

Review of existing smart video surveillance systems capable of being integrated with ADDPRIV project

Deliverable 2.1

GDANSK



Automatic Data relevancy Discrimination for
a PRIVacy-sensitive video surveillance





Automatic Data relevancy Discrimination for a PRIVacy-sensitive video surveillance

SEC-2010.6.5-2 - Use of smart surveillance systems, data protection, integrity and sharing information within privacy rules

D2.1 – Review of existing smart video surveillance systems capable of being integrated with ADDPRIV project

Due date of deliverable: 31-07-2011

Actual submission date: 31-07-2011

Start of project: 01 February 2011

Duration: 36 Months

Lead Contractor for this deliverable: GDANSK

Revision: 1.0

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination level		
PU	Public	+
CO	Confidential, only for members of the consortium (including the Commission Services)	

Revision History

Deliverable Administration and summary		
Project Acronym: ADDPRIV		Grant Agreement no: 261653
Document Identifier: ADDPRIV_20113107_WP2_GDANSK_Scoreboard_R11.pdf		
Leading partner: GDANSK		
Report version: 1.1		
Report preparation date: 28-07-20011		
Classification: Public		
Nature: Report		
Author(s) and contributors: GDANSK, KU		
Status		Plan
		Draft
		Working
	+	Final
		Submitted
		Approved

The ADDPRIV consortium has addressed all comments received, making changes as necessary. Changes to the document are detailed in the change log table below.

Date	Edited by	Status	Changes made
20.07.2011	GDANSK	Beta version	
25.07.2011	KU	Reviewed	
28.07.2011	GDANSK	Corrected	
31.07.2011	ANOVA	Quality check	

Copyright

This report is © ADDPRIV Consortium 2011. Its duplication is allowed only in the integral form for anyone's personal use for the purposes of research and education.

Citation

GDANSK, KU (2011). Deliverable 2.1 – Review of existing smart video surveillance systems capable of being integrated with ADDPRIV. ADDPRIV consortium, www.addpriv.eu

Acknowledgements

The work presented in this document has been conducted in the context of the EU Framework Programme project with Grant Agreement 261653 ADDPRIV (Automatic Data relevancy Discrimination for a PRIVacy-sensitive video surveillance). ADDPRIV is a 36 months project started on February 1st, 2011.

The project consortium is composed by: Anova IT Consulting (ANOVA), Kingston University Higher Education Corporation (KU), Politechnika Gdanska (GDANSK), Lancaster University (ULANCS), Avanzit Tecnologia, S.L. (AVANZIT), Hewlett Packard Italiana Srl (HP), Società Per Azioni Esercizi Aeroportuali Sea SPA (SEA), Renfe Operadora (RENFE) and The Provost Fellows & Scholars Of The College Of The Holy And Undivided Trinity Of Queen Elizabeth Near Dublin (TCD).

More Information

Public ADDPRIV reports and other information pertaining to the project are available through ADDPRIV public website under www.addpriv.eu

Table of contents

Revision History	3
Acknowledgements	5
1. Executive summary	7
2. Introduction	8
3. Existing smart surveillance systems	9
3.1. Commercial system solutions analysis	9
3.2. Commercial smart cameras solutions analysis	30
3.3. Similar projects solutions.....	35
3.4. Surveillance systems summary.....	44
4. Relevant algorithms.....	45
4.1. Object detection.....	46
4.2. Object tracking	49
4.3. Object classification	52
4.4. Event detection.....	57
4.5. Route reconstruction.....	65
5. Potential application of individual algorithms to the ADDPRIV project	68
6. Conclusions	69
7. References	70

1. Executive summary

The purpose of this deliverable is to provide an in depth analysis and cataloguing of existing smart video surveillance solutions which may be relevant for integration with the algorithms developed in ADDPRIV project.

2. Introduction

The ADDPRIV project addresses two main issues within video surveillance. First is the employment of intelligent image processing algorithms for facilitating the work of system operators. This leads to the second issue, related to improving social acceptance of surveillance systems, which can be achieved by limiting the amount of stored recordings only to the most relevant ones, in terms of security issues. Both of these matters are closely related, since the decision about certain video segments of being relevant or non-relevant can be made based on the implemented algorithms. This document is devoted to the description of several security related solutions.

In the first section, existing complete commercial systems as well as smart camera devices are presented. All the indexed system solutions are compared in terms of the offered features and the characteristics of system architecture. This description is followed by an overview of existing smart surveillance cameras which are based on state of the art algorithms for automated intelligent video processing. Also, as a part of Sec. 3, a description of projects related to video surveillance and security matters in general, is presented. Currently ongoing as well as already finished research projects are mentioned.

In turn, in Sec. 4, algorithms which were developed by ADDPRIV partners, in particular by GDANSK and by KU, are described. The presented methods are considered for integration with ADDPRIV systems such as event detection and route reconstruction module.

Finally, a discussion about some selected modules potential applications to the ADDPRIV project is included.

3. Existing smart surveillance systems

In the first part of this overview a number of the most popular commercial intelligent surveillance systems is described. Main solutions which were identified in extensive research of the market as well as systems developed by ADDPRIV Partners in the past, are included. Subsequently, several so called smart surveillance cameras with integrated visual processing modules are presented. This description is followed by an introduction to similar security related projects.

The specifications which are discussed in the following sections contain a number of quotations from producers' technical specifications, advertising materials and other descriptions. To improve clarity of this document, the quotation marks were omitted, thus the paragraphs containing a significant number of citations are shaded with grey background. However, it is assumed that all copyrights for the used descriptions are still owned by the companies or consortia related to the products and research projects named at the beginning of each subsection.

3.1. Commercial system solutions analysis

The solutions described in this section are devoted to commercial systems which involve automatic detection of defined events on the basis of video processing. This could be acquired either by a supplied software or supporting it with some hardware. In general, such systems include several basic devices i.e. for data acquisition, conversion and storage. Additionally the end user software is delivered either as a web interface or more typically as a dedicated application simplifying the system maintenance and/or management. The video processing is implemented based on the advanced technologies which were developed by each company. In some cases, black box devices are offered where each camera (one or more) is plugged in. Alternatively, dedicated high performance servers are delivered to support multiple video streams analysis. The descriptions for several chosen products are shown in the following subsections. They include general system features as well as various system configurations,.

3.1.1. Praetorian

Company website: <http://www.l3praetorian.com/>



Fig. 1. Praetorian system software

The turnkey software provides for swift, protective action in even the most complex and densely populated sites. When deployed, public and private sector operators can benefit from: greater information management and situational awareness; the ability to minimize costs; reduced liability, labour, and risk of loss. As an open-architecture and scalable system, Praetorian integrates with many other surveillance technologies, including commercial off-the-shelf (COTS) PC hardware and software. Praetorian allows operators to see the total picture, and understand real-time threats, so they can act preemptively to stop or contain emerging threats. Praetorian's open-architecture, scalable software improves security by providing:

- Superior situational awareness – Video Flashlight™ immerses the operator in the environment and allows them to follow suspicious activity or objects automatically from camera to camera. Remote Flashlight provides mobile, rapid-deployable, remote solutions – Integrates video, video alarms, or other sensors with man-portable, wireless platform to monitor drug and crime interdiction, domestic violence, special operations, event security, terrorist attacks, and situational awareness at major incidents
- Improved perimeter security – Integration of wired or wireless video, video alarms, fence, radar, and access control inputs into a central monitoring station creates a proactive real-time beyond-the-perimeter situational awareness with VisionAlert™ and Hawk™
- Improved responsiveness – VisionAlert™, tailored analytics, integrates with many other sensors. It reports events as they begin and provides immediate notification of unauthorized activity or threats as they occur
- Force multipliers – Intelligent surveillance frees up resources for other duties until unauthorized activity is in progress and response is required
- Rapid Retrieval & Playback – Synchronized video from multiple cameras available improving forensics capability

Praetorian dramatically improves situational awareness of potential threats, improves response capabilities that minimize operational and training costs, reduce liability and risk of theft/loss, and ultimately improves security.

Video Flashlight™

- Multiple stitched video feeds combined into a common operating picture
- Real-time 3D immersive display of the surveillance environment
- Superior situational awareness, information management, and operational control
- Mobile flashlight ability allows users to access any camera in the network from a PDA or laptop

Video Flashlight™ enables operators to “virtually patrol” security areas easily, moving throughout the 3D environment. Users navigate from indoors to outdoors, and around corners, flying from rooftop to street level, to view both live and recorded video from different perspectives.

System specifications:

Video Flashlight™ is designed to run on Microsoft® Windows 2003 Server and Windows XP Professional operating systems. Typical deployments consist of:

- Praetorian Flashlight™ software
- HP ProLiant DL145 1U Rackmount Server or equivalent work station*
- Dual 2.8GHz AMD Opteron Processors
- 2GB System RAM
- 120GB SCSI Hard Disk Drive
- DVD+RW/CD+RD Combo Drive
- Dual 10/100/1000 baseT NIC
- Redundant Hot-Plug Power Supplies
- NVIDIA 6800-series 256Mb VRAM GPU

Remote Flashlight:

- Mobile devices show 3D display for enhanced situational awareness for mobile forces
- GPS used to determine position of PDA
- PTZ camera control from PDA
- Multiple PDAs work together
- Network connectivity
- Full-screen laptop or remote workstation, client server architecture

VisionAlert™

- Customizable alarm configurations
- Real-time detection for motion, breach, loiter, and left behind object
- Works with many legacy cameras: fixed, PTZ; IP, analog; infrared, EO, thermal, and more

VisionAlert™ offers real-time detection of motion, breach, loitering, left behind objects, and can incorporate many manufacturers' alarms. Through customizable, user-defined alarm configurations, the software delivers users greater situational awareness and provides real-time data for preemptive action. Unlike other video analytic software, VisionAlert™ is capable of programming all four detection categories and multiple alarms on a single camera — providing a virtual surveillance shield. The alarms can be tailored to meet different surveillance needs, such as wide area surveillance, tailgating, objects thrown over a fence, abandoned vehicles, speeding vehicles, zone surveillance, removed objects, and graffiti. VisionAlert™ continuously compensates for camera movement. Environmental factors such as rain, snow, shadows, and swaying branches are all factored out to reduce the rate of false alarms. The software uses polygons, easily drawn by the operator inside a camera's screen-view, to define specific areas of concern for each alarm type.

Individual alarms can be configured, changed, enabled or disabled with a click of a mouse. As an open-architecture platform, VisionAlert™ integrates with many legacy camera technologies such as fixed, Pan/Tilt/Zoom (PTZ); IP, analog; thermal, EO, and more, as well as commercial off-the-shelf (COTS) PC hardware and software.

System specifications:

VisionAlert™ is designed to run on Microsoft® Windows 2003 Server and Windows XP Professional operating systems. Typical deployments consist of:

- Praetorian VisionAlert™ software:

- HP ProLiant DL145 1U Rackmount Server or equivalent work station*
- Dual 2.8GHz Xeon™ Processors
- 2GB System RAM
- 00GB SCSI Hard Disk Drive
- DVD+RW/CD+RD Combo Drive
- Dual 10/100/1000 baseT NIC
- Redundant Hot-Plug Power Supplies
- Winnov 4400VO Video Capture Cards (required only for analog video capture)

Custom architectures and systems for specific environment requirements and applications can also be designed.

3.1.2. GeoVision

Company website: <http://www.geovision.com.tw>



Fig. 2. GeoVision system features

GeoVision offers a diverse product line with industry-leading technologies:

- Digital Surveillance Systems ranging from the cost-effective 20 fps to the leading-edge 480 fps models, choices ranging from BNC to D-Sub, built-in to standalone I/O modules
- Continuous improvement of S/W, H/W compression technology to keep competition edge
- IP surveillance product line
- Integration of IT technology into surveillance system
- Video analysis features for surveillance system
- Digital Surveillance Systems with expandable supports to POS, Central Monitoring Station
- License Plate Recognition Systems
- Access control system
- A complete lineup of security accessories that help security professionals to deliver customized services to users

As to the video analytics a wide range of functionality is introduced, including the following:

- Face detection
- People counting

- Unattended object detection
- Missing object detection
- Camera tampering detection
- Video stabilization
- Crowd detection

3.1.3. Mate

Company website: <http://www.mate.co.il/>

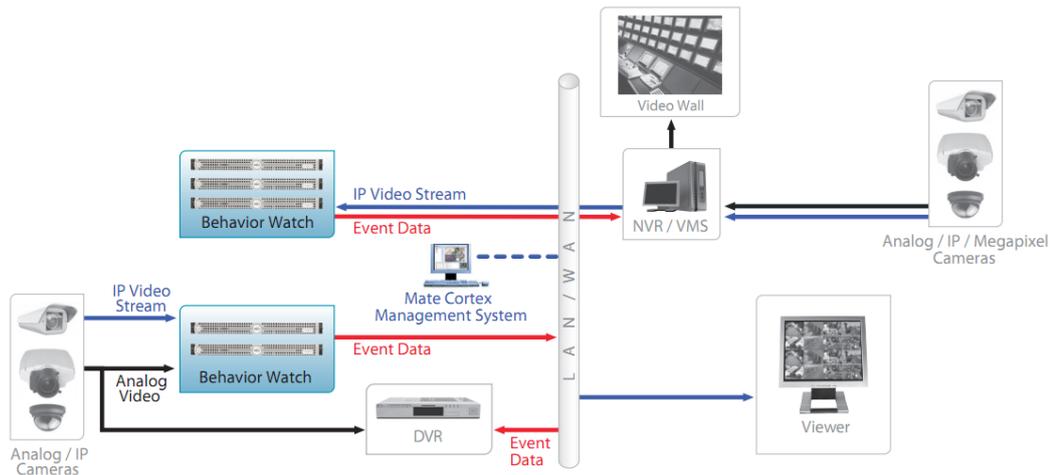


Fig. 3. MATE system architecture

A comprehensive package of integrated and highly scalable products that individually and together provide timely answers to the most pressing security issues:

- Ensuring that review time focuses on highly relevant, real-time events
- Overcoming human tendency toward losing vigilance over time
- Giving security personnel the means to deliver greater value
- Using highly automated and reliable technology
- Applying a system architecture with easy growth potential
- Transferring bi-directional data between remote sites and Control Center

There are many places where an existing security infrastructure cannot be upgraded with 'smart on-site' devices. The answer is to add this capability within the Control Center. The Behaviour Watch collects all video surveillance content inputs and analyses it for potential access violations and other specified behaviours. Only those events of security value are then passed for human intervention.

The Behaviour Watch is equipped to handle a range of surveillance tasks. It detects the difference between video objects and tracks suspect human behaviour. It provides a new level of intelligence that makes alerts more precise and relevant.

3.1.4. Westec

Company website: <http://www.westec.net/>



Fig. 4. WESTEC system operator station

As an industry leader and innovator in the field of Intelligent Video Surveillance, Westec has helped a wide range of businesses in the restaurant, convenience, jewellery and retail industries do a better job of protecting their employees and their profits. The value we can provide extends well beyond safety issues, to operational, merchandizing, productivity and customer service enhancements.

Interactive Video Monitoring Services from Central Command Center

- 24/7 Video Verification to identify alarm events & disturbances
- 2-way audio into sites allowing Westec to intervene & resolve store disruptions
- Video business audits report on operational performance

Remote Access Capabilities

- Web-based portal allows users to view live & recorded video from any Internet-based location
- Provides business owners tremendous management flexibility & efficiency
- Monitor multiple locations at one time

Industry-Leading Software

- Web-based portal allows users to access video from all locations with one-click
- Easy to read reports called dashboards highlight relevant data, making review easy & efficient
- Video is saved for 30 days & is transferable for permanent storage

The core of the Westec video surveillance system is a state-of-the-art Digital Video Recorder (DVR). A feature-rich surveillance platform, the DVR offers far more than traditional CCTV monitoring and analog video recording. The DVR is driven by a powerful software engine that controls digital video display and playback (both in multiplex view) and that also serves as the integration and control point for existing security systems.

Capable of supporting as many as 16 surveillance cameras, the DVR is the central navigation point for the security system. Using the on-screen

interface, you can view live digital video streams, access archived video for multiplex playback, and access system configuration screens.

DVR Features:

- Health Management – The DVR supports health management reporting, which actively communicates the status of critical DVR components throughout the day
- Server Security – The DVR offers an integrated approach to securing the data, overlaying multiple server/network security options
- Enterprise Reporting - Several reporting options are available in the system, including reports that can be generated on-demand and precompiled reports that can be uploaded for online display
- Remote System Access – Remote connection solutions available for the Westec system provide monitoring access options that are the next best thing to being on site in person
- Optional Component Integration – The DVR integrates easily with a broad range of optional hardware and software solutions

3.1.5. Technest

Company website: <http://www.genextech.com/>

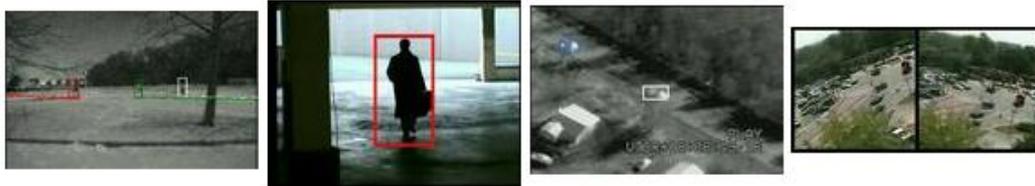


Fig. 5. TECHNEST system video examples

Surveillance systems have been limited and passive for most of the industry's history. Building from years of experience in wide field-of-view imaging and detection algorithms, Technest is now introducing capabilities that far exceed traditional fixed cameras. These innovations are already transforming the way to solve problems. The strength of a surveillance system depends upon the reliability of its algorithms. The quality of a system's algorithms can make a difference between a threat being detected as a critical alarm or a harmless spot of noise. Genex has leveraged its core expertise from research in motion detection, 3D data processing, 3D facial recognition, and image processing to develop a leading set of intelligent surveillance algorithms.

Smart SuiteT algorithms by Genex are a cutting-edge portfolio of advanced video analysis and augmentation modules. Technest team of researchers and product engineers have worked together to take innovative approaches to some of the most challenging problems in image processing.

Solutions currently on the market cannot perform reliably due to numerous influences, including: weather effects, noise distractions, moving platforms, and day/night shifts. Genex is solving these problems by using both new methods and new high speed processing platform to achieve results that were previously unattainable. The Smart SuiteT family of algorithms includes:

- Image Enhancement and Stabilization
- Image Fusion
- Motion Detection

- Object Tracking and Classification
- Target Positioning
- Sound and Vibration Detection
- Custom Detection Zones
- Pan-Tilt-Zoom Control

3.1.6. NUUO

Company website: <http://www.nuuo.com/>

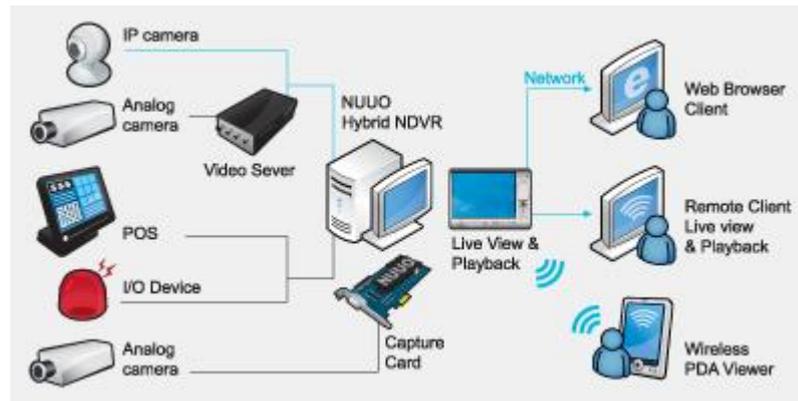


Fig. 6. NUUO system architecture

NUUO PC Based NVR (IP+) is a flexible and cost-efficient solution which can manage IP/Megapixel cameras. NUUO NVR adopts open platform technology supporting as many as 52 brands of IP/Megapixel cameras with more than 1100 models.

NUUO NVR comes with a slew of unique functions including NUUO intelligent video solution (IVS), event detections and intelligent playback search. Other handy tools include intuitive GUI recording schedule, E-map, video enhancement tools, and mobile (3G) support. POS and I/O device integration are nothing short of supply too. NUUO NVR is supported by authentic NUUO Central Management System which is a true monitoring and management solution that supports unlimited number of cameras.

Main console is NUUO NVR's recording server. It can display live video and configure the system. NUUO NVR has an intuitive design which can be learned very quickly. The main functions include the following:

- Stream multiple live videos from IP cameras and IP video servers
- Unique GUI recording schedule
- 6 event detections and 10 instant alarm responses (example, email alert and E-map popup)
- Duplicate channel display with digital PTZ

NUUO Main Console can both display live video and configure the recorder. NUUO playback system makes browsing records smarter and faster through intelligent search. Save video recompresses the image into ASF/AVI format. Video enhancement tool can be used to sharpen, brighten or even grey scale the images. Varies logs keep a record of all the events. Easy to use multi-channel backup application for archiving audio and video locally or remotely. NUUO backup system can also take snapshots of the recordings and store them in a separate location. Remote live viewer client supports 64 channels per monitor and can view up to 128 cameras in dual

monitor systems. Remote live viewer can display live video from multiple NUUO servers (NVR, DVR, NDVR, NVRmini) simultaneously. NUUO remote live viewer features PTZ control, E-map and I/O panel.

3.1.7. MTS

Company website: <http://www.mts-iss.com/>



Fig. 7. MTS system video examples

MTS Intelligent Surveillance Solutions is a complete technology and security design and integration firm specializing in intelligent and integrated digital video surveillance (DVS) and access control solutions.

We incorporate intelligent video features and true internetworking to enable clients to establish a security architecture for cost-effective, unattended surveillance and facility/premise security. MTS provides both strategic and tactical design and implementation services focused on several key markets with specifically focused service offerings.

MTS can provide the following Analytics & Integration capabilities:

- License Plate Recognition
- Face Recognition
- People Counting
- Detect Directional Flow
- Detect Object Removal
- Point of Sale
- Access Control
- Panic & Alarm

MTS believes that a "Safe Community" video surveillance architecture is the responsibility of all community members, from elected officials and police through the citizens of the community. As such, there is an opportunity to share the costs and share the benefits. Increased public safety is desirable in virtually every community yet sometimes difficult to achieve. Video Surveillance offers an affordable and acceptable mechanism to enact improvements in public safety.

3.1.8. MDS

Company website: <http://www.mobiledigitalsecurity.com/>

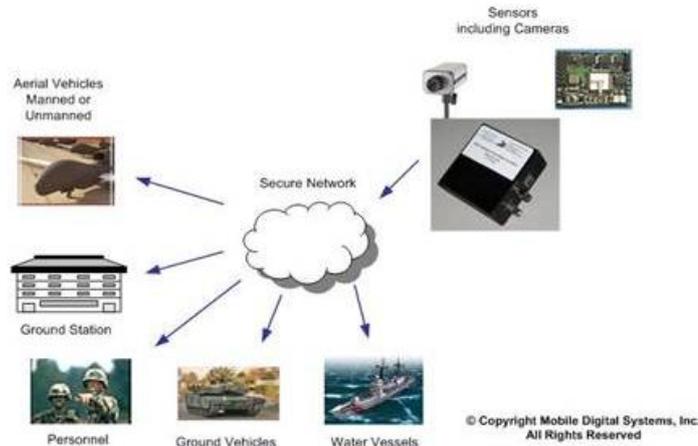


Fig. 8. MDS sample system configuration

Mobile Digital Systems provides Services and Products for Intelligent Detection and Surveillance Solutions for In-Vehicle and Fixed Site applications in Wireless environments for Military, Homeland Security, Transportation, and Public Safety.

MDS' own MDS Intelligent Comprehensive Technology (MDS-ICT) includes Software and Detection Algorithms which operate on COTS / GOTS/ or newly engineered Hardware. MDS offers a complement of Services including Design, Engineering, Development, Installation, and Support.

MDS Intelligent Surveillance System (MDS-ISS) is a Fixed Site embedded system which delivers Motion and Content Analytics on Video and Multiple Sensor Data in Real-Time. MDS-ISS is plug and play which means little or no integration effort.

MDS-ISS is flexible on Customers' preferences. For instance, all information could be stored remotely, and only actual Event information is transmitted. MDS-ISS is appropriate for any critical fixed site, including Sensitive National Security Locations, Water Treatment and Water Supply, Power Stations, and others. Developed with advanced Algorithms, MPEG7 and XML, MDS-ISS is a plug and play solution, requiring little or no integration, is fully networking and internet compatible, and can be configured with Any wireless communications technology.

3.1.9. COGNIMATICS

Company website: <http://www.cognimatics.com/>

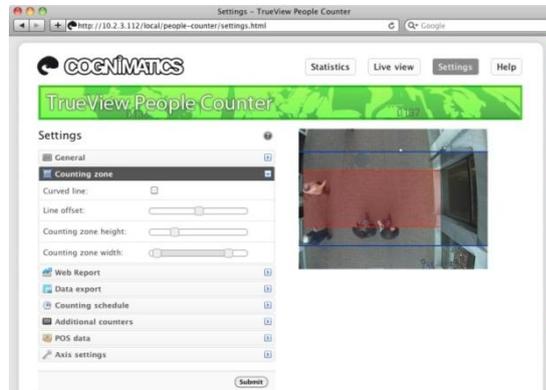


Fig. 9. COGNIMATICS software tool configuration

Cognimatics is a global leader in Intelligent Video and Image Analysis. The company offers efficient and innovative imaging products for Intelligent Retail Surveillance and the Mobile Communications industry. In today's competitive retail environment, success is dependent on a thorough understanding of existing and potential customers. Constantly changing customer preferences, eroding customer loyalty and the inherent complexity of large retail organizations demand increased analysis of the customer behaviour such as:

- Visitor traffic to store
- Where do shoppers go in the store
- What do they look at
- How many are queuing at the checkout

The TrueView™ product suite from Cognimatics is the leading camera based product suite for customer behaviour analysis. Using standard network cameras loaded with Cognimatics leading technology for intelligent video it provides the retail industry with powerful tools for automatic customer behaviour analysis. TrueView People Counter™ is an automated stand-alone, two way people counting system for ceiling mounted cameras, that is powered by Cognimatics' leading patent pending image processing software. When not running TrueView People Counter™, cameras can also be used for standard purposes, like for example surveillance. Furthermore, accessing the camera over IP is very cost efficient for maintenance, where you can manage camera and people counting remotely. This includes downloading or streaming of video remotely. TrueView People Counter™ runs directly on the Axis IP camera, and data can be retrieved for analysis on timed intervals as frequent as one minute. The software is modular and completely autonomous with all counting done on the camera's CPU, requiring no dedicated PC. Manage, analyse and view data from any number of units and from multiple sites, using TrueView Report™ or TrueView Web Report™.

System features:

- Automated system, operated in real time, fully embedded into Axis IP cameras.

- Easy to install and setup
- Maintain people counter remotely over IP, set and monitor parameters, download or stream video
- Seamless integration with TrueView Report™ or TrueView Web Report™
- Push XML counting data automatically from camera to SQL based TrueView Web Report™
- Unlimited number of cameras to a site or portfolio of sites when combined with TrueView Report™ or TrueView Web Report™
- Two way counting: Counts people moving in two directions simultaneously
- Accurate counting even under high density conditions or with baby carriages or trolleys present
- Leading digital image processing minimizes shadow and reflection problems
- Export data from TrueView Report™ to Microsoft Excel (CSV), or as PNG
- Open protocol lets you integrate with data from POS and other systems

3.1.10. Aralia System

Company website: <http://www.araliasystems.com/>

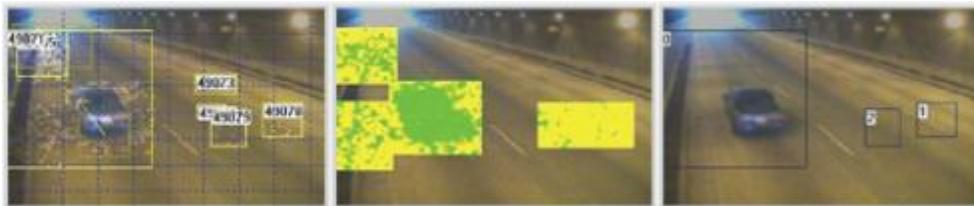


Fig. 10. Aralia system video processing results

Ilex – Advanced Video Surveillance – performs advanced video surveillance. It provides all the functions of video surveillance including analytics, recording, viewing and rapid historical search. Its key features also include full Geographic Information Systems (GIS) and the use of relational databases. Ilex provides all the functions of video surveillance, including analytics, recording, viewing and rapid historical search. It is an advanced, intelligent and scalable video surveillance system that can integrate with existing CCTV infrastructures, as well as the latest digital systems.

Ilex sends video data to a distributed relational database, incorporating real-time cataloguing within video picture, automatic reporting of alarms and rapid SQL-based querying of the entire stored video data. All configuration and video data, including analytic metadata, are stored in the relational databases. The innovative approach permits Ilex to utilise the querying and data processing capabilities of a large scale SQL database. Pioneered by Aralia, this approach gives Ilex a clear advantage over competing systems.

System features:

- Records into a distributed relational database at evidential quality.
- Real-time cataloguing of picture content
- Automatic reporting of alarms

- Video data reduction uses 'intelligent' data thinning
- Rapid SQL-based querying of the entire stored video data
- Object analysis on PTZ cameras
- Ability to use multiple compression technologies
- Runs on any type of platform including blade servers with a SAN or NAS

3.1.11. HASAM

Company website: <http://www.hasam.eu/>



Fig. 11. HASAM system video examples

HASAM develop variety of reliable video surveillance technologies. Video management systems integrated with HASAM command center solutions combined with latest technologies such as thermal imaging systems, video analytic systems and long range laser illuminated cameras provides excellent results in military and home land defence applications. Low cost video surveillance recorders are on other hand very efficient solution for standard security applications in office buildings or housing areas.

3.1.12. Honeywell

Company website: <http://www.honeywell.com>

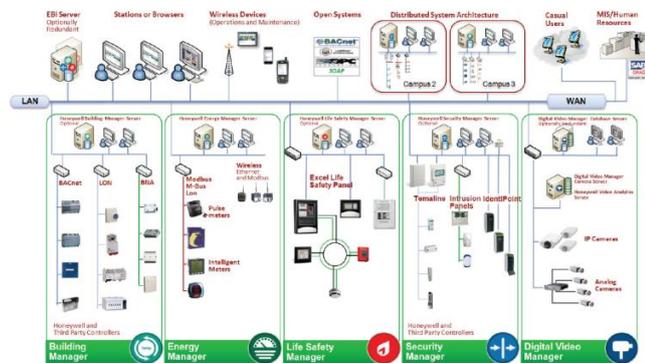


Fig. 12. Honeywell system architecture

Honeywell offers a wide range of fully configurable solutions starting from a typical but complete surveillance systems including digitizers for analog signals, camera controllers, storage servers and user applications for managing the whole system, finishing with smart solutions through the developed software. Honeywell’s Video Analytics product allows for automatic detection, tracking and behaviour classifications the behaviours of people and vehicles as they move through a scene. Three general software solutions are offered:

- Active Alert – a tool for real-time analysis which provides automatic alert triggering on the basis of user definable rules
- Smart Impressions – a piece of software for real-time video analysis which enables the user to acquire the information about the number of people or vehicles entering or existing the monitored scene
- People Counter – a tool which beyond Smart Impressions features offers additional traffic data and the ability to detect camera failures or sabotage

3.1.13. Ipsotek

Company website: <http://www.ipsotek.com>

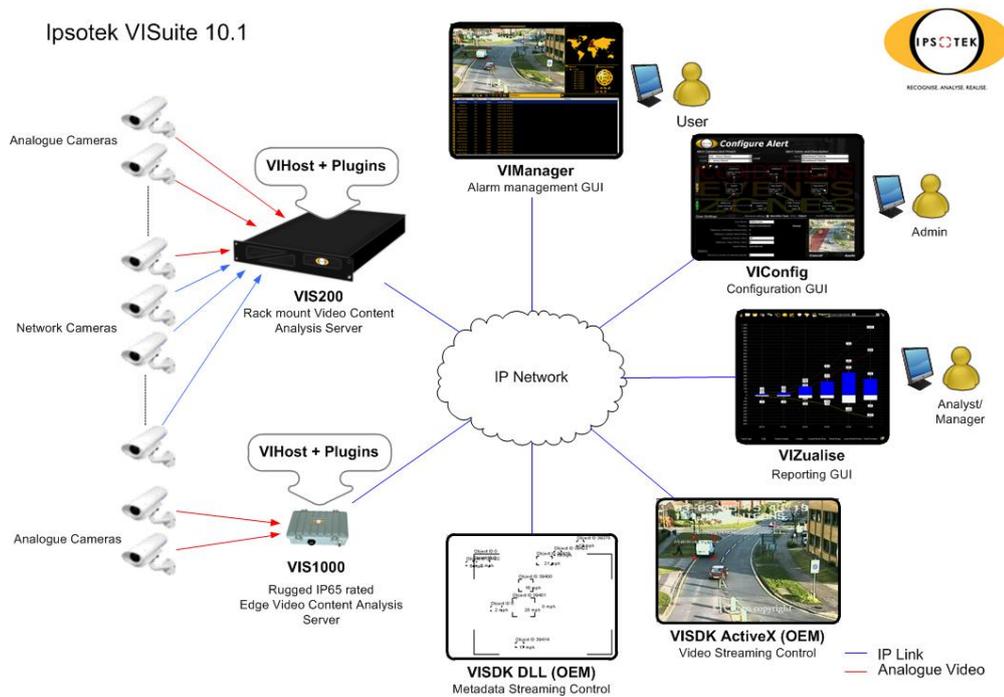


Fig. 13. System components logical relations within Ipsotek product

VISuite 10.1 is a hybrid VCA solution that combines the benefits of both rule and learning based approaches to eliminate the disadvantages of either approach. The solution has a rule based Video Content Analysis engine that has a wide range of VCA tools designed to detect a variety of behaviours. The detected behaviour is then analysed by a statistical filter that relies on the operator feedback to refine the performance of the solution. This way the VISuite solution is operational as soon as the rules have been defined and its performance will continuously improve based on operators’

annotation of the detected alarms. System components logical relations within Ipsotek product

VISuite 10.1 is a scalable Video Content Analysis (VCA) system designed for flexibility, high performance and low false alarm rate. The following are the key features of the product:

- Scenario based event detection that guarantees low false alarm rate.
- Supports both analogue and IP video inputs.
- Combined video and external triggers in one behaviour refining definition of alarms to new levels.
- Detect multiple events per camera and combine triggers from multiple cameras in one alert.
- Simple and efficient configuration procedure using virtual targets for camera calibration and an intuitive drag and drop scenario definition GUI.
- High performance scene specific object classification capability.
- High performance scalable image processing based on the NXP (Philips) Trimedia Digital Signal Processor chip.
- A toolbox of configurable algorithms which allows the system to be applied across a wide range of applications from intruder detection to traffic management.
- Web-hosting service that allows customers to access alarms and statistics from anywhere at any time.
- System architecture that provides seamless integration with 3rd party systems to offer a complete solution to end users.
- Real-Time database functions that allow customers to monitor critical parameters such as site occupancy without the need for restrictive people controls.
- Super-compression of video streams into metadata that fully automates forensic analysis in a fraction of the time needed today.

Distributed processing architecture makes the system robust and less prone to an IT attack, more flexible for installation and maintenance and virtually accessible from anywhere to view real-time alerts and reports.

3.1.14. Vi-System

Company website: <http://www.agentvi.com/>

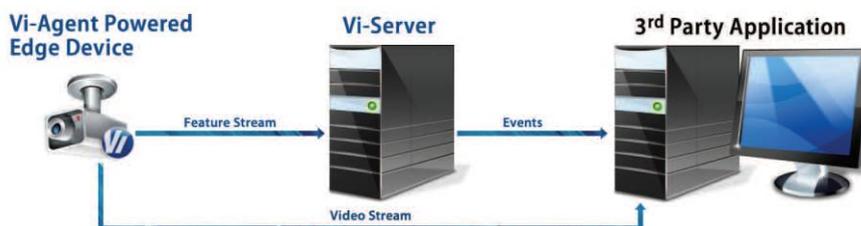


Fig. 14. Vi-System simplified architecture

Vi-System identifies and generates alerts for a variety of user-defined events relating to people, vehicles and static objects, by performing real-time analysis of the video stream. Used for applications such as security, safety and business intelligence, Vi-System offers effective monitoring of multiple video sources in parallel, enabling automatic detections, alerts and responses to events, as they emerge.

Based on Agent Vi's open architecture, pure software approach, Vi-System can be easily integrated with a wide range of edge devices and video management systems, in both new and existing surveillance networks. Vi-System boasts the combined benefits of superior detection performance, high scalability, installation simplicity and ease of use, making it the most advanced, comprehensive and cost effective real-time video analytics solution on the market.

Vi-System offers a wide and flexible range of analytics detection functions related to People, Vehicles and Static Objects. Such functions are available on both fixed and PTZ cameras. Supported features:

- An unlimited number of analytics rules of any kind and combination can be applied to each camera
- The analytics rules can be applied to any number of video channels simultaneously
- Complex detection scenarios can be generated by combining or linking detection rules

3.1.15. Sentient

Company website: <http://sentientai.com/>



Fig. 15. iSentry Active PTZ camera control advantage presentation

Sentient offers two basic iSentry solutions: one devoted to transport and the other to the security surveillance. iSentry Transport Surveillance provides leading edge video analysis module that learns and reports on unusual behaviour in traffic surveillance video. iSentry is ideal for traffic surveillance. Most traffic surveillance video is very repetitive and of little interest. iSentry learns the normal behaviour and will monitor the traffic 24 hours per day. If unusual activity is observed, iSentry can immediately record the event and notify an operator to request their interpretation of the behaviour and the appropriate response. Because iSentry filters traffic video so well, iSentry allows one operator to effectively monitor up to 100 cameras in real-time. iSentry allows you to set up a single camera surveillance system at accident 'black spots' in the road system and record all incidents over a period of 6 months.

System features:

- Detects unusual activity
- Roadway lane Infringement
- Stopped Vehicle and Debris Detection
- Congestion Monitoring
- Multiple camera analysis and management

- Active PTZ camera control
- Reliable outdoor unattended object detection

3.1.16. Citilog

Company website: <http://www.citilog.com>



Fig. 16. Example of object detection from MediaCity solution

Citilog focuses on delivering advanced software-based video monitoring products that address the growing global needs to more efficiently manage traffic, safety and security. Its MediaTunnel, MediaRoad and VisioPaD products are ideal for instant and automated identification of incidents, in real time, across any infrastructure environment: bridges, roadways, secured areas, tunnels and more. The products provide proactive video image processing that identifies incidents when they happen, so there is still time to react and prevent further escalation of the incident or correct incidents to rapidly restore optimum operations, such as traffic flow. The MediaTD product complements this advanced traffic manageability by providing measurable traffic statistical data on roadways and in tunnels to help facilitate future traffic management and planning. MediaCity is Citilog's flagship presence detection product, providing high level identification functionalities for intersections management. It is ideal for any large or main intersection in a city where traffic flow management is critical to efficiency, safety and the security of its users.

MediaIntruder is Citilog's flagship intrusion and security product ideal for a wide variety of environments including, buildings, airports, seaports, railways, and other areas where verifying and authenticating the presences of individuals is critical to safety and security. It provides real time 4D intrusion detection – highlighting the incident on the screen for fast and easy assessment – for asset or resource protection against threats. Citilog's software-based solutions require no additional hardware installation, using existing installed cameras and devices already deployed. Alternatively, Citilog works with technology partners to provide infrastructure hardware as needed. Products are provided through a worldwide network of distributor and reseller partners. Citilog's solutions are in use today at hundreds of sites worldwide for instant incident detection at bridges, roadways and tunnels in some of the world's leading cities such as London, Madrid, Melbourne, New York, Paris, Fort Lauderdale, Hong Kong, Lisbon, Sao Paulo, Shanghai and many other locations.

Another example of the offered version of system is called MediaCity which was created to work under any condition. It provides a lot of

functionalities to dramatically improve safety and reduce congestion at intersections:

- Presence detection feature provides customizable data to traffic controllers about the presence or absence of a vehicle within a specified zone. It detects both moving and stopped vehicles.
- Directional sensitive presence detection to discriminate vehicle detections based on the travelling direction of a vehicle through the detection zone
- Stopped vehicle on a specific lane or in the center of the intersection
- Actual queue length for vehicle stopped at the stop bar
- Average vehicle waiting time at the stop bar
- Pedestrian detection, triggered when a movement by pedestrians is detected within a detection zone

3.1.17. Bosch Intelligent Video Motion Detection (IVMD 1.0)

Company website: <http://products.boschsecuritysystems.eu/>



Fig. 17. Sample detection rules implemented within the Citilog system

IVMD 1.0 is an intelligent digital video motion detector that uses advanced video content analysis to reliably detect moving objects, while suppressing unwanted alarms from spurious sources in the image. The algorithm intelligently adapts to changing lighting and environmental conditions such as rain, snow and leaves blowing in the wind. IVMD 1.0 can easily be configured to select the required sensitive areas of the image, the minimum object size and the motion direction that triggers an alarm, so that it only detects the important moving objects in a scene. Configuration is done by the Configuration Manager with its intuitive plug-in that provides all the necessary tools.

Functions offered with the solution:

- Licensable option for advanced VCA feature for all VIP X and VideoJet X units and Dinion IP and FlexiDome IP cameras
- Robust motion detection
- Background learning algorithm from Bosch's own research group
- Up to 16 independent detector fields for alarm generation
- Detects objects entering or leaving an area (detector field) or loitering within an area
- Object size, speed, and direction filters used to create specific detection rules
- Built-in tamper monitoring detects camera hooding/masking, blinding, defocusing, and repositioning

- Two-way motion direction filter
- Statistics for fine tuning object filters

3.1.18. SISTORE CX EDS

Company website: <http://www.sourcesecurity.com/>

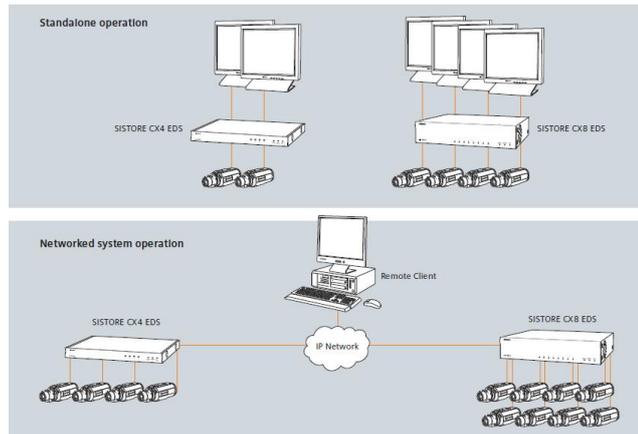


Fig. 18. SISTORE CX EDS system architecture examples

SISTORE CX EDS is a video motion detection system featuring the latest Siemens technology for perimeter protection. Motion detection and object tracking are based on the latest statistical image analysis methods, optimized for use in critical outdoor areas. By using SISTORE CX devices, MPEG4 video recording and transmission as well as a tamper detection function are additionally available. All functionalities can be optimally adapted to the actual environmental conditions using the configuration software. EDS (Enhanced Detection Solution) is a software option that is available for all devices of the SISTORE CX platform. After activation of this function, the standard motion detection functionality is extended. Activating is easily done by entering the requisite license key in the configuration software.

System features:

- Video motion detection system for monitoring critical outdoor areas
- Highest detection rate at low false alarm rate
- Motion detection and object tracking
- Tamper detection function can be activated for each video input
- Software option for SISTORE CX4 and SISTORE CX8
- Easy setup using the configuration software
- Display of courses of motion and object frames
- Simultaneous MPEG4 video recording and transmission
- High scalability over LAN
- Scheduled switching of alarm programs via the IVM software
- Suitable for upgrading existing systems

3.1.19. Summary

In the table below a short summary of the reviewed systems is presented in terms of chosen features. If the information about a specific feature was not available, a question mark is placed in the corresponding table cell. The last column of the comparison contains additional system information. The possible statements are:

- Complete solution – means that a complex system is offered including all the devices and software necessary to build a complete smart video surveillance solution
- Remote system – describes systems where video processing occurs on an external devices which are not a part of the surveillance system directly
- Smart cameras – specifies that smart cameras are utilized within the proposed solution
- Can be remote – defines that access to the system, through a number of mobile devices (i.e. smartphones, tablets), is available

Table 1. Described commercial surveillance systems feature comparison

System \ Feature	Software included	Multiple brands cameras	Modular structure	Remote access	Audio support	Additional info
Praetorian	+	+	+	+	+	complete solution
GeoVision	+	?	+	+	?	complete solution
MATE	+	+	+	+	?	complete solution
Westec	+	-	-	+	+	remote system
Technest	+	-	+	?	?	smart cameras
NUUO	+	+	-	+	+	complete solution
MTS	+	?	+	?	?	-
MDS	+	+	-	+	+	can be remote
COGNIMATICS	+	+	+	+	?	complete solution
Aralia systems	+	+	+	+	+	complete solution
HASAM	+	-	+	?	?	smart cameras
Honeywell	+	?	+	?	?	complete solution
Ipsotek	+	+	+	+	-	complete solution
Vi-System	+	+	+	?	?	-
Sentient	+	+	?	?	?	-
Citilog	+	+	?	?	?	-
IVMD	+	+	+	?	?	-
SISTORE CX EDS	+	+	+	?	?	complete solution

In the context of the ADDPRIV project, the above products could be integrated into the proposed system, as 'Event Detection' components,

which would provide notification of an event. The other ADDPRIV components would then identify the video segments relevant to this event, and store the necessary data using a method that preserves privacy. To decide which product to integrate into a system, a check-list of questions could be used, given below:

- Does the product detect the type of event that is of interest to the end-user (left-luggage, contra-flow, etc)?
- Does the product accept input from cameras that are available at the end-user site?
- Does the product output the event notification in a format that is able to be used by the ADDPRIV system?
- To what extent does the product require configuration and 'fine-tuning' by the end-user, in order to achieve best possible results?
- What is the best possible performance (minimising false positive and false negative alarms) that can be achieved over a representative dataset, recorded at the end-user's site?

Unfortunately, the answers to these questions are not readily available, even from those products that supply quite detailed information on their websites. Answers to the first three questions would require a dialogue with the manufacturer (or distributor). Answers to the last two questions would only be obtained after a trial installation. This creates a problem: significant resources are required to make a meaningful comparative evaluation between products.

Lastly, it has to be mentioned that there is still a large group of other companies offering complete video surveillance system solutions. Some of these other companies are listed below:

- 3VR
- Adaptive Imaging Technologies
- Aimetics
- BiKal
- Briefcam
- BRS Labs
- Bynet
- Camero
- Cernium
- ClickIt
- Emza
- Eptascape
- Genetec
- Geutebrueck
- Intellivid
- ioimage
- Magal Security Systems
- NICE Systems
- ObjectVideo
- Sea-Eye Underwater Technology
- SightLogix
- via:sys
- VideoIQ
- VideoMining
- Vidient
- Vigilant Systems

3.2. Commercial smart cameras solutions analysis

Another approach to application of video analytics to surveillance systems is the employment of so called smart cameras. A smart camera is usually a regular camera equipped with an additional DSP module where video processing is performed. These two elements are typically integrated inside one enclosure. Such cameras can be used within already existing surveillance systems offering the human operator an additional information acquired from the video analytics module. Often, along with this solution producers deliver a dedicated software for the cameras, with an interface allowing communication with the user. In this way the operator can be adequately alerted by displaying additional information about current video stream.

3.2.1. Samsung Techwin

Company website: <http://www.samsungtechwin.com/>



Fig. 19. Samsung Techwin tracking module

Samsung Techwin, world leading imaging technology plays an important role in protecting the safety and happiness of people by providing a comprehensive range of products and complete solutions ranging from city surveillance to the protection of streets, airports, ports, industrial facilities Military installations and B2C.

Samsung Techwin sets a new benchmark in the domestic and international security market by providing higher quality, cleaner images and cutting-edge network functions. Samsung Techwin aims to provide a one.stop security solution that meets the needs of users both now, and into the future, and is committed to becoming the world leading provider of professional security solutions.

3.2.2. Sony Distributed Enhanced Processing Architecture (DEPA)

Company website: [oct09-sony-E_DEPA_whitepaper_Ver10_061221.pdf](#)

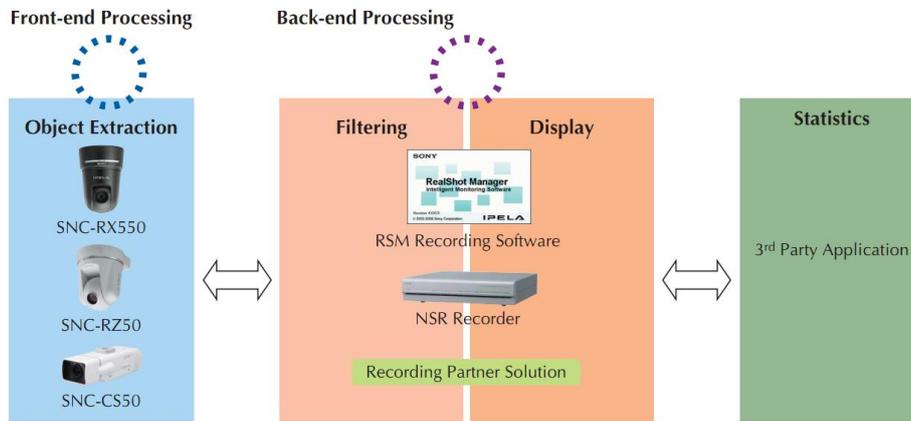


Fig. 20. Sony DEPA proposed solutions

Sony's distributed video analytics is named DEPA, Distributed Enhanced Processing Architecture. The DEPA design divides traditional processing into two separate tasks. Front-end processing is distributed to the endpoints of systems within cameras while back-end processing takes place at the recorder.

Features:

DEPA's Front-End Processing

- Distinguishes objects from environmental noise
- Detects moving and/or stationary objects
- Object information is converted into metadata and then transferred over the network separately from the digital video stream

DEPA's Back-End Processing

- Receives and stores pre-processed object data from cameras
- Extracts objects that match the filtering condition set in recorder
- Displays information; creates alarm responses appropriate to specific conditions

3.2.3. Bosch Security

Company website: <http://www.boschsecurity.com>



Fig. 21. BOSCH smart cameras

Embodied in the recently introduced Dinion 2X Day/Night fixed camera and the Flexidome 2X Day/Night dome camera, Bosch's smart surveillance technology provides customers with a truly advanced level of imaging for the most demanding surveillance applications, according to the company.

The cameras combine a proprietary wide dynamic range CCD sensor (Charged Coupled Device) with a new Bosch designed 20-bit digital signal processor (DSP) with many times the computational power of conventional cameras. These features deliver superior performance with improved detail reproduction and better image quality in low-light conditions. The same level of performance is achieved in the Dinion 2X fixed body cameras and FlexiDome 2X fixed dome cameras. It is also possible to mix and match FlexiDomes and fixed body cameras in an installation to suit the site requirements. In difficult lighting condition, the highly sensitive CCD sensor automatically analyses each image pixel by pixel to reveal details invisible to the human eye, says Bosch.

The new cameras also feature Smart Backlight Compensation (Smart BLC) to optimize light levels for objects of interest in scenes with a bright background. Smart BLC automatically analysis the image and enhances the details to provide the best result, without the need of user intervention. In addition, the cameras have Day/Night capability with infrared contrast for effective surveillance 24 hours a day, with or without IR lighting.

Several tools built into the cameras reduce installation times for technicians. A test pattern generator produces signals to test cables and fault-find other CCTV equipment, while Bilinx allows installers to check camera status, change settings and update firmware from a laptop computer. A host of other features, including six user-programmable modules, a multi-language On Screen Display, built-in smart motion detection and privacy zones, ensure simple set up and operation of the cameras. A high efficiency power supply also improves the operating temperature of the cameras.

3.2.4. Matrix-Vision

Company website: <http://www.matrix-vision.com/>



Fig. 22. Matrix-Vision smart cameras

The new mvBlueLYNX design with its fast PowerPC processor covers video sensor and intelligent camera applications. Easy development and smooth integration within existing networks offers a significant reduction of time-to-market for image processing solutions. The compact, integrated design forms the base for reliable usage in harsh environments.

System features:

Hardware:

- PowerPC - CPU with MMU & FPU
- 32 MB NOR-FLASH (Linux system files 4MB, user area 28 MB. Approx. 40 MB for user available, by using compressed filesystem.)
- 4 MB NAND-FLASH (Bootloader, Kernel, safeboot system, system configuration parameters)
- dig. I/O to PLC and peripherals
- C-mount lens, on request: S-mount, CS-mount
- CCD with trigger input and add. illumination connector
- TV interface on request
- image acquisition with DMA
- power supply 12..24 V DC, min. 6..13 W

Applications:

- Replaces complete PC based image processing systems
- Replaces central image processing systems with many cameras by distributed intelligence
- Replaces div. sensors: light barriers, colour sensors, laser sensors
- 2D /3D – measurement
- Pattern recognition
- OCR, OCV
- Colour control
- Robotics
- Video sensors
- Video compression
- Digital alarm I/Os
- Motion detection
- Image transfer via LAN

3.2.5. National Instruments

Company website: <http://www.ni.com/vision/>



Fig. 23. National Instruments smart cameras

NI Smart Cameras are industrial, high-quality image sensors combined with powerful processors to create rugged, all-in-one solutions for machine vision applications. They are designed to tightly integrate with NI programmable automation controllers and human machine interfaces as well as work with the entire suite of vision algorithms in the NI vision software platform.

The NI 1764 Smart Camera, powered by a 533 MHz PowerPC processor and a 720 MHz Texas Instruments DSP coprocessor, is a real-time target for machine vision. The high-quality Sony CCD image sensor acquires monochrome SXGA (1280 x 1024) resolution images. This camera has four times the resolution of the NI 1722 and 1742 Smart Cameras.

The DSP coprocessor offers improved performance (up to four times) for optical character recognition, pattern matching, code reading, operators, morphology, Nth order filters, and thresholding. You can configure the NI Smart Camera with NI Vision Builder for Automated Inspection (AI), an easy-to-use stand-alone programming environment for creating full visual inspection applications. The NI Smart Camera features a license for Vision Builder AI software that you can use to access the full list of functions to program any NI Smart Camera and to emulate NI Smart Cameras even when the hardware is not physically connected.

If you prefer more customization, you can program the NI Smart Camera with NI LabVIEW software and the NI Real-Time Vision Development Bundle that includes all of the modules needed for programming LabVIEW vision applications on real-time targets.

System features:

Hardware:

- 533 MHz PowerPC processor with 720 MHz Texas Instruments DSP coprocessor
- Monochrome 1280 x 1024 CCD image sensor
- Built-in lighting controller
- Program with LabVIEW Real-Time Module or configure with Vision Builder AI
- Includes Vision Builder AI for programming NI Smart Cameras

3.3. Similar projects solutions

There is a wide range of projects related to the security issues of European Community. They concern state borders, airports, seaports, public transportation monitoring and many more. Since the main aim of this document is the identification of solutions appropriate for the integration with the ADDPRIV system, only several research projects regarding video processing, are included. In addition, the projects in which the ADDPRIV partners are or were involved recently are presented.

Each of the following descriptions contain general statements regarding the problems which are addressed by the corresponding project and the list of main goals to be achieved. Additionally a home page reference is provided, where further information regarding the project could be found.

3.3.1. INDECT

Project website: <http://www.indect-project.eu/>

Duration: 2009/01/01 – 2013/12/31

Project full title: Integrated surveillance of crowded areas for public security



Fig. 24. INDECT project logo

INDECT aims at developing tools for enhancing security of citizens and protecting confidentiality of recorded and stored information. It addresses security issues both in the internet and real world for the purpose of automatic detection of threats related to criminal behaviour, terrorism and violence. The main objectives of the INDECT are:

- To develop a prototype of an integrated, network-centric system supporting activities of police officers
- To provide techniques and tools for direct search of images and video based on watermarked contents and storage of metadata in the form of digital watermarks

The INDECT methodology consists of a three step approach, where first, specific threats or crimes are detected, then the source of these events is identified and in the third step the end user (i.e. police officer) is notified to undertake the final decision.

The main expected results of the INDECT project are:

- Trial of intelligent analysis of video and audio data for threat detection in urban environments
- Construction of a family of device prototypes used for mobile object positioning
- Performing computer-aided detection of threats and targeted crimes in public Internet resources
- Construction of a search engine for rapid semantic search based on watermarking of content related to child pornography and human organ trafficking

- Implementation of a distributed computer system that is capable of effective intelligent processing
- Creation of tools and technology for privacy and data protection during storage and transmission of information using quantum cryptography and new methods of digital watermarking

3.3.2. SECURITY

Project website: <http://www.ppbw.pl/ppbw/1.html>

Duration: 2007/09/01 – 2010/08/31

Project full title: Multimedia system assisting in identification and prevention of delinquency, including violence in schools) and terrorism

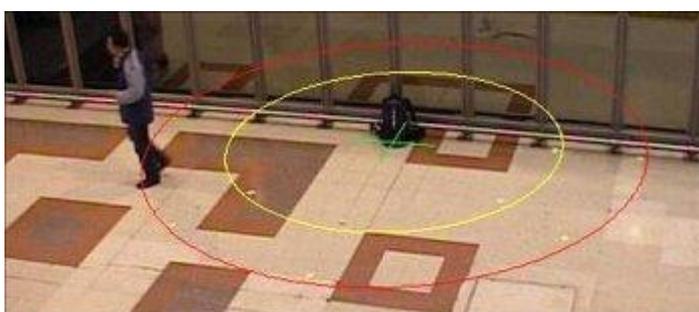


Fig. 25. An example of even detection developed during SECURITY project

The idea of the SECURITY project is to design and develop teleinformatic tools that would supplement the functions of already existing audio and video monitoring systems. This extension is a function of automatic image and sound interpretation, which lets computer systems automatically discover potential threats and generate alerts to appropriate services responsible for public order and security.

3.3.3. SUBITO

Project website: <http://www.subito-project.eu/>

Duration: 2009/01/01 – 2011/07/31

Project full title: Surveillance of unattended baggage and the identification and tracking of the owner



Fig. 26. SUBITO project logo

The SUBITO project is devoted mainly to mass transportation network monitoring where the unattended luggage can be a potential threat. The SUBITO project aims at developing novel video processing technologies which will improve the performance of security surveillance systems employing multiple CCTV systems by achieving the objectives of:

- Autonomously detecting unattended baggage

- Rapidly identifying, locating and tracking the baggage owner

The following scientific objectives are pursued by the project:

- Robust Detection – SUBITO will review and develop novel extensions to this work to develop a stereo based solution to lighting invariant background subtraction for the task of robust abandoned object detection
- Robust and Long Term Tracking – the project will investigate how these problems can be overcome through novel fusion of the outputs of independently run tracking algorithms
- Robust Identification – In SUBITO, robust categorisation and identification of objects will be achieved through fusion of a number of visual cues including size, gait, shape and appearance
- Behavioural Analysis – The project will test the probabilistic models and logic to increase the discriminating power of present day behavioural analysis systems
- Facial Recognition for multi camera surveillance application – The application and advancement of face recognition technology in this field as well as the integration with further complex software components (detection of unattended goods, identification of the owner) are important steps beyond the state of the art
- Sensor Fusion – State of the art probabilistic methods will be explored as part of SUBITO as well as fusing pattern classifiers
- PTZ Cameras – Automatic and robust detection of abandoned goods using a camera network is a complex task as the system is observing a wide 3D scene including occlusions and changes of appearance
- Performance Evaluation – In SUBITO algorithmic robustness will address the development of a novel evaluation methodology (based on appropriately chosen benchmarking criteria) to demonstrate improved object detection, tracking, classification and action recognition capability

3.3.4. ADVISOR

Project website: <http://www-sop.inria.fr/orion/ADVISOR/>

Duration: 2000/01/01 – 2002/12/31

Project full title: Annotated digital video for intelligent surveillance and optimised retrieval



Fig. 27. ADVISOR project logo

The ADVISOR project aims at developing an efficient management system metro operators due to the environmental and economic pressure for increased use of public transport. The ADVISOR system explores advances in computer vision technologies and open network architectures to demonstrate effective use of CCTV for computer-assisted automatic incident detection, content based annotation of video recordings, behaviour pattern analysis of crowds and individuals, and ergonomic human computer

interfaces. The system will provide a set of decision support tools, which will enhance the value of CCTV as an asset for managing public transport operations by:

- Increasing the utility of information from the cameras by annotating the images according to their content
- Reducing the workload on network controllers by automatically alerting them to situations requiring attention
- Adopting a set of open standards to facilitate specification, procurement and testing of advanced CCTV systems

3.3.5. ADABTS

Project website: <https://www.informationssystemsfors.se/~adabts-fp7>

Duration: 2009/08/01 – 2013/07/31

Project full title: Automatic detection of abnormal behaviour and threats in crowded spaces



Fig. 28. ADABTS project logo

ADABTS project aims to facilitate the protection of EU citizens, property and infrastructure against threats of terrorism, crime, and riots, by the automatic detection of abnormal human behaviour. In particular, ADABTS addresses one of the key problems, the definition of abnormal behaviour, by extracting characterizations in realistic security settings based on expert classifications and the analysis of CCTV operator behaviour.

A set of behaviour descriptors are defined and models of threatening behaviour are built. Algorithms are developed that detect pre-defined threat behaviours and deviations from normal behaviour. For accurate and robust detection, data from audio and video sensors will be combined with context information.

3.3.6. PATS

Project website: <http://www.pats-project.eu/>

Duration: 2009/08/01 – 2012/01/31

Project full title: Privacy awareness through security organisation branding

The PATS project aims at increasing privacy awareness across various sectors, from private companies to government agencies with a special focus on the development and usage of CCTV and biometric technologies. The main objective is to promote the idea of taking into consideration privacy issues in development of novel solutions.

3.3.7. ASPIS

Project website: <http://www.aspis-project.eu/>

Duration: 2008/06/01 – 2011/05/31

Project full title: Autonomous surveillance in public transport infrastructure systems



Fig. 29. ASPIS project logo

The ASPIS project aims at the development of a prototype surveillance system based on autonomous, smart monitoring devices that capture data only upon the occurrence of an incident, potentially dangerous for the passengers (like an explosion blast or the triggering of the fire detector). When triggered, these devices propagate the triggering to their neighbouring devices and send an alarm. Successively, they upload the captured data to the central station providing a wide (space and time-wise) coverage of the potentially hazardous incident. Finally, they provide a dedicated bi-directional communication channel between the emergency centre and the affected areas. If, for any reason, they don't succeed to establish communication, they serve as "black boxes", preserving the data until they are physically recuperated by the authorities. This innovative system is meant for the unattended surveillance of public transport (vehicles, stations), maritime transport (ferries or cruise vessels) and other public spaces. It serves primarily for the prompt and reliable situation awareness during the early, most critical emergency phase, thus greatly facilitating the overall crisis response.

3.3.8. SAMURAI

Project website: <http://www.samurai-eu.org/>

Duration: 2009/01/01 – 2011/07/31

Project full title: Suspicious and abnormal behavior monitoring using a network of cameras and sensors for situation awareness enhancement



Fig. 30. SAMURAI project logo

The aim of SAMURAI is to develop and integrate an innovative intelligent surveillance system for robust monitoring of both inside and surrounding areas of a critical public infrastructure. This is achieved by developing:

- Innovative tools and systems for people, vehicle and luggage detection, tracking, type categorisation across a network of cameras under real world conditions

- A hybrid context-aware abnormal behaviour recognition based on a heterogeneous sensor network consisting of both fix-positioned CCTV cameras and mobile wearable cameras with audio and positioning sensors. These networked heterogeneous sensors will function cooperatively to provide enhanced situation awareness
- Innovative tools using multi-modal data fusion and visualisation of heterogeneous sensor input to enable a real-time adaptive behaviour profiling and abnormality detection for alarm event alert and prediction which results in more effective control room operator queries

3.3.9. PROMETHEUS

Project website: <http://www.prometheus-FP7.eu>

Duration: 2008/01/01 – 2010/12/31

Project full title: Prediction and interpretation of human behaviour based on probabilistic structures and heterogeneous sensors



Fig. 31. PROMETHEUS project logo

The overall goal of the project is the development of principled methods to link fundamental sensing tasks using multiple modalities, and automated cognition regarding the understanding of human behaviour in complex indoor environments, at both individual and collective levels. Given the two above principles, the consortium will conduct research on three core scientific and technological objectives:

- Sensor modelling and information fusion from multiple, heterogeneous perceptual modalities
- Modelling, localization, and tracking of multiple people including measurement uncertainty and occlusions
- Modelling, recognition, and short-term prediction of continuous complex human behaviour

3.3.10. ISCAPS

Project website: <http://www.iscaps.reading.ac.uk/>

Duration: 2005/02/01 – 2007/01/31

Project full title: Integrated surveillance of crowded areas for public security



Fig. 32. ISCAPS project logo

The main goal of ISCAPS is to reinforce security for the European citizen and to downsize the terrorist threat by reducing the risks of malicious events. This is undertaken by providing efficient, real-time, user-friendly, highly automated surveillance of crowded areas which are significantly exposed to terrorist attacks. ISCAPS activities breakdown into research on threatening scenarios and operational requirements, research on key technologies, open system architectures, component integration, validation and demonstration. Three types of crowded areas are considered: open, channelled and restricted.

3.3.11. SOBCAH

Project website: <http://www.iosb.fraunhofer.de/servlet/is/18607/>
 Duration: 2006/02/01 – 2007/07/31
 Project full title: Surveillance of borders, coastlines and harbours

The main objective of SOBCAH is to reinforce the security of the European borders through a identification of the key threats relevant to 'green' and 'blue' borders and development of the most suitable architectural solutions using the most advanced sensors and network technologies to address these issues. SOBCAH will maximize the effectiveness of current isolated security systems by fusing and integrating them into high level open architecture to overcome complex systems engineering issues including:

- Integration of incumbent systems with new technologies
- Real time situational awareness
- Data fusion
- Multiple users with multiple views
- Multi-level security with segregated access to controlled information
- Net centric and interoperability

3.3.12. VANAHEIM

Project website: <http://www.vanaheim-project.eu/>
 Duration: 2010/02/01 – 2013/07/31
 Project full title: Video/Audio networked surveillance system enhancement through human-centered adaptive monitoring



Fig. 33. VANAHEIM project logo

The aim of VANAHEIM is to study innovative surveillance components for autonomous monitoring of complex audio/video surveillance infrastructure, such as the ones prevalent in shopping malls or underground stations. The main objectives of the project include:

- Autonomous audio/video data stream modelling

- Human behaviour analysis from sensory-data by real-time monitoring of individual, group and crowd/flow as well as collective behaviour modelling with on-line adaptation

3.3.13. PROBANT

Project website: ftp://ftp.cordis.europa.eu/pub/fp7/security/docs/probant_en.pdf

Duration: 2006/04/01 – 2008/03/31

Project full title: People real-time observation in buildings: assessment of new technologies in support of surveillance and intervention operations

The PROBANT project focuses on the development, integration and validation of technologies enabling operators in crisis intervention and surveillance situations to observe individuals located inside buildings and trace them in real time. The main objectives of the project include:

- Effective detection and real time observation of moving people in closed environments
- Improvement of the quality of information in images derived from raw data
- Improvement of the user-interface features, allowing operators to rapidly understand the images and to take decisions with a high level of confidence
- Provision of more reliable techniques using biometric data to profile and label the moving people
- Implementation of real time wireless transmission of data to remote control centres

3.3.14. PRISE

Project website: ftp://ftp.cordis.europa.eu/pub/fp7/security/docs/prise_en.pdf

Duration: 2006/02/01 – 2008/05/31

Project full title: Privacy enhancing shaping of security research and technology – a participatory approach to develop acceptable and accepted principles for European security industries and policies



Fig. 34. PRISE project logo

The PRISE project provides guidelines and support for security solutions with a particular emphasis on human rights, human behaviour and perception of security and privacy. The key tasks of the project include:

- Developing and testing a set of criteria and guidelines for privacy enhancing security research and technology development
- Elaborating these criteria and guidelines with direct involvement of providers of security technologies, private and public users and implementers, institutions and bodies shaping policies and regulation as well as organisations representing potentially and actually conflicting interests

- Transforming the results into privacy-enhancing development and implementation scenarios of security technologies and measures

3.3.15. PROTECTRAIL

Project website: <http://protectrail.eu/>

Duration: 2010/09/01 – 2014/02/28

Project full title: The railway-industry partnership for integrated security of rail transport



Fig. 35. PROTECTRAIL project logo

The PROTECTRAIL objective is to provide a viable integrated set of railway security solution, by considering the extent of the assets involved, the nature of the possible threats as well as the amount of technical requirements and operational constraints. The main objectives are to:

- Develop an exhaustive common vision of actual and future risks regarding many different assets and regarding and assessing disparity aspects
- Implement asset oriented integrated solutions (sub-mission level) based on mature technologies
- Integrate the asset oriented solutions and demonstrate a global architecture, including modularity and interoperability
- Derive from these results a future design for homogenous security

3.3.16. CARETAKER

Project website: <http://www.ist-caretaker.org>

Duration: 2006/03/01 – 2008/09/30

Project full title: Content analysis and retrieval technologies to apply knowledge extraction to massive recordings



Fig. 36. CARETAKER project logo

CARETAKER project focuses on the extraction of a structured knowledge from large multimedia collections recorded over networks of camera and

microphones deployed in real sites. The produced audio-visual streams, in addition to surveillance and safety issues, could represent a useful source of information if stored and automatically analysed, in urban planning and resource optimization for instance.

CARETAKER models and account for two types of knowledge: on one side, the multi-user knowledge (safety operators, decision makers), represented by their needs, their use-case scenario definition, and their abilities at providing context description for sensory data; on the other side, the content knowledge, characterized by a first layer of primitive events that can be extracted from the raw data streams, such as ambient sounds, crowd density estimation, or object trajectories, and a second layer of higher semantic events, defined from longer term analysis and from more complex relationships between both primitive events and higher-level events.

3.4. *Surveillance systems summary*

Available and known commercial solutions from the area of smart video surveillance were presented in the previous sections. However, without direct access to trial version of systems, the detailed evaluation of their performances is out of the scope of this document. For this reason, these solutions are only briefly described to highlight some of the key characteristics of current state of the art solutions.. The same methodology was applied for describing security related projects. Many of these projects are still ongoing, therefore most of their results are not available yet .Even if the results are available, the utilized methods efficiency is commonly assessed using different criteria to those that will concern the ADDPRIV project. Therefore, this overview is tend mainly to screen the concepts, researched during other security related projects.

4. Relevant algorithms

Video surveillance systems typically use multiple video cameras, transmitting the video signals to a central control room, where a multiplex matrix is used to display some of the images to security personnel. Event detection and recognition requires the perceptual capabilities of human operators to detect and identify objects moving within the field-of-view (FOV) of the cameras and to understand their actions. No matter how vigilant the operators, manual monitoring inevitably suffers from information overload, as a result of periods of operator inattention due to fatigue, distractions and interruptions. In practice, it is inevitable that a significant proportion of the video channels are not regularly monitored, and potentially important events are overlooked. Furthermore, fatigue increases dramatically as the number of cameras in the system is increased. Automating all or part of this process would provide significant benefits, ranging from the capability to alert an operator to these potential events of interest, through to a fully automatic detection and analysis system. However, the reliability of automated detection systems is a very important issue, since frequent false alarms induce scepticism in the operators, who quickly learn to ignore the system.

There are some constraints to the applicability of video analytics algorithms, which can be summarised as follows. Firstly, many such algorithms work on data from a static camera. This allows 'image differencing' techniques (see sections 4.1.1 and 4.1.2) and it also allows relatively straightforward 'calibration' – the definition of the relationship between the camera pixel co-ordinates and the ground or map co-ordinates. This is necessary for e.g. global tracking results (section 4.2). If the camera is not static, then both of these techniques can still be used – but with a more sophisticated approach. Firstly, image differencing techniques need to include method for categorising any observed differences into a) moving objects and b) resulting from movements of the camera. Secondly, to work with moving cameras, calibration methods need to receive information from the telemetry system about the parameters of the movement. This places greater requirements on both the system capabilities and also the automated processing of the geometry to establish the correct calibration. A second constraint on video analytics systems is their capability in crowded scenes. A significant proportion of algorithms use a connected components analysis to estimate trajectories etc. As the scene gets increasingly crowded, the components are increasingly inter-connected, rendering this approach less effective. A third constraint is the available lighting in the scene. With outdoor scenes the lighting is variable, and generally insufficient around night-time. Some indoor scenes, such as airports, are traditionally brightly lit, whereas others (such as railway platforms, eating areas) have lower levels of lighting for economic or cultural reasons.

In this section, algorithms used in automated analysis of visual surveillance data are described and reviewed. The emphasis is placed on methods and techniques used by partners in the ADDPRIV consortium.

In a typical system for event detection, a set of processing stages needs to be performed. First, the relevant objects present in the observed scene need to be detected. In the next step, object classification should be performed which allows distinguishing various object types. Such an operation allows for further analysis depending on the object category. Further processing can lead to event detection which could be associated to some threat related behaviour. This information could be useful for the system operator improving his awareness about security compromising situations. Many more processing modules could be a part of such a system which could simplify and improve operators' work, e.g. face recognition. To

build a feasible smart surveillance system, nowadays, much emphasis is put to prepare robust video processing algorithms. Hence, a large variety of approaches for acquiring similar effects can be found in the literature. The following sections introduce to typical video surveillance processing modules. A general description of their purpose and known methods for accomplishing certain goals, are presented.

4.1. Object detection

In general, object detection is usually the first module in video processing pipeline in smart surveillance systems. This module is designed to detect the relevant objects which are observed in the scene. This goal can be achieved in various ways [1][2].

One of the groups of methods employs so called optical flow algorithms which, depending on specific algorithm, analyse the scene for characteristic point or estimate the motion on the basis of gradient calculated for the whole image [3]. However, this set of methods is in general quite computationally expensive and is not suitable for real time processing of high resolution images and high frame rate videos. Still, some improvements and simplifications to the optical flow calculation algorithm can be found in the literature making it useful for specific tasks [4][5]. This approach is generally a good choice in two cases. First, when a video stream from a non-static camera needs to be analysed. In such situation optical flow is very suitable as it allows to compensate the global motion detecting some local changes[6]. The second case is related to crowded scenes where the information about single objects and their movement is difficult to acquire. By utilizing optical flow based methods for example a general crowd motion trend can be estimated [7].

In the next group of methods, before object detection occurs, a static image (called background model) representing the stationary parts of the observed scene is built [8][9]. To extract the dynamic parts of the processed frame, current image and the background model are compared. The regions which differ much from the background model are marked as moving objects and considered as the foreground. This type of method is usually referred as background subtraction [10]. The simplest way to acquire an adaptive background model is to utilize frame averaging algorithm where the model is a cumulated average of consecutive frames with a set learning factor. Unfortunately, the adaptation offered by such algorithm is insufficient for the use, especially, for typical outdoor conditions. Therefore, some more sophisticated methods are proposed, among which Gaussian Mixture Models based background subtraction is considered to be one of the most popular [11]. This approach allows for a more robust object detection because of higher adaptability and therefore lower sensitivity to global light changes. This method introduces only temporal relations in the procedure of building background model but also a number of algorithms which involve spatio and spatio-temporal relations are proposed in the literature [8][9][12].

Another set of methods detect objects on the basis of their known appearance. This approach is useful in cases where it is not feasible to generate a background image, for example with a moving camera platform or when the background is predominantly obscured by the foreground image. This set of methods require a single image or a set of object representations to be detected [13]. On the basis of some features extracted from the supplied models, the desired object or a part of that object can be detected. Such approach can be useful for example in case of face or registration plate

detection [14]. The most popular algorithm among this group, utilizes Haar-like features and a set of weak cascade classifiers trained on a provided dataset [15]. The contents of the dataset depends on the problem, however a number of already trained classifiers or at least prepared models can be found on the internet. If the colour distribution of the detected object is known, a Mean-shift based approach can be applied [1]. This algorithm, performing image segmentation, and its extension called CAMShift is suitable for real time processing [16]. Like most of the advanced methods it introduces an adaptation procedure and therefore can be quite robust.

4.1.1. Background model subtraction – Gaussian Mixture Models based method

Detection of moving objects is usually the first stage of video processing chain and its results are used by further processing modules. Most video segmentation algorithms usually employ spatial and/or temporal information in order to generate binary masks of objects [17][18][19]. However, simple time-averaging of video frames is insufficient for a surveillance system because of limited adapting capabilities. The solution implemented in the framework employs spatial segmentation for detection of moving objects in video sequences, using background subtraction. This approach is based on modelling pixels as mixtures of Gaussians and using an on-line approximation to update the model [20][21][22][23][24][25][26][27]. This method proved to be useful in many applications, as it is able to cope with illumination changes and to adapt to the background model accordingly to the changes in the scene, e.g. when motionless foreground objects eventually become a part of the background. Furthermore, the background model can be multi-modal, allowing regular changes in the pixel colour. This makes it possible to model such events as trees swinging in the wind or traffic light sequences.

Background modelling is used to model current background of the scene and to differentiate foreground pixels of moving objects from the background [11][17][28].

Object segmentation is supplemented with shadow detection and removal module. The shadow of a moving object moves together with the object and as such is detected as a part of the foreground object by a background removal algorithm. The shadow detection method is based on the idea that while the chromatic component of a shadowed background part is generally unchanged, its brightness is significantly lower [11][29]. Every new pixel recognized as a part of a foreground object during the background subtraction process is checked whether it belongs to a moving shadow. If the current pixel is darker than the distribution, the current pixel is assumed to be a shadow and is considered as a part of the scene background.

Another approach is presented in [23], where the shadow's shape, size, orientation, luminosity, originating position and appearance model is exploited to determine the colour distributions of both the foreground and shadow classes. This is achieved by skeletonisation and spatial filtering process which is developed for identifying components in the foreground segmentation that are most-likely to belong to each class of feature. A pixel classification mechanism is then obtained by approximating both classes of feature data by Gaussian parametric models. This work is further extended in [24] where novel k-nearest neighbour pixel classifier is proposed, which is

applied on pixels previously classed as foreground during detection process in real time.

In the result of background modelling, a binary mask denoting pixels recognized as belonging to foreground objects in the current frame is obtained. It needs to be refined by the means of morphological processing in order to allow object segmentation [11][30]. This process includes finding connected components, removing objects that are too small, morphological closing and filling holes in regions. Additionally, an algorithm for shadow removing from the mask using morphological reconstruction is implemented [31]. The morphological reconstruction procedure involves two binary images: a mask and a marker. In the mask image, all pixels belonging to either the moving object or the shadow have value of one, and all the background pixels have zero value. The marker is obtained by applying an aggressive shadow removal procedure to the object detection result, so that all the shadow pixels are removed. Example results of moving object detection in a single video frame are presented in Fig. 37.

Another challenge for object segmentation is related to detection and removal of ghosts caused by the starting or stopping of objects. Ghosts mainly appear in two cases: In the first case, when a moving object becomes stationary, it will be adapted (merged) into the background, and then, when it starts to move again sometime later, there will be a ghost left behind. In the second case, an existing object that belongs to the background starts to move (e.g. parked vehicle) and will also cause a ghost problem. To tackle this problem [32] compares the similarity between the edges of the detected foreground objects and those of the current frame based on object-level knowledge of moving objects.



Fig. 37. Example results of moving object detection: original video frame (left) and binary mask denoting moving objects (right)

4.1.2. Background model subtraction – other approaches

When two stereo cameras are available, the GMM model can be used to create two background models [25], i.e. a colour intensity background model and a stereo background model. The combination of both models allows performing more robust and effective segmentation even in very crowded environments.

The main disadvantage of GMM is that it does not take into consideration the correlation between neighbouring pixels. In [33] a novel background estimation technique that learns the co variation of grey levels within the incoming images using principal component analysis to generate the eigen-backgrounds, is proposed. Rather than accumulating the necessarily enormous training set, this technique builds and adapts the eigen-model

online. The number of significant modes as well as the mean and covariance of the model as continuously adapted to match the environmental conditions. As a consequence, for each incoming image, a reference frame is hypothesized efficiently to perform background segmentation from a subsample of the incoming pixels. In turn, [34] proposes to take advantage of periodic variation of background appearance over time, which is detected in the temporal frequency domain. This analysis allows estimating of the period for each pixel, and pixels with the same periodicity are grouped into regions. A Markov model can then be constructed, in which each state models the first and second order statistics of the appearance at a given phase of the period. The state values are updated online. As a consequence, the foreground can be segmented accurately from this time-varying background.

4.1.3. Single-image methods for object detection

In some scenarios the estimation of background model may be problematic due to highly dynamic and complex environment. However, when training data is available, the object detection can be performed without need of creating background model.

In recent years several techniques have been proposed to detect categories of objects from a single image. The categories of object include faces [35] and people [36], but also extend to less common sets. When there are multiple categories of object in the same system, it is usually described as an object recognition system (rather than as an object detection system). The PASCAL Visual Object Challenge (2011) includes examples such as bicycle, boat, bus, car, motorbike and train: these are naturally useful in a surveillance context. For instance, the object detection can be achieved by building the cluster boosted tree structure for a multi-view classification based on edgelet features [37] or using predefined 3D models which are combined with local patches of histograms of oriented gradients [27].

4.2. Object tracking

Object tracking usually follows the object detection. This module allows to introduce a relation between objects detected in consecutive frames.

Tracking of human motion is very active research field which has been addressed in various ways. In a case of a single object tracking, [38] uses a limb tracking system based on a 2D articulated model and a double tracking strategy. Its key contribution is that the 2D model is only constrained by biomechanical knowledge about human bipedal motion, instead of being tailored to a specific activity or camera view. This double tracking strategy relies on a Kalman filter for a global position tracking and a set of particle filters for body parts tracking. Since a space of human motion is often very high dimensional, tracking can be performed efficiently in a low dimensional space as shown in [39][40]. This is achieved by the advanced limb correction module in [39], whereas [40] proposes graph-based particle filter to deal also with stylistic variations of motion. Finally, also the condensation algorithm can be adopted to track a single object [41].

To track multiple objects in complex environments, [42] exploits the bank of Kalman filters to track each subject independently in a scene. Then, in order to facilitate data association and track management, a colour model is created for each person like in [25]. Tracking of multiple objects can be enhanced by integration information about self-calibrated ground plane [43] or depth probability density functions [44]. In turn, [45][46] presents a real-time algorithm which is based on blob matching techniques. First, foreground pixels are detected using luminance contrast and grouped into blobs. Then, blobs from two consecutive frames are matched creating the matching matrices. Tracking is performed using direct and inverse matching matrices. This method successfully solves blobs merging and splitting during tracking. Finally [47] combines a several advanced techniques, such as adaptive intensity-plus-chromaticity mixtures of Gaussians, region based representation, Kalman filter, scene model and a Bayesian network, to perform an on-line multi-object tracking in a real-world scenarios from a single fixed camera.

To deal with large occlusions during tracking, [48] uses a constant acceleration motion model to track objects where three predictors are employed simultaneously with a least square correlation stage to select the most likely object position. These three predictions schemas are a α - β tracking schema, a Kalman filtering method and a region segmentation and matching method. Alternatively, [49] proposes a method for estimating the midpoint (or centroid) and bounding box size of each target using a Kalman filter with the measurements of four bounding edges. This structure facilitates the utilisation of incomplete measurements that can arise due to partial occlusion.

Another challenging problem is tracking across multi camera network. The system proposed by [50][51] comprises two processing stages, operating on data from first a single camera and then multiple cameras. The single-view processing includes change detection against an adaptive background and image-plane tracking to improve the reliability of measurements of occluded players. The multi-view process uses Kalman trackers to model the player position and velocity, to which the multiple measurements input from the single-view stage are associated. Multiple cameras with overlapping views can also help with resolving issues of object occlusions. For instance, in [52] tracking is performed in 3D using the Kalman filter together with the ground plane homography constraint.

4.2.1. Kalman filter supported moving object tracking

After the moving objects are found in each consecutive camera frame, movement of each object on the frame-by frame basis is needed. This is the task of an object tracking module. For each new detected moving object, a structure named a tracker is created. The position of the object in the current camera frame is found by comparing the results of object detection (the blobs extracted from the image) with the predicted position of each tracker. The prediction process estimates the state of each tracker from the analysis of the past tracker states. An approach based on Kalman filtering [53][54] was used for prediction of trackers' state.

A relation between a tracker and a blob is established if the bounding box of the tracker covers the bounding box of an object by at least one pixel. There are some basic types of relations possible (Fig. 38), each of them requires different actions to be taken. If a certain blob is not associated with any tracker, a new tracker (Kalman filter) is created and

initialised in compliance with this blob. If a certain tracker has no relation to any of the blobs, then the phase of measurement update is not carried out in the current frame. If the tracker fails to relate to a proper blob within several subsequent frames, it is deleted. The predictive nature of trackers assures that moving objects, whose detection through background subtraction is temporarily impossible, are not “lost” (e.g. when a person passes behind an opaque barrier).

If there is an unambiguous one-to-one relation between one blob and one tracker, this tracker is updated with the results the related blob measurements. However, if there is more than one matching blob and/or tracker, a tracking conflict occurs. The authors proposed the following algorithm for conflict resolving. First, groups of matching trackers and blobs are formed. Each group contains all the blobs that match at least one tracker in the group and all the trackers that match at least one blob in the group. Next, all the groups are processed one by one. Within a single group, all the trackers are processed successively. If more than one blob is assigned to a single tracker, this tracker is updated with all blobs assigned to it, merged into a single blob. This is necessary in case of partially covered objects (e.g. a person behind a post) that causes the blob to be split into parts. In other cases, all the matching blobs are merged and the tracker is updated using its estimated position inside this blob group. This approach utilizes the ability of Kalman trackers to predict the state of the tracked object, provided that it does not rapidly change its direction and velocity of movement, so that the predicted state of the Kalman filter may be used for resolving short-term tracking conflicts. The estimated position is used for updating the tracker position, change of position is calculated using the predicted and the previous states. The predicted values of size and change in size are discarded and replaced by values from the previous filter state, in order to prevent disappearing or extensive growth of the tracker, if its size was unstable before entering the conflict situation. Therefore, it is assumed that the size of the object does not change during the conflict. The vector of parameters used for updating the Kalman tracker during the conflict may be written as:

Possible relations	
	Region without a tracker
	Tracker without a region
	One tracker – one region
	One tracker – many regions
	May trackers – one region
	Many trackers – many regions

Fig. 38. Possible relations between results of object detection (blobs, regions) and trackers

A special case of tracking conflict is related to ‘splitting objects’, e.g. if a person leaves a luggage and walks away. In this situation, the tracker has to

follow the person and a new tracker needs to be created for the luggage. This case is handled as follows [55]. Within each group of matching trackers and blobs, subgroups of blobs separated by a distance larger than the threshold value are found. If there is more than one such subgroup it is necessary to 'split' the tracker: select one subgroup and assign the tracker to it, then create a new tracker for the remaining subgroup. In order to find the subgroup that matches the tracker, the image of the object stored in the tracker is compared with the image of each blob, using three measures: colour similarity, texture similarity and coverage. The descriptors of the blob are calculated using the current image frame. The descriptors of the tracker are calculated during tracker creation and updated each time the tracker is assigned to only one blob (no conflict in tracking).

After the conflict resolving is done, the tracking procedure finishes with creating new trackers for unassigned blobs and removing trackers to which no blobs have been assigned for a defined number of frames. The process is repeated for each camera frame, allowing for tracking the movement of each object.

Fig. 39 presents an example of object tracking with conflict resolving, using the procedure described here. A person passes by a group of four persons walking together. During the conflict, positions of both objects are estimated using the Kalman filter prediction results. When these two objects become separated again, assignment of trackers to blobs is verified using the colour, texture and coverage measures. As a result, both objects are tracked correctly before, during and after the conflict occurs.

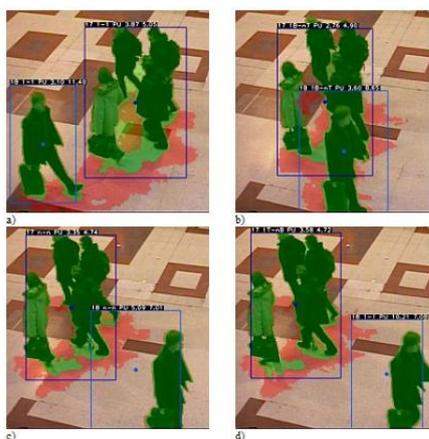


Fig. 39. Example of object tracking and conflict resolving results; green – detected moving objects, red – shadows of objects (images from the PETS 2006 database)

4.3. Object classification

An essential part of a smart surveillance system is the objects classification module. Distinguishing between objects allows for more detailed analysis of object behaviours. If the objects present in the analysed material can be assigned to a specified class, further detected actions or other applied operations could be done depending on the type assigned to the object.

Object classification in video surveillance systems is not an easy task mainly due to variation of camera viewpoints. Several methods are proposed to solve this problem which can be divided into two general groups: feature-based and motion-based [56]. Feature based approach utilizes the information about the object spatial characteristics commonly related to shape or texture descriptors. One of the specific methods uses the object real dimension to decide about its type. This approach considers possible perspective differences between cameras but requires a calibration procedure to be applied [57][58]. Such process utilizing Tsai calibration method is described in detail in Sec. 4.3.1.

Other approaches, exploit more advanced properties acquired from the detected objects. Although, in this case, no calibration is required, a large set of prepared object models is needed for proper classifier training. For the purpose of model and object parameterization, a set of various features can be used. The most common parameters present in the literature include SIFT descriptor [59][60][61] and a number of contour and region based shape descriptors [56][62][63]. Still, many other description methods, for example with wavelet coefficients [64], can be found in the literature. A particular method based on the object shape and its contour description is presented deeper in the following subsections [65].

Object categorization is performed by the built classifier on the basis of the calculated feature vector. Again a number of different approaches for the classification process exist. Feature-based methods applying various distance metrics as well as machine learning algorithms, like Support Vector Machines (SVM) and Artificial Neural Networks (ANN), can be found in the literature [56][59][66][67].

Another approach found in the literature employs Motion History Images and Recurrent Motion Images as a method not only for action recognition but for object classification process as well [68]. These methods base on the property that some objects, like cars, are more rigid in comparison to other objects, like people [66]. Therefore, by analysing the spatio-temporal characteristics, it is possible to distinguish between these objects.

4.3.1. Dimension based object classification

Dimension based object classifier is an example of classifier which utilizes a set of thresholds for the purpose of assigning objects to a proper category, i.e. Human or Vehicle. Due to the perspective distortion in a typical camera image, it is not possible to use the size of an object measured in image pixels for classification purposes, but the width and the height of the object in physical units (i.e. metres or inches) has to be known. Therefore, a conversion between the camera coordinates and the world (physical) coordinates has to be defined and this can be achieved by means of the camera calibration procedure. Various calibration methods were proposed, but Tsai's method [57] is one of the most popular and accurate ones and this method was implemented in the described system.

Tsai's calibration method requires marking several points in the area observed by the camera and measuring their positions relative to each other, both in the real world and in the computer image (Fig. 40). One of these points is usually set as an origin of the world coordinate system. The calibration algorithm processes pairs of the world coordinates and the image coordinates of each point, in order to calculate 11 conversion coefficients,

describing translation and rotation of the camera relative to the world coordinates, camera lens distortions, camera focal length, etc. [57].

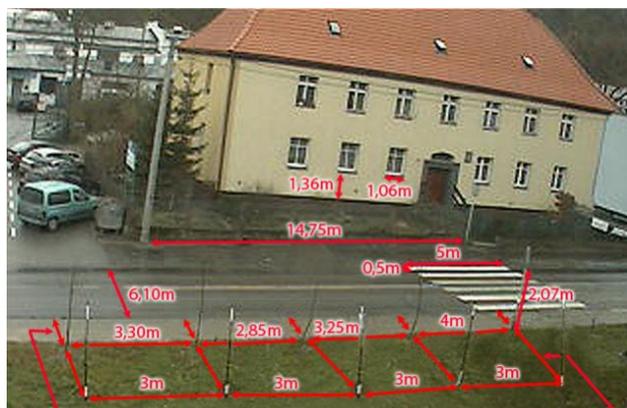


Fig. 40. Calibration points used in the experiments

For the purpose of the experiments, estimated physical width and height of the object are the only parameters used for classification. Other descriptors (related to object's movement and appearance) were purposely omitted in this version of the algorithm in order to validate the estimation procedure. In this test system, three classes of objects were created and the rules of object classification according to their averaged estimated physical width and height were as follows:

- Humans (H) – objects having width in range $0.5m \leq w \leq 2.5m$ and height in range $1m \leq h \leq 3m$
- Big trucks (T) – objects having width $w > 6m$ and height $h > 4.5m$
- Small trucks (S) – objects having width $4.5m \leq w \leq 6m$ and height $3m \leq h \leq 4.5m$
- Passenger cars (C) – objects having width $3m \leq w \leq 4.5m$ and height $1.5m \leq h \leq 3m$
- Other vehicles (V) – objects having width greater than 2.5m and not belonging to any other vehicle classes
- Unknown objects (U) – all other objects

The threshold values were chosen experimentally by processing the video recordings and examining the estimated sizes of various moving objects.

The rules listed above are of course greatly simplified and they will not allow for accurate object classification in complex real scenarios. However, they are sufficient for testing the proposed system for object classification based on their estimated size. The results of the simulations are discussed further in the paper. This simplified classification system can be extended by adding new object classes and additional, more complex classification rules, based not only on the physical size of objects, but also using other descriptors, related to movement and appearance of the tracked objects.

For the purpose of experiments, a stationary camera was mounted on the roof of a building. The camera image shows a fragment of the street with sidewalks. The camera axis is almost perpendicular to the path of vehicles movement. The calibration points that were measured included posts placed at the site for the purpose of camera calibration, as well as landmarks such as road sign posts, small trees, sidewalk borders, etc. (Fig. 40).

During experiments the size estimation is performed if the object is fully in the image frame (the bounding box of the object does not touch the image frame margin), the object is not in conflict with any other moving object (no overlapping with other objects) and the size of the bounding box is larger than the imposed threshold (estimation error for distant, small objects may be very large, because a small difference in object size in pixels results in a large difference in estimated physical size).

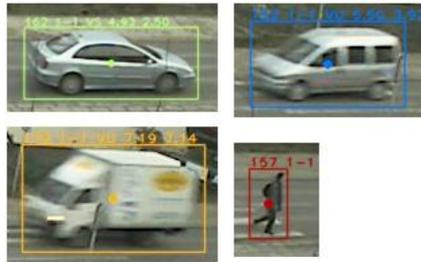


Fig. 41. Moving objects whose size is estimated (car [C], small truck [S], big truck [T], human [H])

4.3.2. Shape based object classification

Video surveillance cameras can be placed variously: on top of the building, attached to a pole, hung under the ceiling. This implies that the same object observed from different cameras can be represented by a different silhouette. Moreover, the object itself can rotate around its own axis creating more silhouette variations. This shows the complexity of the considered problem [65].

Shape based object classification is a pattern recognition problem and as one it requires a set of objects for a proper classifier training. To provide different datasets for specific camera observation angles, 3D object models were prepared. These models represent two basic classes: Vehicles and Human.

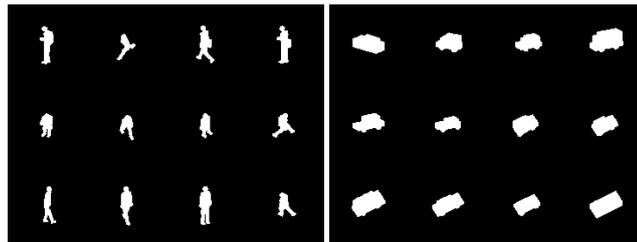


Fig. 42. Sample 3D models projections representing various vehicles and human positions observed from a specified angle

For the purpose of further parameterization and training process those 3D models are projected at the defined viewpoint horizontal angles with a vertical angle interval set to 15° (Fig. 42). This produces a large number of vehicle and human silhouette images for a defined horizontal angle. Carried out experiments show that three such datasets corresponding to 20°, 40° and 60° are enough to cover the horizontal angles referring to the most popular camera orientations. To provide fewer errors during classification process, additional negative dataset is created representing shapes which do not belong to any of the previous classes. The parameterization process lowers significantly the amount of data representing an object mask and

unifies its description. Object classification in video surveillance systems needs to be performed at low computational cost. Commonly used methods chosen for object classification problem include parameters as: fill ratio, compactness, Hu invariant moments, etc. Parameterization utilized in the implemented method is presented in Fig. 43 is used. In the first stage of parameterization input binary mask is resized to defined dimensions (i.e. 50x50 pix) constraining proportions. Resizing the mask allows the parameterization process to be invariant to object scale.

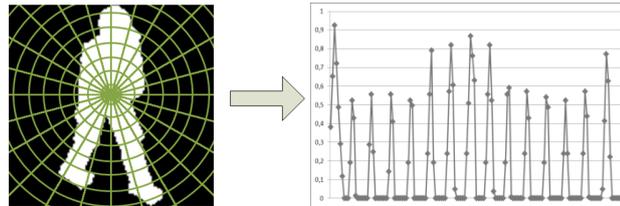


Fig. 43. Visualization of the parameterization process. Vertical axis corresponds to the normalized pixel count and the horizontal axis refers to vector features

Next, the feature vector is built. Center of gravity (CoG) of the analysed mask is calculated and in relation to this point angle and distance bins are created. This way, using 10 distance bins and 20 angle bins (18° interval) a vector of 200 features is generated. Before proceeding to the classification process every created parameter vector is normalized to the same range (0,1).

For the purpose of classification process Support Vector Machines were chosen considering . Its goal is to find a hyper plane which divides two sets of data with the biggest margin between them. The SVM classifier used utilizes the RBF which is said to be the best first choice. Optimal SVM cost and gamma parameters are set during training utilizing the grid-search method explained in the literature [69]. For the training and classification process One vs. All algorithm is used. Therefore in case of 3 classes, two binary classifiers are built (Fig. 44). To obtain classification probabilities a statistical model is built utilizing the algorithm implemented in Libsvm library [70].

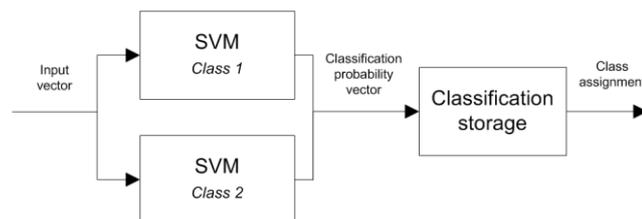


Fig. 44. Classification process scheme

The classification process can be divided into two separate parts (Fig. 44). First, the features of all objects present in the currently processed frame are classified by the SVM classifiers. As the result a vector containing each class assignment probability is created. The vectors are stored during object lifetime. To include the relation of object classification occurring in consecutive frames the object type is set to a class whose average probability is the highest. If none of the probabilities reaches 0.5 threshold the object is automatically assigned to the Other class. Averaging makes the

classification process robust to temporal shape distortions and objects occlusions.

Another issue is that the objects entering or exiting the scene are partially cut what can lead to classification errors. To solve this problem a small border around the analysed scene is set. The border is presented in Fig. 45 as the darkened area around the frame. Only the objects whose outline is entirely inside this border are further analysed.



Fig. 45. Video frames from the test recording. Detected object classification begins when it enters the scene entirely (right)

4.3.3. 3D wire frame models based object classification

Similar approach, to the one described in Sec 4.3.2 is presented in the literature [22][26], where instead of using SVM, a classifier generates a hypothesis of a vehicle or pedestrian being present in the scene by using corresponding 3D models, which are placed onto the scene's ground plane and projected to the camera view. A match measure is calculated for every hypothesis by comparing the model with the image silhouette. Every model is placed on a grid of positions on the ground plane to produce the match measure for every silhouette. The highest measure indicates the most likely position of the vehicle given the silhouette. The highest match measures of different classes are compared to make a decision about the class of a silhouette. Silhouettes with consistently low match measures for all classes are rejected as ambiguous. To use the 3D models, cameras are calibrated by means of a map and five corresponding points with the image. These 3D models are further enhanced in [27] by integrating local patches which are derived from histograms of oriented gradients. In contrast to shape based classification, this allows performing more robust appearance based classification and detection.

4.4. Event detection

In recent years, image-processing solutions have been proposed to automatically detect incidents and make measurements on video images from CCTV cameras, relieving the staff in control rooms of much of the difficulty in finding out where interesting events are happening. Events of particular interest of CCTV operators include abnormal stationarity, queuing, intrusion detection, loitering [71], unattended luggage detection [17][55] as well as closely related problems such as action recognition [72][73][74][75][76][77][78]. Some of the above events could be detected on the basis of a number of user predefined rules others require more sophisticated methods involving behaviour analysis[79][80][81].

In general event detection approaches can be divided into two groups [82][83]. First case, can be presented as an unsupervised system which builds a kind of probabilistic behaviour patterns. This way, the anomalies which do not fit the estimated model, can be detected [84]. The main advantage of this approach is that no user input, regarding the events, is required and that it adapts dynamically to the changing conditions. At the same time, it offers a low control of the detected events as they are not strictly defined. Therefore a number of non-relevant situations can be detected if some situations are rare in a general context. The second, more frequent case is when a rule based approach is applied. Here, a number of atomic events like: *object stopped*, *object left the scene*, *object entered area*, are defined. On the basis of these elementary events, complex rules can be built that allow for specific situation detection [85][86][87][88][97].

In respect to the analysed objects the event detection problem can concern objects in general or specifically people [83]. In the second case, a single person, groups and crowds needs to be considered. Various events are possible to be detected in each of these cases. For the crowd, in general, its flow-based specific activity and some anomalies from main trend can be detected. Regarding single people, primitives-based event detection as well as more complex behaviour recognition can be applied.

Several methods for various events detection are presented in the following subsections.

4.4.1. Rule based event detection

The task of the event detection module is the discovery and interpretation of events occurring in the camera images. One of the approaches for event detector is testing the rules that describe the events, using the parameters obtained from the object tracking and the object classification modules.

The event detection module may be divided into two functional parts, performing the detection and the analysis of the events, respectively [58]. The first part is the low-level event detector (LLED), situated lower in the system hierarchy. The LLED uses parameters describing the object's position, physical size, class, statistics, etc. for the detection of low level-events, such as an object entering or leaving the screen, object that stopped or started moving, etc. The task of the high level event detector (HLED) is the interpretation of detected low-level events and detection of complex events, such as a car parking in the observed area, a person getting into a building, etc. Both parts of the event detector have their own set of rules. The rules for HLED are defined using terms that are similar to description of events in the natural language, therefore the HLED is positioned as closest to the user interface in the system architecture.

The simple example of the LLED is the module evaluating the position of the object relative to the screen, the class to which the object was assigned to by the object classification module, and the current direction and velocity of the object's movement. Using these descriptors, the following set of rules for detection of low-level events may be formulated:

- The object is entering the screen, crossing its border. The position of the object and direction of its movement are detected. The

detected event may be logged using a description like “an object entered the screen at position (x,y) from the left side”.

- The object appeared in the area inside the screen (not crossing the border), e.g. car leaving the parked area, a person leaving the door through the building. Ex-ample of event description: “a new object appeared in the screen at position (x,y)”.
- The object left the screen, crossing its border, e.g. a car driving away from the observed area.
- The object disappeared in the screen, but not leaving the screen margin, e.g. a parked car.
- The object stopped moving, e.g. a car waiting at the traffic lights.
- The object resumed its movement, e.g. a bus leaving the bus stop.
- The object was assigned to a class by the classification module, e.g. “object No. 113 was assigned to a class ‘human’”.
- The object entered a defined area. For example, an operator may define a ‘forbidden zone’ in the image. If an object belonging to a selected class enters this area, an event is detected, e.g. “object No. 113 of class ‘human’ entered zone No. 2”.

Detection of the low-level events should not be performed using parameters obtained from the analysis of the current camera image frame only, because short-term variations and errors in image content analysis may result in erroneous detection of events. For example, an error in object tracking may result in an overestimation of the object’s width and height for just one camera frame, but it would result in detection of a non-existing event. Therefore, it is important to perform event detection on post-processed (e.g. averaged) parameters describing the moving objects.

While the LLED rules operate only on the parameters of tracked objects, the HLED rules also interpret the series of consecutive low-level events in order to detect the high-level events. In other words, the high-level rules describe what the events detected by the LLED mean. The high level rules may be very simple, for example:

IF an object of class ‘*human*’ disappeared in the area described as ‘*the door of the building*’ THEN ‘*a person entered the building*’

In a practical system, the high-level rules are often much more complex, as they interpret a number of detected low-level events occurring in a certain period. This kind of complex rules may be defined by the operator using an available set of the low-level events or they may be selected from a predefined set. The implementation of the HLED system requires a logical structure (data-base) that stores the detected low-level events, together with their time signatures and screen positions, and a rule interpretation algorithm that browses the detected events and evaluates all possible rules. The rule interpretation system may operate on a system of simple deterministic rules or it may utilize more complex methods, such as fuzzy logic, that provide much more possibilities for processing of the rules, especially the high-level ones. Examples of typical high-level events detected in the automatic video content analysis systems, are presented below:

- Intrusion detection – an object of a defined class, entering the ‘forbidden zone’, e.g. a person trying to cross the motorway, a car in the ‘no entry’ area, a person climbing the wall, etc.
- Traffic regulations violation – e.g. a car ignoring the red light, a car parking in the restricted zone, a private car in the bus lane, etc.

- Abandoned luggage – an object that is dropped by a person and left without any person in its vicinity for a defined time (important e.g. for airport security)
- Theft – an object disappearing out of its usual position (e.g. a picture removed from the wall in a museum)
- Loitering – a person moving around in the observed area for a defined time
- Vandalism – detecting the act of destroying objects (e.g. demolishing a bus stop, painting a graffiti)
- People counting – counting a number of persons in the observed area (airports, museums, sport arenas, etc.) and raising an alarm if the number of persons exceeds the threshold

The main problem in event detection is that the system has to cope with the situations in which the event occurs, while the important objects are not visible in the camera image. For example: the abandoned luggage needs to be detected even if the person was obscured by another object while they were dropping the luggage on the floor. Therefore, an advanced processing of detection results gathered from analysis of a large number of frames is required in order to interpret all the important events. Even more complex analysis procedures are needed when the task of the system is to detect the potential events, before they would occur. Such a system needs to analyse the behaviour of the moving objects and compare it to some patterns of typical object's actions in order to predict what will happen next. The discussed systems often implement complex models of behaviour of humans and other moving objects. Specifically, modelling human behaviour, together with pose estimation algorithms, helps to detect and interpret complex events such as fights, assaults, thefts, riots, etc.

It is evident that the system for automatic event detection can never reach a 100% accuracy in practical applications. Therefore, the system has to be carefully tuned up in order to obtain a satisfactory balance between false positive results (detecting an event that did not occur) and false negative ones (missing an important event). In practical applications, a number of false negative results should be kept as low as possible, so that no important events are missed by the system. A certain level of false positive results is acceptable, as long as it does not make the system operator to ignore the system alerts, because they would think that the system is overreacting.

4.4.2. Unattended luggage detection

The content of this subsection is a specific case of rule based event detection described in Sec. 4.4.1 with some additional processing. It shows the case of abandoned luggage detection.

As it was already stated, for the efficient event detection in video stream, it is convenient to divide the event detection module into two parts, namely a low-level and a high-level event detector. The former part operates directly on the results of video analysis. The input to this module are data on detected moving objects: their static parameters (size, appearance, shape, class) and dynamic parameters (position, velocity, direction of movement, shape variability, etc.). Using these data, the low-level event detection module identifies the basic events, such as:

- Object appearing in the camera view or disappearing
- Object that stopped or started moving

- Object entering or leaving a specified area inside the camera view
- Object crossing a defined barrier

The detected low-level events are then passed to the high-level module part. After analysis and interpretation of recently detected low-level events and moving object data, the module detects complex, high-level events. Usually, the actual event that should be detected, e.g. unattended luggage, is of a high-level type. The rule detecting this event comprises of checking whether defined low-level events have been detected.

For the unattended luggage detection several low-level rules have to be defined, using data from object detection, tracking and classification modules [17][55]. The first condition is that a person has to leave a luggage. This is detected if an object classified as 'person' splits into two objects: 'person' (now without luggage, but with the same tracker still assigned to it) and 'luggage' (which remains stationary and receives a new tracker). The object splitting is handled by the tracking module, with conflict resolving algorithm implemented. The second rule checks if the luggage remains inside the defined detection area. With the third rule, the module tests if distance between the person and the luggage exceeds some threshold value. The fourth rule tests whether the person does not return to the left luggage for a defined period. Finally, a high-level rule is needed in order to test whether during a number of recent camera frames, all of the relevant low-level events occurred. If this is the case, a high-level event is detected and the notification is sent to the system. All the rules described here, written in natural language using IF...THEN clauses, are presented in Fig. 46. These rules are implemented in the event detection framework, using the algorithm shown in Fig. 47.

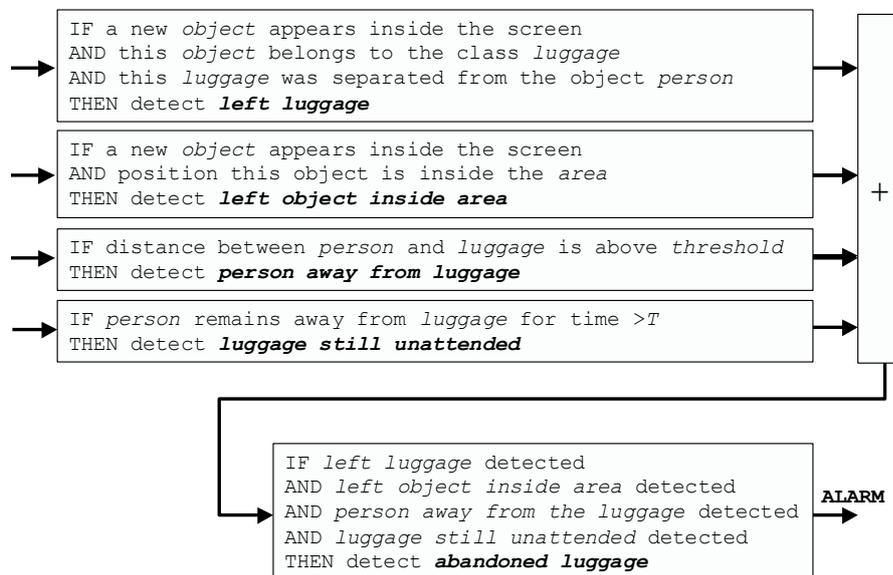


Fig. 46. Rules for detection of the unattended luggage

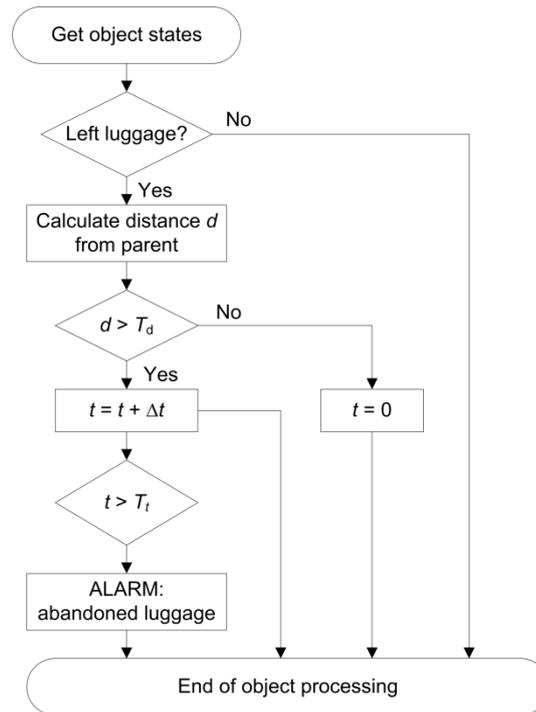


Fig. 47. Algorithm for unattended luggage detection

An important problem with detection of left and removed objects is that due to the nature of the background subtraction algorithm, leaving an object in the scene causes the same effect as removing an object that was a part of the background (e.g. a luggage that remained stationary for a prolonged time). In both cases, a new tracker is created, containing either a left object or a 'hole' in the background (remaining after the object was taken). The detection system has to decide whether the detected object was left or taken. The proposed method works by examining the content of the newly created tracker, using the edge detection algorithm (Canny method) [55]. If the object is left, its edges are located close to the object border, while no distinct edges are present in case of the taken object (provided that a background is sufficiently smooth). First, the grey-scale image B of the object (blob) and its mask M (having non-zero values for pixels belonging to the blob and zero values otherwise) are processed by the Canny detector in order to find the edges. The results of edge detection in the mask is processed by morphological dilation using a 7×7 structuring element in order to increase a detection margin. The result is image $R1$. Next, this result is combined with the result of edge detection in the image B using the logical AND operator, then it is dilated using the same structuring element, yielding the result $R2$. Finally, a measure D of difference between the object and the background is calculated as a ratio of non-zero pixels in $R1$ and $R2$. If the blob represents a left object, D is expected to be significantly larger than for a removed object. Therefore, the analysed object is classified as a left one if D exceeds a defined threshold. Fig. 48 presents an example of the procedure described above, for left and removed object cases (threshold = 0.6).

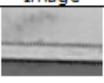
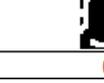
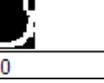
	Left object		Taken object	
	Image	Mask	Image	Mask
Object				
Edge detection				
Dilation				
Combined result				
Value of D	0.810		0.203	

Fig. 48. Example of differentiation of left and removed objects

The preliminary version of the event detection framework was tested for detection of unattended luggage at the Poznan-Lawica airport in Poland. It was found that if the data provided by object detection and object tracking are accurate, the detection works as expected (Fig. 49). However, the number of errors increases rapidly in conditions that cause the detection and tracking modules to produce erroneous data. For object detection, it happens in adverse lighting conditions (deep shadows, brilliant highlights, reflections on the floor, etc.). The errors in object tracking are caused by inability of the Kalman-based procedure to ensure proper object tracking in case of large number of moving objects (persons at the busy airport) and high number of frequent tracking conflicts. All errors from object detection and tracking are propagated to the event detection module and result in misdetections. Therefore, the main focus of further development should be on improving the tracking procedure. It is expected that more robust tracking procedure will provide sufficiently more accurate data on moving objects to the event detection module, which in turn increase the accuracy of event detection.





Fig. 49. Example of unattended luggage detection

4.4.3. Action recognition

Vision-based human action recognition is a high level process of image sequence analysis. Any robust action recognition system should be able to generalise over variations of style, view and speed within one class and distinguish between actions of different classes. This is usually achieved by learning so called action models from pre-acquired training datasets. Subsequently, these action representations are used for classification of unseen action instances.

The popular approach for modelling action is referred as Bag of Words [72][73][74] which models an action as a large visual vocabulary (dictionary, codebook) of discriminative code words. This visual dictionary is formed by the vector quantization of local feature descriptors extracted from images using for instance the k-means algorithm. A sequence of images is summarised by the distribution of code words from the fixed codebook by computing a histogram of code word occurrences based on the assignment of local descriptors. Action classification is performed by constructing a feature vector for video based on the defined dictionary to relate “new” descriptors in query images to descriptors previously seen in training.

Action video sequences are very high dimensional because of the human motion complexity. However, different instances of the given action reside only in a part of the entire feature space. This subspace can be considered as a nonlinear manifold embedded in a space of image frames. As a result, the discriminative and low dimensional manifold of the action can be discovered by a dimensionality reduction process. For instance, a view dependent action recognition can be performed by applying temporal Lapacian Eigenmaps on training videos to extract intrinsic characteristic of the action followed by a nearest neighbour classification schema in the obtained low dimensional space [75]. This concept is further extended in [76] and [77] to perform view independent action recognition. Alternatively, the temporal extent of action can be represented using Hidden Markov Model and the low dimensional Self Organizing Map [78], so the subsequent inference of action label can be performed in probabilistic manner.

4.5. Route reconstruction

Route reconstruction is one of the methods which is of particular interest of ADDPRIV project. Several approaches to this issue have already been proposed in literature. They depend on the camera arrangement within the multi camera system. A number of problems needs to be tackled for a robust route reconstruction. For instance, spatio-temporal camera relations need to be discovered and differences in images from various cameras should be taken into account for the purpose of proper object description.

In the following subsections, a number of approaches for object tracking in multi camera systems found in the literature, is briefly presented. After this description, a specific and already implemented solution, introducing the basic idea of route reconstruction is shown with more details.

4.5.1. Path recovery of a Disappearing Target in a Large Network of Cameras

The described method is based on given spatio-temporal topology of the monitoring network [89]. Whole system is described as a graph which is used to search and re-recognise the same object in particular cameras (as the object is moving). The most important thing in this method, is the object recognition algorithm, because it is crucial for the effectiveness of the described method. Proposed network model assumes two categories for object state: hidden and visible for cameras. Features of the objects are stored as parameter vectors and their similarity is the main input for re-recognition algorithm. Two phases are suggested in proposed method:

- Modified particle filter modelling of hidden (unobserved by cameras) places, where objects are not seen in any FOV – particle filter is chosen (instead of Kalman filter) because process of tracking many object is non-linear one and noise is expected.
- Finding the shortest way in state space (graph defined by the first phase)

Tracking the object in the system is supported by Markov chain which represents the object over the reconstructed path. Well-trained model of the network allows for more accurate tracking.

4.5.2. Appearance Modelling for Tracking in Multiple Non-overlapping Cameras

This method presents the approach to route reconstruction on the basis of brightness transfer functions (BTF) defined between pairs of cameras [90]. Authors empirically proves that these BTFs are contained in small subspace of the space of all possible BTFs, for particular pair of cameras. In order to obtain this subspace, learning process must be carried out. Next BTFs are used for mapping brightness of one camera to another. In relation to object appearance and illumination of the scene, some assumptions are required to be done: scene is illuminated by white light, object gives only diffusive light (no specular reflectance). Trained system decides if the analysed object representations from two different cameras are related to the same object, which is true if the brightness transfer function of this object lies in the trained (learnt) subspace. Another parameter supporting the decision is colour of the tracked object. Result of

using BTFs and colours can be joint in order to get better effectiveness. Described experiment assumes using only one (red) component as brightness. Effectiveness of the utilized method is very encouraging and in the mentioned experiment reaches 99% of accurate decision.

4.5.3. Bridging the Gaps between Cameras

This method uses a great amount of video data and displays "activity maps" utilizing temporal correlation between pairs of cameras and similarity of particular object [91]. Places of entry and exit points in cameras FOVs are also detected and used to recognize entry and exit events. In places where the camera network is "blind", virtual states are created (added). Conditional probabilities define rate of transitions between particular pair of states and on this basis, topology of the network is discovered. Obtained topology is spatio-temporal one. Entire method can operate unsupervised and doesn't need any calibration of the cameras. This method can generate redundant connection between cameras.

4.5.4. Continuously Tracking Objects Across Multiple Widely Separated Cameras

The proposed method of tracking object between non-overlapped FOVs of cameras uses two types of information acquired from monitoring system [92]. The first is appearance of the observed object and the second is spatio-temporal information. Two-dimensional histograms are used: histogram of colour for whole object (1st dimension) and histograms of colour for a part of the object, for example: head, torso and legs (2nd dimension). The spatio-temporal information are acquired on the basis of transition time distribution. This distribution function is modelled as mixture of Gaussian distributions. In this method three types of Gaussian distribution are chosen and each distribution describes the way of walking. Some people walk slowly other people very quickly. Each of these Gaussian distribution is weighted in order to fit the mixture of Gaussian distributions to the histogram of transition time between particular pair of cameras. Weights in the proposed model are estimated by "expectation maximization" algorithm. Experiments performed by authors of this paper validate correctness of used method.

4.5.5. Tag and Track System Across Multiple Cameras Network

This system is designed for tracking an object across multiple non-overlapped cameras in real time [93]. It is based on a probabilistic framework which integrates information from a variable number of heterogeneous modules in real time. These informative clues include:

- Information about a target (e.g. appearance and position) obtained from tracking modules, working on single and calibrated cameras
- Information from people re-identification modules, working across cameras
- The prior knowledge about scene, which origins from the camera network topology

The process starts with the initialisation, i.e. a manual selection of individual by the operator. The objective of system is to follow this chosen target automatically across different non-overlapped cameras. This is achieved by taking advantage of a few advanced technologies to tackle challenging problems of static and dynamic occlusions, crowded scenes, poor video quality and real-time operation.

First, foreground detection exploits the presence of many periodically-changing background elements in indoor scenes (escalators, scrolling advertisements, flashing lights). In general, it detects and models pixels exhibiting periodic changes in colour, and uses this information to predict the pixel colour in subsequent frames in order to improve on foreground detection [34]. Then the extracted moving foreground is fed into the Kalman Filter to perform a single camera tracking. For the problem of data association between cameras in network, a novel colour correction method is proposed to allow a robust appearance comparison between targets [94]. It can learn differences in camera colour responses up to second-order statistics; the algorithm is completely unsupervised, and can automatically detect whether it has gathered enough information for a reliable colour correction. In turn, the camera network layout is learnt automatically based on an activity-based semantic scene model [95]. In the first step, regions of interests, such as entry/exit zones, junctions, paths, and stop zones, are determined from motion tracks. Then, the topological relations between them are expressed in probabilistic manner using a Bayesian belief network.

Finally, a probabilistic approach fuses all heterogeneous information coming from the described modules [96]. The framework finds a target identity that is globally most likely to be the one intended by the operator. The solution is re-computed at each frame, using only the state of the system at the previous frame, thus avoiding the computational overhead of optimising a long track and allowing the multi-camera tracker to work in real-time.

5. Potential application of individual algorithms to the ADDPRIV project

The methods described in Sec. 4 include mainly the algorithms developed by, the Kingston University or the Gdansk University of Technology. Since prototypes of these algorithms are designed to work in real time, all of the presented methods could be potentially integrated in the designed system. Some of the presented system modules are related to the smart surveillance system core functionality, e.g. object detection, whereas others can be seen as higher level processing units for conceptual analysis like route reconstruction. The choice of certain modules should be done on the basis of End User requirements specified in the Deliverable 2.2.

If a similar functionality could be achieved employing a variety of described algorithms, a proper selection should be done during the system development, considering applied methods efficiency and existing test-bed conditions.

6. Conclusions

The aim of the ADDPRIV project is to enhance the privacy of individuals by extending functionality of surveillance systems. The catalogued commercial solutions show that a number of various smart video surveillance systems already exist. Still, none of them offers the full functionality developed within the ADDPRIV project. The main project objective, in terms of algorithmical solutions, is to prepare a route reconstruction algorithm (RRA) which application may lead to lowering the amount of privacy sensitive data stored. The RRA module is considered to be a vital part of smart surveillance system. At the same time, the development of a set of basic video processing methods is not a part of this project. Therefore, benefiting from some Partners' earlier experience in the image processing area, their previously developed algorithms can be integrated within the designed system. These algorithms can provide the functionality of event detection which could potentially trigger the RRA module. Besides the methods described in details in Sec. 4, the Partners have some additional algorithms at their disposal, which could also be used in the project upon several adjustments, helping to achieve project requirements. An important part of work to be done during the project scope is the integration of those existing solutions with the newly created system modules, including the route reconstruction module, the intelligent data management solutions with secure data deletion and access management.

7. References

- [1] B. Karasulu, "Review and evaluation of well-known methods for moving object detection and tracking in videos", Journal of aeronautics and space technologies, vol. 4, July 2010
- [2] K. Kim, L. S. Davis, "Object detection and tracking for intelligent video surveillance", Multimedia Analysis, Processing & Communications, SCI 346, pp. 265–288, 2011
- [3] S. N. Tamgade, V. R. Bora, "Motion Vector Estimation of Video Image by Pyramidal Implementation of Lucas Kanade Optical Flow", 2nd International Conference on Emerging Trends in Engineering and Technology, pp. 914–917, 2009
- [4] P. Lenz, J. Ziegler, A. Geiger, "Sparse scene flow segmentation for moving object detection in urban environments", IEEE Intelligent Vehicles Symposium, pp. 926–932, 2011
- [5] H. Lategahn, W. Derendarz, T. Graf, B. Kitt, J. Effertz, "Occupancy grid computation from dense stereo and sparse structure and motion points for automotive applications", IEEE intelligent Vehicles Symposium University of California, June 2010
- [6] N. Thakoor, J. Gao, H. Chen, "Automatic object detection in video sequences with camera in motion", Acivs, 2004
- [7] E. L. Andrade, S. Blunsden, R. B. Fisher, "Hidden Markov models for optical flow analysis in crowds", 18th International Conference on Pattern Recognition, pp. 460–463, Hong Kong, 2006
- [8] S. S. Cheung, C. Kamath, "Robust techniques for background subtraction in urban traffic video", Visual Communication and Image Processing, 2004
- [9] J. M. Milla, S. L. Toral, M. V. Barrero, F. Barrero, "Computer Vision Techniques for Background Modelling in Urban Traffic Monitoring", InTech, ISBN: 978-953-307-100-8, 2010
- [10] L. M. Fuentes, S. A. Velastin, "Foreground segmentation using luminance contrast", SIPA 2001
- [11] P. Dalka, "Detection and segmentation of moving vehicles and trains using Gaussian mixtures, shadow detection and morphological processing", Machine Graphics and Vision, vol. 15, no. 3/4, pp. 339–348, 2006
- [12] A. Shimada, R. Taniguchi, "Hybrid Background Model using Spatial-Temporal LBP", Advanced Video and Signal Based Surveillance, 2009
- [13] A. L. C. Barczak, "Feature-based rapid object detection: from feature extraction to parallelisation", PhD Thesis, 2007
- [14] C. N. Khac, J. H. Park, H.-Y. Jung, "Face detection using variance based Haar-like feature and SVM", World Academy of Science, Engineering and Technology, 2009
- [15] R. Lienhart, J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", IEEE International Conference on Image Processing, 2002
- [16] A. R. J. François, "CAMSHIFT Tracker Design Experiments with Intel OpenCV and SAI", IRIS Technical Report IRIS-04-423, University of Southern California, 2004
- [17] A. Czyżewski, G. Szwoch, P. Dalka, P. Szczuko, A. Ciarkowski, D. Ellwart, T. Merta, K. Łopatka, L. Kulasek, J. Wolski, "Multi-Stage Video Analysis Framework", Video Surveillance, Weiyao Lin (Ed.), ISBN: 978-953-307-436-8, InTech, 2011

- [18] H. Li, K. Ngan, "Automatic video segmentation and tracking for content-based applications", *IEEE Communication Magazine*, vol. 45, no. 1, pp. 27–33 (2007)
- [19] J. Konrad, "Videopsy: Dissecting visual data in space time", *IEEE Communication Magazine*, vol. 45, no. 1, pp. 34-42, 2007
- [20] A. Elgammal, D. Harwood, L. Davis, "Non parametric model for background subtraction", *Lecture Notes in Computer Science*, vol. 1843, pp. 751-767, 2000
- [21] C. Stauffer, W. Grimson, „*Learning patterns of activity using real-time tracking*“, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000
- [22] N. Buch, J. Orwell, S. A. Velastin, "Detection and Classification of Vehicles for Urban Traffic Scenes", *Visual Information Engineering*, 2008
- [23] J. Renno, J. Orwell, D. Thirde, G. A. Jones, "Shadow Classification and Evaluation for Soccer Player Detection", *BMVA 2004*
- [24] J. Renno, J. Orwell, G. A. Jones, "Evaluation Of Shadow Classification Techniques For Object Detetion And Tracking", *ICIP 2004*
- [25] D. Bloisi, L. Iocchi, D. N. Monekosso, P. Remagnino, "A Novel Segmentation Method Fof Crowded Scenes", *VISAPP 2009*
- [26] N. Buch, J. Orwell, S. A. Velastin, "Urban Road User Detection and Classification using 3D Wire Frame Models", *IET Computer Vision Journal*, 2010
- [27] N. Buch, J. Orwell, S. A. Velastin, "3D Extended Histogram of Oriented Gradients (3DHOG) for Classification of Road Users in Urban Scenes", *BMVC 2009*
- [28] A. Czyżewski, P. Dalka, „*Visual traffic noise monitoring in urban areas*“, *International Journal of Multimedia and Ubiquitous Engineering*, vol. 2, no. 2, pp. 91–101, 2007
- [29] T. Horprasert, D. Harwood, L. Davis, "A statistical approach for real-time robust back-ground subtraction and shadow detection", *Proc. of IEEE Frame Rate Workshop*, pp. 1–19, Kerkyra, Greece, 1999
- [30] E. Dougherty, R. Lotufo, "Hands-on morphological image processing", *SPIE Press*, 2003
- [31] L. Xiu, J. Landabasso, M. Pardas, "Shadow removal with blob-based morphological reconstruction for error correction", *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. II-729–732, Philadelphia, March 2005
- [32] F. Yin, D. Makris, S. A. Velastin, "Time efficient ghost removal for motion detection in visual surveillance systems", *Electronics Letters*, 2008
- [33] J. Rymel, J. Renno, D. Greenhill, J. Orwell, G. A. Jones, "Adaptive Eigen-Backgrounds For Object Detection", *ICIP 2004*
- [34] A. Colombo, V. Leung, J. Orwell, S. A. Velastin, "Markov Models of Periodically Varying Backgrounds for Change Detection", *IET 2007*
- [35] Viola, Jones, "Robust Face Detection", *International Journal of Computer Vision*, vol. 57, no. 2
- [36] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", *International Conference on Computer Vision and Pattern Recognition*, June 2005
- [37] D. Simonnet, S. A. Velastin, "Pedestrian detection based on Adaboost algorithm with a pseudo-calibrated camera", *IPTA 2010*
- [38] A. Moutzouris, J. Martinez-del-Rincon, M. Lewandowski, J-C. Nebel, D. Makris, "Human Pose Traking In Low Dimensional Space Enhanced By Limb Correction", *ICIP, 2011*

- [39] J. Martinez-del-Rincon, D. Makris, C. Orrite-Urunuela, J-C. Nebel, "Tracking Human Position and Body Parts Using Kalman and Particle Filters Constrained by Human Biomechanics", IEEE Transactions on Systems Man and Cybernetics 2010
- [40] J. Martínez del Rincón, J-C. Nebel, D. Makris, "Graph-based Particle Filter for Human Tracking with Stylistic Variations", BMVC 2011
- [41] J. Martínez-Sotoca, F. J. Ferri, J. Gutierrez, P. Remagnino, G. A. Jones, "Approximated tracking of multiple non-rigid objects using adaptive quantization and resampling techniques", VIIP 2001
- [42] D. Bloisi, L. I. Iocchi, L. Marchetti, D. N. Monekosso, P. Remagnino, "An Adaptive Tracker for Assisted Living", AVSS 2009
- [43] J. Renno, J. Orwell, G. A. Jones, "Learning Surveillance Tracking Models for the Self-Calibrated Ground Plane", BMVC 2002
- [44] D. Greenhill, J. Renno, J. Orwell, G. A. Jones, "Occlusion analysis: Learning and utilising depth maps in object tracking", Image and Vision Computing, 2007
- [45] L. M. Fuentes , S. A. Velastin, "People tracking in surveillance applications", Image and Vision Computing, 2006
- [46] L. M. Fuentes , S. A. Velastin, S.A. "Tracking People for Automatic Surveillance Applications", PRIA 2004
- [47] M. Xu, T. Ellis, "Augmented tracking with incomplete observation and probabilistic reasoning", Image and Vision Computing, 2006
- [48] E. Corvee, S. A. Velastin, G. A. Jones, "Occlusion tolerant tracking using hybrid prediction schemes", Automatica Sinica, Special Issue on Visual Surveillance of Dynamic 2003
- [49] M. Xu, T. Ellis, S. J. Godsill, G. A. Jones, "Visual tracking of partially observable targets with suboptimal filtering", IET Computer Vision, 2011
- [50] M. Xu, J. Orwell, L. Lowey, D. J. Thirde, "Architecture And Algorithms For Tracking Football Players With Multiple Cameras", IEE Proceedings - Vision, Image and Signal Processing 2005
- [51] M. Xu, J. Orwell, G. Jones, "Tracking football players with multiple cameras", ICIP, 2004
- [52] J. Black, T. Ellis, "Multi camera image tracking", Image and Vision Computing, 2006
- [53] G. Welch, G. Bishop, "An introduction to the Kalman filter", Technical report, TR-95041, Department of Computer Science, Univ. of North Carolina, Chapel Hill, 2006
- [54] A. Czyżewski, P. Dalka, "Examining Kalman filters applied to tracking objects in motion", 9th International Workshop on Image Analysis for Multimedia Interactive Services, pp. 175–178, Klagenfurt, May 2008
- [55] G. Szwoch, P. Dalka, A. Czyżewski, "A framework for automatic detection of abandoned luggage in airport terminal", 3rd International Symposium on Intelligent and Interactive Multimedia: Systems and Services, Baltimore, July 2010
- [56] Y. Dedeoğlu, B. U, Töreyn, U. Güdükbay, A. E. Çetin, "Silhouette-based method for object classification and human action recognition in video", Computer and Information Science: Multimedia Systems and Applications, pp. 64–77, 2006
- [57] R. Tsai, "A Versatile Camera Calibration Technique For High Accuracy 3d Machine Vision Metrology Using Off-The-Shelf TV Cameras And Lenses", IEEE J. Robotics Automat., vol. RA-3, no. 4, pp. 323–344, 1987
- [58] G. Szwoch, P. Dalka, A. Czyżewski, "Objects Classification Based on Their Physical Sizes for Detection of Events in Camera Images", NTAV/SPA, pp. 15–20, Poznań, 2008

- [59] T. Deselaers, G. Heigold, H. Ney, "*Object classification by fusing SVMs and Gaussian Mixtures*", Pattern Recognition Journal, vol. 43, Issue 7, July 2010
- [60] B. Epshtein, S. Ullman, "*Feature hierarchies for object classification*", 10th IEEE International Conference on Computer Vision, vol. 1, pp. 220–227, 2005
- [61] S. Nakajuma, A. Binder, U. Brefeld, K.-R. Müller, M. Kawanabe, "*Non-sparse feature mixing in object classification*", Information Processing Society of Japan, 2009
- [62] C.-W. Chong, P. Raveendran, R. Mukundan, "*The scale invariants of pseudo-Zernike moments*". Pattern Analysis Applications, pp. 176–184, 2003
- [63] S. Oh, Y. Lee, K. Hong, K. Kim, K. Jung, "*View-point Insensitive Human Pose Recognition using Neural Network*", Proc. of World Academy of Science, Engineering and Technology, vol. 34, 2008
- [64] G. Yu, "*Fast wavelet-based visual classification*", Proc. IEEE International Conference on Pattern Recognition, 2008
- [65] D. Ellwart, A. Czyżewski, "*Viewpoint independent shape-based object classification for video surveillance*", 12th International Workshop on Image Analysis for Multimedia Interactive Services", 2011
- [66] O. Javed, M. Shah, "*Tracking and object classification for automated surveillance*", Proceedings of 7th European Conference on Computer Vision – Part IV, 2002
- [67] S. Ullman, E. Sali, M. Vidal-Naquet, "*A fragment-based approach to object representation and classification*", Lecture Notes in Computer Science, vol. 2059/2011, pp. 85–100, 2001
- [68] S. Johnsen, A. Tews, "*Real-Time Object Tracking and Classification Using a Static Camera*", Proceedings of the IEEE ICRA 2009, Workshop on people detecting and tracking, Kobe, Japan, May 2009
- [69] C-W. Hsu, C-C. Chang, C-J. Lin, "*A practical guide to support vector classification*", Technical report, Department of Computer Science, National Taiwan University, July, 2003
- [70] C-C. Chang, C-J. Lin, "*LibSVM: a library for Support Vector Machines*", Technical report, Department of Computer Science, National Taiwan University, February, 2009
- [71] L. Khoudour, D. Aubert, S. A. Velastin, V. Leung, J. Orwell, "*Video-based detection of specific events in public transport networks*", Studia Informatica Universalis, 2010
- [72] C. Schuldt, I. Laptev, B. Caputo, "*Recognizing Human Actions: A Local SVM Approach*", ICPR, 2004
- [73] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, "*Behavior Recognition via Sparse Spatio-temporal Features*", ICCCN 2005
- [74] J-C. Nebel, M. Lewandowski, J. Thevenon, F. Martinez, S. A. Velastin, "*Are current monocular computer vision systems for human action recognition suitable for visual surveillance applications?*", ISVC 2011
- [75] M. Lewandowski, J. Martinez-del-Rincon, D. Makris, J-C. Nebel, "*Temporal Extension of Laplacian Eigenmaps for Unsupervised Dimensionality Reduction of Time Series*", ICPR 2010
- [76] M. Lewandowski, D. Makris, J-C. Nebel, "*View and Style-Independent Action Manifolds for Human Activity Recognition*", ECCV 2010
- [77] M. Lewandowski, D. Makris, J-C. Nebel, "*Probabilistic Feature Extraction from Time Series using Spatio-Temporal Constraints*", PAKDD, 2011
- [78] F. Martinez-Contreras, C. Orrite-Urunuela, E. Herrero-Jaraba, H. Ragheb, S. A. Velastin, "*Recognizing Human Actions Using Silhouette-based HMM*", AVSS, 2009

- [79] Y. Ke, R. Stukthankar, M. Herbert, "*Volumetric features for video event detection*", International Journal of Computer Vision, vol. 88, no. 3, pp. 339–362, 2010
- [80] Y. Ke, R. Sukthankar, M. Hebert, "*Spatio-temporal Shape and Flow Correlation for Action Recognition*", Visual Surveillance Workshop, 2007
- [81] Y. Ke, R. Sukthankar, M. Herbert, "*Event detection in crowded videos*", IEEE 11th International Conference on Computer Vision, pp. 1–8, Rio de Janeiro, 2007
- [82] C. Piciarelli, G. L. Foresti, "*Surveillance-oriented event detection in video streams*", IEEE Intelligent Systems, vol. 26, issue 3, pp. 32–41, 2011
- [83] L. Ballan, M. Bertini, A. D. Bimbo, L. Seidenari, G. Serra, "*Event detection and recognition for semantic annotation of video*", Computer Science, Multimedia Tools and Applications, vol. 51, no. 1, pp. 279–302, 2011
- [84] Y.-L. Tian, A. Hampapur, L. Brown, R. Feris, M. Lu, A. Senior, C.-F. Shu, Y. Zhai, "*Event detection, query, and retrieval for video surveillance*", Artificial Intelligence for Maximizing Content Based Image Retrieval, 2008
- [85] U. K. Kotikalapudi, "*Abnormal Event Detection in Video*", A project Report, Supercomputer Education Research Center, Indian Institute of Science, July 2007
- [86] F. Jiang, J. Yuan, S. A. Tsafaris, A. K. Katsaggelos, "*Anomalous video event detection using spatiotemporal context*", Computer Vision and Image Understanding, vol. 115, issue 3, pp. 323–333, 2011
- [87] S. Guler, W. H. Liang, I. A. Pushee, "*A Video Event Detection and Mining framework*", Conference on Computer Vision and Pattern Recognition Workshop, pp. 42–49, 2008
- [88] K. Ellingsen, "*Salient Event-detection In Video Surveillance Scenarios*", Proc. of 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams, pp. 57–63, 2008
- [89] A. Lev-Tov, Y. Moses, "*Path Recovery of a Disappearing Target in a Large Network of Cameras*", ICDSC 2010, ACM, pp. 57–64, 2010
- [90] O. Javed, "*Appearance Modeling for Tracking in Multiple Non-overlapping Cameras*", IEEE International Conference on Computer Vision and Pattern Recognition, IEEE, pp. 26–33, 2005
- [91] D. Makris, T. Ellis, J. Black, "*Bridging the Gaps between Cameras*", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), IEEE, pp. 200–205, 2004
- [92] Y. Cai, W. Chen, K. Huang, T. Tan, "*Continuously Tracking Objects Across Multiple Widely Separated Cameras*", ACCV 2007, Springer, pp. 843–852, 2007
- [93] A. Colombo, "*Tracking People across Multiple Cameras in the Presence of Clutter and Noise*", PhD Thesis 2011
- [94] A. Colombo, V. Leung, J. Orwell, S. A. Velastin, "*Colour constancy techniques for re-recognition of pedestrians from multiple surveillance cameras*", M2SFA2 2008
- [95] D. Makris, T. Ellis, "*Learning Semantic Scene Models From Observing Activity in Visual Surveillance*", IEEE Transactions on Systems, Man, and Cybernetics 2005
- [96] J. Annesley, V. Leung, A. Colombo, J. Orwell, S. A. Velastin, "*Fusion of multiple features for identity estimation*", IET 2006

- [97] Y. Ke, R. Sukthankar, M. Herbert, "*Efficient visual event detection using volumetric features*", Proc. Of International Conference on Computer Vision, 2005