. Mr. Markus Sihvonen proofread the publication.

4.3 Publication 3

“Scalable Video Transmission for a Surveillance System” was published at the 5th International Conference on Information Technology: New Generations, which was held in Las Vegas, Nevada, USA, 7–9 April 2008. The research problem of the publication is to present a solution that will ultimately reduce the quantity of video information required to be transmitted the security personnel while keeping all the required information that allows the security personnel to be fully aware of the surveyed indoor area. The solution, called Area of Interest (AoI), is a distributed scalable video transmission subsystem for a surveillance system which concentrates on decreasing the amount of video information transmitted to the end user equipped with a mobile device. The video is processed by the Video Surveillance Intelligent Platform (VSIP) to define the main images. The main image are of the indoor area under fixed video surveillance. The AoI system analyses the output of the VSIP’s images and eXtended Markup Language (XML) image information. The AoI system is able to define and extract the essential information, e.g., a tracked individual, and it transmits only this image to the end-user. First, the AoI transmits the entire image of the indoor area to the mobile device of the end user. Then, the AoI system transmits only the isolated tracked objects’ images to the mobile device. The end user’s device portrays the scaled portrait images of the targeted object on top of the background image. The AoI system attempts to reduce the size of the video images transmitted to a smart phone over a wireless network and to keep the understanding of a tracking situation. The operability of the constructed prototype indicates that this attempt was successful. The author is the main writer of the publication and the innovator of all new solutions researched in the publication. Mr. Lassi Lehikoinen had the responsibility of implementing the prototype solution. Dr. Francois Bremond contributed to the innovation of the AoI system and provided the innovation in the usage of VSIP’s output.

4.4 Publication 4

“Sensor Data Collection of the Single Location Surveillance Point System” was published at the 7th International Conference on Computer and Information Science, which was held in Portland, Oregon, USA, 14–16 May 2008. It will appear as “Collection and Transmission of Sensor Data in the Single Location

Surveillance Point System" in the Journal of Information Technologies and Control. The research problem of the publication is to present a solution to the information collection and transmission of the SLSP, which intends to ease the collection of information from a surveillance point and to reduce the amount of abundant information presented to the surveillance personnel, by automatically collecting sensor data and providing automatically derived information. The SLSP contains a decisive amount of sensors that collect readings from a single location, which is the surveillance point. The SLSP system contains the following realized sensors: a fingerprint sensor, a video camera, an audio sensor, and a network analyzing monitor. The sensors are located in an indoor region. Each sensor automatically collects information from its environment. Each sensor automatically routes its crude sensor data to a session server, which handles the connections among the components. The session server routes the crude sensor information to the logical decision making server. The LDMS automatically derives the situation at the surveillance point based on the received sensor data. The intention is to deduct the situation which is emerging in the surveyed area based on the received raw data from the sensors. By deriving the situation of a surveyed area, the surveillance personnel may use refined information valid to occurring events of the surveyed area. This branch of the SLSP intends to ease the collection of data from a surveillance point and reduce the amount of information presented to the surveillance personnel by automatically gathering sensor data and providing automatically derived information to the surveillance personnel's end-device. The operability of the constructed SLSP system prototype indicates that these attempts were achieved. The author is the main writer of the publication and the innovator of all new solutions researched in the publication. Mr. Johannes Oikarinen and Mr. Mikko Nieminen had the responsibility of implementing the prototype solution. Mr. Mikko Lindholm contributed to the specification of the LDMS component and contributed the state of the art research part of the publication.

4.5 Publication 5

"Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder" was published at the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, which was held in Valencia, Spain, September 29 – October 4 2008. The research problem of the publication is to present a solution to perform and distribute of logical decisions of the SLSP system. The deduction of logical decisions and distributing the alarms to surveillance personnel and, when required, automatically positioning the video recorder to the location of an appropriate alarm of the SLSP intends to 1) ease the deduction of alarms regarding an indoor surveillance point, and to 2) reduce the amount of excessive information presented to the surveillance personnel. The SLSP system contains a decisive amount of sensors that collect data from a single location, which is the surveillance point. The SLSP system includes the following realized sensors: a fingerprint sensor attached to a door with an electronic lock, a video camera, an audio sensor, and a network analysing monitor. The sensors are situated in an indoor region. Each sensor collects information from its area. Once the raw data has been gathered from the sensors and transmitted to Logical Decision Making Server (LDMS) by the session server, the LDMS automatically performs logical deductions based on the data received from the sensors. The logical deductions create: 1) information for end users or 2) control messages to sensors. Based on the alarms, the LDMS can order instructions to the video recorder. The LDMS distributes the logical deductions to the human security administrator of the Security Manager Server (SMSU) and/or the end devices of the nomadic guards. The operability of the constructed SLSP system prototype indicates that this attempt was successful. The author is the main writer of the publication and the innovator of all new solutions researched in the publication. Mr. Mikko Lindholm contributed to the specification of the LDMS component and contributed the state of the art research part of the publication. Mr. Johannes Oikarinen and Mr. Mikko Nieminen had the responsibility of implementing the prototype solution.

4.6 Publication 6

“Testing and Validation of a Multi-sensor Distributed Surveillance System” was published at the International Caribbean Conference on Devices, Circuits and Systems, which was held in Cancun, Mexico, 28–30 April 2008. The research problem of the paper was to present a solution to the testing and validation procedure, accompanied with the description of the testing tools and their usage of the SLSP system. The main attempt in testing and validating the SLSP system consisted of using a sufficient testing and validation process accompanied with sufficient tools to test the SLSP system. The testing and validation process was explained and testing was executed by utilizing two tools. The TCR (Test Case Runner) tool was specifically built for the testing purposes of the SLSP system. A proprietary tool, Nethawk's EAST (Environment for Automated Systems Testing) IMS (IP (Internet Protocol) Multimedia Subsystem) simulator, was utilized in testing. The operability of the constructed prototype accompanied with its successful testing and validation indicates that this attempt was successful. The author is the main writer of the publication and the innovator of all new solutions researched in the publication. Mr. Miao Luo, Mr. Johannes Oikarinen and Mr. Mikko Nieminen had the responsibility of implementing the prototype solution. Mr. Johannes Oikarinen is the chief developer of the TCR tool. Mr. Miao Luo contributed to the authoring of the testing procedure.

5. Validation

To address the modern problems of surveillance systems, a Single Location Surveillance Point (SLSP) system has been defined. The SLSP system addresses two specific requirements: the requirement of mobility and the requirement of ubiquitousness. The mobile requirement contains the reduction of superfluous information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction. The sensors survey and obtain information from a single mutual location. This area is the surveillance point. The SLSP architecture consists of a decisive amount of miscellaneous sensors, a session server, a logical decision making server, a security manager server & UI, and a decisive amount of end devices, e.g., smart phones. Table 14 presents a review of the conclusion, results, and validation according to the publications.

The realized sensors consist of a biometrical sensor, an audio sensor, a video recorder, and a network activity sensor. The sensors monitor their immediate environment, which is called the surveillance point, and transmit knowledge about it to the session server. The session server routes the information to the logical decision making server. The logical decision making server collects all the information from the various sensors and performs logical deductions from the collected information. These logical deductions indicate different situations of the surveillance point. The logical deductions are transmitted to the security manager server, at which a human security administrator resides via the session server. The human administrator can transmit information to the end devices, e.g., smart phones. The end devices are registered to the SLSP system. The nomadic security personnel patrol the premises, or they can be dispatched to the area which is under surveillance. The nomadic security personnel can receive the information on their end devices, e.g., smart phones.

The SLSP system decreases the amount of unnecessary information that otherwise would have to be handled by the human security administrator and the nomadic security personnel. Deductions based on the sensor information are made automatically and they are informed to the security manager server. The human security administrator and the nomadic security personnel will not be overflowed with unnecessary information. Due to the nature of the surveillance information, it is urgent to transmit only the most important knowledge as rapidly as possible. The security administrator server can be used to alert the patrolling nomadic security personnel of emergencies instantaneously. This can be done by the security manager administrator ordering the distribution of critical information automatically and directly from the session server over a wireless network to the nomadic security personnel's end devices, e.g., smart phones. This will make the reception of the raw sensor information and logical deductions quicker at the end device, instead of having the information first being routed to the security manager server and the human security administrator deciding on what information to transmit to the end devices. Another option is for the security manager administrator to select the received information, e.g., crude sensor information and/or logical deductions, which the security manager administrator wants to route to the nomadic security personnel's end devices over a wireless network. The decrease of redundant information through situation deduction attains the mobile and the ubiquitous requirements.

Based on the SQoS subsystem of publication 2, the initial and optimum settings for the video stream for the Nokia 6680 were following: bit-rate 64 kbps, frame size QCIF and frame rate 30 fps. When the receiving end begins the reception of the video stream, it commences measuring the bit-rate. Samples are obtained once per every ten seconds. The leader agent compares the values measured from the video camera to the values transmitted by the receiving end. If a discrepancy is detected, the video stream is downscaled. If the resulting discrepancy is more than half, the video stream's bit-rate is decreased by dividing it by two. If the receiving end maintains its bit-rate consistently for one minute, the video stream's bit-rate is then multiplied by two. [71]

An example of downscaling is the following. Once the video streaming begins and before the downscale occurs, the average bit-rate for the video stream reported by the Nokia 6680 is 60 kbps. Once the downscaling happens, the bit-rate is halved in the sending end, by adjusting the bit-rate to 32 kbps. The bit-rate drops and the end device reports the bit-rate as 13 kbps at the moment of the decrease. After the downscaling the Nokia 6680 reports the average bit-rate as 32 kbps. [71]

An example of upscaling is the following. The initial bit-rate is 32 kbps, the frame rate is 30 fps and the frame size is QCIF. To upscale the video stream, the receiving end has to uphold and indicate a steady bit-rate for the duration of one minute. Before the upscaling is issued, the receiving end indicates the average 32 kbps as the bit-rate for the received video stream. After one minute has elapsed, the sending end issues upscaling by doubling the bit-rate, i.e., by setting it to 64 kbps. The frame size is retained as QCIF and the frame rate is 30 fps. The average bit-rate for the video stream indicated by the receiving end is 32 kbps after the upscaling. [71]

Based on the AoI subsystem of publication 3, the decreases in the amount of transmitted image information is the following. If the image contains one object, which is tracked, and the tracked image is transmitted to the end device, then the average size of the transmitted image is 2.8% in comparison to the whole image. If the image contains two objects, which are tracked, and the tracked image is transmitted to the end device, then the average size of the transmitted image is 7.4% in comparison to the whole image. If the image contains three objects, which are tracked, and the tracked image is transmitted to the end device, then the average size of the transmitted image is 8.5% in comparison to the whole image.

The conclusive result of this dissertation is materialization that the concept of the SLSP prototype system verifies new innovations. First, the dissertation focuses on the theory of architectural improvements for a distributed multi-sensor intelligent surveillance system against mobile and ubiquitous requirements. Finally, in addition to theory, the materialization of concept technology experimentations is brought up in the dissertation. The innovations are tested and verified by the experimental SLSP system. The created surveillance system is reflected against the mobile and ubiquitous requirements of the end users of the surveillance system. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction.

One main discovery was a large amount of complementary information that could be assembled with multiple sensors. Individual sensors can gather specific data from the environment. This data can be used to perform deductions of the surveyed area. The deducted information from individual sensors is valuable and tenable. Incidents and situations can be derived by the SLSP system based on individual sensor data. Another completely significant factor is the capability of deducing events based on the combined data. The usage of a vast amount of sensors results in collecting a significantly higher amount of events. These events provide more accurate and novel events that cannot be collected with individual sensors or the usage of multiple sensors provide complementary and strengthening data to existing event notifications. The electronic lock sensor and the door status sensor collaboration proved to be extremely suitable. If the electronic lock indicated that the lock was activated, i.e., the door was locked, and the door status sensor indicated that the door was open, then an alarm was automatically raised. Both of these sensors individually indicating that the lock is activated and the door is open would not necessarily raise an alarm.

The testing and validating process of the SLSP dictated an important innovation. A distributed multi-sensor intelligent surveillance system has specific requirements for testing and validating the system. The defined testing and validating process includes component testing and integration testing with two notable tools. The testing and validating process has component-level and integration-level testing and validating. The utilized testing tools were chosen specifically according to needs of a distributed multi-sensor intelligent surveillance system.

As a reflection of the technical evolution of intelligent surveillance systems, as shown in Table 2 on page 28, the subsequent diagram (Figure 3) illustrates the main aspects of each generation of surveillance systems, comprising of 1GSS, 2GSS, and 3GSS.

The main focal point of 1GSS is image distribution and reception. 2GSS concentrates on enhanced video surveillance, including detection and tracking functionalities. 3GSS consists of multiple sensors, distribution of information, data fusion and intelligence, and presents design as one of the major challenges for 3GSS. [9]

Table 14. Review of conclusion, results, and validation according to publications.

	Conclusions and results	**Validation**
Publication 1	• Architecture to decrease the amount of excessive information that otherwise would be handled by the human security administrator and the security personnel.	• Constructive analysis method • Successful experimentation in the SLSP platform
Publication 2	• Determination of a scalable quality of service middleware, which improves the control of the video transmission over a mobile system.	• Constructive analysis method • Successful experimentation in the SLSP platform
Publication 3	• Definition of a distributed scalable video transmission subsystem, that reduces the amount of video information transmitted to the mobile device.	• Constructive analysis method • Successful experimentation in the SLSP platform
Publication 4	• Definition of the sensor data collection and sensor data distribution of SLSP.	• Successful experimentation in the SLSP platform
Publication 5	• Definition of conduct and distribution of SLSP's logical decisions.	• Successful experimentation in the SLSP platform
Publication 6	• Distinction of testing and validating procedures and process of the SLSP system.	• Successful experimentation in the SLSP platform

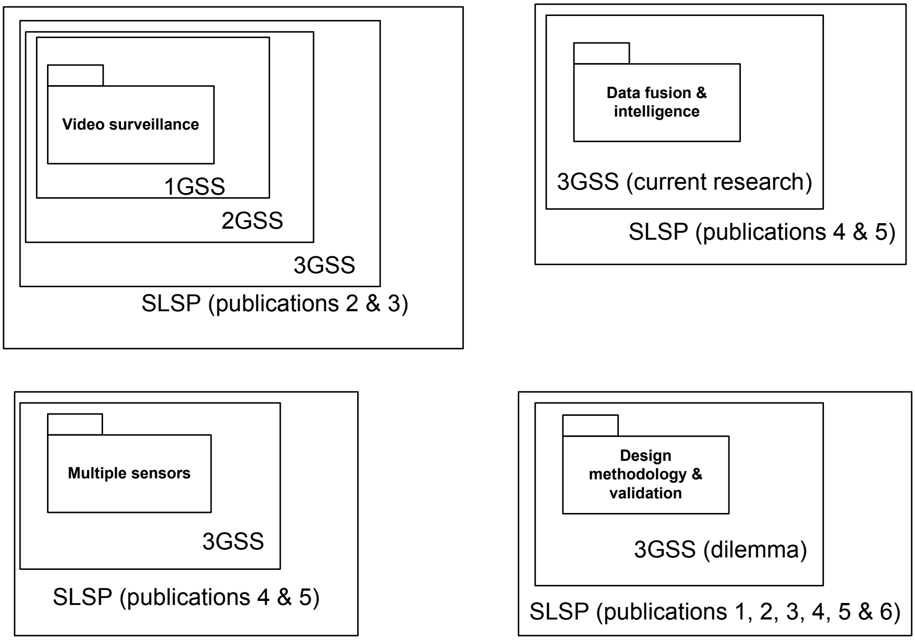

Video surveillance
1GSS
2GSS
3GSS
SLSP (publications 2 & 3)
Data fusion & intelligence
3GSS (current research)
SLSP (publications 4 & 5)
Multiple sensors
3GSS
SLSP (publications 4 & 5)
Design methodology & validation
3GSS (dilemma)
SLSP (publications 1, 2, 3, 4, 5 & 6)

Figure 3. Rendition of 1GSS, 2GSS, 3GSS in comparison to SLSP.

The conceptualization of video surveillance is addressed by each generation of surveillance systems (1GSS, 2GSS, and 3GSS) and, in addition, the SLSP prototype contributes to video surveillance. This is presented especially in publications 2 and 3. 3GSS addresses issues regarding multiple sensors, which is a topic of the SLSP prototype, in particular publications 4 and 5 contribute to this issue. Data fusion and intelligence are properties currently being examined in 3GSS. It is addressed by the SLSP prototype, which is especially presented in publications 4 and 5. 3GSS entails at least two substantial dilemmas, namely, the design and validation difficulties. The SLSP publications in reference to design issues are 1, 2, 3, 4, and 5. Publication 6 presents a solution to the validation difficulties of a complex 3GSS design.

5.1 The answer to the requirement of ubiquitousness

The novel innovation of the SLSP system contains the decrease of excessive information, which is delivered to the security administration. This is the answer to the requirement of ubiquitousness. This is achieved with the SLSP system automatically raising alarms based on simple logical rules.

The intention of the SLSP system is not to exclude the human from the information system loop, but to use one of his most unique and strongest qualities, his dynamicity. The SLSP attempts to focus on the static and common processes that occur normally and which can be reliably and consistently determined. These include distinguishing the direction of an audio event which exceeds a certain threshold level; indications of door, lock and fingerprint sensor malfunctions or abnormalities; and automatically raising alarms and/or directing video recording to previously mentioned events. These events are associated with time requirements. The time requirements contain aspects regarding open and closed hours of the surveyed area, e.g., different thresholds for raising alarms when a shopping mall is open during office hours and during the time when the area is closed. Time requirements include the recurrence of events, e.g., if an individual attempts to open a door multiple times unsuccessfully, then this might be an indication of a perpetrator attempting to enter the area illegitimately. Drawing the attention of the security administration to the abnormal or suspicious activities results in these events being noticed easily and swiftly. Highlighting significant information effaces the requirement of showing excessive information.

Based on the previous innovation, two other innovations followed: Sensor data collection and sensor data distribution, and the performance and distribution of logical decisions. The sensor data is collected by the session server, a central server that addresses the communications of the SLSP system, and transmitted to the LDMS. The LDMS performs logical decisions based on rules and delivers the deductions to the session server. The session server in turn distributes the logical decisions to the suitable receiver, i.e., the SMSU, and if ordered, the mobile devices of the nomadic security personnel.

5.2 The answer to the requirement of mobility

Another important innovation was the scalable video transmission subsystems of the SLSP: SQoS and AoI. This is the answer for the mobile requirement. The SQoS system scales the quality of the transmitted video at the source based on the QoS information received from the mobile devices. The AoI system distributes only the images of the tracked objects instead of an image of the entire area under video surveillance. Both systems reduce the amount of information transmitted across a wireless network.

The SQoS subsystem delivers adaptive video transmission from the video recorder to the mobile devices. This quality of the rendered image is controlled to transmit the video stream at an appropriate pace to the end user's mobile device. If the network is congested, it is crucial to transmit the visible video stream to the end users as rapidly as possible, and thus the quality of video stream can be reduced.

The AoI subsystem includes a person-tracking functionality. The transmission of the entire image of a surveyed area is considerably larger than an image of the changed image. Therefore the AoI subsystem deducts the changed image and transmits them to the mobile device of the end user. Both these subsystems improve the transmission rates of video stream or images. Hence, the mobile devices of the end users receive the most important video information as quickly as possible.

5.3 Outline of the SLSP architecture

This chapter illustrates a brief outline of the SLSP architecture in the form of a package diagram. A more detailed and thorough itemization and resolution is presented in the publications declared in Chapter 4 "List of original publications". The main components of the SLSP architecture are the following: the biometrical sensor, the video sensor, the audio sensor, the network activity monitoring sensor, the session server, the LDMS, the SMSU and the end device.

The biometrical sensor transmits the information of the fingerprint sensor and the statuses (open/closed) of the door and electronic lock to the session server. The video sensor transmits all the video information to the session server. The audio sensor transmits all the bearings of the sound location and volumes of the audio events. The network activity monitoring sensor transmits all the information relevant to the devices attempting to access the SLSP environment. The session contains a message buffer handling the messages received from or transmitted to the sensors. The session server also contains a message buffer for handling the messages received from or transmitted to the LDMS, SMSU, and end device. The session server contains a logical component for processing the incoming and outgoing messages. The LDMS contains components for receiving, interpreting, handling, and transmitting messages to and from the session server. The LDMS contains logical components and tables for deducting and issuing alarms. The SMSU receives all the sensor information transmitted from the sensors. The SMSU also receives the instantaneously and logically deducted alarms that originate from the LDMS. These sections of information are all ultimately displayed to the human operator of the SLSP system at the SMSU. The end device is also capable of receiving the instantaneously and logically deducted alarms that originate from the LDMS. These pieces of information may also be displayed to the nomadic guard with the end device. The package diagram of the SLSP architecture is presented in Figure 4.

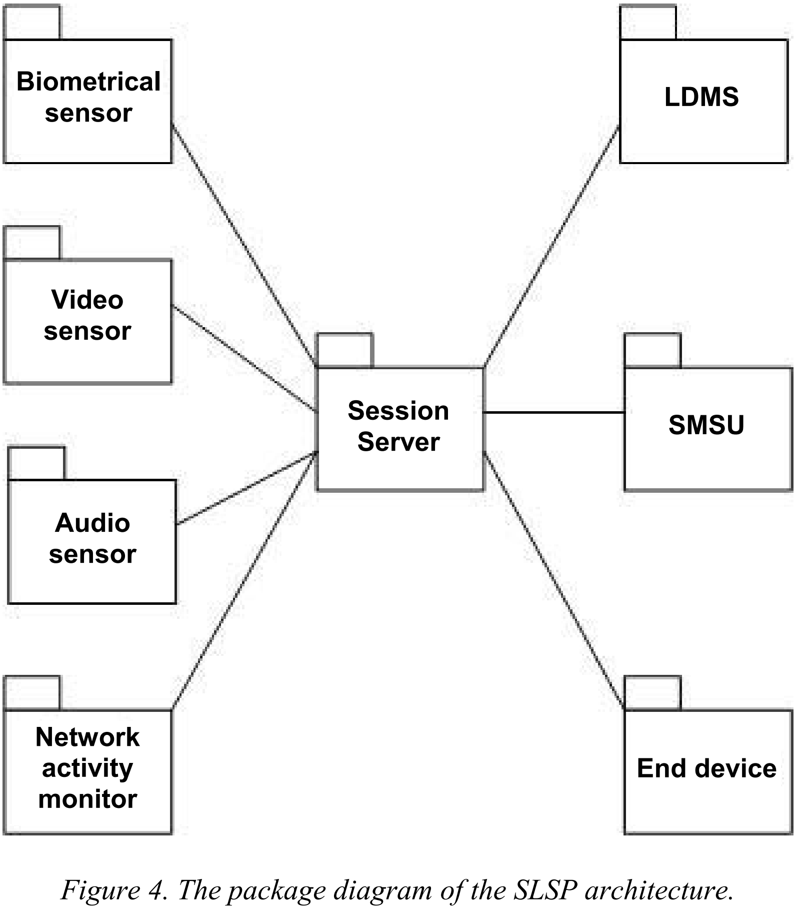


Figure 4. The package diagram of the SLSP architecture.

6. Conclusions

This section concludes the dissertation by presenting the summary of the results, the limitations of results, and outlining the future research. The summary of the results forms a conclusion to the research question and reviews how the research question was answered in the papers and in the dissertation summary. The limitations of the results pronounce the validity and applicability of the results. Future research section emphasizes both the incomplete and the most robust areas of the dissertation and highlights a future research plan to complement and continue the work.

6.1 Summary of the results

This dissertation defined architectural improvements for a distributed multi-sensor intelligent surveillance system. The created surveillance system was reflected against the mobile and ubiquitous requirements of the end users of the surveillance system. The mobile requirement contains the reduction of excessive information distributed to the end user. The ubiquitous requirement consists of sensor data fusion and situation deduction. The automatic collection of sensor data, performing logical decisions based on the sensor information, and the distribution of the logical decisions to security administration personnel reduces the amount of excessive information presented to the security administration personnel. The logical decisions attempt to form simple and reliable deductions of the surveyed environment at real time. Even though the SLSP system is not capable of distinguishing all abnormal or suspicious activities, it can be used as an improvement to existing surveillance systems. The new innovated solutions were evaluated in the SLSP prototype system. The SLSP system contains distributed middleware components that present a distributed multi-sensor intelligent surveillance system. The SLSP system enables the usage of miscellaneous subsystems. Some of the innovations experimented with SLSP system are applicable to commercial applications. **The research question: Define architectural improvements to 3GSS that**

1) **allow the utilization of mobility for security personnel (comprising the *mobile requirement*),**
2) **allow ubiquitous utilization for wireless security personnel (comprising the *ubiquitous requirement*).**

The answer to the research question is that the architectural improvements are 1) the distribution of logical deductions and video information rapidly to the mobile security personnel, and 2) conducting automatically logical deductions based on sensor data and automatically informing security personnel when required.

The research question was itemized into two sub-questions, of which the first research sub-question studied in this dissertation was the following.

To what extent are distributed multi-sensor intelligent surveillance systems collecting data from a public location and transmitting intelligent information to surveillance administrators examined and resolved by modern science? Categorized into five segments, video surveillance, audio surveillance, data fusion, architecture and communication, and testing surveillance systems.

The answer to the first research sub-question is presented in Chapter 2 of this dissertation "2. A review of existing distributed multi-sensor intelligent surveillance systems". Current research strongly focuses on video surveillance to distinguish events happening in the surveyed area. There are studies relevant to specific sensors, but there is a considerable omission of combining numerous sensor information and establishing deductions based on the combined information. Another omission in research is forming automatic logical decisions based on a large amount of sensor data.

The second research sub-question studied in this dissertation was the following.

How to collect, correlate and analyse automatically distributed data resulting from specific devices in public locations, and provide the security personnel important, accurate information instantaneously by the means of a distributed multi-sensor intelligent surveillance system according to the mobility and ubiquitousness requirements?

The answer to the second research sub-question is presented in the six scientific publications and in the dissertation summary. Automatic collection, correlation, and analysis of dispersed data resulting from specific devices in public locations, and provision of important, accurate information instantaneously to security personnel by the means of a distributed multi-sensor intelligent surveillance system is possible by presenting a prototype solution for the defined research objective. The prototype SLSP platform implemented the collection, correlation, and analysis of distributed data resulting from specific devices in a public location, and the provision of important, precise information instantaneously to the security personnel. This capability belongs to the answer of the **ubiquitous requirement**. The sensors, i.e., the video recorder, audio sensor, biometrical sensor, and network activity monitor, of the SLSP platform collect data from their environment. The data is transmitted to the session server, which addresses the data and information distribution of the SLSP system. The LDMS forms logical deductions based on the data received from the sensors. The logical deductions are transmitted to the SMSU, where the security surveillance administrator is, and, if stated necessary by the operator of the SMSU, the logical deductions may also be distributed to the nomadic security personnel's mobile devices. The SLSP prototype includes subsystems, i.e., the AoI and SQoS, to transmit the video information as quickly as possible to the security personnel from the video source. These subsystems belong to the answer of the **mobile requirement**.

The high-level architecture of the SLSP system is presented in Publication 1, which illustrates the basic components and their interrelations.

Publications 2 and 3 concentrate on defining the SQoS and AoI subsystems of the SLSP platform respectively. Both subsystems reduce the quantity of information transmission of the video stream to the mobile device. The SQoS addresses the problem regarding the connection from the video recorder to the mobile device by reducing the amount of information transmitted through decreasing the quality of the video stream but keeping its understandability. The AoI decreases the image transmission required from a video tracking subsystem to the mobile device. The tracked images from the video tracking subsystem are selected and transmitted to the mobile device. The results of both papers were derived from the implemented prototype.

Publications 4 and 5 specify the definition of the sensor data collection and sensor data distribution of SLSP, and the definition of performing and distributing SLSP's logical decisions, respectively. The results of both publications were derived from the implemented prototype.

Publication 6 defines the testing and validating procedures and process of the SLSP system. The results were derived through testing and validating the implemented SLSP system.

In reflection to the original research question, the architectural improvements are illustrated in Figure 4, the package diagram of the SLSP architecture, in Section 5.3 Outline of the SLSP architecture. The responses to this research problem are highlighted in Section 5.1 The answer to the requirement of ubiquitousness and 5.2 The answer to the requirement of mobility. The answer to the ubiquitousness requirement is to reduce the amount of information distributed to the security personnel through raising alarms automatically. The answer to the mobile requirement is to reduce the transmitted video information over a wireless channel to the security personnel. The resolution to the mobile requirement and to the ubiquitousness requirement responds to allowing mobile security personnel and to allowing ubiquitous wireless personnel. These two aforementioned items, as a consolidation, resolve allowing remote surveillance security personnel. Resolving the mobility and ubiquitousness requirements ultimately results in proposing architectural improvements to 3GSS.

6.2 Limitations of the results

The research results have three clear and undeniable limitations. The first limitation is in regard to the specific sensors used in the SLSP system. The realized sensors that were used in the SLPS system were the biometrical sensor, the video recorder, the audio sensor, and the network activity sensor. Any additional sensor would have a considerable impact on the session server and the LDMS. Also, the SMSU and the end devices would require changes to be able to address the new sensor information. The second limitation is the fact that the SLSP is a single location surveillance point, therefore it is not applicable for large-scale surveillance by definition. The SLSP system is functional for only the surveillance of a single location. The third limitation is that the SLSP system functions only as a closed network. A WLAN access point is required. The sensors, servers and end devices all have the SLSP-specific interfaces and APIs for communication inside the SLPS platform. An indoor location was selected for the omission of weather and changing light conditions. A public location was used to collect an assembly of miscellaneous individuals in an accessible area.

This condition brings forward a broad variety in alarm situations. The SLSP system serves as a point of beginning for the development of distributed multi-sensor intelligent surveillance systems. The intention was not to construct a thorough surveillance system. In addition, the aim was to highlight a design of a distributed multi-sensor intelligent surveillance system. The design was validated and verified by testing the implemented SLSP system. With respect to the analytical section of the dissertation, an additional limitation comes from the interpretive research strategy employed. The outcome is based on the researcher's evaluation.

The ensuing aspects were omitted from the research. This dissertation does not address any issues regarding the GUI or UI of the end user. Video codecs and image compression techniques are also omitted from this study.

6.3 Future research

This dissertation consisted two research steps. Due to a lack of a review of the literature of the existing distributed multi-sensor intelligent surveillance systems approaches, the initial step of this dissertation was to revise and evaluate the state of the existing research of distributed multi-sensor intelligent surveillance systems. This analysis is useful for practitioners, who do not have the time to go through a vast quantity of literature. For scholars, such a literature analysis reveals what regions of distributed multi-sensor intelligent surveillance systems have been considered, and to which requirements future research is most important.

The analysis revealed that only a few of the existing studies on distributed multi-sensor intelligent surveillance systems are theoretically grounded. Knowledge of the underlying theoretical background would assist practitioners and scholars in understanding why a particular distributed multi-sensor intelligent surveillance system approach is expected to have the desired impact on distributed multi-sensor intelligent surveillance systems. The result of the analysis revealed that the majority of the existing distributed multi-sensor intelligent surveillance system approaches do not offer empirical evidence on their practical efficiency.

The future research will begin with extending the thoroughness of the SLSP system. As stated in the limitations, the SLSP consists of a realization of the following sensors: the biometrical sensor, the video recorder, the audio sensor, and the network activity monitor. A natural evolution is the addition of multiple sensors. In reference to the previous statement, the enhancement of the SLSP system to a completely thorough and exhaustive surveillance system is precise. The following characteristics form an important improvement to the existing SLSP system: 1) to improve the intelligence in the security systems, 2) to develop evolutionary security system equipment, 3) to model the security system's applicability to authentic and exhaustive use case scenarios, and 4) to insert an additional configurability layer, targeted at the system's flexibility, according to the needs of the end user systems. The intelligence of the SLSP can naturally be extended to deduct more accurate, more exhaustive, and more complicated scenarios and logical deductions. Sensors will also follow a natural growth in the information and range they provide. To achieve the utmost supremacy of a surveillance system, it must be applied on authentic and exhaustive use case scenarios. In reference to the previously mentioned statement, the commercialization of the future work needs to be considered. This system must have a layer or layers of software that are customizable according to purchaser of the surveillance system. Each client is declared to have his own specific needs for the surveillance system. The future research topics in this chapter will belong to the knowledge on the subject and also establish fresh topics for experimentation of new ideas.

References

[1] Velastin, S.A., Boghossian, B.A., Lo, B.P.L., Sun, J. and Vicencio-Silva, M.A. PRISMATICA: Toward Ambient Intelligence in Public Transport Environments. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 35, No. 1, January 2005, pp. 164–182.

[2] Collins, R.T., Lipton, A.J., Fujiyoshi, H. and Kanade, T. Algorithms for Cooperative Multisensor Surveillance. Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1456–1477.

[3] Reiter, M. and Rohatgi, P. Homeland Security Guest Editor's Introduction. IEEE Internet Computing, November/December 2004, pp. 16–17.

[4] Walls, J.G., Widmeyer, G.R. and El Sawy, O.A. Building an Information Systems Design Theory for Vigilant EIS. Information Systems Research 3(1)1992, pp. 36–59.

[5] Petrushin, V.A., Gang W., Ghani, R. and Gershman, A.V. Multiple Sensor Indoor Surveillance: Problems and Solutions. IEEE Workshop on Machine Learning for Signal Processing, 28–30 Sept. 2005. Pp. 349–354.

[6] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S., Senior, A., Shu, C.-F. and Tian, Y.L. Smart Video Surveillance. IEEE Signal Processing Magazine, March 2005, pp. 38–51.

[7] Fong, A.C.M. and Hui, S.C. Web-based intelligent surveillance system for detection of criminal activities. Computing and Control Engineering Journal, December 2001, pp. 263–270.

[8] Bramberger, M., Doblander, A., Maier, A., Rinner, B. and Schwabach, H. Distributed Embedded Smart Cameras for Surveillance Applications, Computer. Published by the IEEE Computer Society, February 2006. Pp. 68–75.

[9] Valera, M. and Velastin, S.A. Intelligent distributed surveillance systems: a review. IEE Proc. – Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 192–204.

[10] Hugues, J., Pautet, L. and Kordon, F. Contributions to middleware architectures to prototype distribution infrastructures. Proceedings of 14th IEEE International Workshop on Rapid Systems Prototyping, June 9–11, 2003. Pp. 124–131.

[11] Kamouskos, S. Supporting nomadic users within virtual private networks. Service Portability and Virtual Customer Environments. IEEE, 1 Dec. 2000. Pp. 128–133.

[12] Bass, L., Clements, P. and Kazman, R. 1998. Software architecture in practice. Reading, Massachusetts: Addison-Wesley. 452 p. ISBN 0-201-19930-0.

[13] IEEE-1471. 2000. IEEE recommended practice for architectural descriptions of software-intensive systems. New York: IEEE. 23 p.

[14] Shenker, S. and Wroclawski, J. Network Element Service Specification Template. IETF Network Working Group, Request for comments: 2216, September 1997. URL: http://www.apps.ietf.org/rfc/rfc2216.html.

[15] Järvinen, P. 1997. The New Classification of Research Approaches. In: Zemanek, H. (ed.) The IFIPPink Summary – 36 years of IFIP. IFIP, Laxenburg, Austria. Pp. 124–131.

[16] Järvinen, P. 2000. Research Questions Guiding Selection of an Appropriate Research Method. Proceedings of the 8th European Conference on Information Systems (ECIS 2000).

[17] Webster, J. and Watson, R.T. 2002. Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly 26(2): xiii–xxiii.

[18] Regazzoni, C.S., Ramesh, V. and Foresti, G.L. Scanning the Issue/Technology Special Issue on Video Communications, Processing, and Understanding for

Third Generation Surveillance Systems. Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1355–1367.

[19] Castanedo, F., Patricio, M.A., Garcia, J. and Molina, J.M. Extending Surveillance Systems Capabilities Using BDI Cooperative Sensor Agents. VSSN'06, October 27, 2006, Santa Barbara, U.S.A. Pp. 131–138.

[20] Li, S.-T., Hsieh, H.-C., Shue, L.-Y. and Chen, W.-S. PDA Watch for Mobile Surveillance Services. Proceedings of the IEEE Workshop on Knowledge Media Networking (KMN'02), 2002, IEEE.

[21] Ho, W. K.-H., Cheuk, W.-K. and Lun, D. P.-K. Content-Based Scalable H.263 Video Coding for Road Traffic Monitoring. IEEE Transactions on Multimedia, Vol. 7, No. 4, August 2005.

[22] Pham, H. and Xie, M. A Generalized Surveillance Model with Applications to Systems Safety. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, Vol. 32, No. 4, November 2002.

[23] Detmold, H., Dick, A., Falkner, K., Munro, D.S., van den Hengel, A. and Morrison, R. Middleware for Video Surveillance Networks. MidSens'06, November 27 – December 1, 2006, Melbourne, Australia. Pp. 31–36.

[24] Pavlidis, I., Morellas, V., Tsiamyrtzis, P. and Harp, S. Urban Surveillance Systems: From the Laboratory to the Commercial World. Proceedings of the IEEE, Vol. 89, No. 10, October 2001.

[25] Ott, R., Gutierrez, M., Thalmann, D. and Vexo, F. Advanced Virtual Reality Technologies for Surveillance and Security Applications. VCRIA 2006, Hong Kong, 14–17, June, 2006. Pp. 163–170.

[26] Petrushin, V.A., Shakil, O., Roqueiro, D., Wei, G. and Gershman, A.V. Multiple-Sensor Indoor Surveillance System. Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06), IEEE, 2006.

[27] Valencia-Jimenez, J.J. and Fernandez-Caballero, A. Holonic Multi-agent Systems to Integrate Multi-sensor Platforms in Complex Surveillance.

Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06), IEEE, 2006.

[28] Atrey, P.K., Kankanhalli, M.S., and Jain, R. Timeline-based Information Assimilation in Multimedia Surveillance and Monitoring Systems, VSSN'05, November 11, 2005, Singapore. Pp. 103–112.

[29] Cucchiara, R. Multimedia Surveillance Systems. VSSN'05, November 11, 2005, Singapore. Pp. 3–10.

[30] Valera, M. and Velastin, S.A. Real-time Architecture for a Large Distributed Surveillance System. The Institution of Electrical Engineers, IEE, Stevenage, 2004.

[31] Attwood, C.I. and Watson, D.A. Advisor – Socket and See: Lessons Learnt in Building a Real-time Distributed Surveillance System. Thales Research & Technology (UK) Ltd., UK, 2004.

[32] Velastin, S.B. Special Section on Intelligent Distributed Surveillance Systems. IEE Proc. – Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 191.

[33] Greiffenhagen, M., Comaniciu, D., Niemann, H. and Ramesh, V. Design, Analysis, and Engineering of Video Monitoring Systems: An Approach and a Case Study. Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1498–1517.

[34] Bartolini, F., Tefas, A., Barni, M. and Pitas, I. Image Authentication Techniques for Surveillance Applications. Proceedings of the IEEE, Vol. 89, No. 10, October 2001.

[35] Makris, D. and Ellis, T. Learning Semantic Sense Models from Observing Activity in Visual Surveillance. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol. 35, No. 3, June 2005, pp. 397–408.

[36] Trivedi, M. M., Gandhi, T. L. and Huang, K. S. Homeland Security Distributed Interactive Video Arrays for Event Capture and Enhanced Situational Awareness. IEEE Intelligent Systems, September/October 2005.

[37] Müller, K., Smolic, A., Dröse, M., Voigt, P. and Wiegand, T. 3-D Construction of a Dynamic Environment with a Fully Calibrated Background for Traffic Scenes. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 4, April 2005.

[38] Desurmont, X., Bastide, A., Chaudy, C., Parisot, C., Delaigle, J.F. and Macq, B. Image analysis architectures and techniques for intelligent surveillance systems. IEE Proc. – Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 224–231.

[39] Foresti, G.L., Micheloni, C., Snidaro, L., Remagnino, P. and Ellis, T. Active Video-Based Surveillance System. IEEE Signal Processing Magazine, March 2005, pp. 25–37.

[40] Micheloni, C., Foresti, G.L. and Snidaro, L. A Network of Co-operative Cameras for Visual Surveillance. IEE Proc. – Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 205–212.

[41] Bowden, R. and KaewTraKulPong, P. Towards automated wide area visual surveillance: tracking objects between spatially-separated, uncalibrated views. IEE Proc. – Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 213–223.

[42] Kreucher, C., Kastella, K. and Hero III, A.O. Multitarget Tracking Using the Joint Multitarget Probability Density. IEEE Transactions on Aerospace and Electronic Systems, Vol. 41, No. 4, October 2005.

[43] Hu, W., Tan, T., Wang, L. and Maybank, S. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, Vol. 34, No. 3, August 2004, pp. 334–352.

[44] Kumar, P., Ranganath, S., Weimin, H. and Sengupta, K. Framework for Real-Time Behaviour Interpretation from Traffic Video. IEEE Transactions on Intelligent Transportation Systems, Vol. 6, No. 1, March 2005.

[45] Bremond, F., Thonnat, M. and Zuniga, M. Video Understanding Framework for Automatic Behaviour Recognition. Behaviour Research Methods, Vol. 28, No. 3, August 2006, pp. 416–426.

[46] Caricotte, C., Desurmont, X., Ravera, B., Bremond, F., Orwell, J., Velastin, S.A., Obodez, J.M., Corbucci, B., Palo, J. and Cernocky, J. Toward Generic Intelligent Knowledge Extractions from Video and Audio: The EU-funded CARETAKER Project. The IET Conference on Imaging for Crime Detection and Prevention (ICDP 2006), London, Great Britain, June 13–14, 2006. Pp. 470–476.

[47] Maier, A., Rinner, B., Schriebl, W. and Schwabach, H. Online Multi-Criterion Optimization for Dynamic Power-Aware Camera Configuration in Distributed Embedded Surveillance Clusters. Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA'06), IEEE, 2006.

[48] Korshunov, P. and Ooi, W.T. Critical Video Quality for Distributed Automated Video Surveillance. MM'05, November 6–11, 2007, Singapore. Pp. 151–160.

[49] May, A., Teh, J., Hobson, P., Ziliani, F. and Reichel, J. Scalable Video Requirements for Surveillance Systems. The Institution of Electrical Engineers, IEE, Stevenage, 2004. Pp. 17–25.

[50] Frescura, F., Giorni, M., Feci, C. and Cacopardi, S. JPEG2000 and MJPEG2000 Transmission in 802.11 Wireless Local Area Networks. IEEE Transactions on Consumer Electronics, Vol. 49, No. 4, November 2003.

[51] Stanacevic, M. and Cauwenberghs, G. Micropower Gradient Flow Acoustic Localizer. IEEE Transactions on Circuits and Systems – I: Regular Papers, Vol. 52, No. 10, October 2005, pp. 2148–2157.

[52] Julian, P., Andreou, A.G., Riddle, L., Shamma, S., Goldberg, D.H. and Cauwenberghs, G. A Comparative Study of Sound Localization Algorithms for Energy Aware Sensor Network Nodes. IEEE Transactions on Circuits and Systems – I: Regular Papers, Vol. 51, No. 4, April 2004.

[53] Smeaton, A.F. and McHugh, M. Towards Event Detection in an Audio-Based Sensor Network. VSSN'05, November 11, 2005, Singapore. Pp. 87–94.

[54] Aarabi P. Self-Localizing Dynamic Microphone Arrays. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and reviews, Vol. 32, No. 4, November 2002.

[55] Wald, L. A European proposal for terms of reference in data fusion. International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 7, 1998, pp. 651–654.

[56] Hall, D.L. Challenges in Data Fusion: Dirty Secrets, Current State of Technology and a Research Roadmap. Associate Dean for Research School of Information Sciences and Technology, February 28, 2005.

[57] Steinberg, A.N., Bowman, C.L. and White, F.E. Revisions to the JDL Data Fusion Model. Proc. SPIE Vol. 3719, Sensor Fusion: Architectures, Algorithms, and Applications III, Dasarathy, B.V.(ed.), March 1999. Pp. 430–441.

[58] Steinberg, A.N. An approach to threat assessment. 8th International Conference on Information Fusion, Vol. 2, 25–28 July 2005. 8 p.

[59] Jaeger, C. Security risk assessment methodology for communities (RAM-C). 38th Annual International Carnahan Conference on Security Technology, 11–14 Oct. 2004. Pp. 106–110.

[60] Blasch, E. and Plano, S. Proactive decision fusion for site security. 8th International Conference on Information Fusion, Vol. 2, 25–28 July 2005. 8 p.

[61] Blasch, E. and Plano, S. DFIG level 5 (user refinement) issues supporting situational assessment reasoning. 8th International Conference on Information Fusion, Vol. 1, 25–28 July 2005. 9 p.

[62] Hall, D.L. Perspectives on the fusion of image and non-image data. Proceedings 32nd Applied Imagery Pattern Recognition Workshop, 15–17 Oct. 2003. Pp. 217–220.

[63] Pavlidis, I., Morellas, V., Tsiamyrtzis, P. and Harp, S. Urban surveillance systems: from the laboratory to the commercial world. Proceedings of the IEEE Volume 89, Issue 10, Oct. 2001, pp. 1478–1497.

[64] Nelson, C.L. and Fitzgerald, D.S. Sensor fusion for intelligent alarm analysis. Aerospace and Electronic Systems Magazine, IEEE, Vol. 12, Issue 9, Sept. 1997, pp. 18–24.

[65] Newman, A.R. Confidence, pedigree, and security classification for improved data fusion. Proceedings of the Fifth International Conference on Information Fusion, Vol. 2, 8–11 July 2002. Pp. 1408–1415.

[66] Ming, L., Xie, G., Li, H. and Yang, L. Research on Remote Intelligent Surveillance Using Wireless Network. Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21–23, 2006, Dalian, China. Pp. 4518–4521.

[67] Kreimer, J. Effectiveness-Analysis of Real-Time Data Acquisition and Processing Multichannel Systems. IEEE Transactions on Reliability, Vol. 51, No. 1, March 2002.

[68] Yang, H., Xie, L. and Xie, F. Research on Cluster Remote Video Surveillance System. 2006 IEEE International Conference on Industrial Informatics. Pp. 1171–1174.

[69] Marseguerra, M., Zio, E. and Podofillini, L. Optimal Reliability/Availability of Uncertain Systems via Multi-Objective Genetic Algorithms. IEEE Transactions on Reliability, Vol. 53, No. 3, September 2004.

[70] Avritzer, A., Ros, J.P. and Weyuker, E. Reliability Testing of Rule-Based Systems. IEEE Software, September 1996.

[71] Oikarinen, J., Räty, T., Luo, M. and Sihvonen, M. Quality of Service Management in ASEMA System for Unicast MPEG 4 Video Transmissions. 33rd Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2007, 28–31 August, 2007. Pp. 191–202.

PUBLICATION 1

High-Level Architecture for a Single Location Surveillance Point

Proceedings of the Third International Conference on Wireless and Mobile Communications, ICWMC'07. Guadeloupe, French Caribbean, 4–9 March, 2007.

HIGH-LEVEL ARCHITECTURE FOR A SINGLE LOCATION SURVEILLANCE POINT

Tomi Räty
Software Architectures and Platforms
VTT Technical Research Centre of Finland
Oulu, Finland
tomi.raty@vtt.fi

Abstract—**The Single Location Surveillance Point (SLSP) is a distributed multi-sensor surveillance software system. It comprises of an arbitrary amount of sensors that collect readings from a single location, which is the surveillance point. Each sensor transmits its crude sensor data to a session server, which handles the connections between the components. The session server routes the crude sensor information to the logical decision making service. The logical decision making server automatically deducts the situation at the surveillance point based on the received sensor information. The logical decision making server informs the security manager server of the situation at the surveillance point. The security manager server's user interface displays essential information about the surveillance point to a human security administrator. The security manager server can transmit information to the nomadic security personnel's smart phones over wireless networks. The SLSP system decreases the amount of redundant information that otherwise would be handled by the human security administrator and the security personnel. This goal is achieved with a prominent architecture between the components and a server that conducts automatic decision making. The research is based on the constructive method of the related publications and technologies and the results are derived by the abstractive analysis of the available material. This paper illustrates the high-level architecture of the SLSP system.**

Keywords-component; Mobile & Wireless applications & services; Media and content distribution over wireless networks; Multi-sensor surveillance system

I. INTRODUCTION

Recent progress in computing, communication, and sensor technology are inciting the development of multiple new applications. This trend is apparent in pervasive computing, sensor networks, and embedded systems. During the past two decades, surveillance systems have been an area of vehement research. Recently, considerable research efforts have been concentrated on video-based surveillance systems, especially for public safety and transportation systems. [1]

The increasing demand for safety and security has resulted in more research in constructing more efficacious and intelligent automated surveillance systems. A future challenge is to develop a wide-area distributed multi-sensor surveillance system which has robust, real-time computer algorithms able to execute with minimal manual reconfiguration on variable applications. These systems should be adaptable enough to automatically accommodate and endure with the changes in the environment, such as lighting, scene geometry or scene activity. The system should be expandable; hence it should be based on standard hardware and exploit plug-and-play technology. [10]

To address the contemporary vicissitudes of surveillance systems, we have defined a Single Location Surveillance Point (SLSP) system and its architecture. The sensors survey and obtain information from a single mutual location. This area is the surveillance point. The SLSP architecture comprises of an arbitrary amount of miscellaneous sensors, a session server, a logical decision making server, a security manager server, and an arbitrary amount of end-devices, e.g., laptops, desktops, and smart phones.

The realized sensors consist of a biometrical sensor, an audio sensor, a video recorder, a network activity sensor, and an alternative video and motion sensor. The sensors monitor their immediate environment, which is called the surveillance point, and transmit knowledge about it to the session server. The session server routes the information to the logical decision making server. The logical decision making server collects all the information from the various sensors and performs logical deductions from the obtained information. These logical deductions indicate different situations of the surveillance point. The logical deductions are transmitted to the security manager server, at which a human security administrator resides via the session server. The human administrator can transmit information to the end-devices, e.g., laptops, desktops, and smart phones. The end-devices are registered to the SLSP system. The nomadic security personnel patrol in the premises, or they can be dispatched to the area, which is under surveillance. The nomadic security personnel can receive the information on their end-devices, e.g., smart phones, over wireless networks. The SLSP system decreases the amount of redundant information that otherwise would have to be handled by the human security administrator and the nomadic security personnel. Deductions based on the sensor

Proceedings of the Third International
Conference on Wireless and Mobile Communications (ICWMC'07)
0-7695-2796-5/07 $20.00 © 2007 IEEE

information are made automatically and they are informed to the security manager server. The human security administrator and the nomadic security personnel will not be inundated with superfluous information. Due to the disposition of the surveillance information, it is vital to transmit only the most important erudition as rapidly as possible. The security administrator server can be used to alert the patrolling nomadic security personnel of emergencies instantaneously. This can be conducted by the security manager administrator ordaining the distribution of critical information automatically and directly from the session server over a wireless network to the nomadic security personnel's end-devices, e.g., smart phones. This will make the reception of the crude sensor information and logical deductions quicker at the end-device, instead of having the information first being routed to the security manager server and the human security administrator deciding on what information to transmit to the end-devices. Another option is for the security manager administrator to select the received information, e.g., crude sensor information and/or logical deductions, which the security manager administrator wants to route to the nomadic security personnel's end-devices over a wireless network.

The conference paper is presented in the ensuing manner. First, the two most common and fundamental surveillance structures, video surveillance and audio surveillance, are presented. This is followed by a concise description of the current development of surveillance systems. Then Single Location Surveillance Point is presented in detail. This is followed by a comparison between the SLSP and theoretical paradigms presented. Finally, the conclusion recapitulates this conference paper.

II. VIDEO SURVEILLANCE STRUCTURES

According to Foresti et al., the importance of video surveillance techniques has augmented significantly since the latest terrorist incidents. Safety and security have become critical in numerous public areas, and there is a designated need to enable human operators to remotely monitor activity across large environments, e.g. shopping malls. Modern video-based surveillance systems utilize real-time image analysis techniques for efficacious image transmission, colour image analysis, event-based attention focusing, and model-based sequence comprehension. [3]

Trivedi et al. proclaim that the video surveillance activity has significantly augmented recently. Earlier work addressed mostly with single stationary cameras, but the current trend is to active multicamera systems. These systems provide multiple advantages over single camera systems. This includes multiple overlapping views for procuring 3D information and plying occlusions, multiple non-overlapping cameras for covering vast tracts, and active pan-tilt-zoom (PTZ) cameras for discerning object details. [8]

A. Generations of surveillance systems

Bramberger et al. state that cameras can be equipped with a high-performance onboard computing and communication infrastructure, coalescing video sensing, processing, and communications in an individual embedded device. By offering access to multiple views via the cooperation among individual cameras, networks of embedded cameras can possibly support more complex and demanding applications, containing smart rooms, surveillance, tracking, and motion analysis, than an individual camera. [1]

Video-based surveillance systems have developed in the three generations. First-generation surveillance systems utilized analogue paraphernalia throughout the plenary system. Analogue closed-circuit television cameras captured the observed scene and transmitted the video signals over analogue communication lines to the central back-end systems, which rendered and archived the video data. [1]

Second-generation surveillance systems employ digital back-end components, enabling real-time automated analysis of the incoming video data. Hence, an automated event detection and alarm raising substantially augmented the content of simultaneously monitored data and the plenary surveillance system's quality. [1]

Third-generation surveillance systems have finalized the digital transformation. In these systems, the video signal is converted into the digital domain at the cameras, which transmit the video data through a computer network, for instance a local area network. The digital cameras can also directly compress the video data to conserve bandwidth. The back-end and transmission systems of a third-generation surveillance system have also augmented their functionality. For instance, they employ intelligent hubs to gather the video data, accumulate the information from different cameras, and transmit it to the video archive and the operators. [1]

B. Smart cameras

Modern processor technology enables the implementation of smart cameras, which directly execute highly sophisticated video analysis. These smart cameras integrate video sensing, video processing, and communication into an individual embedded device. They are designed as reconfigurable and resilient processing nodes with self-configuration, self-monitoring, and self-diagnosis capabilities. Smart cameras maintain the prevailing paradigm shift from a central to a distributed control surveillance system. The main motivation for this shift is augmenting the surveillance system's functionality, availability and autonomy. Smart cameras are key components of these novel surveillance systems, because they offer adequate performance for onboard video processing and distributed control. These surveillance systems can respond autonomously to alterations in the system's environments and to detected events in the monitored scenes. [1]

III. AUDIO SENSOR STRUCTURES

Accurate and robust localization and tracking of acoustic sources is of interest to a variety of applications in surveillance, multimedia, and hearing enhancement. Miniaturization of microphone arrays incorporated with acoustic processing further augments the utility of these systems, but poses challenges to achieve precise localization performance due to abating aperture. For surveillance, acoustic emissions from ground vehicles offer a facilely detected signature, which can be employed for unobtrusive and passive tracking. [7]

An integrated miniature sensor array with localization and communication capability could be maintained as a low-cost, low-power small autonomous node in network configuration distributed over a vast region. This results to a higher localization performance in distributed sensing environments bypassing the requirement for excessive data transfer and fine-grain time synchronization among nodes, with low communication bandwidth and low complexity. Additional improvement can also be attained by fusion with other data modalities, such as video. [7]

IV. CURRENT DEVELOPMENT OF SURVEILLANCE SYSTEMS

As the personal computing era evolves into a ubiquitous computing one, there is a requisite for a world of completely connected devices with inexpensive wireless networks. Enhancements in wireless network technology interfacing with emanating microsensors predicated on MEM (Micro-Electro-Mechanical) technology is enabling sophisticated, yet inexpensive, sensing, storage, processing, and communication capabilities to be unobtrusively embedded into the everyday physical world. Embedded web servers can be utilized to connect the physical world of sensors and actuators to the virtual world of information, utilities, and services. Consequently, a rush of research activity has begun in the sensor networks domain, particularly in wireless ad hoc sensor networks. Even though many of the sensor technologies are not novel, some physical and technological barriers of performing wireless communications have confined the viability of such devices in the past. Some of the advantages of the newer, more capable sensor nodes are their abilities to establish large-scale networks, implement sophisticated protocols, decrement the amount of communication (wireless) required to execute tasks by distributed and local calculations, and implement intricate power saving modes of operation depending on the environment, the application, and the state of the network. [5]

Valera & Velastin presented the state of deployment of intelligent distributed surveillance systems, including a revision of contemporary image processing techniques, which are employed in different modules that constitute part of surveillance systems. Reviewing these image processing tasks, it has discriminated research areas that need to be scrutinized further, such as adaptation, data fusion, and tracking methods in a co-operative multi-sensor environment, extension of techniques to distinguish complex activities and interactions between detected objects. In terms of communication or integration between different modules, an examination of new communication protocols and the creation of metadata standards are required. It is vital to consider ameliorated means of task distribution that optimize the use of central, remote facilities, and data communication networks. One of the facets that will be inherent in the future for the development of distributed surveillance systems is the definition of a framework to design distributed architectures well established in the systems engineering best practice. [10]

Advances in information and communication technologies can potentially offer considerable improvements in the management of public places pertaining to safety and security. These include technologies, for instance, digital storage of video, transmission of video/audio streams over wired and wireless networks, etc. Discriminating features of the deployment of technology to maintain surveillance in modern urban environments includes large, geographically dispersed facilities and hierarchical multi-agency management structures. These are then construed into requirements of robust image processing, distribution, scalability, and usability. Video surveillance applications need to be real-time, because there is security requirement considering minimum time constraints. Video surveillance must low delay and timing constraints for processing. [2] & [11]

V. THE SINGLE LOCATION SURVEILLANCE POINT ARCHITECTURE AND COMPONENTS

The Single Location Surveillance Point comprises of three individual domains as in Figure 1. These three domains are 1) the Surveillance Domain, which comprises of an arbitrary amount and variety of sensors, and 2) the Security Administration and Surveying Domain, which comprises of the session server to which the sensors transmit their information and the logical decision making server, and the end-devices, e.g., the laptops, desktops and smart phones 3) the Security Personnel Management domain, which is intended for conducting security personnel from a remote and centralized location. This domain also provides an interface to the human security administrator.

Initially, sensors transmit their information to the session server. The session server transmits the crude sensor information to the logical decision making server. The logical decision making server is responsible of transmitting its logical deductions regarding the surveillance point to the security manager server through the session server. Then the security manager server can transmit orders, with the help of the human security administrator, to the security personnel, e.g., the laptop, desktop, and/or smart phones end-devices.

The data flow from the sensors of the Surveillance Domain is primarily the ensuing: 1) the crude sensor information is conveyed from the sensors to the session server, 2) the session server transmits all the crude sensor information it receives to the logical decision making server, 3) after performing the its automatic logical calculations, the logical decision making server transits its deductions to the security manager server and/or the nomadic security personnel via the session server. Then the human security administrator can issue orders to the end-devices, or even re-route the deduction information and/or crude sensor information, e.g., video footage, it receives to the end-devices.

The session server acts as an interface from which crude sensor information can be procured. The session server entails two fail-safe mechanisms: 1) if the crude sensor information cannot be distributed to the logical decision making server, then the session server will transmit the crude sensor information to the security manager server, and 2) if the crude sensor information cannot be distributed to the security manager server either, or if the session server is ordained by the security manager server to transmit the crude sensor information directly to the end-devices, then the session server will transmit the crude sensor information to the end-devices.

The Testing Environment is utilized during the Surveillance Domain's and the Security Administration and Survey Domain's development phases. The test server is primarily employed to test the session server on behalf of the network activity monitor. The test server may also be utilized to test the session server by providing artificial sensor information on behalf of the sensors.

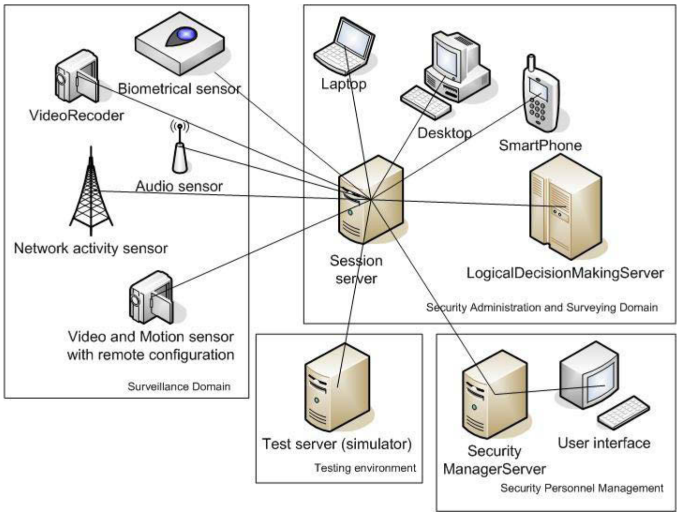


Figure 1. Single Location Surveillance Point.

A. *The Surveillance Domain*

The actual biometrical sensor will be a veritable fingerprint sensor hardware device accompanied with its software module stored on the session server. The network activity monitor surveys the current situation of the wireless network activity in its local environment. Its software module will also be stored on the session server.

The purpose of the video recorder is to deliver constant, pristine, and immaculate video stream to the receiving party. The video recorder utilized is the Axis 213 PTZ network camera. The audio sensor will provide real-time sound identification and localization with multiple microphones.

The video and motion sensor with remote configuration (VMSRC) comprises of a stationary video camera and a motion detector. The VMSRC can transmit the information obtained from its sensors to a remote receiver, the logical decision making server. The VMSRC provides an alternative and direct route from the VMSRC to the logical decision making server. The VMSRC entails the capability of creating a wireless and secure connection, e.g. GSM/GPRS, between the VMSRC and the logical decision making server.

The main functions of the Surveillance Domain's sensors are twofold: 1) Collect information from its ambit, and 2) Transmit the collected information to the session server of the Security Administration and Surveying Domain.

B. *The Security Administration and Surveying Domain*

The Security Administration and Surveying Domain comprises of the session server, the logical decision making server, and an arbitrary amount of end-devices., e.g., laptops, desktops, and smart phones. An end-device is reputed to be a device from which security personnel may view data related to security erudition. The session server handles information collection from the sensors of the Surveillance Domain. The gathered information may be either directly transmitted to end-devices and/or the security manager server, or transmitted to the logical decision making server for further information processing. If the information is transmitted to the logical decision making server, then the information from an individual sensor, or the consolidation of the sensor information, will processed further to deduct intricate information. The refined knowledge resulting from the logical decision making server is then transmitted to the session server. The session server conducts the ultimate decision of what end-devices and whether to transmit the security manager server the acquired refined knowledge. The session server handles/interacts with the sensors. The session server must store the information transmitted by the sensor accompanied with the date and timestamp. This information is stored to be reviewed later by security officials. The logical decision making server must have a modular structure to be expandable for new possible sensors. Additional sensors need to be attachable to the session server and the logical decision making server.

Out of the three end-devices illustrated in the SLSP system, i.e., the laptop, the desktop, and the smart phone, the smart phone will have the highest priority. Wireless technologies are deemed as interesting by contemporary research and science, for instance by Sagiraju et al. [6], Kaplan [4], and Tseng et al. [9] all of whom have conducted research pertaining wireless technology and surveillance/safety systems.

The main functions of the Security Administration and Surveying Domain's components, a.k.a. the session server, the logical decision making server, and the end-devices, are the ensuing:

For the session server: 1) Collection of information from the sensors of the Surveillance Domain, 2) Transmission of the collected information from the sensors to the logical decision making server, 3) Retrieval of the processed and refined information from the logical decision making server, 4.1) Transmission of the processed and refined information from the logical decision making server to the Security Personnel Management’s security manager server, 4.2) If the logical decision making server is unavailable, then the session server will transmit the crude sensor information to the security manager server and 4.3) If the security manager server is unavailable or the human security administrator has ordained the session server to automatically distribute the crude sensor information and/or logical deductions to the end-devices, then the session server will transmit the crude sensor information and/or logical deductions to the end-devices, and 5) Perform as an interface to (and from) the assorted sensors.

For the logical decision making server: 1) Collection of information provided by the session server, 2) Refinement and processing of gathered information, either based on the particular sensor it originally resulted from and/or based on the consolidation of the sensors, and, 3) Transmission of the refined and processed information to the Server for the distribution to the appropriate recipients.

For the end-devices: 1) Displaying/illustrating the plebeian or refined information to the security personnel, which has been received from 1.1) the Security Personnel Management's security manager server, but if unavailable 1.2) the logical decision making server, but if unavailable, or if ordained by the human security administrator of the security manager server then 1.3) the information received directly from session server.

C. Testing environment

The test server can induce different sensor information to the session server, and different test cases can be executed from the test server. This will conducted in a manner that the session server will postulate it is receiving input information from the sensors of the Surveillance Domain. Initially, the Test server will be used to test the pseudo-functionality on behalf of the network activity monitor, but can be elaborated to produce input regarding the other sensors of the Surveillance Domain.

The main functions of the Testing environment's component, a.k.a. the Test Server (simulator), are the following test functionalities: 1) Propagating artificial sensor information to the session server, and 2) Receiving responses from the session server.

D. Security Personnel Management

The Security Personnel Management comprises of the user interface and the security manager server. The Security Personnel Management is used to conduct and coordinate security personnel from a remote location. The end-devices of the security personnel are the end-devices of the Security Administration and Surveying Domain, e.g., the laptop, desktop and smart phones. The SLSP focuses chiefly on ambulating security personnel equipped with smart phones. The security personnel ambulate in their own patrol region. The information that the security management server receives from the Surveillance Domain, i.e., sensors, can be either crude or processed by the logical decision making server. The administrator can verbally ordain instructions to the security personnel through the user interface to the smart phones of the security personnel. The human security administrator at the user interface of the security manager server may also re-route the information it receives from the logical decision making server or even the crude sensor information, e.g., video footage, to the security personnel's smart phones. The human security administrator may also ordain the session server to automatically forward the crude sensor information and/or the logical deductions of the logical decision making server automatically.

The main functions of the Security Personnel Management's components are the ensuing:

For the security manager server: 1) Reception of processed information from the logical decision making server or if unavailable 2) The reception of the crude sensor information from the session server, 3) Routing of orders and communication data between the end-devices and the user interface, and 4) Transmitting crude sensor and/or logical deductions to the end-devices, either by ordaining the session server to transmit this information automatically or by allowing the human security administrator to selectively choose what crude sensor and/or logical deduction information to transmit to the end-devices.

For the user interface: 1) Displaying processed information received from the logical decision making server or crude sensor information from the session server, and 2) Receiving orders from the human security administrator and distributing them to the smart phones and/or the session server through the security manager server.

VI. A BRIEF SUMMARY OF SLSP SOLUTIONS COMPARED TO THE THEORETICAL PARADIGMS

The Single Location Surveillance Point system achieves the requirement of Foresti et al. by enabling human operators to remotely monitor activity across large environments. Trivedi et al. stated that multicamera systems are a current trend, the SLSP utilizes this approach, but in a different manner. The SLSP focuses on a single and remote point of surveillance, but with multiple sensors.

Bramberger et al. argued that in third-generation surveillance systems, the video signal is converted into the digital domain at the cameras, which transmit the video data through a computer network, for instance a local area network. The SLSP system is predicated on a similar functionality, the crude sensor information is distributed in a digital data format to a network. This network comprises of the session server, logical decision making server, security manager server, and the end-devices, e.g., the laptop, desktop and smart phone end-devices.

Bramberger et al. introduce systems to gather the video data, accumulate the information from different cameras, and transmit it to the video archive and the operators. Stanacevic & Cauwenbergh declare that additional improvement can also be attained by fusion with other data modalities, such as video. In addition, Megerian et al announce that some of the benefits of the newer, more capable sensor nodes are their abilities to establish large-scale networks, implement sophisticated protocols, decrement the amount of communication (wireless) required to execute tasks by distributed and local calculations, etc. Valera and Velastin declare that certain research areas need to be examined further, such as data fusion. The SLSP system handles all of the aforementioned issues. It entails resembling methods of operation, instead of only collecting video data, the SLSP system's session server culls multi-sensor information and transmits it to the logical decision making server. The logical decision making server performs automated deductions based on the crude sensor information and transmits its deductions to the security manager server at which the human security administrator resides. The obtained crude sensor information can be stored at the session server for any possible use required by authorities or administrators.

Bramberger et al. also proclaim that smart cameras maintain the prevailing paradigm shift from a central to a distributed control surveillance system. Additionally, Bramberger et al. promulgate that these surveillance systems can respond autonomously to alterations in the system's environments and to detected events in the monitored scenes.

The SLSP attains both requirements, the surveillance control in the SLSP is distributed over the logical decision making server and the security manager server. In addition, the fail-safe systems of the session server will coerce crude sensor information to secondary and tertiary recipients, if the currently primary recipient fails to receive the crude sensor information. The logical decision making server performs deductions automatically based on the inputs from the sensors.

Megerian et al. declare that wireless sensor networks provide a viable alternative to numerous existing technologies. Desurmount et al. and Velastin denote that wireless technologies provide advantages in surveillance systems. The SLSP system utilizes wireless connections. The crude sensor information may be transmitted over a wireless connection between the sensors and the session server. Also, the data connections between the session server and the logical decision making server, the logical decision making server and the security manager server, and between the security manager server and the end-devices may be wireless. Only the connection between the security manager server and the smart phone end-device is mandated to be wireless.

VII. Conclusion

This paper has illustrated the architecture of a high-level distributed multi-sensor surveillance system. The main advantage of the SLSP system is that it is a distributed surveillance system, which utilizes multiple sensors in automatically deducting the situation of the monitored point. The processed information is distributed to the security administration manager and the end-devices, e.g., smart phones, of the security personnel over a wireless network. In detail, the distribution of processed and refined critical information is transmitted to the session server from the sensors. Then the crude sensor information is transmitted to the logical decision making server. Both the logical deductions of the logical decision making server and the crude sensor information of the sensors can be transmitted to the security administration manager and the nomadic security personnel, which can only be reached over a wireless network. A fundamental advantage of the SLSP system is that it reduces the amount of redundant information which is delivered to the human security administrator and the nomadic security personnel. The SLSP system informs the human security administrator and the nomadic security personnel about situations that require intervention. The SLSP distributes the most vital information to the appropriate human users, i.e., the human security administrator and the nomadic security personnel, as quickly as possible. The current disadvantage of the SLSP is that it monitors only a single location. The SLSP can be applied to various locations at which automatic surveillance is desirable, e.g., entrances of buildings.

References

[1] Bramberger, M., Doblander, A., Maier, A., Rinner, B., & Schwabach, H.: Distributed Embedded Smart Cameras for Surveillance Applications, Computer, Published by the IEEE Computer Society, February 2006, pp. 68-75.

[2] Desurmont, X., Bastide, A., Chaudy, C., Parisot, C., Delaigle, J.F., and Macq, B.: Image analysis architectures and techniques for intelligent surveillance systems, IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, April 2005.

[3] Foresti, G.L., Micheloni, C., Snidaro, L., Remagnino, P., & Ellis, T.: Active Video-Based Surveillance System, IEEE Signal Processing Magazine, March 2005, pp. 25-37.

[4] Kaplan, L.M.: Local Node Selection for Localization in a Distributed Sensor Network, IEEE Transactions on Aerospace and Electronic Systems, Vol. 42, No. 1, January 2006, pp. 136-146.

[5] Megerian, S., Koushanfar, F., Potkonjak, M., & Srivastava, M.B.: Worst and Best-Case Coverage in Sensor Networks, IEEE Transactions on Mobile Computing, Vol. 4, No. 1, January/February 2005, pp. 84-92.

[6] Sagiraju, P.K., Agaian, S., & Akopian, D.: Reduced complexity acquisition of GPS signals for software embedded applications, IEE Proc.-Radar Sonar Navig., Vol. 153, No. 1, February 2006, pp. 69-78.

[7] Stanacevic, M. & Cauwenberghs, G.: Micropower Gradient Flow Acoustic Localizer, IEEE Transactions on Circuits and Systems-I: Regular Papers, Vol. 52., No. 10, October 2005, pp. 2148-2157.

[8] Trivedi, M. M., Gandhi, T. L., & Huang, K. S.: Homeland Security Distributed Interactive Video Arrays for Event Capture and Enhanced Situational Awareness, IEEE Intelligent Systems, September/October 2005, pp. 58-66.

[9] Tseng, Y.-C., Lin, T.-Y., Liu, Y.-K., & Lin, B.-R.: Event-Driven Messaging Services Over Integrated Cellular and Wireless Sensor Networks: Prototyping Experiences of a Visitor System, IEEE Journal on Selected Areas in Communications, Vol. 23, No. 6, June 2005, pp. 1133-1145.

[10] Valera, M. & Velastin, S.A.: Intelligent distributed surveillance systems: a review, IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, April 2005, pp. 192-204.

[11] Velastin, S.B.: Special Section on Intelligent Distributed Surveillance Systems, IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, April 2005.

PUBLICATION 2

A Scalable Quality of Service Middleware System with Passive Monitoring Agents over Wireless Video Transmission

Proceedings of the Sixth International
Conference on the Quality of Information and
Communications Technology, QUATIC 2007.
Lisbon, Portugal, 12–14 September, 2007.
© 2007 IEEE.
Reprinted with permission from the publisher.

A Scalable Quality of Service Middleware System with Passive Monitoring Agents over Wireless Video Transmission

Tomi Räty, Johannes Oikarinen, and Markus Sihvonen
VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-Oulu FIN,
{Tomi.Raty, Johannes.Oikarinen, Markus.Sihvonen}@vtt.fi

Abstract

We have constructed a Scalable Quality of Service middleware system, which contains a monitoring user agent client, a monitoring user agent server and a leader agent. A network camera sends video transmission to the smart phone. The video transmission transits through a Scalable Quality of Service server. The monitoring user agent client resides in the smart phone. The monitoring user agent server and leader agent reside in the Scalable Quality of Service server. Both monitoring user agents monitor the video transmission's bit-rate. The monitoring user agents transmit their evaluation to the leader agent. Then the leader agent deducts whether to ordain the network camera to scale the Quality of Service values down or up. The research problem of the paper is to determine a middleware, which improves the control of the video transmission over a mobile system. Our innovative theories are the Scalable Quality of Service middleware system's architecture, passive monitoring paradigm, and calculation and deduction methods. The theories endeavor to optimize the video transmission rate to a smart phone over a wireless network. The operability of the constructed prototype indicates that this endeavor is attained. The research is based on the constructive method of the related publications and technologies and the results are derived by the implemented Scalable Quality of Service middleware system.

1. Introduction

Accompanied with the fast growth of wireless networks and the huge success of Internet video, wireless video services are expected to be broadly deployed in the near future. As assorted types of wireless networks are converging into all IP networks, i.e., the Internet, it is crucial to scrutinize video delivery over the wireless Internet. The recent trends in the development of real-time Internet applications and the quick growth of mobile systems attest that the future Internet architecture will be required to support assorted applications with various Quality of Service (QoS) requirements. QoS support is a multidisciplinary topic comprising of several contingents, compassing from applications, terminals, networking architectures to network management, business models, and ultimately the end users. [10]

Middleware has stemmed as a key architectural component in supporting distributed applications. The role of middleware is to render a unified programming model to application writers and to mask problems of heterogeneity and distribution. The importance of the topic is reflected in the increasing visibility of industrial standardization activities such as Microsoft's Distributed Component Object Model (DCOM) and Object Management Group's (OMG) Common Object Request Broker Architecture (CORBA). [4]

The Scalable Quality of Service (SQoS) middleware system we have developed endeavors to solve the predicament of fluctuating QoS over a wireless network. In practice, the system is targeted at a plenary connection between a Symbian OS 8.0 smart phone and a network camera. The video transmission of the network camera is routed through a SQoS server to the smart phone. During the initial video transmission from the network camera through the SQoS server to the ultimate destination, the smart phone, the bit-rate of the utilized protocol, the Real-time Transport Protocol (RTP), over the wireless network is evaluated. The SQoS comprises of three software agent components: 1) the monitoring user agent client residing in the smart phone, 2) the monitoring user agent server residing in the SQoS server, and 3) the leader agent residing at the SQoS server. The monitoring user agent server and the leader agent are integrated together. The evaluation is conducted on a consolidation of these two monitoring user agents, i.e., the monitoring user agent client and the monitoring user agent server. Once the evaluation has been concluded, the leader agent scales the quality of the network camera's video transmission according to the evaluations of the agents and the deductions calculated by the leader agent. The paradigm according to which the monitoring user agents execute is called the passive monitoring paradigm. In this paradigm, the monitoring user agents collect all the information and they distribute the information directly to the leader agent without refining the information.

In detail, the SQoS system components comprise of the monitoring user agent client, residing in the smart phone, the monitoring user agent server, residing in the SQoS server, the leader agent, residing in the SQoS server, and the network camera, which can be conducted through its open API (Application Protocol Interface). The monitoring user agent server and the leader agent are integrated together. The applied protocol for video transmission in the implemented SQoS system is the RTP. The smart phone utilizes a GPRS (Global System for Mobile Communications) or WCDMA (Wideband Code Division Multiple Access) connection to the SQoS server and ultimately to the network camera over the wireless network.

The innovative theories, which we have created, are the Scalable Quality of Service middleware system's architecture, passive monitoring paradigm, and calculation and deduction methods. The intent of the theories is to optimize, or to improve, the video transmission rate to a smart phone over a wireless network. The operability of the constructed SQoS middleware system prototype indicates that this endeavor is attained.

The structure of this paper is the following. First a general overview of contemporary QoS middleware systems is presented, including detailed information regarding scaling the QoS. Then a concise presentation of monitoring agents is rendered. This is followed by a presentation of the implemented SQoS middleware system, containing detailed information regarding the architecture of the SQoS middleware system, the monitoring paradigm of the SQoS middleware system, and the calculation method utilized for analyzing the current bit-rate in the SQoS middleware system. The conclusion summarizes the paper.

2. Quality of Service middleware systems

Middleware systems have emerged in recent years to support applications in ubiquitous and heterogeneous computing environments. Nahrstedt et al. introduce four key aspects of a QoS-aware middleware system: QoS-specification to enable the description of application behavior and QoS parameters; QoS translation and compilation to construe specified application behavior into candidate application configurations for different resource conditions; QoS setup to appropriately choose and instantiate a certain configuration; and QoS adaptation to accommodate to runtime resource fluctuations. [8]

A new generation of distributed applications, e.g. telemedicine and e-commerce applications, is being deployed in ubiquitous and heterogeneous computing environments. These applications are expected to deliver adaptive and adequate QoS, which attempts to be accepted by common users. This presents a challenge in supporting QoS specification, setup, and enforcement for these applications. [8]

2.1 The End-to-End QoS

To provision end-to-end QoS with an end-system solution, the video applications should be aware of and accommodative to the variation of the network condition in the wireless Internet. This accommodation comprises of network adaptation and media adaptation. The network adaptation refers to how many network resources, e.g., bandwidth and battery power, a video application should employ for its video content, i.e., to design an adaptive media transport protocol for video delivery. The media adaptation conducts the bit-rate of the video stream predicated on the predicted available bandwidth and adjusts error and power control behaviors according to the varying wireless Internet conditions. [10]

Intermittent loss and excessive delay have a negative impact on perceived video quality, and these are typically caused by network congestion. A congestion control mechanism is required at the end systems to reduce packet loss and delay. For conferencing and streaming video, congestion control typically takes the form of rate control. Rate control endeavors to minimize the possibility of network congestion by matching the rate of the video stream to the available network bandwidth. [10]

Delivering real-time behavior can be perceived as provisioning an adequate QoS, in which quality in this scope indicates guaranteed bandwidth with low and bounded network-inducted jitter and latency. Many real systems are complicated and contain different subsystems that function sporadically. Some real systems tolerate operational mode changes according to the environment stimuli. Other real systems reconfigure themselves dynamically according to online requirements update. There are real systems that require handling a variable number of requests from other subsystems or environments, e.g., mobile robots operating in dynamic environments. In these exemplars, the level of resources employed in the system may fluctuate dynamically and static resource allocation policies become insufficient. For efficiency reasons, and for cost reasons, these emerging applications require online changes to the communication requirements. [9]

2.2 QoS-aware middleware system

Solutions have been proposed for setting up and enforcing QoS in IP or asynchronous transfer mode networks (ATM), in operating system (OS) kernels, and in applications themselves. While network and OS-level solutions offer native and general QoS support, they may not be quickly and easily deployed in a large scale and for all new applications. Application-level solutions, such as adaptive or layered video coding, may be applicable only to a distinct application domain. [8]

Recently, various solutions at the middleware layer have been introduced, which are located between the applications and OS kernels. In comparison, middleware solutions provision more flexibility when assisting new applications in ubiquitous computing environments. Nahrstedt et al. propose their solution to the QoS specification, setup, and enforcement at the middleware layer. The middleware of Nahrstedt et al. cooperates easily with prevailing solutions at the OS, network, and application levels. Even when the OS or the network functions as best effort, rather than QoS-enabled, the middleware system can still assist applications with QoS adaptations. The solution of Nahrstedt et al. spans from the QoS specification and translation in the development phase of an application to QoS setup and adaptation at runtime. Nahrstedt et al. believe that these capabilities are intrinsic to any QoS-aware middleware system. [8]

QoS-aware middleware systems have emanated to help a new spectrum of applications that require QoS in heterogeneous and ubiquitous computing environments. Nahrstedt et al. have demonstrated that by applying an application component model, it is feasible to provision end-to-end application QoS through QoS-aware middleware systems, by: 1) creating appropriate QoS specifications, 2) interpreting and compiling multiple application configurations for the commensurate application to be executed in heterogeneous environments, 3) choosing an appropriate configuration and discovering the participating application components, and 4) adapting QoS at numerous levels with different granularities in case of QoS degradations. [8]

At runtime, after the QoS setup has been performed, the QoS-aware middleware may perform QoS adaptation during the execution of an application. The end-to-end resource assignment for every application configuration is the minimum assignment, predicated on the lowest acceptable QoS delivered by this configuration. During the execution, the delivered application QoS should be dynamically modified according to the actual resource availability. Hence, runtime QoS adaptation is required. [8]

2.3 QoS negotiation

Abdelzaher et al. propose a mechanism for QoS (re)negotiation as a method to assure graceful degradation in cases of overload, failures, or violation of pre-runtime postulations. This mechanism allows clients to express in their service requests a spectrum of QoS levels they can accept from the provider and perceived utility of receiving service at each of these levels. As a result, the application designer will be able to express acceptable compromises in QoS and their relative benefit/cost as deducted from application domain knowledge. [1]

Complex distributed applications, located on the heterogeneous environment, need to be flexible and accommodate to the QoS variations in their end-to-end execution. They must indicate several important properties. First, they must accept and endure resource scarcity to a certain minimum bound and can improve their performance if given a larger share of resources. Second, when QoS variations happen, they are willing to sacrifice and trade off quality of less critical parameters for the quality of critical parameters. [7]

2.4 The passive method in network performance monitoring

Passive measurements are chiefly utilized to observe actual traffic patterns in the network but can be employed for network performance monitoring. Traffic monitoring needs continuous collection of data and monitoring of links at full load. This can be problematic on very high-speed links, because it demands computing resources. The quality of analyzed information depends on the integrity and granularity of collected data. [3]

3. Monitoring agents

Agents have existed for many years, but recently they have become increasingly popular. Part of the reasons regarding the popularity of agent paradigms are their modularity, flexibility, and general applicability to a vast ambit of vicissitudes. Their recent increase in popularity is partially because of technological developments in distributed computing and the emergence of object-oriented programming models. Amendments in distributed computing technologies have augmented the need for paradigms, such as agents, that can model distributed problem solving. Object-oriented programming has emanated concepts into the main stream that conduct structuring agent-based approaches. [6]

An important use of agents is to monitor processes. Monitoring tasks, e.g. surveying gauges in nuclear power plants, supervising patients at intensive care, and conducting satellites from ground stations, tend to be quite tedious for the people doing the monitoring most of the time. When the monitoring tasks are not tedious, there is far too much occurring to heed everything, and mistakes are easily imputed. Agents can assist in these types of tasks, because they do not become bored when nothing occurs, and during crises they can assist in managing information overload. [6]

A middleware infrastructure, utilizing distributed software agents, is developed to provision services for high-performance programming environments and applications in clusters and networked heterogeneous systems. The agents amend expandability, enabling the number of machines involved to be augmented facilely by provisioning services that contain job distribution, monitoring, and controlling for the system. This offers flexibility and ease of managing the various resources available. Additionally, the distributed agents offer the required runtime support for the parallel and distributed applications developed at the application level. [2]

Monitoring agents can apply the conventional QoS metrics, such as traffic loss (e.g. in lost bytes per sent byte), erroneous packets (e.g. in lost packets per sent packet), throughput (e.g. in bytes per second), goodput, one-way delay, round-trip delay, and jitter (e.g. in milliseconds). The agents typically employ only the packet headers, e.g. the packet length field, and the timestamp offered by the interface to the service. [5]

4. The SQoS middleware system over a wireless video transmission

When a client and server have formed a connection, and the server transmits video data to the client, the QoS of the connection may fluctuate. There are multiple reasons for a connection fluctuating, for instance, network congestion causing jitter, delay or loss of packets. To alleviate these problems, the quality of video may be accommodated to suit the current status of the wireless network connection. In the implemented SQoS middleware system, this contains three options of actions: 1) downgrading the quality of the video data transmission at the source, if the quality requirements of the client are transgressed, 2) upgrading the quality of the video data transmission at the source, if the quality requirements of the client are surpassed, or 3) maintaining the current video data transmission at its current level. The decision of downgrading, upgrading or maintaining the data transmission is conducted by the leader agent, e.g., a user agent that conducts and controls the quality level of the video data transmission based on the QoS information of the connection, received from the monitoring user agent client and the monitoring user agent server. The implemented leader agent is an integration of the monitoring user agent server and the leader agent. The initial predilections, according to which the SQoS server adapts its video transmission, are the default preferences of the SQoS server. This initial method is then refined by the leader agent's calculations based on the information collected, or evaluations, of the monitoring user agent client and the monitoring user agent server.

Once the connection has been formed, the video data transmission from the SQoS server to the smart phone client may commence. The SQoS middleware system contains the centralized architecture, which comprises of distributed agents. These agents are the monitoring user agent client, the monitoring user agent server and the leader agent. In the centralized architecture, there is a central server, where the leader agent resides. In the centralized architecture, the centralized point contains the logic, viz. the leader agent, for discriminating the decision-making policy and conducting the level of the video data transmission of the network camera. This is the responsibility of the leader agent. In the centralized architecture, the leader agent resides in the SQoS server. The monitoring user agent client and the monitoring user agent server transmit their collected information to the leader agent individually.

The SQoS middleware system entails a model for collecting QoS information and conducting QoS levels: the passive, viz. unapt, model. In the passive model, the monitoring user agent client and monitoring user agent server transmit all the information they collect to the leader agent. The leader agent performs the calculations. In addition, based on the deduction of the calculations, the leader agent informs the network camera of the desired values regarding the level of transmission. The SQoS middleware system also entails a fashion of calculation, the proactive fashion. In the proactive fashion, the calculations of QoS are compared to the average values before the leader agent informs the network camera of modifying the quality level of video transmission. The leader agent must contain the average measurements in the proactive fashion.

The detailed devices utilized in the implementation of the SQoS middleware system are the Nokia 6680, as the smart phone, the Dell Optiplex GX150, as the SQoS server, and the Axis 213 PTZ, as the network camera. The smart phone is able to connect to the camera, and to the SQoS server containing the camera's stored video stream, via GPRS/WCDMA technologies. The network camera and the SQoS server are reachable from the global internet.

4.1 Architecture of the SQoS middleware system

QoS information surveyed by the monitoring user agent client and the monitoring user agent server is transited to the leader agent. The monitoring user agent client and monitoring user agent server transmits all the QoS information to the leader agent. The monitoring user agent client resides on end-user device, i.e., the smart phone, and the monitoring user agent server resides on the server, i.e., the SQoS server, which initially provides the stream. The leader agent is contained in a central server, i.e., the SQoS server. In the implemented SQoS middleware system, the monitoring user agent server and leader agent are integrated together, and hence denoted with the appellation leader agent. The leader agent performs the QoS calculations and decides of the change in the QoS. The leader agent always makes the ultimate order to the transmitting server, i.e., the network camera, about modifying the transmission's quality and quantity. The video transmissions are conveyed with the RTP protocol. Based on the bit-rate deviation assessment, the quality of the video transmission can be modified. The quality and the quantity of the transmittable video content can be refined. This is achieved by the leader agent ordaining the network camera through the HyperText Transport Protocol (HTTP) API of the network camera. The QoS of the connection is monitored by evaluating the RTP packets' bit-rate.

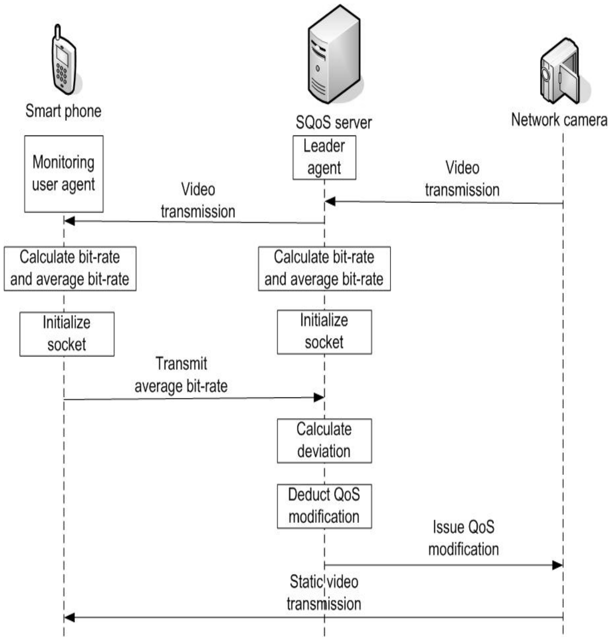


Figure 1. SQoS architecture and main sequence.

In figure 1, the initialization and the analysis of the QoS is performed by a leader agent residing at a central server. Figure 1 illustrates intricately the individual physical components of the SQoS middleware system. The smart phone contains the software component called the monitoring user agent. The SQoS server contains the software component called the leader agent, which entails the monitoring user agent server. The network camera begins to transmit video to the smart phone through the SQoS server. Both monitoring agents, the monitoring user agent client and the monitoring user agent server, begin to evaluate the connection. The monitoring agents, i.e., the monitoring user agent client and the monitoring user agent server, calculate the bit-rate and the average bit-rate. Then the monitoring user agent client and leader agent initialize their appropriate sockets for transmitting information between themselves. Once the sockets are open, the monitoring user agent transmits the average bit-rate to the leader agent. The monitoring user agent server transmits its average bit-rate internally to the leader agent. The leader agent calculates the deviation between the calculated bit-rate averages. If the calculations require it, the leader agent issues the network camera to scale the video transmission. The network camera proceeds to transmit the video with the new values. This video transmission persists with the same quality until the SQoS middleware system or the video transmission desists.

The initial setup is handled through the SQoS server. During the video transmission, the bit-rates collected by the monitoring user agent client and monitoring user agent server are transmitted to the central server's leader agent. The leader agent then informs the network camera to refine the video transmission’s quantity and quality pertaining to the smart phone and the wireless network.

Bit-rate measurements against time of the SQoS system were taken with Ethereal. Figure 2 illustrates the bit-rate in kilobytes per second (kbps) against time in the downscaling scenario. At the beginning, the SQoS system is already active and the video transmission commences. The monitoring user agent and server constantly transmit their surveyed bit-rate information to the leader agent. The bit-rate at the beginning of the video transmission to the Nokia 6680 was 64 kbps. When the leader agent received the information from the monitoring user agents, it concluded to downscale the video transmission bit-rate to 32 kbps. The monitoring user agents retain transmitting the bit-rate information. The frame size utilized was QCIF (Quarter Common Intermediate Format), viz. 176x144 pixels.

The upscaling scenario is illustrated in figure 3. The bit-rate at the beginning of the video transmission to the Nokia 6680 was 32 kbps. When the leader agent received the information from the monitoring user agents, it concluded to upscale the video transmission bit-rate to 64 kbps. The monitoring user agents retain transmitting the bit-rate information. The frame size utilized was QCIF (Quarter Common Intermediate Format), viz. 176x144 pixels.

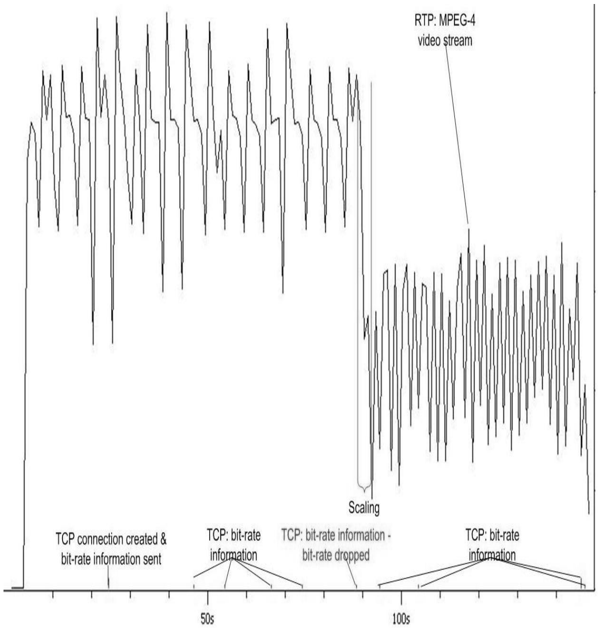


Figure 2, downscaling on the Nokia 6680, bit-rate against time.

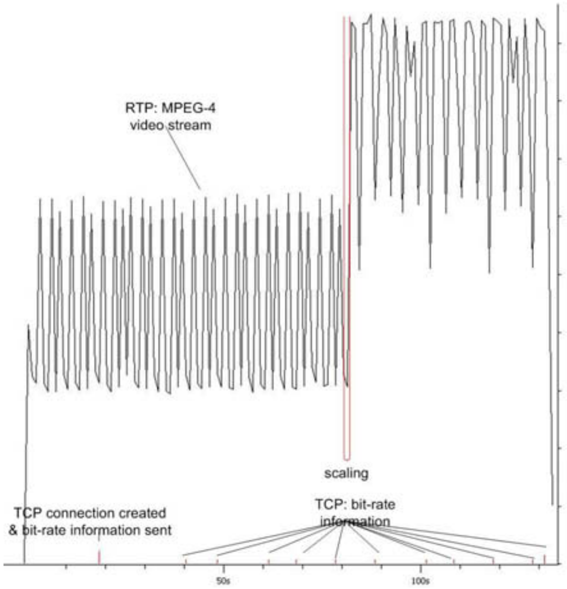


Figure 3, upscaling on the Nokia 6680, bit-rate against time.

4.2 Passive monitoring paradigm of the SQoS middleware system

The passive monitoring paradigm is used for processing information gathered by the monitoring agents. In the passive case, the monitoring client agent and the monitoring server agent perform a monitoring paradigm, which does not contain any decision-making or logical refining of the QoS. The decision-making and logical refining is conducted by the leader agent. The passive monitors merely gather the information and transmit it to the location of the leader agent, which performs the calculations and decision-making of regulating the QoS. Passive monitors do not perform any complex logical actions. Asgari et al. state that passive monitoring is chiefly utilized in traffic observation.

4.3 Calculation and deduction methods of the SQoS middleware system

The calculation method is based on assessing the bit-rate of the video being streamed. The solution uses network level traffic monitoring in measuring the amount of data transmitted through network adapter and a particular network connection at the monitoring user agent. Samples that describe the amount of data received are acquired once per second. The video bit-rate, measured in kilobytes per second, is calculated from these samples and the ten second average video bit-rate is transmitted to the leader agent once every ten seconds. The leader agent measures the actual video bit-rate and calculates ten seconds average, which is compared to the average value sent by the monitoring user agent, while an end-user is viewing the video. Depending on the degree of deviation between these two averages, the video quality is downscaled accordingly by the leader agent. The video quality upscale occurs when the average sent by the monitoring user agent maintains at a steady level for a certain period.

5. Conclusion

The QoS values pertaining to video transmissions over wireless networks are demanding. In this paper, we have illustrated the innovative theories of the Scalable Quality of Service middleware system's architecture, passive monitoring paradigm, and calculation and deduction methods. The Scalable Quality of Service middleware system has been implemented. The functionality of the middleware system indicates that the video transmission rate from a network camera to a smart phone over a wireless network has improved.

In comparison to Nahrstedt et al.'s four key aspects of their QoS-aware middleware system, the SQoS middleware system initiates itself according to a sound configuration and QoS adaptation runtime fluctuations are handled. Nahrstedt et al. proclaim that a middleware solution offers flexibility and adaptability for new applications. Abdelzaher et al. attest that a QoS negotiation method can be utilized for graceful degradation and that the flexibility and adaptability are required characteristics. The SQoS middleware system is targeted precisely at the middleware layer. Zhang et al. state that video applications should be aware of the network conditions over the wireless Internet. The SQoS interpolates the awareness and modification of these network conditions into the SQoS system middleware, which is distributed among discrete agents regarding the video transmission. Hence, the SQoS system middleware and its agent architecture is in line with Hayes et al.'s, Al-Jaroodi et al.'s, and Gunter & Brown's edicts of monitoring agents' suitability for monitoring, including QoS metrics, and their aptness for edifying a middleware infrastructure.

The constructed Scalable Quality of Service middleware system comprises of a monitoring user agent client, a monitoring user agent server and a leader agent. A network camera is used to convey video transmission to the smart phone. The video transmission passes through a Scalable Quality of Service server. The monitoring user agent client is located in the smart phone. The monitoring user agent server and leader agent are integrated together and it is located in the Scalable Quality of Service server. Both monitoring user agents survey the video transmission's bit-rate. The monitoring user agents send their evaluation to the leader agent. Then the leader agent derives whether to issue a request to the network camera to scale the Quality of Service values down or up.

The innovative theories, which we have created, are the Scalable Quality of Service middleware system's architecture, passive monitoring paradigm, and calculation and deduction methods. The intent of the theories is to optimize, or improve, the video transmission rate to a smart phone over a wireless network. The operability of the constructed SQoS middleware system prototype indicates that this endeavor is attained.

6. References

[1] Abdelzaher, T.F. et al., "QoS Negotiation in Real-Time Systems and Its Application to Automated Flight Control", IEEE Transactions on Computers, Vol. 49, No. 11, November, 2000, pp. 1170-1183.

[2] Al-Jaroodi, J. et al., “Middleware Infrastructure for Parallel and Distributed Programming Models in Heterogeneous Systems”, IEEE Transactions on Parallel and Distributed Systems, Vol. 14, No. 11, November, 2003, pp. 1100-1111.

[3] Asgari, A. et al., “Scalable Monitoring Support for Resource Management and Service Assurance”, IEEE Network, November/December, 2004, pp. 6-18.

[4] Coulson, G., et al., “Supporting Mobile Multimedia Applications Through Adaptive Middleware”, IEEE Journal on Selected Areas in Communications, Vol. 17, No. 9, September, 1999, pp. 1651-1659.

[5] Gunter, M. & Braun, T., “Internet Service Monitoring with Mobile Agents”, IEEE Network, May/June, 2002, pp. 22-29.

[6] Hayes, C.C., “Agents in a Nutshell - A Very Brief Introduction”, IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, January/February, 1999, pp. 127-132.

[7] Li, B. & Nahrstedt, K., “A Control-Based Middleware Framework for Quality-of-Service Adaptations”, IEEE Journal on Selected Areas in Communications, Vol. 17, No. 9, September, 1999, pp. 1632-1650.

[8] Nahrstedt, K., et al., “QoS-Aware Middleware for Ubiquitous and Heterogeneous Environments”, IEEE Communications Magazine, November, 2001, pp. 140-148.

[9] Pedreiras, P., et al., “FTT-Ethernet: A Flexible Real-Time Communication Protocol That Supports Dynamic QoS Management on Ethernet-Based Systems”, IEEE Transactions on Industrial Informatics, Vol. 1, No.3, August, 2005, pp. 162-172.

[10] Zhang, Q., et al., “End-to-End QoS for Video Delivery Over Wireless Internet”, Proceedings of the IEEE, Vol. 93, No. 1, January, 2005, pp. 124-134.

PUBLICATION 3

Scalable Video Transmission for a Surveillance System

Proceedings of the Fifth International Conference on
Information Technology: New Generations, ITNG 2008.
Las Vegas, Nevada, USA, 7–9 April, 2008.

Scalable Video Transmission for a Surveillance System

Tomi Räty
VTT Technical Research Centre of Finland

Lassi Lehikoinen
VTT Technical Research Centre of Finland

Francois Bremond
Institut National de Recherche en Informatique et en Automatique

Abstract

The Area of Interest (AoI) is a distributed scalable video transmission subsystem, for a surveillance system, which concentrates on decrementing the amount of video information transmitted to the end-user equipped with a mobile device. The video information is processed by the Video Surveillance Intelligent Platform (VSIP) to discriminate the essential images of the indoor area under stationary video surveillance. The AoI system analyzes the output of the VSIP's images and eXtended Markup Language (XML) image information. The AoI system is able to define and extract the essential information, e.g., a tracked individual, and it transmits only this image to the end-user. First, the AoI transmits the entire image of the indoor area to the mobile device of the end-user. Then, the AoI system transmits only the secluded tracked objects' images to the mobile device. The end-user's device portrays the scaled portrait images of the targeted object on top of the background image. The AoI system endeavors to decrease the size of the video images transmitted to a smart phone over a wireless network and to retain the comprehension of a tracking situation. The operability of the constructed prototype indicates that this endeavor is attained. The research is based on the constructive method of the related publications and technologies and the results are derived by the implemented AoI system.

Key Words- Multimedia communication, cooperative information systems, intelligent sensors, mobile communication, and multimedia systems.

1. Introduction

Video surveillance has become a ubiquitous aspect of the modern urban landscape [1]. Video surveillance is a significant market [2]. The systems must be network-connected, entail multiple cameras, and the complete system has to be reliable and robust [2]. Video surveillance applications must be real-time, which entail low delay and timing constraints for processing [2]. Target detection and tracking is a fundamental technology to develop real-world computer vision systems [3].

The Area of Interest (AoI) system comprises of the AoI server and the AoI client. The AoI server resides in a desktop and the AoI client resides in a surveillance personnel's mobile device. The AoI server utilizes images and eXtended Markup Language (XML) files, containing image information, that are received from Video Surveillance Intelligent Platform (VSIP). The images are snap-shots from a stationary camera of a surveyed indoor area. The XML files contain information about the tracked entities of the surveyed indoor area. This information includes the location of the tracked entity on the snap-shot image. During the initial transmission from the AoI server to the AoI client, one entire image of the surveyed indoor area is transmitted. This image is utilized as a background image by the AoI client and it is displayed on the security personnel's end-device. After the first image transmission, the AoI server distinguishes the tracked image from each image received from the VSIP. The AoI server extracts the tracked object from each image and the extracted image is transmitted to the AoI client. Upon reception of an extracted image, the AoI client displays the extracted image on the correct location, i.e., where the tracked object actually resides, of the background image. This procedure of extracting the tracked object by the AoI server, transmitting it to the AoI client, and displaying the extracted image at the correct location of the background image is conducted until the AoI system is shut down. By extracting the tracked object from the image and forming an extracted image decreases the amount of information required to be transmitted to the end-device.

The intent of the AoI system is to ultimately abate the quantity of information required to be transmitted to the security personnel while retaining all the required information for the security personnel to be fully cognizant about the surveyed indoor area. The operability of the constructed AoI system prototype indicates that this endeavor is attained.

The structure of this paper is the following. First a general overview of contemporary video monitoring is presented. Then a concise presentation of the VSIP is rendered. This is followed by a presentation of the implemented AoI system, containing detailed information regarding the structure of the AoI system and the transmission paradigm of the AoI system. The conclusion denotes the image subsidization in examples and summarizes the paper.

2. Video monitoring

Video monitoring typically deploys multiple video cameras, channeling video signals to a central monitoring room, where multiplexing is utilized to render a subset of the images to security personnel [1]. Modern video-based surveillance systems utilize real-time image analysis techniques for efficacious image transmission and event-based attention focusing [4].

Object tracking is an essential task for many applications in the region of video surveillance. Every detected object is tracked and their trajectories are analyzed to derive their movement in the scene. Detected objects are recognized and their behavior is analyzed to verify if state is potentially dangerous or normal. [5]

2.2. Situation awareness

The key factor to security is situation awareness, which requires information and spans multiple scales of space and time [6]. Multi-scale techniques evoke a completely novel region of research, in addition to challenges in performance modeling and evaluation [6]. Visual surveillance in dynamic scenes endeavors to detect and track certain objects from image sequences, and to understand object behaviors [7]. The goal of visual surveillance is to achieve the plenary surveillance task as automatically as possible [7].

2.3. Middleware

Enabling a group of video surveillance algorithms to cooperate in the monitoring of a large surveillance network presents substantial challenges [8]. Middleware can assist with the general aspects of video surveillance network construction, containing support for communication and computation [8]. The drawbacks in many current video surveillance systems contain lower Quality-of-Service (QoS) in video transmission. [9]

Researchers concentrate mainly on the vicissitude of content comprehension, e.g., detecting and tracking. They have not heeded the scalability of video surveillance systems. They typically utilize a centralized architecture and posit the required system resources. [10]

Digital surveillance systems disclose restrictions regarding delay and visual quality that pose demands on the video codec. Flexible composition of the compressed video data is required. In a large surveillance system, the digital network enables interconnected LANs with distinct bandwidths and QoS. [11]

3. Video Surveillance Intelligent Platform

A demanding problem in the domain of computer vision and artificial intelligence is video comprehension. The first step utilizes typically extensive usage of methods for data analysis while the second step conducts structural analysis of the symbolic data collected at the antecedent step, as Figure 1 illustrates. [12]

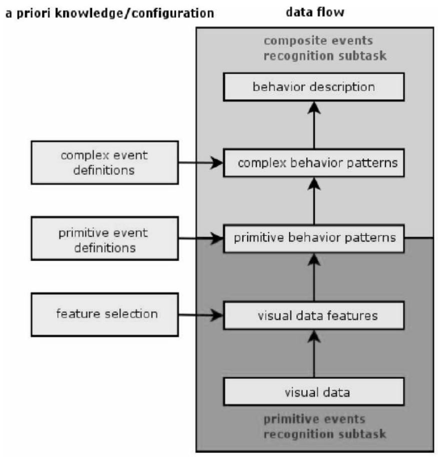

a priori knowledge/configuration
data flow
composite events recognition subtask
behavior description
complex event definitions
complex behavior patterns
primitive event definitions
primitive behavior patterns
feature selection
visual data features
visual data
primitive events recognition subtask

Figure 1. A generic architecture of a video comprehension system. [12]

This approach is available as a platform for image sequence comprehension named VSIP. VSIP has been developed by the research group ORION at INRIA (Institut National de Recherche en Informatique et en Automatique), Sophia Antipolis. VSIP is a generic environment for amalgamating algorithms for processing and analysis of videos which enables to combine and exchange miscellaneous techniques at different stages of the video comprehension process. VSIP is oriented to assist developers depicting their own scenarios and systems capable of monitoring behaviors, dedicated to specific applications. [12]

VSIP elicits primitive geometric features, such as areas of motion. Objects are then recognized and tracked. At the second level the events, in which the detected objects participate, are discriminated. To perform this task, an event description language is used. [12]

4. Area of Interest system

The intention of the AoI (Area of Interest) system is to transmit merely the important dynamic video image information, gathered from a surveillance point by a stationary camera, to a mobile device. The static information, a.k.a. the environment background of the video images, is transmitted only once during inception. The system sequesters the essential objects from .jpg images and sends the separated objects to a mobile device. In the Figure 2, a moving object is traced from a surveillance camera perspective. The traced object is extracted from the background as depicted in the Figure 3. Finally, the traced object and its deployment pixel coordinates, meaning the coordinates of the traced object in the whole image, are transmitted to a mobile device which displays it at the right location on the screen. The end-device view of the extracted image placed on the background image is presented in Figure 4. When the AoI system is executing, the outcome at the mobile device is a continuous video stream in which only the traced dynamic areas are received from the server via a network connection and merged into the background. The mobile device is the Nokia N95 phone, equipped with S60 3rd Edition SDK for Symbian OS 9.2, Supporting Feature Pack 1. The server's software is implemented in Windows XP OS utilizing Microsoft .NET Framework SDK v2.0.

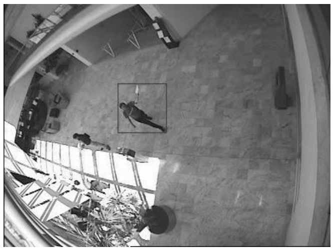


Figure 2. A traced dynamic object segregated by a square.

Figure 3. A traced dynamic object extracted from the static background.

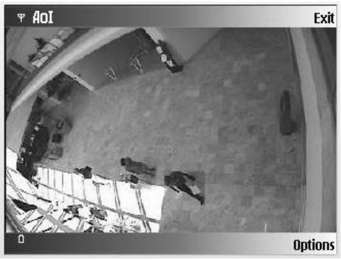


Figure 4. View from the end-device. The object images are placed on the background image.

4.1. The server structure of the AoI system

The software components of the AoI server are delineated in Figure 5. The main components are the following: 1) AoIMain, 2) ImageSender, 3) XMLParser and 4) ImageProcessor. The AoIMain component is the main executable, which has the responsibility of controlling all the components. The ImageSender component is employed for establishing TCP/IP socket communication with AoI client(s) and sending images through the connection. The XMLParser component is responsible for parsing the .xml file containing object tracing information. The ImageProcessor component has the responsibility of separating traced objects from the background environment. The separation is exerted with the .jpg images.

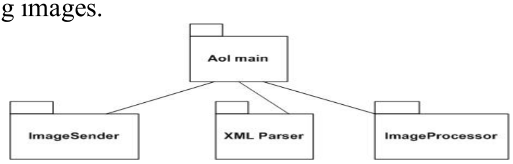


Figure 5. The component diagram of the AoI server.

After the activation of the AoI server, initialization procedures are executed. All the appropriate class instances are created and a listening socket is established for a remote client connection. Once a client has connected to the server, the server's 'Start' method is called and the main functionalities begin. If the server is transmitting objects for the first time, the environment background image is sent by the server. The server parses the frame elements from the .xml file, which contains the object tracing data. After that, the image objects are respectively separated from an image file. Next, the separated objects are transmitted to the client through the socket connection. This procedure is repeated until the .xml file is processed completely. When the file is processed to the end, the AoI server is suspended and de-initialized.

4.2. The client structure of the AoI system

The software components of the AoI client are rendered in Figure 6. The main components are the ensuing: 1) UI, 2) ImageConverter, 3) ImageViewer and 4) SocketCommunicator. The UI component is used for user interactions. Additionally, the UI component controls the ImageConverter, ImageViewer and SocketCommunicator. The ImageConverter converts the 8-bit image descriptors, received from the server, into bitmaps which are displayable on the end-device's screen. The ImageViewer component is responsible for joining the bitmaps of the separated image objects and the background environment bitmap into one merged view, which may be displayed on the screen of the end-device. The SocketCommunicator component is employed in receiving images and control messages from the AoI server via TCP/IPv4 socket connection.

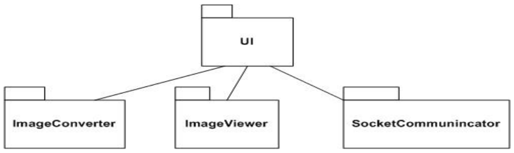


Figure 6. The component diagram of the AoI client.

After the launch of the AoI client application, initialization procedures are exerted. All the appropriate class instances are created and a socket connection to the AoI server is formed. Once the client has connected to the server, it receives a background image and an instruction message indicating if the background was successfully transmitted. Then the background is displayed on the screen. Next, the client receives separated image objects and coordinate instruction messages from the server. The separated images are displayed on top of the background image. When the client receives an instruction message indicating the frame was sent successfully, the client clears the screen from the previous separated objects, and the client prints the background image again to the screen. If a human user presses the "Exit" button from the user interface, the application exits, de-initializes and shuts down.

4.3. The AoI server's main execution sequence

The AoI server reads the .xml file, separates traced dynamic objects from the images and transmits objects to the AoI client. See Figure 3 for an example of a transmitted object. As preconditions to the AoI server execution sequence, the AoI client has already connected to the server and the AoI server has XML file with the object tracing data and the corresponding .jpg image files.
Description of sequence's events, illustrated in Figure 7:
1. The Start() function is called when an AoI client has connected to the server.
1.1 The OpenXMLFile() function opens the .xml file for reading operations.
1.2 The ParseNextXMLFrameContent() function parses next frame element from the .xml file and stores the content to a CXMLFrameContent object.
1.2.1 The CXMLFrameContent object reference is returned.
1.3 The SeparateObjectsFromImage() function is called. The function secludes traced dynamic objects from a .jpg image. The function receives the CXMLFrameContent object reference as a function parameter.
1.4 The SendObjectsToClient() function transmits separated dynamic objects images to the AoI client. Additionally, after transmitting an image, the server transmits a deployment coordinate instruction to the client. When all the objects and coordinate instructions are transmitted, the server sends an instruction to the client denoting that all the objects of the frame are sent.

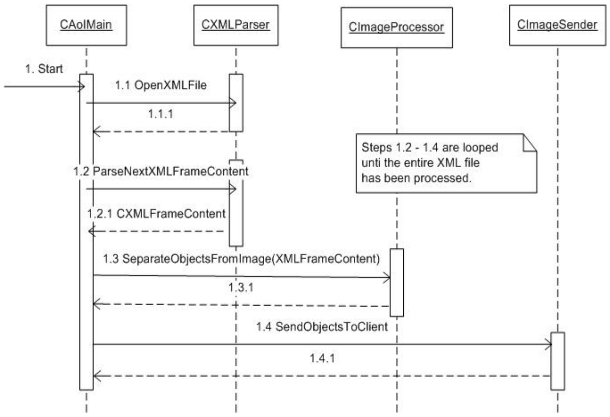

CAoIMain
CXMLParser
CImageProcessor
CImageSender
1. Start
1.1 OpenXMLFile
1.1.1
Steps 1.2 - 1.4 are looped unti the entire XML file has been processed.
1.2 ParseNextXMLFrameContent
1.2.1 CXMLFrameContent
1.3 SeparateObjectsFromImage(XMLFrameContent)
1.3.1
1.4 SendObjectsToClient
1.4.1

Figure 7. The AoI server execution sequence diagram.

The outputs of the AoI server are the image files of the separated dynamic objects. The separated image files are deleted from the AoI server's file system after the sending has executed successfully. There are two notable exceptions regarding: 1) an addition to the step 1.4, when transmitting image objects for the first time to the AoI client, the background environment image and a notification instruction of successful sending are sent. After the first time, the background is not sent; and 2) if the .xml file in not processed completely, sequence restarts from the step 1.2 after the execution of step 1.4. If the .xml file is processed entirely, Start() method returns.

4.4. The AoI server's main execution sequence

The AoI client receives TCP packets from the AoI server. The packets are parsed and addressed appropriately. First, a background image is received. Then an arbitrary amount of separated images are received. In the steps 1 – 2.1.2, a background image is received from the AoI server. These steps are performed only once. In the steps 3 – 4.1.2, the separated image objects are received from the server with appropriate coordinate instruction messages. These steps are looped until the AoI server stops, i.e., the xml file has been read to the end. As a precondition to the AoI client execution sequence, the AoI server is running.

Description of sequence's events, illustrated in Figure 8:
1. The MessageReceived() callback is called by the CCommunicator class instance when a message is received from the connected socket. When receiving messages for the first time, the content of the message is added to the background image buffer. Since the background image can be large, the MessageReceived() function is called often before the whole image is completely in the image buffer. The AoI server transmits a notification when the background image is successfully sent to a client. Upon reception of the notification, step 1.1 is executed.

1.1 The ConvertDesL() function is called to convert the serialized 8-bit background image buffer into bitmap image which can be displayed on the devices screen.
2. The ConversionComplete() function is called after the image conversion is ready.
2.1 The SetBitmap() function sets the converted bitmap for the CAoIAppView object.
2.1.1 The Draw() function is called. It displays the converted bitmap, in this case the background image, on the end-device's screen. Now the background is displayed successfully on the screen. Then the server begins to send the separated image objects.
3. The MessageReceived() callback is called by CCommunicator class instance. It is called as many times as required until the whole separated image is stored into the image buffer. When the AoI server has transmitted the whole separated image, it transmits a coordinate instruction message to a client. This message contains the deployment coordinates for the separated images and the image length for a validation check at the client side.
3.1 The ParseXCoordinate() is called to parse the X coordinate from the coordinate instruction message. This value is stored into a variable.
3.2 The ParseYCoordinate() is called to parse the Y coordinate from the coordinate instruction message. This value is stored into a variable.
3.3 The ParseImageLength() is called to parse the image length from the coordinate instruction message. This value is stored into a variable.
3.4 The SetImageTopLeftX() function is called to set the X coordinate value for the CAoIAppView class instance.
3.5 The SetImageTopLeftY() function is called to set the Y coordinate value for the CAoIAppView class instance.
3.6 The ConvertDesL() function is called to convert the serialized 8-bit image buffer into a bitmap image which can be displayed on the devices screen.
4. The ConversionComplete() function is called after the image conversion is ready.
4.1 The SetBitmap() function can be called now, when the coordinates of the separated image has set to the CAoIAppView class instance.
4.1.1 The Draw() function is called. It draws the separated image object at the correct coordinates. The image is added onto the background image.

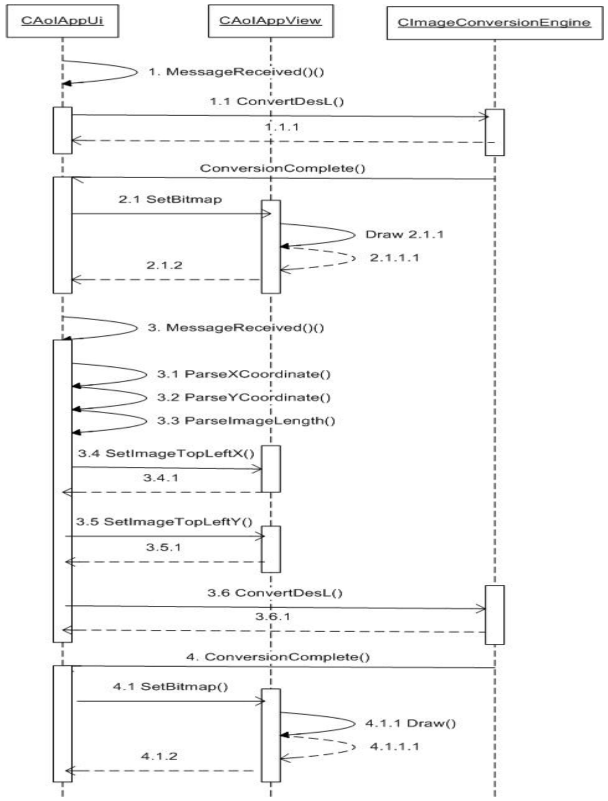


Figure 8. The AoI client execution sequence diagram.

There are two notable exceptions regarding: 1) if the AoI server is not running, then the client initialization fails and application does not start; and 2) in step 3, if the AoI client receives an instruction message indicating that a whole frame content has been transmitted, the background image is printed on the display again before adding the separated objects into it. In this manner, the screen is "cleared" from the previous separated objects.

5. Conclusion

Video surveillance is an important branch in the field of surveillance. With the utilization of advanced video surveillance tools, such as VSIP, it is possible to distinguish images of tracked objects. By abating the amount of image information that needs to be transferred, the images can be transmitted faster to the end users, e.g., surveillance personnel. We have illustrated the implemented design and communication how this endeavor is attained with the AoI system.

The information and structure of the AoI system was modeled on recent journals and conference papers regarding video surveillance and monitoring. There are theories demanding real-time reactivity, low delay and timing constraints from Desurmont et al. and the importance of situation awareness according to Hampapur et al. which includes challenges in performance modeling and evaluation. The AoI system attempts to reduce the real-time challenges by subsiding the amount of image information that needs to be transmitted. May et al. deem that there is a requirement for the flexible composition for the compressed video data, the AoI system decreases the amount of image information transmitted. The AoI applies to the May et al.'s restrictions of surveillance applications regarding delay, complexity, security, visual quality and QoS predicaments by scaling the image size. This also applies to the drawbacks of lower QoS in video transmission declared by Yan et al. Korshunov et al. state that enough research hasn't been contributed on scalability of video surveillance systems. These typically utilize a centralized architecture and posit availability of all the required system resources, such as computational power and network bandwidth. The AoI system endeavors to transmit as little image information as possible while retaining the quality of the prominent image information.

The AoI system comprises of the AoI server and the AoI client. The AoI server processes the images received from the VSIP tool and accompanied with the VSIP tool's XML an extraction of the tracked object is performed. The AoI server transmits the images of the tracked objects to the AoI client. The AoI client receives the entire background indoor image in the first transmission from the AoI server. After the first transmission, the AoI server only transmits the images of the tracked objects. The AoI client updates the tracked object image onto the initially received background image. The image sizes of the transmitted tracked objects are subsided in comparison to their entire and original image sizes. For instance, the entire size of Frame003 is 21 814 bytes and the size of the images containing the extracted objects are 924 and 690 bytes. In Frame103, the size of the complete image is 21 834 bytes and the size of the image containing the extracted objects is 624 bytes. In Frame186, the size of the complete image is 21 994 bytes and the size of the images containing the extracted objects are 498, 644, and 736 bytes respectively.

The intent of the AoI system is to ultimately subside the quantity of information required to transmit the security personnel while retaining all the required information for the security personnel to be fully aware about the surveyed indoor area. The operability of the constructed AoI system prototype indicates that this endeavor is attained.

6. Acknowledgement

We would like to sincerely thank the EU-CAVIAR project for the utilization of the images and XML files that were employed in this publication.

12. References

[1] Makris, D. and Ellis, T.: Learning Semantic Sense Models from Observing Activity in Visual Surveillance, *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 35, No. 3, June 2005, pp. 397-408.

[2] Desurmont, X., Bastide, A., Chaudy, C., Parisot, C., Delaigle, J.F., and Macq, B.: Image analysis architectures and techniques for intelligent surveillance systems, *IEE Proc.-Vis. Image Signal Process.*, Vol. 152, No. 2, April 2005, pp. 224-231.

[3] Matsuyama, T. and Ukita, N.: Real-Time Multitarget Tracking by a Cooperative Distributed Vision System, *Proceedings of the IEEE*, Vol. 90, No. 7, July 2002, pp. 1136-1150.

[4] Foresti, G.L., Micheloni, C., Snidaro, L., Remagnino, P., and Ellis, T.: Active Video-Based Surveillance System, *IEEE Signal Processing Magazine*, March 2005, pp. 25-37.

[5] Micheloni, C., Foresti, G.L., and Snidaro, L.: A Network of Co-operative Cameras for Visual Surveillance, *IEE Proc.-Vis. Image Signal Process.*, Vol. 152, No. 2, April 2005, pp. 205-212.

[6] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S., Senior, A., Shu, C.-F., and Tian, Y.L.: Smart Video Surveillance, *IEEE Signal Processing Magazine*, March 2005, pp. 38-51.

[7] Hu, W., Tan, T., Wang, L., and Maybank, S.: A Survey on Visual Surveillance of Object Motion and Behaviors, *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 34, No. 3, August 2004, pp. 334-352.

[8] Detmold, H., Dick, A., Falkner, K., Munro, D.S., van den Hengel, A., and Morrison, R.: Middleware for Video Surveillance Networks, MidSens'06, November 27-December 1, 2006, Melbourne, Australia.

[9] Yang, H., Xie, L., and Xie, F.: Research on Cluster Remote Video Surveillance System, 2006 IEEE International Conference on Industrial Informatics.

[10] Korshunov, P. & Ooi, W.T.: Critical Video Quality for Distributed Automated Video Surveillance, MM'05, November 6-11, 2007, Singapore.

[11] May, A., Teh, J., Hobson, P., Ziliani, F., and Reichel, J.: Scalable Video Requirements for Surveillance Systems, 2004, The Institution of Electrical Engineers, printed and published by the IEE, Michael Faraday House, Six Hills Way, Stevenage, SG1, 2AY.

[12] Bremond, F., Thonnat, M., and Zuniga, M.: Video Understanding Framework For Automatic Behaviour Recognition, Behaviour Research Methods, Volume 28, Number 3, August 2006.

PUBLICATION 4

Sensor Data Collection of the Single Location Surveillance Point System

Proceedings of the Seventh International Conference
on Computer and Information Science, ICIS 2008.
Portland, Oregon, USA, 14–16 May, 2008.
© 2008 IEEE.
Reprinted with permission from the publisher.

Sensor Data Collection of the Single Location Surveillance Point System

T. Räty, J. Oikarinen, M. Nieminen and M. Lindholm

Abstract—**The Single Location Surveillance Point (SLSP) is a distributed multi-sensor surveillance software system. It contains an arbitrary amount of sensors that collect readings from a single location, which is the surveillance point. The SLSP system contains the following realized sensors: a fingerprint sensor, a video camera, an audio sensor, and a network analyzing monitor. The sensors are located in an indoor region. Each sensor automatically collects information from its environment. Each sensor automatically routes its crude sensor data to a session server, which handles the connections among the components. The session server conveys the crude sensor data to the logical decision making service. The logical decision making server (LDMS) automatically derives the situation at the surveillance point based on the received sensor data. The intention is to deduct the situation which is transpiring in the surveyed area based on the received crude data from the sensors. By deriving the situation of a surveyed area, the surveillance personnel may utilize refined information cogent to occurring events of the surveyed area. This branch of the SLSP intends to facilitate the collection of data from a surveillance point and decrement the amount of superfluous information and rendered to the surveillance personnel, by acquiring automatically sensor data and providing automatically derived information to the surveillance personnel's end-device. The operability of the constructed prototype indicates that this endeavor is attained. The research is based on the constructive method of the related publications and technologies and the results are derived by the implemented branch of the SLSP system.**

Index Terms—**Multi-sensor systems, intelligent systems, and middleware architectures and techniques**

I. Introduction

RECENT progress in computing, communication, and sensor technology are inciting the development of multiple new applications [1]. This trend is apparent in pervasive computing, sensor networks, and embedded systems [1]. Homeland security is an inherent concern for governments worldwide, which must protect their people and the critical infrastructures [2]. Information technology can assist in mitigating risk and enable effective responses to disasters of natural of human origin. [2]

The creation of a distributed automatic surveillance system by developing multi-camera or multi-sensor surveillance systems, and fusion of information procured across cameras, or by creating an integrated system is also an active region of research. A distributed multi-agent approach may provide numerous benefits. Intelligent co-operation between agents may enable the use of less expensive sensors, therefore a large number of sensors may be deployed over a larger area. Robustness is augmented, because even if some agents fail, others remain to perform the mission. Performance is more resilient, there is a distribution of tasks at miscellaneous locations between groups of agents. [3]

The increasing demand for safety and security has resulted in more research in constructing more efficacious and intelligent automated surveillance systems. A future challenge is to develop a wide-area distributed multi-sensor surveillance system which has robust, real-time computer algorithms able to execute with minimal manual reconfiguration on variable applications. [3]

The SLSP (Single Location Surveillance Point) system is a distributed multi-sensor surveillance system. It includes multiple sensors, constituting of a fingerprint sensor, a video camera, an audio sensor, and a network analyzing monitor. The sensors are located in an indoor area for surveillance. Each sensor acquires from its environment and transmits the crude data to the session server. The fingerprint sensor transmits access information each time a fingerprint is read. The video camera transmits visual data of the surveyed area. The audio sensor transmits aural data of the surveyed area. The network analyzing monitor views the SLSP network and transmits data apposite to the network and the devices attached to it. The session server handles all the connections among the components of SLSP. The session server transmits the received data from the sensors to the LDMS (Logical Decision Making Server). The LDMS automatically deducts the surveillance point's situation predicated on the data it receives from the sensors routed by the session server. The deductions are transmitted to the surveillance personnel's end-device.

The intent of the SLSP system is to ultimately collect automatically sensor data and transmit it to the LDMS for automatic logical decision making of the surveyed area for security personnel. The operability of the constructed SLSP system prototype indicates that this endeavor is attained.

Manuscript received November 22, 2007.

T. Räty is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, phone: +358 20 722 2131; fax: +358 20 722 2320, e-mail: tomi.raty@vtt.fi

J. Oikarinen is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: johannes.oikarinen@vtt.fi

M. Nieminen is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: mikko.nieminen@vtt.fi

M. Lindholm is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: mikko.lindholm@vtt.fi

The structure of this paper is the following. First a general overview of contemporary surveillance systems is presented. Then a concise presentation of the multi-sensors is rendered. A general presentation of logical decision making is evoked next. This is ensued by a presentation of the implemented branch of SLSP system, containing detailed information regarding the structure of the automatic collection of sensor data, transmission of sensor data and logical deductions performed on the sensor data. The conclusion summarizes the paper.

II. Surveillance Systems

Ameliorating the smart cameras with additional sensors could transform them into a high-performance multi-sensor system. By consolidating visual, acoustic, tactile, or location-based information, the smart cameras become more sensitive and can transmit more precise results. This makes the results more applicable widely. [1]

The increasing demand for safety and security has resulted in more research in constructing more efficacious and intelligent automated surveillance systems. The fundamental goals that are to offer good scene comprehension, surveillance information and utilizing low cost standard components. Spatially distributed multi-sensor environments render interesting possibilities and challenges for surveillance. Recently, there has been some investigation of data fusion techniques to tolerate with information sharing pertaining to erudition resulting from different types of sensors. [3]

Regazzoni et al. concentrate on solutions that are predicated on a stronger integration of techniques for multi-sensor data acquisition, communications, and processing. Especially the problem of remote surveillance of unattended environments has received enlarging attention, for instance monitoring of indoor and outdoor environments like banks, supermarkets, car etc. [4]

Multi-sensor systems can capitalize from processing the same type of information obtained by sensors of different type, e.g., video cameras, microphones, etc., on the same monitored area. Appropriate processing techniques and new sensors offering the real-time information associated to different scene characteristics can assist both to augment the size of monitored environments and to enhance performances of alarm detection in regions monitored by more sensors. [4]

A. Architecture and middleware

Typically, control in an automatic surveillance system has been centralized, with a topology or configuration in which sensors are branches of a central node. The captured data is transmitted to the central node where it is processed and decisions on how to act are committed. This architecture is conceptually simple, but inflicts many dilemmas related to scalability, bottlenecks and robustness, which are substantial regarding the hierarchical rigidity of this architecture to unanticipated variations. [5]

In the contemporary generation of surveillance systems, in which a multiple asynchronous and miscellaneous sensors are employed, assimilation of the information procured from them to derive the events from the environment is an important and challenging research problem. The issue of information assimilation is vital, because the information procured from multiple sources when assimilated offers more precise inferences of the environment than individual sources. [6]

Most of the new research activities in surveillance are exploring larger dimensions, such as distributed video surveillance systems, systems with multimedia streams, including audio, video, and sensors signals, surveillance and biometric systems [7]. Due to the availability of more advanced and powerful communications, sensors, and processing units, the architectural choice can potentially become extremely variable and flexibly customized to procure a vied performance level. The system architecture commences to delineate a key factor. [4]

A further evolution is the integration among surveillance networks predicated on sensors of either different types [4]. Enabling a group of video surveillance algorithms to cooperate in the monitoring of a large surveillance network presents essential challenges [8]. Middleware can assist with these general aspects of video surveillance network construction, containing both support for computation and for communication. [8]

Typically, middleware offers miscellaneous transparencies that help to simplify application development, including distribution/location transparency. In context-aware systems, middleware is also needed to ease adaptation predicated on context information. This requires support for the ensuing tasks: 1) collecting context information from sensors and other sources, 2) management and integration of context information, 3) reasoning support and context querying, and 4) support for decisions regarding adaptations. [9]

III. Multi-sensors

There are immediate needs for automated surveillance systems in commercial, military applications, and law enforcement. Video data currently are employed only retrospectively as a forensic tool, thus losing its primary benefit as an active, real-time medium. What is required is an incessant 24-hour monitoring of surveillance video to alert security officers to a burglary in progress, while there still is time to circumscribe the criminal offence. [10]

A. Sensor and event definitions

A sensor refers to the processing directly consorted with a physical transducer (camera, microphone, fire detector, etc.) or actuator and is directly meaningful to an operator and employed in a geographical representation of the site. A sensor can measure or detect one or multiple events. Depending on the processing capability of the sensor, such events can be either simple, e.g., the door has been opened, or complex, e.g., the person at coordinates x,y has been there for 22 minutes. Simple sensors will be typically associated with only one event, e.g., fire detected, but it is feasible for a sensor to be capable of detecting a number of events. Events can be of different types, such as an alarm (an incident has occurred), measurement (a continuous quantity such as the amount of people in an area), or status (system information such as power failure). Finally, the concept of event groups entails the idea that what a user might consider as an event can be a combination or aggregation of evidence captured by one or multiple sensors. [11]

B. *Multi-sensors in data fusion*

Typically, surveillance systems are composed of numerous sensors, e.g., camera, radar, to obtain data from each target in the environment [12]. These systems encounter two types of dilemmas: 1) fusion of data, It is related to the combination of data from discrete sources in an optimal manner, and 2) management of multiple sensors, it presumes that the previous predicament is solved, and it conducts optimizing the global management of the joint system through the application of individual operations in every sensor [12]. Networked sensors can collaborate to process and conduct deductions from the obtained data and provide the user with access to continuous or selective observations of the environment [13].

IV. Logical Decision Making

Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources [14]. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application [14]. One, concise way to define sensor and data fusion is the following. Sensor Fusion is "Data Fusion from Multiple Sensors (same or different sensor types)". Data Fusion is "Combining information to estimate or predict the state of some aspect of the world". [15]

Applications for multi-sensor information fusion (IF) require analysis of how these systems will be deployed and utilized. Increasingly complex scenarios arise, requiring more intelligent and efficient reasoning strategies. Essential to information reasoning is decision making (DM) which requires pragmatic knowledge representation for user interaction. [16]

Situation assessment is a concept of how people become aware of things happening in their environment [16]. Design of decision support systems requires an understanding of both the fusion processes and the DM processes. Important aspects of fusion include timeliness, mitigation of uncertainty, and output quality [16]. Data fusion involves the use of multiple data, often from multiple sources, to estimate or predict the state of some aspect of reality, e.g., estimation/prediction of the state of individual(s), which are treated as if they were independent of the states of other entities [17].

The goal of multi-sensor fusion is to achieve inferences about the observed environment or situation that cannot be achieved by a single sensor or source of information. Information about the observed situation is combined to achieve high-level inferences. Multiple techniques may be used to achieve these high-level inferences [18].

Multi-sensor data fusion can provide solutions to problems that are characterized by intensive and diverse sensor information. It can be defined as the process of integrating raw and processed data into some form of meaningful inference that can be used intelligently to improve the performance of the system beyond the level that any one of the components of the system separately could achieve [19].

V. An Introduction to the SLSP system

The information collection of the Single Location Surveillance Point comprises of two individual domains as in Figure 1. These two domains are 1) the Surveillance Domain, which comprises of the fingerprint sensor, video recorder, audio sensor and the network activity monitor, and 2) the Security Administration and Surveying Domain, which comprises of the session server to which the sensors transmit their information, the LDMS and the security personnel's end-device, which is the Nokia N95 smart phone.

The data flow from the sensors of the Surveillance Domain is primarily the ensuing: 1) the crude sensor data is conveyed from the sensors to the session server, 2) the session server transmits all the crude sensor data it receives to the LDMS, 3) the LDMS performs the its automatic logical calculations, based on the crude sensor data provided by the sensors and 4) the logical calculations are routed by the session server to the security personnel's end-device.

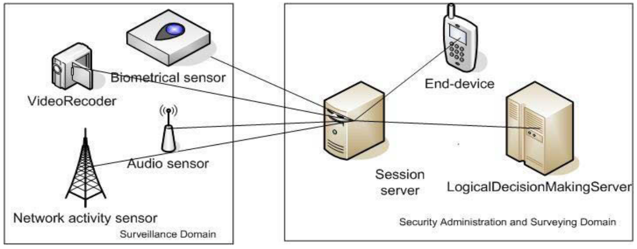


Figure 1, the information collection architecture of the SLSP system.

The information collection of the SLSP comprises of:

1. the sensors gathering data from their environment, composing of the fingerprint sensor, audio sensor, video recorder, and a network activity monitor,
2. the session server, which receives data from the sensors and delivers the data to the LDMS,
3. the LDMS, which receives the sensor data from the session server and conducts logical deductions based on this information, and
4. the end-device of the security personnel, to which the deductions of the LDMS can be transmitted. In the SLSP system, the end-devices utilized were Nokia N95 smart phones.

VI. The sensors of the SLSP system

The utilized sensors of the SLSP system are independent or proprietary devices. They compose of the fingerprint sensor, the audio sensor, the video recorder, and the network activity monitoring sensor. Elaborate descriptions of the sensors are presented in the ensuing subchapters.

A. *The fingerprint sensor*

The SLSP system contains a fingerprint sensor, which is Deltabit's Gatekeeper fingerprint recognition product. The product registers fingerprints at a door, and informs of access permitted or access denied based on the access rights predicated on the pertaining fingerprint. The fingerprint sensor distributes the access rights of an individual's fingerprint to the session server.

B. *The audio sensor*

The SLSP system entails an audio sensor, which monitors the environment for threatening sound events (audio events that exceed a threshold level indicating an alarming event) and provide the bearing of the sound location. SLSP's audio sensor comprises of the implemented components of the audio interface recorder, signal processing library, and applications and libraries for transport protocol and streaming server functionality. The surveyed data of the audio sensor, comprising the volume of the audio event and the bearing of the direction from which the audio event occurred, is distributed to the session server.

C. *The video recorder*

The video recorder conveys the video transmission to the session server. The video recorder is the Axis 213 PTZ video camera. The video recorder conveys the video transmission to the session server. This is the crude sensor information of the video recorder.

D. *The network activity monitoring sensor*

The SLSP system's network activity monitor is Nethawk's M5 traffic analyzer. The network activity monitor surveys all the IP-level network traffic inside the SLSP. The network activity monitor observes the data, both the amount and type, traversing in the network and the devices. The devices generating the traffic must be registered to the SLSP system, unregistered devices are not allowed. To monitor the devices, the network activity monitor implements a "watchdog" property, which ensures that the device transmitting is a valid device. If not, the device's MAC address is reported to the session server. The watchdog property relies on capturing DHCP protocol packets that are utilized in the network to allocate an IP address for the device. Each device connecting to the network transmits DHCP request to the DHCP server providing IP address in a response. To implement the watchdog property each DHCP message is analyzed and cross-referenced to the MAC address table in the network activity monitor. The report is transmitted to the session server.

VII. THE SESSION SERVER OF THE SLSP SYSTEM

The session server consists of a single component, the SessionServer, containing the main logic of the session server. To communicate with the servers and the sensors in SLSP system, the session server uses network libraries in communication contained inside the NetworkInterfaces component. The component diagram of the session sever is delineated in Figure 2. To procure the video, from the video recorder, the Darwin streaming server is used. The Darwin streaming server is component of the session server but it operates independently. The purpose of the VLC video player is to handle the storing of the video stream into permanent file repository. The interface of the VLC video player is a command line interface, implemented with system library of C programming language. Both VLC and Darwin streaming server are separate applications and they are used to obtain and deliver the video stream and recorder video clips. The data that is conveyed with TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) between sensors and LDMS are in textual format. The video data is conveyed over RTSP (Real-Time Streaming Protocol) / RTP (Real-time Transport Protocol) and the conveying is handled with Darwin Streaming Server.

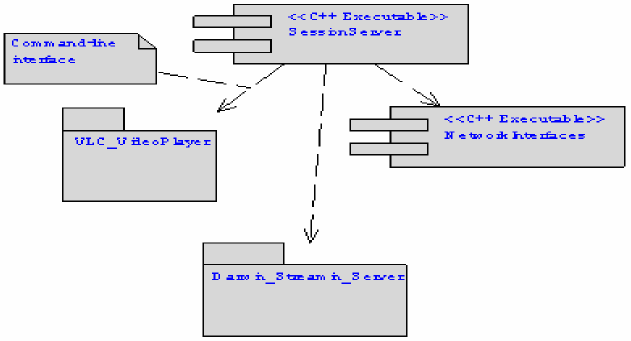


Figure 2, the session server component diagram.

A. *The streaming in the Session Server*

The streaming concerning VLC and Darwin functions in the manner depicted in figure 3. The VLC and Darwin are subcomponents of the session server but are separate applications. The VLC is utilized to store the video clips into permanent file repository. The sequence begins when a request to store video clip is received. Then the VLC starts to stream and store the video stream. Once the video stream is stored, the created video clip is saved under Darwin for streaming. Darwin handles the delivery of live video stream to the end-devices. Darwin receives media requests from the end-devices and transmits the video stream from the video recorder to the end-device.

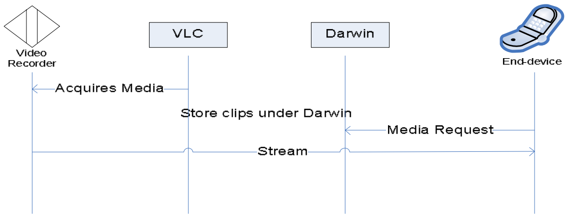


Figure 3, streaming from a video recorder to the end-device.

B. *The main functionality the Session Server*

The session server's main functionality is presented in Figure 4. CSessionServerMain is the main class of the session server. The class conducts the initial startup procedures and handles messages that were received through interfaces. The CSensorInterface class handles the socket level connection to a sensor. A ThreadSocket is run in a separate thread. The ThreadSocket class implements a socket that is capable of receiving and sending through one socket interface. The CDeviceInterface class handles the socket level connection to an end-device. A ThreadSocket is also run in a separate thread.

The CSessionServerMain contains a main loop performing receive buffer checks, message handling and pinging sensors and servers periodically. Also CSensorInterface and CDeviceInterface classes function inside this main loop controlled by the CSessionServerMain. Once a remote interface, either for an end-device or a sensor, is initialized and set to listen, another loop is initiated in each ThreadSocket. This loop performs the actual receiving and sending operations periodically by transmitting everything from the send buffer and conveying received messages to receive buffer. On the sensor side, an equal amount of ThreadSocket classes (sockets) are created to each sensor. A socket is created for every end-device and LDMS.

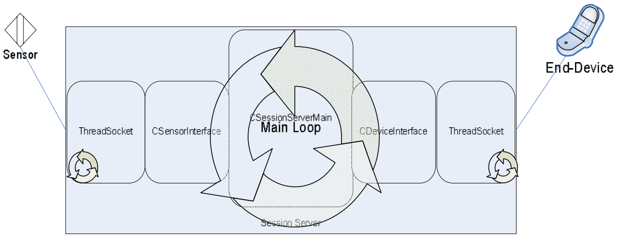


Figure 4, main functionality of the session server.

VIII. The LDMS of the SLSP system

The LDMS conducts the logical decisions of the SLSP. It receives sensor data from the session server. Sensor data is the crude data from the sensors, e.g., "warning sound level" is audio sensor data from the audio sensor which denotes that sound level has exceeded a certain decibel level. Mere data is crude data from the sensors, e.g., video recorder transmits video stream to session server. This type of data is video data. The LDMS conducts logical deductions according to sensor data and according to rules designed to handle sensor data.

The deductions produce information of the surveyed area. Information created by the LDMS creates alarms of different security levels. Alarms may contain the location of the alarm event and a short description of the alarm. The lowest security level (GREEN) only attests that everything is normal. Caution (YELLOW) indicates situations that are potentially hazardous. Hazards (RED) denotes that there is a situation that requires attending.

The LDMS subsystem subcomponents are illustrated in figure 5. LdmsMain is the main executable, which is responsible for initializing the server. The SenderReceiver component plies sending and receiving messages to/from the session server and other SLSP components. This is done by using the NetworkInterface component which implements the required transfer protocols. The RequirementHandler component encompasses the main logical functions for executing requirements, definitions of which are parsed from a locally stored XML file using the RequirementXmlParser. In this context, "requirement" refers to the logical deduction requirements found in the LDMS requirement specification. Based on the XML file's requirement definitions, the LDMS can respond to events and trigger its own timed requirements. DbHandler subcomponents addresses the database operations needed by the LDMS.

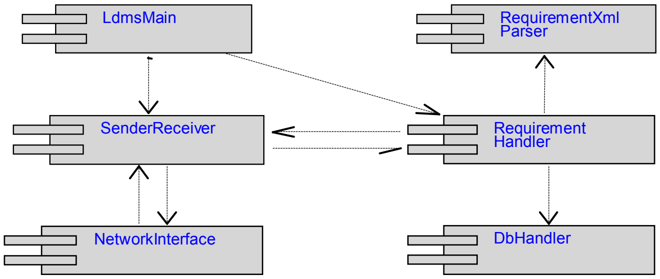


Figure 5, the component diagram of the LDMS.

An example of the XML requirement structure used to illustrate an LDMS logical decision making requirement, in this case In-Fingerprint-Sensor-Access-Granted-Opening-Hours, is shown in Figure 6. The requirement has two conditions in order to execute its actions: the operating mode must be opening hours, and the fingerprint sensor must send a notification of an authorized entry through the session server. If these conditions are met, two actions are executed: notifying the session server about the authorized entry during opening hours, and incrementing the fingerprint sensor tally.

```
<req>
 <req-name>In-Fingerprint-Sensor-Access-Granted-Opening-Hours</req-
name>
...
 <req-type>sensor event</req-type>
 <cond-list>
  <cond>
   <cond-source>LDMS</cond-source>
   <cond-type>hours</cond-type>
   <cond-data>opening hours</cond-data>
  </cond>
  <cond>
   <cond-operator>AND</cond-operator>
   <cond-source>FingerprintSensor</cond-source>
   <cond-type>reply</cond-type>
   <cond-title>fingerprint sensor status</cond-title>
   <cond-data>Access ok</cond-data>
  </cond>
 </cond-list>
 <act-list>
  <act>
   <act-method>notify</act-method>
   <act-target>SessionServer</act-target>
   <act-data>access granted during opening hours</act-data>
  </act>
  <act>
   <act-method>requirement</act-method>
   <act-target>ldms</act-target>
   <act-data>In-Fingerprint-Sensor-Keep-Tally</act-data>
  </act>
 </act-list>
</req>
```

Figure 6, an example of LDMS's XML file.

IX. Conclusion

Multi-sensor surveillance accompanied with automatic logical decision is an important branch in the field of surveillance. With the utilization of advanced sensors, i.e., a fingerprint sensor, an audio sensor, a video recorder, and a network activity monitor, it is possible to automatically form deductions of a surveyed indoor area. We have illustrated the implemented design and communication how this endeavor is attained with the sensor data collection and transmission system of the SLSP system.

The information and structure of the SLSP system was modeled on recent journals and conference papers regarding surveillance, especially focusing on multi-sensors, architecture and middleware, and logical decision making. Predicated on Valera and Velastin's approach, we utilize a multi-agent approach. As evoked by Velara and Velastin, we have constructed an intelligent multi-sensor surveillance system, which also utilizes low cost standard components. In accordance to Cucchiara et al., Regazzoni et al., and Bramberger et al., we have consolidated acoustic and visual information and, in addition, a fingerprint sensor and a network activity monitor. Based on Valera and Velastin's argument of data fusion on multi-sensor information, we perform task in our SLSP system. We utilize multi-sensor data acquisition, communications and processing, as stated by Regazzoni et al. In comparison to Valencia-Jimenez and Fernandez-Caballero's statement on architecture, we have employed a centralized architecture. Atrey et al. deem that multiple sensors should be used and the information should be derived from their data. This is our approach in the SLPS system. The SLSP system's middleware is established on a sound middleware to support information collection, information management and decision making, as its importance was indicated by Hardian and Detmold et al. In relation to Velastin et al.'s proclamation of different events and sensors that can elicit alarms, the SLSP system's LDMS supports both simple and complex (timed) events and alarms are raised accordingly, either by an individual sensor or from a consolidation of multiple sensors. Pertaining to Castanedo et al.'s and Tabar et al.'s remarks of the vitality of data fusion, data management and networked sensors, the fusion of data and management of multiple sensors was successfully achieved in the SLSP system with networked sensors. The LDMS performs sensor fusion, data fusion, situation assessment and decision making, as defined by Hall, Blasch et al., Steinberg, and Nelson and Fitzgerald.

The information collection and transmission of the SLSP intends to facilitate the collection of information from a surveillance point and to decrement the amount of superfluous information rendered to the surveillance personnel, by procuring automatically sensor data and providing automatically derived information. These two antecedently depicted aspects are the main endeavors of the SLSP's information collection architecture. The SLSP system's multiple sensors collect sensor data from their ambit and transmit the data through to the session server. Thus, the sensor data is automatically collected from a surveyed point, then the data is transmitted to the LDMS by the session server. The LDMS automatically deducts situations based on the sensor data. The session server transmits the LDMS deductions to the security personnel's end-devices. The operability of the constructed SLSP system prototype indicates that these endeavor is attained.

References

[1] Bramberger, M., Doblander, A., Maier, A., Rinner, B., and Schwabach, H.: Distributed Embedded Smart Cameras for Surveillance Applications, *Computer*, February 2006, pp. 68-75.

[2] Reiter, M. and Rohatgi, P.: Homeland Security Guest Editor's Introduction, *IEEE Internet Computing*, November/December 2004, pp. 16-17.

[3] Valera, M. and Velastin, S.A.: Intelligent distributed surveillance systems: a review, *IEE Proc.-Vis. Image Signal Process.*, Vol. 152, No. 2, April 2005, pp. 192-204.

[4] Regazzoni, C.S., Ramesh, V., and Foresti, G.L.: Scanning the Issue/Technology Special Issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems, *Proceedings of the IEEE*, Vol. 89, No. 10, October 2001, pp. 1355-1367.

[5] Valencia-Jimenez, J.J. & Fernandez-Caballero, A.: Holonic Multi-agent Systems to Integrate Multi-sensor Platforms in Complex Surveillance, Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06), IEEE, 2006.

[6] Atrey, P.K., Kankanhalli, M.S., and Jain, R.: Timeline-based Information Assimilation in Multimedia Surveillance and Monitoring Systems, VSSN'05, November 11, 2005, Singapore.

[7] Cucchiara, R.: Multimedia Surveillance Systems, VSSN'05, November 11, 2005, Singapore.

[8] Detmold, H., Dick, A., Falkner, K., Munro, D.S., van den Hengel, A., and Morrison, R.: Middleware for Video Surveillance Networks, MidSens'06, November 27-December 1, 2006, Melbourne, Australia.

[9] Hardian, B.: Middleware Support for Transparency and User Control in Context-Aware Systems, MDS'06, November 27 – December 1, 2006, Melbourne, Australia.

[10] Collins, R.T., Lipton, A.J., Fujiyoshi, H., and Kanade, T.: Algorithms for Cooperative Multisensor Surveillance, *Proceedings of the IEEE*, Vol. 89, No. 10, October 2001, pp. 1456-1477.

[11] Velastin, S.A., Boghossian, B.A., Lo, B.P.L., Sun, J., and Vicencio-Silva, M.A.: PRISMATICA: Toward Ambient Intelligence in Public Transport Environments, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 35, No. 1, January 2005, pp. 164-182.

[12] Castanedo, F., Patricio, M.A., Garcia, J., and Molina, J.M.: Extending Surveillance Systems Capabilities Using BDI Cooperative Sensor Agents, VSSN'06, October 27, 2006, Santa Barbara, U.S.A.

[13] Tabar, A. M., Keshavarz, A., and Aghajan, H.: Smart Home Care Network using Sensor Fusion and Distributed Vision-based Reasoning, VSSN'06, October 27, 2006, Santa Barbara, U.S.A.

[14] Wald, L.: A European proposal for terms of reference in data fusion., International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 7, 651-654, 1998.

[15] Hall, D.L.: Challenges in Data Fusion: Dirty Secrets, Current State of Technology and a Research Roadmap, Associate Dean for Research School of Information Sciences and Technology, February 28, 2005.

[16] Blasch, E. and Plano, S.: DFIG level 5 (user refinement) issues supporting situational assessment reasoning, Information Fusion, 2005 8th International Conference on, Volume 1, 25-28 July 2005.

[17] Steinberg, A.N.: An approach to threat assessment, Information Fusion, 2005 8th International Conference on, Volume 2, 25-28 July 2005.

[18] Hall, D.L.: Perspectives on the fusion of image and non-image data, Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd, 15-17 Oct. 2003.

[19] Nelson, C.L. and Fitzgerald, D.S.: Sensor fusion for intelligent alarm analysis, *Aerospace and Electronic Systems Magazine*, IEEE, Volume 12, Issue 9, Sept. 1997 Page(s):18 – 24.

PUBLICATION 5

Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder

Second International Conference
on Mobile Ubiquitous Computing, Systems,
Services and Technologies, UBICOMM 2008.
Valencia, Spain, September 29 – October 4, 2008.
Accepted for publication.

Distributing Essential Logical Deductions to Surveillance Personnel and a Video Recorder

Tomi Räty
VTT Technical Research Centre of Finland
P.O. Box 1100
90571 Oulu, Finland
+358 8 551 2131
tomi.raty@vtt.fi

Mikko Lindholm
VTT Technical Research Centre of Finland
P.O. Box 1100
90571 Oulu, Finland
+358 8 551 2225
mikko.lindholm@vtt.fi

Mikko Nieminen
VTT Technical Research Centre of Finland
P.O. Box 1100
90571 Oulu, Finland
+358 8 551 2057
mikko.nieminen@vtt.fi

Johannes Oikarinen
VTT Technical Research Centre of Finland
P.O. Box 1100
90571 Oulu, Finland
+358 8 551 2057
johannes.oikarinen@vtt.fi

ABSTRACT

The Single Location Surveillance Point (SLSP) is an indoor distributed multi-sensor surveillance software system. It encompasses an arbitrary amount of sensors that collect data from a single location, which is the surveillance point. The ensuing sensors are realized: a fingerprint sensor attached to a door with an electronic lock, a video camera, an audio sensor, and a network analyzing monitor. Each sensor collects information from its ambit. Once the crude data has been acquired from the sensors and transmitted to Logical Decision Making Server (LDMS) by the session server, the LDMS automatically performs logical deductions based on the data received from the sensors. The logical deductions create: 1) information for end users or 2) control messages to sensors. Based on the alarms, the LDMS can ordain instructions to the video recorder. The LDMS distributes the logical deductions to the human security administrator of the Security Manager Server (SMSU) and/or the end devices of the nomadic guards. The SLSP system provide the surveillance personnel refined information cogent to occurring events of the surveyed area. The SLSP system intends to decrement the amount of superfluous information rendered to the surveillance personnel, by providing automatically derived information. These two antecedently depicted facets are the main endeavors of the SLSP system distribution of logical deductions. The operability of the constructed prototype indicates that this endeavor is attained. The research is based on the constructive method of the related publications and technologies and the results are derived by the implemented branch of the SLSP system.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems – *intelligent agents,* and I.2.3 [**Deduction and Theorem Proving**]: Deduction (e.g., natural, rule-based).

General Terms

Design, Experimentation.

Keywords

Middleware infrastructures for event-based computing, information systems, surveillance systems, and multi-sensor systems.

1. INTRODUCTION

Traditionally, surveillance systems in private and public environments approaches rely on the installation of wide-area closed-circuit television (CCTV) [23]. CCTV requires a relatively small amount of operators to incessantly monitor a significant number of cameras and other devices [23]. Video surveillance has become a ubiquitous aspect of the modern urban landscape, situated in a vast variety of environments including shopping malls, government buildings, and commercial premises [15].

The inherent limitation in the effectiveness of CCTV surveillance systems is the cost of offering adequate human monitoring cover for what is a considerably tedious task. Consequently, CCTV tends to be employed as a reactive tool and the perception that a public transport operator is in charge of its space is lost if no response is procured when a problem incurs. The proactive approach is desirable, in which the likelihood of events can be recognized automatically to guide the attention and action of the human operators in charge of conducting a transport network. It is vital to perform this in a manner that conceives surveillance systems as decision-support tools for human operators to address with complex and vast environments. [22]

There are immediate needs for automated surveillance systems in commercial, military applications, and law enforcement. Mounting video cameras is inexpensive, but locating available human resources to survey the output is expensive. What is

Conference'04, Month 1–2, 2004, City, State, Country.

required is an incessant 24-hour monitoring of surveillance video to alert security officers, while there still is time to circumscribe the criminal offence. [3]

Homeland security is an inherent concern for governments worldwide, which must protect their people and the critical infrastructures that uphold them. Information technology plays a significant role in such initiatives. It can assist in mitigating risk and enable effective responses to disasters of natural of human origin. [17]

The SLSP (Single Location Surveillance Point) system is a distributed indoor multi-sensor surveillance system. It contains multiple sensors, constituting of a fingerprint sensor, a video camera, an audio sensor, and a network analyzing monitor. The sensors are located in an indoor area for surveillance. Each sensor acquires from its environment and transmits the crude data to the session server. The session server handles all the connections among the components of SLSP. The session server transmits the received data from the sensors to the LDMS (Logical Decision Making Server). The LDMS automatically deducts the surveillance point's situation predicated on the data it receives from the sensors routed by the session server. The LDMS may issue instructions automatically to the video recorder based on its logical deductions, e.g., for the video recorder to be directed to a location of an event discriminated by another sensor. The deductions are transmitted to the human surveillance operator of the SMSU (Security Manager Server) and to the end device of a nomadic guard.

The intent of the SLSP system is to ultimately conduct logical decisions automatically based on received sensor data of the surveyed area and transmit the logical deductions to the human surveillance operator and possible nomadic guards of the surveyed indoor area. The operability of the constructed SLSP system prototype indicates that this endeavor is attained.

The structure of this paper is the ensuing. First a general overview of contemporary surveillance systems is presented, followed by a corollary of situation awareness and real-time threat, integrated multi-sensors surveillance systems, security personnel issues, middleware, logical decision making, and the definition of a sensor and an event. An introduction of the SLSP system is then presented. This is followed by detailed descriptions of the SLSP's internal domains, consisting of the surveillance domain, the security administration and surveying domain, and the security personnel management. A comparison of the SLSP system to the state-of-the-art is performed and the conclusion summarizes the paper.

2. SURVEILLANCE SYSTEMS

Video monitoring usually deploy multiple video cameras, channeling video signals to a central monitoring room, where multiplexing is utilized to render a subset of the images to security personnel. Event detection and recognition use the perceptual capabilities of a human operator to discern objects moving within the field-of-view (FOV) of the cameras and to conclude their actions. However vigilant the operators, manual monitoring inevitably suffers from information overload, which results in periods of operator inattention due to fatigue, distractions, and interruptions. Automating all or part of this process would obviously offer dramatic benefits, ranging from a capability to alert an operator of potential event of interest, through a completely automatic detection and analysis system. [15]

CCTV devices have played a crucial role in the management of public places pertaining to safety and security. The explosion in the amount of cameras that must be monitored, the accruing costs of offering monitoring personnel and the limitations of human operators to uphold sustained levels of concentration severely circumscribe the efficaciousness of these systems. Alternatively, subsequent advances in information and communication technologies can potentially offer considerable improvements. The deployment of technology to maintain surveillance is used in modern urban environments. [23]

The distinction between surveillance for indoor and outdoor applications exists because there are differences in the design at the architectural and algorithmic implementation levels. The topology of the indoor environments is different from the outdoor environments. To survey a vast region implicates geographical distribution of paraphernalia and a hierarchical structure of the personnel who handle security. [21]

3. SITUATION AWARENESS AND REAL-TIME THREAT DETECTION

The key to security is situation awareness. Awareness requires information, which spans multiple scales of time and space. To offer comprehensive, non-intrusive situation awareness, it is vital to ply the challenge of multi-scale, spatiotemporal tracking. From the perspective of real-time threat detection, it is a known fact that human visual attention decreases below acceptable levels even when trained personnel are assigned to visual monitoring. [10]

Intelligent remote monitoring systems allow users to survey sites from significant distances. These systems exert rapid and efficacious remedial actions to be executed immediately once a suspicious activity is detected. An alert system can be employed to warn security personnel of impending vicissitudes and numerous sites can be concurrently monitored. This substantially abates the load of the security personnel. With the decreasing cost of computational power and advancement in Internet technologies, implementation of a web-based security surveillance system becomes a considerable option to the traditional manually operated systems. Streaming technology enables video servers to transmit content in a subsequent stream, which can be decoded and played back shortly after it has been received by the client contraption. This is the preferred mode of operation. [6]

4. INTEGRATED MULTI-SENSOR SURVEILLANCE SYSTEMS

Most of the new research activities in surveillance are exploring larger dimensions, such as distributed video surveillance systems, heterogeneous video surveillance systems accompanied with fixed, PTZ, and active cameras, multi-spectral camera systems, systems with multimedia streams, including audio, video, and sensors signals, surveillance and biometric systems [4]. Spatially distributed multi-sensor environments render interesting possibilities and challenges for surveillance [21]. Recently, there has been some investigation of data fusion techniques to tolerate with information sharing pertaining to erudition resulting from different types of sensors [21]. The communication facets within separate parts of the system play a crucial role, with particular challenges either due to bandwidth constraints or the asymmetric disposition of the communication [21].

A surveillance system should be complete and it should enable data accessibility for direct alarm raising needs to include: a user oriented mechanism, a sufficiently extended amount of functionalities adequate to offer a spatial surveillance support appropriate for the task (completeness), and an alarm raising mechanism fulfilling real-time alarm raising user requirements (real-time response). [16]

Especially each functionality should be associated with the ensuing: a computational epitome of a detection method appropriate to distinguish events of interest from available signal representations (computability), and an appropriate selection of sensors to provide data required for detecting events of interest (multimodal sensorial support). Multi-sensor systems can capitalize from processing either the same type of information obtained by sensors of different type, e.g., video cameras, microphones, etc., on the same monitored area. [16]

Accurate and robust localization and tracking of acoustic sources is of interest to a variety of applications in surveillance, multimedia, and hearing enhancement. Miniaturization of microphone arrays incorporated with acoustic processing further augments the utility of these systems, but poses challenges to achieve precise localization performance due to abating aperture. For surveillance, acoustic emissions from ground vehicles offer a facilely detected signature, which can be employed for unobtrusive and passive tracking. [19]

Considering the nature of an event that is desirable to detect, the content of information created is more than just visual information. Many of the significant events from a monitoring point of view are accompanied by audio information, which would be useful to scrutinize. The significance of these events is not provided only by their semantic information, but by their temporal context. With audio, one can have event detection on a graded scale, from minor events to abnormal sounds. It is possible to detect outlier audio events utilizing simple forms of analysis. [18]

When an event is detected in both of the systems, CCTV and audio, the chances of it being a significant event enlarges. By expanding the compass of information available to the system, the precision of the operation can be improved. The purpose of an audio sensor network would be to assist the end user to percolate through data and return the points of interest. This would not be conducted by adding an overwhelming amount additional data, but by drawing attention to the data already obtained, but not might find. [18]

5. INTEGRATED MULTI-SENSOR SURVEILLANCE SYSTEMS

An individual human operator cannot efficaciously monitor a vast area by viewing dozens of monitors displaying raw video output. Maintaining track of people, vehicles, and their interactions across a vast area is an arduous task for a human observer. It certainly cannot be done efficiently by viewing a wall of video screens with each displaying a disparate sensor view. [3]

Visual surveillance and monitoring (VSAM) systems constantly becoming stronger factors in prevention and reduction of criminal offences and in the enhancement of efficient management of resources, e.g., traffic management and subway monitoring [7]. A generic surveillance and security system is composed of three essential parts: data acquisition, information analysis, and on-field operation [14]. Any surveillance system requires means to monitor the environment and obtain data in the form of video, still images, audio, etc [14]. Such data is to be processed and analyzed by a human, a computer or a consolidation of both at a command centre [14]. An administrator can decide on performing an on-field operation to put the environment back into a situation considered as normal [14]. On-field control operations are conducted by on-field agents who require effective communication channels to retain a close interaction with the command centre [14].

Data acquisition is conducted by means of a set of video cameras. Information analysis is the integral part of a surveillance system. To provide an appropriate response to a given incident within reasonable timing, all the information of the entire situation, must be gathered in one distinct location. On-field operation is the result of decisions exerted at the control centre and require a team of surveillance agents to control the situation on the ground. Common communication devices contain pagers, headsets, etc. Recent security studies and initiatives have indicated the importance of permanent multimodal communication. Surveillance agents require efficacious means of communication with the commander. Interoperable communications are vital to maximize efficacy of on-field agents. Security personnel review their wireless video systems for critical incident information. The need for providing elaborate real-time information to the surveillance agents has been identified and is being addressed by the research community. [14]

A multimedia surveillance system should be a surveillance system capable of providing distilled video, images and sounds of the monitored environment. It also gathers, processes in real-time, correlates and addresses multimedia data resulting from different sources. Multimedia surveillance systems can ameliorate visual data with audio streams and information resulting from other sensors. In vast distributed environments, the exploitation of networks of small cooperative sensors should substantially improve the surveillance capability of few higher levels sensors, such as cameras. [4]

Typically, surveillance systems are composed of numerous sensors to obtain data from each target in the environment. These systems encounter two types of dilemmas: 1) fusion of data, it is related to the combination of data from discrete sources in an optimal manner, and 2) management of multiple sensors, presuming that the previous predicament is solved, and it conducts optimizing the global management of the joint system through the application of individual operations in every sensor. [2]

6. MIDDLEWARE

Due to the availability of more advanced and powerful communications, sensors, and processing units, the architectural choice in 3rd Generation Surveillance Systems (3GSS) can potentially become extremely variable and flexibly customized to procure a vied performance level. The system architecture commences to delineate a key factor. [16]

Video surveillance networks are a class of sensor networks with multiple purports including the protection of major facilities from terrorism and other threats. During routine operation, automated surveillance software needs to be substantially autonomous to relieve human operators of active involvement. When a threat is detected, a phase change occurs, and human operators need to be able to exert control over the parts of the network in which the threat is extant. [5]

Typically, middleware offers miscellaneous transparencies that help to simplify application development, including distribution/location transparency. In context-aware systems, middleware is also needed to ease adaptation predicated on context information. This requires support for the ensuing tasks: 1) collecting context information from sensors and other sources, 2) management and integration of context information, 3) reasoning support and context querying, and 4) support for decisions regarding adaptations. The support for adaptation and include management of rules and/or user preferences that are utilized to distinguish how the context-aware system will respond to the available context information. [11]

7. LOGICAL DECISION MAKING

In the contemporary generation of surveillance systems, in which multiple asynchronous and miscellaneous sensors are employed, assimilation of the information procured from surveillance systems to derive the events from the environment is an important and challenging research problem. Information assimilation refers to the process of consolidating the sensor and non-sensor information using the context and past experience. The issue of information assimilation is vital, because the information procured from multiple sources when assimilated offers more precise inferences of the environment than individual sources. [1]

A concise way to define sensor and data fusion is the following. Sensor fusion is “data fusion from multiple sensors (same or different sensor types)” [9]. Data fusion is “combining information to estimate or predict the state of some aspect of the world” [9]. Data fusion involves the use of multiple data, often from multiple sources, to estimate or predict the state of some aspect of reality, e.g., estimation/prediction of the state of individual(s), which are treated as if they were independent of the states of other entities [20].

Numerous multi-sensor systems have been developed to collect, process, and disseminate image and non-image data. The goal of multi-sensor fusion is to achieve inferences about the observed environment or situation that cannot be achieved by a single sensor or source of information. Information about the observed situation is combined to achieve high-level inferences. [8].

Machine intelligence demands techniques which can transform incomplete, inconsistent, or imprecise data provided by one sensor into more useful information by fusing it with data provided by other sensors [13]. Multi-sensor data fusion can provide solutions to problems that are characterized by intensive and diverse sensor information [13]. Fusion makes a synthesis of input data appropriate to a decision maker’s need for meaningful information [13]. An automatic system should also integrate this sensory data with contextual and domain information provided by humans to maintain a coherent logical picture of the world. [12]

8. WHAT IS A SENSOR AND AN EVENT

A sensor refers to the processing directly consorted with a physical transducer (camera, microphone, fire detector, etc.) or actuator and is directly meaningful to an operator and employed in a geographical representation of the site. A sensor can measure or detect one or multiple events. Simple sensors will be typically associated with only one event, e.g., fire detected, but it is feasible for a sensor to be capable of detecting a number of events. Events can be of different types, such as an alarm (an incident has occurred), measurement (a continuous quantity such as the amount of people in an area), or status (system information such as power failure). [22]

9. INTRODUCTION TO THE SLSP SYSTEM

The SLSP comprises of three individual domains as in Figure 1. These three domains are 1) the Surveillance Domain, which comprises of an arbitrary amount and variety of sensors, and 2) the Security Administration and Surveying Domain, which comprises of the session server to which the sensors transmit their information and the logical decision making server, and 3) the Security Personnel Management domain, which is intended for conducting security personnel from a remote and centralized location. This domain also provides an interface to the human security administrator and the end-devices, e.g., smart phones.

Initially, sensors transmit their information to the session server. The session server transmits the crude sensor information to the LDMS. The LDMS is responsible of transmitting its logical deductions regarding the surveillance point to the SMSU through the session server. Then the SMSU can transmit orders, with the help of the human security administrator, to the nomadic guards, e.g., end-devices.

The data flow from the sensors of the Surveillance Domain is primarily the ensuing: 1) the crude sensor information is conveyed from the sensors to the session server, 2) the session server transmits all the crude sensor information it receives to the LDMS, 3) after performing the its automatic logical calculations, the LDMS transits its deductions to the SMSU and/or the nomadic guards via the session server. Then the human security administrator can issue orders to the end-devices, or even re-route the deduction information and/or crude sensor information, e.g., video footage, it receives to the end-devices.

The session server acts as an interface from which crude sensor information can be procured.

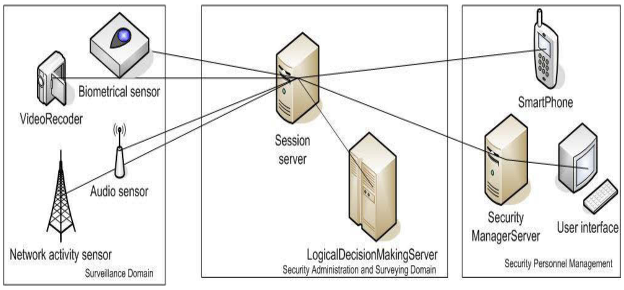


Figure 1, the high-level structure of the SLSP system.

10. SURVEILLANCE DOMAIN OF THE SLSP SYSTEM

The Surveillance Domain comprises of the discrete sensors utilized in the SLSP system. The employed sensors are the ensuing: the biometrical sensor, the video recorder, the audio sensor, and the network activity monitor. The sensors are each proprietary or independent devices. Each sensor will be briefly elaborated in the following subchapters.

10.1 The Biometrical Sensor

The biometrical sensor of the SLSP system is a fingerprint sensor, which is Deltabit's Gatekeeper fingerprint recognition product. The product registers fingerprints at a door, and transmits the access information derived from the access rights based on the access rights of the fingerprint. The fingerprint sensor distributes the access information of an user's fingerprint to the session server. The information apropos to reading a fingerprint may entail an access granted, access denied, etc. notification. The Gatekeeper product also transmits information apposite to the door's electronic lock. The information relevant to registering the status of the door's electronic lock may entail a door locked, door unlocked, door open, door closed, etc. notification. The information of the fingerprint readings and electronic lock status is transmitted to the session server.

10.2 The Audio Sensor

The audio sensor of the SLSP system monitors the environment for threatening sound events. A threatening event is discriminated as an audio event that exceeds a pre-defined threshold of volume. The audio sensor indicates the bearing of the sound location. The surveyed data of the audio sensor, comprising the volume of the audio event and the bearing of the direction from which the audio event occurred, is distributed to the session server.

10.3 The Video Recorder

The video recorder of the SLSP system is the Axis 213 PTZ video camera. The video recorder transmits the video transmission to the session server. This information is the crude sensor information of the video recorder.

10.4 The Network Activity Monitoring Sensor

The network activity monitor is Nethawk's M5 traffic analyzer. The network activity monitor surveys all the IP-level network traffic inside the SLSP. The network activity monitor observes the data, both the amount and type, traversing in the network and the devices. The devices generating the traffic must be registered to the SLSP system, unregistered devices are not allowed. To monitor the devices, the network activity monitor implements a "watchdog" property, which ensures that the device transmitting is a valid device. If not, the device's MAC address is reported to the session server. The watchdog property relies on capturing DHCP protocol packets that are utilized in the network to allocate an IP address for the device. Each device connecting to the network transmits DHCP request to the DHCP server providing IP address in a response. To implement the watchdog property each DHCP message is analyzed and cross-referenced to the MAC address table in the network activity monitor. The report is transmitted to the session server.

11. SECURITY ADMINISTRATION AND SURVEYING DOMAIN

The Security Administration and Surveying Domain composes of the session server and the LDMS. The session server is responsible for transmitting all the messages of the SLSP environment to the appropriate receivers. This includes all the components connected to the session server, i.e., the biometrical sensor, the video recorder, the audio sensor, the network activity monitor, the LDMS, the SMSU and the end-device. The LDMS is responsible for deriving logical deductions based on the data received from the surveillance point's sensors. The LDMS receives the crude sensor information from the session server. The conclusions of the deductions are transmitted to the session server, which sends them to the SMSU, and based on the SMSU's instructions, possibly to the smart phone. The LDMS may automatically issue panning and zooming instructions to the video recorder based on the logical deductions of certain events, e.g., when a exceptionally loud audio event happens the LDMS automatically issues the video recorder to pan and zoom in the direction of the loud audio event.

11.1 The Session Server

The session server consists of a single component, which contains the main logic of the session server. To communicate with the servers and the sensors in SLSP system, the session server uses network libraries for communication. To obtain the video, from the video recorder, the Darwin streaming server is used. The Darwin streaming server is component within the session server but it operates independently. The purpose of the VLC video player is to address video stream storing into a permanent file repository. Both the VLC and Darwin streaming server are separate applications and they are used to procure and transfer the video stream and recorder video clips.

The streaming sequence is illustrated in Figure 2. The sequence begins when a request to store video clip is received. Then the VLC commences to stream and store the video stream. Once the video stream is stored, the created video clip is saved under Darwin to be streamed. Darwin manages the delivery of live video stream to the end-device/SMSU. Darwin receives media requests from the end-device/SMSU and transports the video stream from the video recorder to the end-device/SMSU.

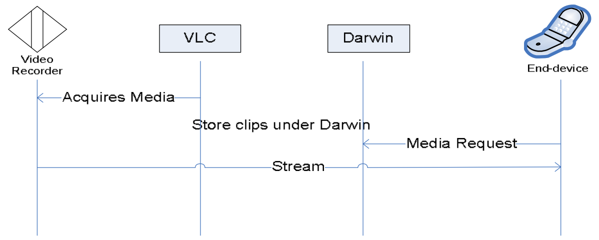


Figure 2, streaming sequence from a video recorder to the end device.

The data that is conveyed with TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) between sensors and LDMS are in textual format. The video data is transferred over RTSP (Real-Time Streaming Protocol) / RTP (Real-time Transport Protocol) and the transportation is handled with Darwin Streaming Server.

The session server's main functionality is rendered in Figure 3. The CSessionServerMain class is the main class of the session server. The class administers the initial startup procedures and addresses messages that were received through interfaces. The CSensorInterface class controls the socket-level connection to a sensor. A ThreadSocket is executed in a separate thread. The ThreadSocket class implements a socket that handles receiving and sending through one socket interface. The CDeviceInterface class controls the socket-level connection to an end-device. A ThreadSocket is executed in a separate thread.

The CSessionServerMain class entails a main loop that performs receive buffer checks and message handling. The CSensorInterface and CDeviceInterface classes function inside this main loop that is managed by the CSessionServerMain. When a remote interface, either from an end-device or a sensor, is initialized and set to listen, another loop is initiated in each ThreadSocket. This loop conducts the actual and periodical receiving and sending operations by transmitting everything from the send buffer and conveying received messages to receive buffer. An equal amount of ThreadSocket classes (sockets) are generated for each sensor of the SLSP system. A socket is also generated for every end-device and LDMS.

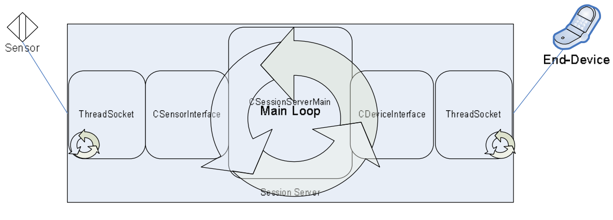


Figure 3, the session server's main functionality.

11.2 The LDMS

The LDMS receives sensor data from the Surveillance Domain sensors. The received sensor data is the crude data from the sensors. The LDMS makes logical deductions predicated on sensor data and according to rules discriminated to process sensor data.

The logical deductions of the LDMS produce 1) information for end users, i.e., the operators of the SMSU and the end-device, or 2) control messages to sensors. Logical deductions, created by the LDMS, generates alarms of different security levels. Alarms may include the location of the event and a short depiction of the event. The lowest security level (GREEN) only indicates the security personnel that everything is normal. The intermediate security level (YELLOW), indicates a heightened level of security. The highest security level (RED), indicates the maximum level of security.

Some events require different conditions for opening hours and after hours. Opening hours are when public can enter the premises, e.g., a shopping mall, under surveillance by the SLSP system. After hours is when the premises, e.g., shopping mall, is closed, typically during the night and holidays but occasionally also during a state of emergency. The information regarding opening hours and after hours is received from the session server. The default value is after hours.

The LDMS is capable of reacting to its own logical decisions. An important feature of the LDMS is to automatically issue instructions to the video recorder to pan and zoom to a certain direction. When an alarm is distinguished by the LDMS occurs, it may request the video recorder service for monitoring the location of the event. Examples of an event causing the LDMS to issue the video recorder to pan and zoom is an unusually loud audio event, to which the video recorder will be directed, or a failed entrance attempt at the biometrical sensor after hours. The monitoring continues long enough for security personnel to observe the situation.

11.2.1 Security Level Occurrences (SLOs)

Security level Occurrences (SLOs) comprise of 1) events in the surveillance domain, 2) session server commands or 3) replies to the LDMS status/data requests from the sensors or session server. Usually, the SLO comes to logical decision making server based on an event, which has been collected by a sensor, from the surveillance point. A status or data request to surveillance domain sensor is a SLO, because every request must have a reply. If this reply is not received, the LDMS performs a logical deduction that the requested sensor is inoperative. The session server handles the direct communication between the session server and the LDMS, therefore there is no direction connection or communication between the LDMS and the sensors.

SLOs, which are connected to physical events, are detected by the sensors of the surveillance domain. Typically, a SLO is an event of YELLOW or RED security level, but also some events with GREEN security level can indicate SLO. For instance, an unauthorized attempt to enter the premises during after hours, informed by biometrical sensor, is a RED security level occurrence. During the opening hours an unauthorized attempt to enter the premises, informed by the biometrical sensor, constitutes to a YELLOW SLO, because the probability of criminal action is considered smaller. A status request to biometrical sensor is always a GREEN SLO, but if no reply is received from the sensor within 10 seconds, during after hours, the biometrical sensor is deemed inoperative. Then this event is considered to be a RED SLO.

When a SLO occurs, the LDMS always notifies the session server of it. Commonly, an action ensues this notification. The most typical action is to pan the video recorder to the probable area where the SLO occurred. GREEN and YELLOW SLOs are retained for 5 minutes in the Security Level Table (SLT). RED SLOs are retained in the for 10 minutes in the SLT.

When LDMS receives a SLO, it needs the security level occurrence administrator. The security level occurrence administrator is the procedure by which the LDMS recognizes a SLO and links it to proper requirement.

11.2.2 Security Level Table (SLT)

The LDMS retains a SLT of the YELLOW and RED SLOs. Table 1 indicates the structure of an SLT event. Any separate event that denotes YELLOW or RED SLO must be added to the SLT. The SLOs in connection with SLT can mean a single event, a security level combination (see subsequent subchapter “Security level combination”) or a special combination (see subsequent subchapter “Special combination”). Some GREEN instances of events that can trigger a YELLOW or RED SLOs are placed in the SLT. The SLO event can have a SLT type of INDIVIDUAL, CONTINUOUS or it may be without a type.

The INDIVIDUAL SLO event, which is placed in the SLT type, is associated to a distinguishably separate instance of a SLO. For instance, entrance to a secured premises granted by the biometrical sensor or a loud sound of the duration of a short period (possibly a gunshot), detected by the audio sensor are INDIVIDUAL. Events that have no direct connection to security threats are neither INDIVIDUAL nor CONTINUOUS and they are never placed in the SLT. These requirements include, e.g., a response to sensor data denoting a normal situation.

A CONTINUOUS SLT type event contains multiple consecutive events during a short duration of time which are perceived as one continuing event. For instance, the biometrical sensor’s door status indicator may denote that the door is open for five minutes. If the status of the door is requested once every second, then there will be 300 separate occurrences of “Door-Open”. These 300 individual SLOs are fused to one “Door-Open” SLO.

Table 1, the content of the SLT.

SLT content
1. the instance of the event connected to the SLO
2. the security level of the event (GREEN, YELLOW, RED)
3. the type of the event (INDIVIDUAL or CONTINUOUS)
4. the event’s priority in accordance to the video recorder service
5. the time of the first occurrence of the event
6. the time of the latest occurrence of the event (only for the CONTINUOUS type)

An SLO event is effaced from the SLT based on either 1) exceeding a time limit (5 minutes or 10 minutes) or 2) there are already 20 instances of this same SLO event in the SLT. When YELLOW or RED SLOs of events are removed, the session server must be informed. In practice, this notification indicates that the SLO is not considered active anymore.

11.2.3 Security Level Combination

Two or more single requirements and/or special combinations are required to compose security level combination. A security level combination is a consolidation of SLOs that have RED and/or YELLOW security levels and they have resulted from different surveillance domain sensors. The events must also reside in the SLT at the same time. Distinct security level combinations can comprise of 1) at least two RED security level occurrences from different sensors, 2) one RED and at least one YELLOW SLO from different sensors, or 3) at least three YELLOW SLOs from different sensors. Security level combinations are always considered CONTINUOUS. They are perceived as continuous events and complex security threats, not as separate events occurring successively and stochastically.

The security level combinations of the sensors in the Surveillance Domain differ from a single event occurring of any individual sensor. Security level combinations indicate that there may be a complicated security threat progressing. The urgency for security personnel to react can be the same for any single SLO, but the procedure how to react with security level combinations may be more comprehensive. For example, a simultaneously inoperative video recorder and a biometrical sensor may indicate a premeditated burglary.

11.2.4 Security Combination

A special combination is an event that requires another event, a.k.a. a “triggering” event (e.g. no reply for sensor status request), to occur in order to act or complete its action. The security level of the triggering event may be GREEN. For this reason, the security level table must also include some events with GREEN security levels.

For instance, the special combination Logical-Decision-Making-Server-In-Biometrical-Sensor-Inoperative-Opening-Hours indicates that the biometrical sensor is inoperative currently, during opening hours. There must be a reaction to a possible threat, if the LDMS’s status request to the biometrical sensor, Logical-Decision-Making-Server-In-Biometrical-Sensor-Status-Request, has not received a status reply within a time limit. This may be an indication that the biometrical sensor has been broken down. It may also indicate that there has been an temporary and harmless disruption in the communications network that does not require any specific action. This special combination is considered as a YELLOW security level occurrence of CONTINUOUS type.

The Logical-Decision-Making-Server-In-Biometrical-Sensor-Inoperative-Opening-Hours is a type of a special combination that can function only if the LDMS checks at intervals, when the latest status reply has been received and when the previous status request has been transmitted. This check is performed each time a status request has been sent.

11.2.5 Video Recorder Service Request Table (VRSRT)

Many SLOs will result in automatically utilizing the video recorder when an alarm occurs. For instance, if an loud sound (a possible gunshot) occurs, the LDMS automatically instructs the video recorder to pan and zoom in the direction of the loud sound event. Multiple service requests to the video recorder may occur simultaneously, the LDMS must know which video recorder service request to be execute. Therefore, the LDMS must uphold a video recorder service request table (VRSRT). These requests:

1. must be properly organized according to their priority,

2. must be deleted from the table when serviced, either fully or partially,

3. the request interrupted by a request with higher priority must be suspended properly,

4. requests pending for excessive duration must be discarded from the table and,

5. therefore, the table must be automatically revised every second.

If high priority SLO occurs, the LDMS suspends the current service request at once and enables utilization of the video recorder. The previous service request is discarded from the table, because after the high priority SLO's service request has been handled, there is probably no need to revive the previous service request. It is most likely that the previous SLO that evoked the service request has already ceased to exist. However, this practice does entail the possibility of incurring a security risk. In the future, a rule-based inference could ordain if an old request or a suspended request should be serviced after the high-priority service request has been completed.

11.2.6 The Main Structure of the LDMS

The LDMS component-level structure is rendered in Figure 4. LdmsMain is the main executable, which has the responsibility for initializing the server. The SenderReceiver component administrates the message transmission and reception to and from the session server. This functionality is performed by the NetworkInterface component, which implements the transfer protocols. The RequirementHandler component includes the main logical functions for executing requirements. These requirements are parsed from a locally stored XML file using the RequirementXmlParser. In this context, "requirement" refers to the logical deduction requirements located in the LDMS requirement specification. Based on the XML file's requirement definitions, the LDMS responds to events and trigger its own timed requirements, for instance sensor status checks. DbHandler subcomponents handle the database operations required by the LDMS. The database contains the SLT for maintaining SLOs and the VRSRT for maintaining a list of video monitoring requests.

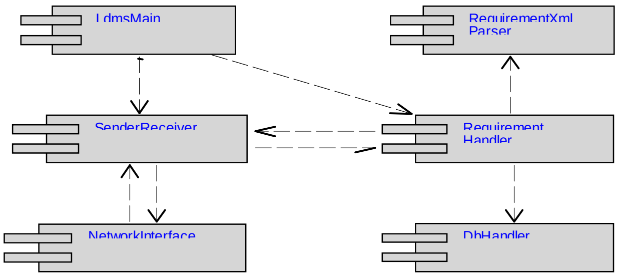


Figure 4, the component-level diagram of the LDMS structure.

An example of the XML requirement structure illustrates an LDMS logical decision making requirement, in this case In-Fingerprint-Sensor-Access-Granted-Opening-Hours. This is shown in Figure 5. The requirement has two conditions to execute its actions: the operating mode must be opening hours, and the fingerprint sensor must send a notification of an authorized entry through the session server. If these conditions are met, two actions are exerted: notifying the session server about the authorized entry during opening hours, and incrementing the fingerprint sensor tally.

```
<req>
 <req-name>In-Fingerprint-Sensor-Access-Granted-Opening-Hours</req-name>
...
 <req-type>sensor event</req-type>
 <cond-list>
  <cond>
   <cond-source>LDMS</cond-source>
   <cond-type>hours</cond-type>
   <cond-data>opening hours</cond-data>
  </cond>
  <cond>
   <cond-operator>AND</cond-operator>
   <cond-source>FingerprintSensor</cond-source>
   <cond-type>reply</cond-type>
   <cond-title>fingerprint sensor status</cond-title>
   <cond-data>Access ok</cond-data>
  </cond>
 </cond-list>
 <act-list>
  <act>
   <act-method>notify</act-method>
   <act-target>SessionServer</act-target>
```

```
    <act-data>access granted during opening hours</act-data>
  </act>
  <act>
    <act-method>requirement</act-method>
    <act-target>ldms</act-target>
    <act-data>In-Fingerprint-Sensor-Keep-Tally</act-data>
  </act>
 </act-list>
</req>
```

Figure 5, an example of LDMS's XML file.

12. SECURITY PERSONNEL MANAGEMENT

The Security Personnel Management composes of the SMSU and the end-device. The human operator of the SMSU initializes and shutdowns the SLSP system. The SMSU receives all the logical deductions from the LDMS and receives all the crude sensor information from the sensors of the Surveillance Domain through the session server. The end-device is the smart phone of the nomadic guard residing the area under surveillance.

12.1 The Security Manager Server

The SMSU receives all the logical deductions conducted by the LDMS. The SMSU may also receive all the crude information resulting from the sensors. The human security administrator may ordain the session server to transmit refined information from the LDMS and/or crude information from the sensors directly to the end-devices.

The human security administrator of the SMSU handles the initialization of the SLSP system. The SMSU forms a connection with the session server. The session server handles the initiation of the sensors and the LDMS. Then the end-device may register to the SLSP system. The human security administrator of the SMSU initiates the shutdown of the SLSP system. The session server handles the shutdown process regarding the sensor and the LDMS. The session server also deregisters the end-device during the shutdown process.

The SMSU provides the ensuing services:

- The User interface enabling/disabling the SLSP service, including enabling/disabling servers and sensors in SLSP.
- The views to display crude sensor information, e.g., video stream.
- The User interface for conducting the information that the end-devices receive, i.e., crude sensor information and LDMS deductions.
- The User interface provides an interface for the human administrator to communicate with end-device users, i.e., transmission/reception of textual messages.

12.1.1 The Main Structure of the SMSU

The SMSU consists of two components. The User Interface entails all the controls and views of the SMSU. To connect with session server the SMSU utilizes the NetworkInterface component, which offers socket level connections. The SMSU, including its user interface and their relationships with other components are described in the Figure 6. The SMSU controls allow the human operator of the SMSU to initialize and shutdown the session server and receive information.

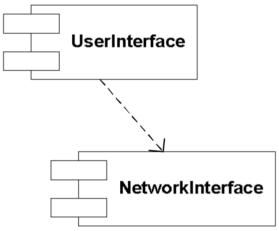


Figure 6, the component diagram of the SMSU.

12.2 The End Device

The end devices are perceived as the devices that are wielded by nomadic security guards. In the SLSP environment, the end-devices are Nokia N95 smart phones The end device has a direct connection to the session server. The end device registers to the session server. The session server upholds the session to the end-device. The end device provides a view for viewing video stream from video recorder. It also contains a view to register to the SLSP system, and thus the session server.

12.2.1 The Main Structure of the End Device

The end device comprises of the ensuing components. The UI component is the interface to the nomadic security guard. The UI entails a display for viewing messages from the LDMS. The UI has a display for viewing video from the video recorder. The Engine component incorporates event handlers for receiving and sending messages. The Engine component handles the initialization of the Communicator and Streamer components. The Streamer component administrates the video playback of the stream. The administration and playback of the actual video data is done by Symbian's Multi Media Framework included in the smart phone's APIs. The Communicator component is for receiving messages from the LDMS. The SERKET end device components and their relationships with other components are described in the Figure 7.

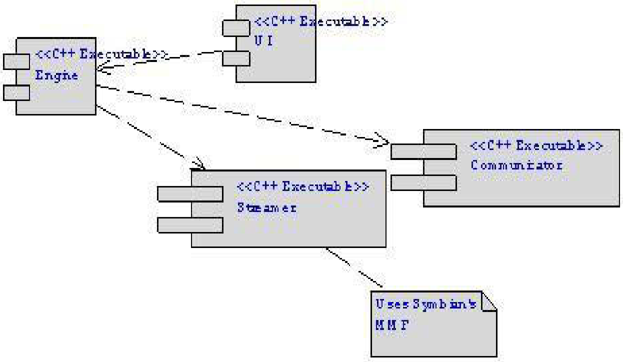


Figure 7, the component diagram of the end device.

13. COMPARISON OF THE SLSP SYSTEM TO THE STATE OF THE ART

The information and structure of the SLSP system was modeled on recent journals and conference papers regarding surveillance, especially focusing on multi-sensors systems, architecture and middleware, security personnel and logical decision making. The intent of the SLSP system is to address tedious surveillance task of human monitoring as [6] denotes. The SLSP aims at the proactive approach of [22], by providing a system entailing decision support to human operators. As according to [21], the SLSP automates the process of human surveillance without removing the human factor. The SLSP system is focused on indoor surveillance. The system contains a distinction between the most common end-users, as defined by [21], by incorporating the sedentary human surveillance operator and the nomadic guard.

In reflection to [10], the SLSP system addresses the vitality of situation awareness and real-time threat detection by immediately transmitting potential threats to the security personnel in the form of alarms. The system utilizes video streaming, the importance of which was emphasized by [6]. The SLSP system is a heterogeneous surveillance system as [4] defines. The SLSP system incorporates [16]'s ideas of acquiring complete information for surveillance support with multiple sensors delivering the information rapidly to surveillance personnel.

The LDMS addresses computability of identifying interesting events and the sensors of the SLSP system were selected according to their appropriateness. The system abides to [14]'s three essential parts of a generic surveillance and security system. Data is acquired from an indoor area. The LDMS performs the information analysis, in addition to the human security administrator conducing his own deductions. The on-field operation comprises of the automatic panning of the video recorder, and the actions of the human surveillance administrator and nomadic guards. The SLSP system coincides to [4]'s definition of multimedia surveillance system. The SLSP system is capable of collecting, processing, correlating and addressing multimedia data from multiple complimentary sensors. The system retains the human in the loop, therefore the SLSP system provides assistance to the human operator, as [5] and [13] attest. The SLSP system resolves the four aspects elicited by [11]. The system is capable of collecting data from multiple sources, is capable of managing the context information through delivery, entails reasoning support through the LDMS, and offers support for decision adaptations through the SMSU being able to ordain the information to be distributed to the end-device.

The SLSP system employs an audio sensor to discriminate the location of an audio event, as stated by [19]. The audio information ameliorates the utilization of video, as [18] indicates. We also utilize a scale regarding the distinction of a normal audio event from an abnormal audio event, advised by [18]. The SLSP system utilizes information assimilation of multi-sensor data to provide precise information as [1], [15], [12], and [8] indicate. The LDMS employs the utilization of multiple data sources to evaluate the situation of the surveillance point, as stated by [20] and [13].

14. CONCLUSION

Multi-sensor surveillance systems equipped with automatic logical decision is an important branch in the field of surveillance. With the utilization of multiple and diverse sensors, i.e., a fingerprint sensor, an audio sensor, a video recorder, and a network activity monitor, it is possible to automatically form deductions of a surveyed indoor area. We have illustrated the implemented design and communication how this endeavor is attained by automatically conducting logical decisions based on crude sensor data and distributing the deductions to the human end-users of the SLSP system, i.e., the human security administrator and the nomadic guard. In addition, the SLSP system can automatically react to its own alarms and issue instructions to one of its own sensors.

The deduction of logical decisions and distributing the alarms to surveillance personnel and, when required, automatically positioning the video recorder to the location of an appropriate alarm of the SLSP intends to 1) facilitate the deduction of alarms regarding an indoor surveillance point, and to 2) abate the amount of superfluous information rendered to the surveillance personnel. These two antecedently depicted aspects are the main endeavors of the SLSP's logical deductions and information distribution architecture. The SLSP system's sensor data is collected from a surveyed point, then the data is transmitted to the LDMS. The LDMS automatically deducts alarms based on the sensor data. The LDMS may issue instructions automatically to certain sensors based on its logical deductions, e.g., for the video recorder to be directed to a location of an event discriminated by another sensor. The deductions are transmitted to the human surveillance operator and to the end-device of a nomadic guard. The operability of the constructed SLSP system prototype indicates that these endeavor is attained.

15. REFERENCES

[1] Atrey, P.K., Kankanhalli, M.S., and Jain, R. 2005. Timeline-based Information Assimilation in Multimedia Surveillance and Monitoring Systems. VSSN'05, (November 11, 2005), Singapore, 103-112.

[2] Castanedo, F., Patricio, M.A., Garcia, J., and Molina, J.M.. 2006. Extending Surveillance Systems Capabilities Using BDI Cooperative Sensor Agents. VSSN'06, (October 27, 2006), Santa Barbara, U.S.A., 131-138.

[3] Collins, R.T., Lipton, A.J., Fujiyoshi, H., and Kanade, T. 2001. Algorithms for Cooperative Multisensor Surveillance. Proceedings of the IEEE, Vol. 89, No. 10, (October 2001), 1456-1477.

[4] Cucchiara, R. 2005. Multimedia Surveillance Systems. VSSN'05, (November 11, 2005), Singapore, 3-10.

[5] Detmold, H., Dick, A., Falkner, K., Munro, D.S., van den Hengel, A., and Morrison, R. 2006. Middleware for Video Surveillance Networks. MidSens'06, November 27-(December 1, 2006), Melbourne, Australia, 31-36.

[6] Fong, A.C.M. and Hui, S.C. 2001. Web-based intelligent surveillance system for detection of criminal activities.

Computing and Control Engineering Journal, (December 2001), 263-270.

[7] Greiffenhagen, M., Comaniciu, D., Niemann, H., and Ramesh, V. 2001. Design, Analysis, and Engineering of Video Monitoring Systems: An Approach and a Case Study. Proceedings of the IEEE, Vol. 89, No. 10, (October 2001), 1498-1517.

[8] Hall, D.L. 2003. Perspectives on the fusion of image and non-image data. Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd, (15-17 Oct.) 2003, 217 – 220.

[9] Hall, D. L. 2005. Challenges in Data Fusion: Dirty Secrets, Current State of Technology and a Research Roadmap. Associate Dean for Research School of Information Sciences and Technology, (February 28), 2005.

[10] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S., Senior, A., Shu, C.-F., and Tian, Y.L. 2005. Smart Video Surveillance. IEEE Signal Processing Magazine, (March 2005), 38-51.

[11] Hardian, B. 2006. Middleware Support for Transparency and User Control in Context-Aware Systems. MDS'06, (November 27 – December 1, 2006), Melbourne, Australia, (4-10?)

[12] Makris, D. and Ellis, T. 2005. Learning Semantic Sense Models from Observing Activity in Visual Surveillance. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 35, No. 3, (June 2005), 397-408.

[13] Nelson, C.L. and Fitzgerald, D.S. 1997. Sensor fusion for intelligent alarm analysis. Aerospace and Electronic Systems Magazine, IEEE, Volume 12, Issue 9, (Sept. 1997), 18 – 24.

[14] Ott, R., Gutierrez, M., Thalmann, D., and Vexo, F. 2006. Advanced Virtual Reality Technologies for Surveillance and Security Applications. VCRIA 2006, Hong Kong, (14-17, June 2006), 163-170.

[15] Petrushin, V.A., Gang, W., Ghani, R., and Gershman, A.V. 2005. Multiple Sensor Indoor Surveillance: Problems and Solutions. Machine Learning for Signal Processing, 2005 IEEE Workshop on, (28-30 Sept. 2005), 349 – 354.

[16] Regazzoni, C.S., Ramesh, V., and Foresti, G.L. 2001. Scanning the Issue/Technology. Special Issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems, Proceedings of the IEEE, Vol. 89, No. 10, (October 2001), 1355-1367.

[17] Reiter, M. and Rohatgi, P. 2004. Homeland Security Guest Editor's Introduction. IEEE Internet Computing, (November/December 2004), 16-17.

[18] Smeaton, A.F. and McHugh, M. 2005. Towards Event Detection in an Audio-Based Sensor Network. VSSN'05, (November 11, 2005), Singapore, 87-94.

[19] Stanacevic, M. and Cauwenberghs, G. 2005. Micropower Gradient Flow Acoustic Localizer. IEEE Transactions on Circuits and Systems-I: Regular Papers, Vol. 52., No. 10, (October 2005), 2148-2157.

[20] Steinberg, A.N. 2005. An approach to threat assessment. Information Fusion, 2005 8th International Conference on, Volume 2, (25-28 July) 2005.

[21] Valera, M. and Velastin, S.A.2005. Intelligent distributed surveillance systems: a review. IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, (April 2005), 192-204.

[22] Velastin, S.A., Boghossian, B.A., Lo, B.P.L., Sun, J., and Vicencio-Silva, M.A. 2005. PRISMATICA: Toward Ambient Intelligence in Public Transport Environments. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 35, No. 1, (January 2005), 164-182.

[23] Velastin, S.B. 2005. Special Section on Intelligent Distributed Surveillance Systems. IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, (April 2005), 191.

PUBLICATION 6

Testing and Validation of a Multi-sensor Distributed Surveillance System

Proceedings of the Seventh International Caribbean
Conference on Devices, Circuits and Systems,
ICCDCS 2008. Cancun, Mexico, 28–30 April, 2008.

Testing and Validation of a Multi-sensor Distributed Surveillance System

T. Räty, M. Luo, J. Oikarinen, and M. Nieminen

Abstract—The Single Location Surveillance Point (SLSP) is an distributed indoor multi-sensor surveillance system. It incorporates an arbitrary amount of sensors that collect data from a single location, which is the surveillance point. Each sensor collects information from its ambit. Once the crude data has been acquired from the sensors and transmitted to Logical Decision Making Server (LDMS) by the session server, the LDMS automatically performs logical deductions based on the data received from the sensors. The LDMS transmits the logical deductions to the human security administrator of the security manager server and/or the end-devices of the nomadic guards. Complicated surveillance systems possess rigid requirements regarding testing and validation. The main endeavor in testing and validating the SLSP system comprised of utilizing an adequate testing and validation process accompanied with sufficient tools to test the SLSP system. The testing and validation process is explained and testing is executed by utilizing two tools. The TCR (Test Case Runner) tool was specifically built for the testing purposes of the SLSP system. A proprietary tool, Nethawk's EAST (Environment for Automated Systems Testing) IMS (IP (Internet Protocol) Multimedia Subsystem) simulator, was utilized in testing. The operability of the constructed prototype accompanied with its successful testing and validation indicates that this endeavor is attained. The paper presents the testing and validation procedure, accompanied with the depiction of the testing tools and their utilization. The research is based on the constructive method of the related publications and technologies and the results are derived through testing and validating the implemented SLSP system.

***Index Terms*—Software testing, software verification and validation, information systems, and multi-sensor systems.**

I. Introduction

VIDEO monitoring systems are increasingly being employed in medium-scale shopping centers and in small shops. The intelligence functionality in modern systems presents the inherent issue of validation of the intelligent components to verify that the alarm generation software fulfills the user requirements. [1]

The usual scenario in an industrial research and development unit developing vision systems is that a customer discriminates a system specification and its requirements. The engineer then construes these requirements to a system design and validates that the system design fulfills the user-specified requirements. [2]

The SLSP system is a distributed indoor multi-sensor surveillance system. It entails multiple sensors, constituting of a biometrical sensor, a video camera, an audio sensor, and a network analyzing monitor. The sensors reside in an indoor area for surveillance. Each sensor acquires data from its environment and sends the crude data to the session server. The session server administrates all the connections pertaining to the components of SLSP. The session server transmits the received sensor data to the LDMS. The LDMS automatically deducts the surveillance point's situation predicated on the sensor data it receives from the session server. The deductions are sent to the human surveillance operator of the security manager server and to the end device of a nomadic guard.

A distributed indoor multi-sensor surveillance system imposes specific challenges to testing and validation. The testing and validation process of a distributed indoor multi-sensor surveillance system is explained. The tests were conducted with two different testing tools. The TCR tool, which was constructed for our testing purposes, concentrates on unit/component testing. The proprietary tool, Nethawk's EAST simulator, was utilized to conduct integration/release testing. The operability of the constructed SLSP system prototype and its successful validation and test execution indicates that this endeavor is attained.

The structure of this paper is the ensuing. First a general overview of contemporary surveillance systems is presented. This is followed by a depiction of the challenges regarding the validation and testing of multi-sensor surveillance systems. A concise corollary of the SLSP system is presented next. Then we present our testing and validation process. A presentation of the TCR and EAST follows. The conclusion summarizes the paper.

Manuscript received December 19, 2007.

T. Räty is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, phone: +358 20 722 2131; fax: +358 20 722 2320, e-mail: tomi.raty@vtt.fi

M. Luo is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: miao.luo@vtt.fi

J. Oikarinen is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: johannes.oikarinen@vtt.fi

M. Nieminen is with VTT Technical Research Centre of Finland, P.O. Box 1100, 90571-FIN Oulu, Finland, e-mail: mikko.nieminen@vtt.fi

II. Surveillance Systems' Generations

Electronic video surveillance systems belong to three generations. Basic scientific discoveries allowed surveillance video devices to be progressively developed. CCTVs (close circuit TV systems) can be considered as the beginning point for online surveillance. First generation surveillance systems (1GSS) (1960-80) fundamentally extend human perception capabilities in a spatial sense. Video data from a collection of cameras viewing remote scenes are presented to the human operators after analogue communication of the video signal. The enhanced resolution of video cameras and the availability of low-cost computers are two fundamental breakthroughs for video processing and detection of events. This technological evolution corresponds to second generation surveillance systems (2GSS) (1980-2000). The main endeavor of third generation surveillance systems (3GSS) is to offer "full digital" solutions to the design of surveillance systems, beginning at the sensor level, up to the presentation of mixed symbolic and visual erudition to the operators. The main objective of full digital 3GSSs is to ease the efficacious data communication, management, and extraction of events in real-time video from a large collection of sensors. [1]

III. Challenges of Validating Surveillance Systems

The design, development, and deployment of 3GSS systems in the real world are influenced by miscellaneous factors, including the validation that the system designed is according to user requirements. A substantial pitfall in incorporating these intelligence functions in real-world systems is the inability to test and validate these systems under a variety of use cases. Testing and validation of these systems is costly and tedious, because of the manual labor required in validation. [1]

The design process composes of two basic steps. These are 1) the choice of the system architecture and the modules, and 2) the statistical analysis and validation of the system to check if it fulfills user requirements. In real life, the system design and analysis phases usually follow each other in a cycle until the engineer procures a design and a suitable analysis that satisfies the user specifications. [2]

Vu et al. present a modeling framework for the visualization and simulation of automatic video interpretation. This framework, called test framework, must be adequately supple (configurable) for testing the different configurations of the interpretation system. This test framework will be an efficacious tool for the developers and for the experts of the application domain, e.g., agents of security. [3]

When the implications of failures are considerable, the analyst cannot be fulfilled with 'average' predictions, but must have assurance that the required performance is achieved [4]. With the development of modern technologies and the evolution of the industrial society, sundry engineering systems have been developed and are becoming more intricate [5]. Many of such systems, including national defense systems, require frequent surveillance by a certified individual, who must identify potential problems and retain the systems performing satisfactorily [5].

Pavlidis et al. state that good laboratory technology should be supported by profound knowledge of the business, market, and user realities to become a success. Considering the practical realities, Pavlidis et al. states that a prototype should be developed and tested in an actual environment. This would proof its ultimate suitability. [6]

IV. Challenges of Testing Surveillance Systems

Testing of individual modules is called unit testing. Integration testing comprised of rerunning the unit test cases after the system was completely integrated. For feature testing, which is also called system testing, testers developed test cases predicated on the system's requirements. They chose adequate test cases for every expected result to occur. One subphase of load testing is stress testing. Stress testing comprises of verifying the software's ability to ply heavy loads for short periods without crashing. [7]

Advancements in sensor, communications and storage capacities render it more facile to gather large corpora of multimedia material. Carincotte et al. concentrate on the extraction of structured knowledge from multimedia collections recorded over a network of camera and microphones. The multimedia streams they induce, in addition to surveillance and safety issues, could possibly delineate a useful source of information if stored and automatically analyzed. The ultimate goal of the Carincotte et al. was to examine current and novel technologies, by assessing them in a real test case. [8]

V. Testing and Validating the SLSP system

The SLSP is an indoor distributed multi-sensor surveillance software system. It contains an arbitrary amount of sensors that cull data from a single location, which is the surveillance point. Each sensor distributes its crude sensor data to a session server, which administrates the connections between the components. The session server conveys the crude sensor information to the LDMS. The LDMS automatically deducts the situation at the surveillance point based on the received crude sensor information. The LDMS dispatches information of the situation at the surveillance point to the security manager server. The security manager server's user interface displays inherent information about the surveillance point to a human security administrator. The security manager server can issue information transmission from the session server to the security personnel's smart phone. The SLSP system abates the amount of redundant information that would be handled by the human security administrator. The SLSP system is illustrated in Figure 1.

The main chain comprises of the session server, the security manager server, LDMS and end device.

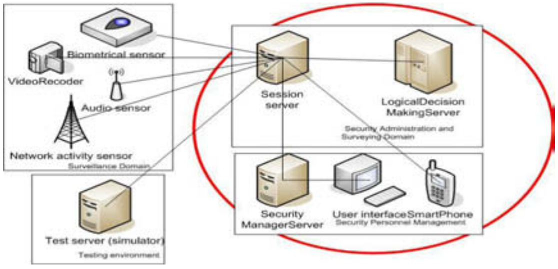


Figure 1, the Single Location Surveillance Point (SLSP) with the main chain circled in red.

The Single Location Surveillance Point comprises of three individual domains, which are required for SLSP to be functional. These three domains are 1) the Surveillance Domain, which comprises of an arbitrary amount and variety of sensors, 2) the Security Administration and Surveying Domain, which comprises of the session server to which the sensors transmit their information and the LDMS, and 3) the Security Personnel Management domain, which comprises of the security manager server, its UI (User Interface), and the end device to which the appropriate information is ultimately dispatched.

The Testing Environment is utilized during the Surveillance Domain's, the Security Administration and Survey Domain's and the Security Personnel Management’s development phases. The test server may be utilized to test the servers either directly or indirectly. The session server is tested directly by the test server. The LDMS, security manager server and end device are tested indirectly through the session server by the test server. The test server providing artificial sensor information on behalf of the other sensors, if required and viable. The realization of the test server is Nethawk’s EAST IMS simulator.

The main chain comprises of the session server, the security manager server, LDMS, and the end device. The integration/release testing chain comprises of the sensor in question with the main chain. I.e., the integration/release testing chain of the video recorder comprises of the video recorder and the main chain. The integration/release testing chain must be tested with sensor in question bi-directionally, if applicable. All the sensors transmit information to the session server. Bi-directional indicates that the sensor also receives information originating from the main chain, e.g., the biometrical sensor receiving a status request from the LDMS.

Testing the SLSP system is divided into two distinctive testing levels, as illustrated in Figure 2: 1) unit/component testing, and 2) integration and release testing. In regard to testing the entire system, the integration and release testing is divided into subcategories according to sensor/sub-system: 1) main chain, 2) the biometrical sensor, 3) audio sensor, 4) video recorder, and 5) network activity monitor.

The main chain is tested individually. Nethawk’s EAST simulator may be utilized to test the main chain. Nethawk’s EAST simulator is capable of executing tests from the sensor perspective and verifies that the main chain performs correctly according to artificial sensor input.

After the integration/release testing of each individual sensor or sub-system has been conducted, the integration/release testing of the entire SLSP is administrated. Testing partial segments of the SLSP is permitted in prefatory phases. Testing the entire SLSP includes executing all the tests of each individual sensor or subsystem and additional tests that require multiple sensors and/or subsystem to form consolidated tests.

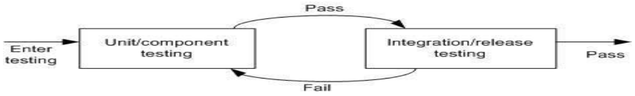


Figure 2 depicts the testing levels in which integration/release test cases are executed after component test cases.

Completed integration/release test cases result in a software release with descriptions of possible software errors. Detected errors should be corrected for the next software release. The other test activities, such as test planning, test preparation and designing test, are executed in parallel between the testing levels.

A. *Unit/component testing process in detail*

The purpose of the unit/component testing is to test individual functions and classes within the component to locate possible memory leaks, check error handling, and to test the majority of the branches of the system. The units/components shall be tested utilizing both valid and invalid values and states.

Unit/component testing is performed by the component designer or person responsible for testing. The testing procedure is informal containing debug testing in both the emulator and device environments. The TCR tool was developed for testing some particular components of the SLSP system. After the component has cleared the unit/component testing process without any major or known malfunctions the component can be integrated as a part of the SLSP system.

B. *Integration and release testing process in detail*

The integration and release testing process is similar to the unit/component testing process in which individual units/components are tested. In the integration and release testing, the focus is on the interfaces between the components and the testing is conducted for either the whole system or large compositions of the system. The integration should exerted in layers instead of attempting to integrate all the components as a whole system at once. The integration testing is executed first by testing the main chain. Then each sensor and sub-system is included individually to the SLSP and tested. Early tests were performed in an emulated environment, retaining the emphasis always on actual devices. For instance, the S60 3rd Edition SDK for Symbian

OS Supporting Feature Pack's emulator was utilized.

During testing, a test report, which is part of the test cases document, is maintained by inserting the outcome of each selected test case. If outcome is negative, indicating that the system contains errors, the reason and outcome must be described in the test report. Once all test cases are successfully completed a release is created. If release contains errors, they should be corrected by the next release.

C. *Testing environment*

The testing environments comprise of Windows XP computers running the different servers' software and the Nokia N95 mobile client. The mobile client software is tested with the emulator software and the actual hardware. This is applicable to both unit/component testing and integration/release testing. The integration/release testing of the sensor requires the utilization of the hardware in question. I.e., performing the integration/release tests of the video recorder requires the utilization of the actual video recorder.

The sensor hardware composition comprises of the following: 1) Deltabit Gatekeeper Lite (biometrical sensor), 2) AXIS 213 PTZ (video recorder), 3) 4 x AKG C562CM microphones, AKG B29L power supply, Edirol UA-101 audio interface, and 60cm X 60 cm panel (comprisal of the audio sensor), and 4) Nethawk M5 Analyser (network activity monitor).

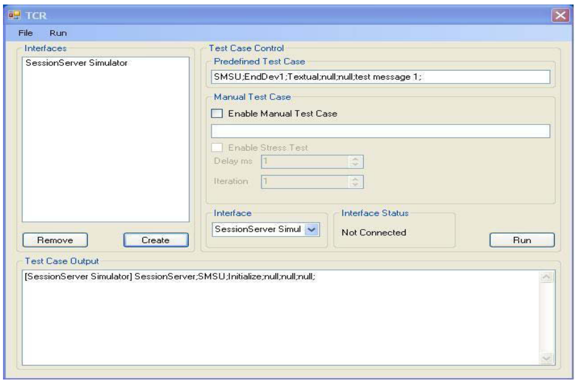
Integration and release testing is mainly conducted by the user through user interface, i.e., through the user interface of the security manager server and the end device. This applies at least with launching the software and activating some of the test cases. Some of the tests might be executed in the background automatically, for example in cases in which sensors produce information for the main chain to handle and to which to respond appropriately. Testing the SLSP system utilizes the Nethawk EAST simulator and the TCR tool, which was constructed for testing the SLSP system.

VI. Conducting tests with the TCR tool

Tests are conducted with the TCR tool. The TCR tool was built for testing the SLSP system. The tool is capable of simulating sensors, servers or end devices perceived as a part of the SLSP.

The TCR tool enables its users to establish either listening or connecting socket interfaces, through which the test data is being relayed to the test target. In the SLSP test environment, the test data is textual and can be either loaded from a text file or typed manually to the user interface of the TCR tool. The outcome of the test, providing that the test case results in a reply, can be espied from the output screen of the tool.

Figure 3 illustrates the main user interface of the TCR tool. The interface area enables a user to establish listening or connecting interfaces. Only one interface can be employed to transmit test cases to the tested systems at a time, but all the interfaces are capable of receiving data from tested systems in parallel.



Interfaces: SessionServer Simulator

Predefined Test Case: SMSU;EndDev1;Textual;null;null;test message 1;

☐ Enable Manual Test Case

☐ Enable Stress Test

Delay ms: 1

Iteration: 1

Interface: SessionServer Simul

Interface Status: Not Connected

Test Case Output: [SessionServer Simulator] SessionServer;SMSU;Initialize;null;null;null;

Figure 3, the TCR's main interface.

The Interface configurator dialog is launched from the main interface. The dialog requires name, address and port for the interface to be created. The user must define the interface's type, either listening or connecting. Antecedent to initiation of the interface, the validity of the IP address is verified and if the interface is connecting type, then the existence of the IP address is also verified. To conduct tests, at least one interface must be created, which is connected to the tested system or to which the tested system connects. Test cases, either manual or predefined, are executed by selecting the interface and pressing the "Run" button of the user interface.

The test case control displays tests cases that are executed. The predefined test case region displays test cases loaded from a file. In the figure 3's panel "Test Case Output", the format of test cases is illustrated. A single test case is discriminated in one line of the file. Predefined test cases are loaded by utilizing "Open Test Case" option from the "File" menu. To execute a test case, either manual or predefined, the interface through which the test case is transmitted must be chosen from the drop down list. The manual test case needs to be typed or a predefined test case needs to be loaded before performing tests.

The manual test case enables the user to type and execute test cases manually. The manual test case contains a stress test function. With the stress test function, it is feasible to perform a manual test case 1 to 1000 times with intermission varying from 1ms to 1s.

After executing either a manual or predefined test case, the TCR tool waits for the reply of the test case. The reply is displayed in the output window of the main user interface. If a predefined test is executed, then the next test case is loaded, if a reply is received to the test case. The user can transit to the next test case, if no reply is received, by selecting the "Cancel wait" option. This is also applicable for the manual test case option. Predefined test cases may be skipped before execution by selecting "Skip Current" option.

VII. Testing components with the TCR tool

In many cases, to test functionalities, messages must be transmitted over the network to the SUT (System Under Test). Functionalities of some systems may be tested through the user interface provided, these systems are the security manager server, the end device and the network activity monitor. An example of a test case is illustrated in Table 1.

Table 1, example of a test case.

Test system	Security Manager Server
Test case group	Initialization of the Security Manager Server
Test case identification number	SMS-Init
Test objective	Launch the Security Manager Server
Test procedure	Ensure that a socket connection to the Session Server is created at start-up
Test summary	OK

A. *Testing the Session Server*

Testing of the session servers internal functionalities lies on receiving messages from the network through the session server's interfaces. Some of the session server's interfaces are static. This indicates that the configuration remains unchanged and unalterable. These interfaces include the interface for the security manager server and the interface for the end device to register. Sensor interfaces are defined in the initialization file.

B. *Testing the Security Manager Server*

Half of the tests for the security manager server can be executed from the UI of the security manager server. To perform some of the tests completely from the UI, some input from the network is required. Therefore, the TCR tool is appropriate for testing the security manager server's functionalities. The TCR tool needs to create a listening interface to which the security manager server connects.

C. *Testing the end device*

The tests for the end device are similar to the security manager server's tests. The UI can be tested, but a network connection is required to discern if the messages resulting from the UI utilization are transmitted. To test the end device, the TCR tool can be employed by establishing two interfaces to which the end device can connect. One interface simulates the session server's register interface and another interface conducts the assignments of the actual interface to which the rest of the messages, such as sensor registration change, etc., are transmitted.

VIII. Conducting tests with Nethawk's EAST IMS simulator

Nethawk's EAST IMS simulator is a test automation and traffic generation tool, which enables users to emulate or simulate network elements in the telecommunications network. EAST offers a programmable testing platform, including test creation, execution and reporting, to simulate miscellaneous test scenarios. Both manual and automatic tests can be developed and executed. EAST offers capabilities for regression testing and load testing. [9]

A. *A test case for a RTP session*

A RTP (Real-Time Protocol) streaming session is elaborated to impart the usage of EAST. The fundamental idea is that EAST simulates a RTP server, which offers a video or audio streaming session. This allows the application under test to negotiate with RTP server to establish a streaming session. EAST is mainly employed for testing against protocol standards, therefore the protocol-based servers must be activated and simulated on EAST. In this example, a simple RTP session is created that playbacks the video stream locally. Once the "regression run" command is selected from the UI, a TC (Test Case) Runner will be displayed. The "Play" button will instigate testing. The user may select different types of views to check the progression of the session.

B. *A test case for the initialization of the Session Server*

This test case utilizes the EAST to simulate the security manager server that transmits an initialization message to the session server. The sequence of the session server's initialization messaging process is expressed as the ensuing three steps: 1) the session server is started, 2) the initialization message is received, and 3) the session server is halted. In detail, the session server is activated. Then the server receives the initialization message. If the message is appropriate, the interface is registered.

The key step for the antecedent test case is described in EAST as following. To transmit the initialization message to session server from the EAST tool, the resource for EAST must be defined. In this case, an EAST resource for TCP transmission is required. First, the TCPSocket is defined and assigned an appropriate IP address and port.

Then the "Test Case Editor" needs to be launched. The Test Independent Objects (TIO) are building blocks that are located in the Toolbars of the Logical Editors and enable users to designate the state machine that delineates the logic of the test case. The TIOs are independent of any particular protocol, network element, or test scenario. The TIOs need to be selected into the panel. The ones selected are "Start", "Variable", "Resource", "ASCII Send", and "End" (Success). Figure 4 illustrates the view of the Test Case Editor.

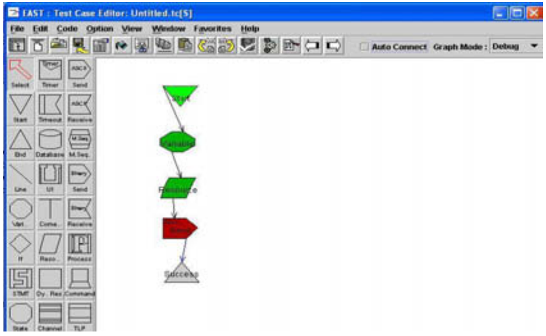
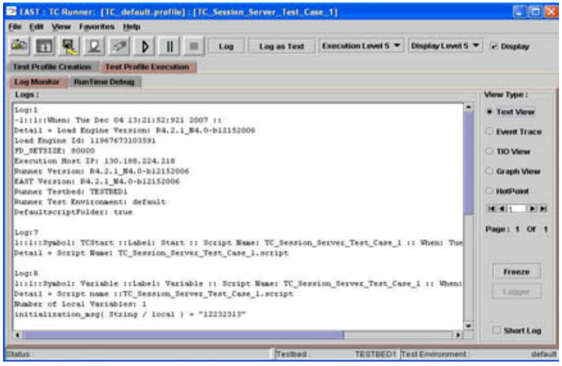


Figure 4, the view of the Test Case Editor accompanied with the selected TIOs.

The “Variable” TIO is filled in with the ensuing information: {Type: String, Name: Initialization_msg, Value: “<SessionServer; SMSU; Initialize; null; null; null;>”, Scope: Local}. The “Resource” TIO is given the local identifier name “TCPSocket” and the Logical Address is set to TCPSocket, which has been already created in EAST’s configuration. The “ASCII Send” TIO is set to “initialisation_msg”, which has been already specified in “Variable”. Hence, the registration message is transmitted to the session server by employing the buffer defined in “Variable”. This is conducted by such means because the message is manual data (unformatted), unlike the ones defined by protocol standards, such as RTP. A code check, supported by the EAST tool, should be executed to verify that the setup is correct. Then the tool’s “Regression Run” function should be executed. This launches the TC Runner. By pressing “Play”, the test is executed. Results are displayed on the Log Monitor, as illustrated in Figure 5.



```
Log:1
-1::1::When: Tue Dec 04 13:21:52:921 2007 ::
Detail = Load Engine Version: R4.2.1_N4.0-b12152006
Load Engine Id: 11967673103591
FD_SETSIZE: 80000
Execution Host IP: 130.188.224.218
Runner Version: R4.2.1_N4.0-b12152006
EAST Version: R4.2.1_N4.0-b12152006
Runner Testbed: TESTBED1
Runner Test Environment: default
DefaultscriptFolder: true

Log:7
1::1::Symbol: TCStart ::Label: Start :: Script Name: TC_Session_Server_Test_Case_1 :: When: Tue
Detail = Script Name: TC_Session_Server_Test_Case_1.script

Log:8
1::1::Symbol: Variable ::Label: Variable :: Script Name: TC_Session_Server_Test_Case_1 :: When:
Detail = Script name ::TC_Session_Server_Test_Case_1.script
Number of Local Variables: 1
initialization_msg( String / local ) = "12232313"
```

Figure 5, the results of the test case are displayed in the Log Monitor.

IX. Conclusion

The information and structure of the SLSP system’s validation and testing processes were modeled on recent journals and conference papers regarding surveillance, especially on multi-sensor surveillance systems. [1] declare that an inherent pitfall of real-world systems reside in the inability to test and validate the systems. [6] announce that a prototype should be developed and tested in an actual environment. The validation and testing processes of the SLSP system approach the real-world testing and validation, by constructing a testing environment that simulates a real-world environment with independent testing tools that can completely simulate the sensor outputs of the SLSP system. The sensors themselves are not tested, because they are proprietary products. These testing tools enable artificial inputs to the main chain that are difficult, or completely infeasible, to emulate in the real environment. Through these comprehensive testing and validation processes, we are able to attain considerable and thorough coverage. [3] created a test framework for achieving diverse configurations for testing the system. The validation and testing processes of the SLSP system contained configurable test cases and the simple execution of regression tests. As according to [7], we utilized unit/component and integration/release testing. The TCR tool enables the employment of stress testing. We utilize also multimedia information as [8], and we endeavor to exploit the multimedia knowledge in the SLSP system.

Multi-sensor surveillance systems entail strict requirements regarding testing and validation. We have illustrated our validation process, comprising of both unit/component and integration/release testing procedures. The tests were executed with two tools, the TCR, which was built specifically for testing the SLSP system, and Nethawk’s EAST simulator, which is a proprietary tool. With our testing and validation processes, accompanied with the tools, we were able discriminate the operability of the constructed SLSP prototype. This attests that our endeavor to successfully formulate a testing and validation procedure, accompanied with tools, is attained.

References

[1] Regazzoni, C.S., Ramesh, V., and Foresti, G.L.: Scanning the Issue/Technology Special Issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems, Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1355-1367.

[2] Greiffenhagen, M., Comaniciu, D., Niemann, H., and Ramesh, V.: Design, Analysis, and Engineering of Video Monitoring Systems: An Approach and a Case Study, Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1498-1517.

[3] Vu, V.-T., Bremond, F., and Thonnat, M.: Human Behaviour Visualisation and Simulation for Automatic Video Understanding, The 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2002, WSCG 2002, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 4-8, 2002.

[4] Marseguerra, M., Zio, E., and Podofillini, L.: Optimal Reliability/Availability of Uncertain Systems via Multi-Objective Genetic Algorithms, IEEE Transactions on Reliability, Vol. 53, No. 3, September 2004.

[5] Pham, H. & Xie, M.: A Generalized Surveillance Model with Applications to Systems Safety, IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, Vol. 32, No. 4, November 2002.

[6] Pavlidis, I., Morellas, V., Tsiamyrtzis, P., and Harp, S.: Urban Surveillance Systems: From the Laboratory to the Commercial World, Proceedings of the IEEE, Vol. 89, No. 10, October 2001.

[7] Avritzer, A., Ros, J.P., and Weyuker, E.: Reliability Testing of Rule-Based Systems, *IEEE Software*, September 1996, 0740-7459/96/$05.00 © 1996 IEEE.

[8] Caricotte, C., Desurmont, X., Ravera, B., Bremond, F., Orwell, J., Velastin, S.A., Oboez, J.M., Corbucci, B., Palo, J., and Cernocky, J.: “Toward Generic Intelligent Knowledge Extractions from Video and Audio: The EU-funded CARETAKER Project”, The IET conference on Imaging for Crime Detection and Prevention (ICDP 2006), pp. 470-476, London, Great Britain, June 13-14, 2006.

[9] Nethawk Ltd.: EAST user manual release 4.2: Introduction to EAST. Nethawk, Sept 2006.

Series title, number and
report code of publication

VTT Publications 691
VTT-PUBS-691

Author(s)
Räty, Tomi

Title

Architectural Improvements for Mobile Ubiquitous Surveillance Systems

Abstract

Surveillance systems have begun to be integrated into the common lives of humans and improved surveillance systems will spread even further. Systems manufacturers will continue to provide powerful surveillance systems with different aspects from single sensors to an abundance of different intelligent sensors. These different devices will have the ability to deliver a large variety of information to either remote or local surveillance personnel for immediate utilization or for extracting information about occurred events. The objective of this dissertation is to analyse the reduction of excessive information delivered to security personnel and the immediate delivery of essential alarms to security personnel by refining a design of a distributed multi-sensor intelligent surveillance system. The surveillance system created is reflected against the mobile and ubiquitous requirements of the end users of the surveillance system. The mobile requirement contains the reduction of excessive information distributed to the end user, a.k.a., the service personnel. The ubiquitous requirement consists of sensor data fusion and situation deduction. This dissertation uses a constructive research method, in which the results are validated by technical implementation and experimentation against mobile and ubiquitous requirements.

The major results of this dissertation are the prototype implementations of the Single Location Surveillance Point (SLSP) system. It consists of a selective amount of sensors that collect readings from a single location, which is the surveillance point. Each sensor transmits its crude sensor data to a session server, which handles the connections between the components. The session server routes the crude sensor information to the logical decision making service. The logical decision making server automatically deducts the situation at the surveillance point based on the received sensor information. The logical decision making server informs the security manager server of the situation at the surveillance point. The security manager server's user interface displays essential information about the surveillance point to a human security administrator. The security manager server can transmit information to the nomadic security personnel's smart phones over wireless networks.

ISBN
978-951-38-7113-0 (soft back ed.)
978-951-38-7114-7 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN	Project number
VTT Publications 1235-0621 (soft back ed.) 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)	19166

Date	Language	Pages
September 2008	English	105 p. + app. 55 p.

Name of project	Commissioned by

Keywords	Publisher
surveillance systems, mobile ubiquitous surveillance systems, Single Location Surveillance Point, logical decision making servers, multi-sensor systems, intelligent surveillance systems, video surveillance, video analysis, Quality of Service, audio surveillance, system architecture	VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland Phone internat. +358 20 722 4520 Fax +358 20 722 4374

VTT PUBLICATIONS

678 FUSION Yearbook. Association Euratom-Tekes. Annual Report 2007. Eds. by Seppo Karttunen & Markus Nora. 2008. 136 p. + app. 14 p.

679 Salusjärvi, Laura. Transcriptome and proteome analysis of xylose-metabolising *Saccharomyces cerevisiae*. 2008. 103 p. + app. 164 p.

680 Sivonen, Sanna. Domain-specific modelling language and code generator for developing repository-based Eclipse plug-ins. 2008. 89 p.

681 Kallio, Katri. Tutkimusorganisaation oppiminen kehittävän vaikuttavuusarvioinnin prosessissa. Osallistujien, johdon ja menetelmän kehittäjän käsityksiä prosessin aikaansaamasta oppimisesta. 2008. 149 s. + liitt. 8 s.

682 Kurkela, Esa, Simell, Pekka, McKeough, Paterson & Kurkela, Minna. Synteesikaasun ja puhtaan polttokaasun valmistus. 2008. 54 s. + liitt. 5 s.

683 Hostikka, Simo. Development of fire simulation models for radiative heat transfer and probabilistic risk assessment. 2008. 103 p. + app. 82 p.

684 Hiltunen, Jussi. Microstructure and superlattice effects on the optical properties of ferroelectric thin films. 2008. 82 p. + app. 42 p.

685 Miettinen, Tuukka. Resource monitoring and visualization of OSGi-based software components. 2008. 107 p. + app. 3 p.

686 Hanhijärvi, Antti & Ranta-Maunus, Alpo. Development of strength grading of timber using combined measurement techniques. Report of the Combigrade-project – phase 2. 2008. 55 p.

687 Mirianon, Florian, Fortino, Stefania & Toratti, Tomi. A method to model wood by using ABAQUS finite element software. Part 1. Constitutive model and computational details. 2008. 51 p.

688 Hirvonen, Mervi. Performance enhancement of small antennas and applications in RFID. 2008. 45 p. + app. 57 p.

689 Setälä, Harri. Regio- and stereoselectivity of oxidative coupling reactions of phenols. Spirodienones as construction units in lignin. 104 p. + app. 38 p.

690 Mirianon, Florian, Fortino, Stefania & Toratti, Tomi. A method to model wood by using ABAQUS finite element software. Part 2. Application to dowel type connections. 55 p. + app. 3 p.

691 Räty, Tomi. Architectural Improvements for Mobile Ubiquitous Surveillance Systems. 2008. 105 p. + liitt. 55 p.

Julkaisu on saatavana

VTT
PL 1000
02044 VTT
Puh. 020 722 4520
http://www.vtt.fi

Publikationen distribueras av

VTT
PB 1000
02044 VTT
Tel. 020 722 4520
http://www.vtt.fi

This publication is available from

VTT
P.O. Box 1000
FI-02044 VTT, Finland
Phone internat. + 358 20 722 4520
http://www.vtt.fi

ISBN 978-951-38-7113-0 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7114-7 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

VTT PUBLICATIONS 691 Architectural Improvements for Mobile Ubiquitous Surveillance Systems Tomi Räty