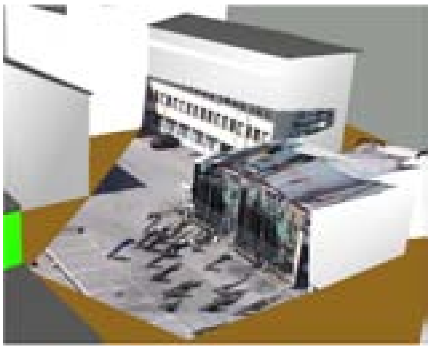
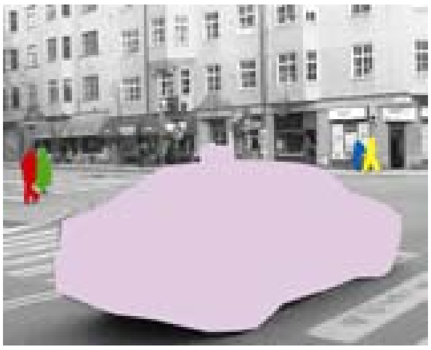
New Systems for Urban Surveillance

Hedvig Sidenbladh, Jörgen Ahlberg, Lena Klasén





FOI-R--1668--SE
ISSN 1650-1942

Sensor Technology
User report

June 2005

FOI is an assignment-based authority under the Ministry of Defence. The core activities are research, method and technology development, as well as studies for the use of defence and security. The organization employs around 1350 people of whom around 950 are researchers. This makes FOI the largest research institute in Sweden. FOI provides its customers with leading expertise in a large number of fields such as security-policy studies and analyses in defence and security, assessment of different types of threats, systems for control and management of crises, protection against and management of hazardous substances, IT-security an the potential of new sensors.

FOI
Defence Research Agency
Sensor Technology
P.O. Box 1165
SE-581 11 Linköping

Tel: +46-8-5550 3000
Fax: +46-8-5550 3100

www.foi.se

New Systems for Urban Surveillance

FOI-R--1668--SE
ISSN 1650-1942

Sensor Technology
User report

June 2005

Issuing organization FOI – Swedish Defence Research Agency Sensor Technology P.O. Box 1165 SE-581 11 Linköping	**Report number, ISRN** FOI-R--1668--SE	**Report type** User report
	Research area code 4. C4ISTAR	
	Month year June 2005	**Project no.** I3475
	Sub area code 42 Above water Surveillance, Target acquistion and Reconnaissance	
	Sub area code 2	
Author/s (editor/s) Hedvig Sidenbladh Jörgen Ahlberg Lena Klasén	**Project manager** Jörgen Ahlberg	
	Approved by Mattias Severin	
	Sponsoring agency FOI	
	Scientifically and technically responsible Tomas Chevalier	
Report title New Systems for Urban Surveillance		

Abstract (not more than 200 words)

The growing number of surveillance sensors in our society present new challenges, both in terms of data overview and of personal integrity preservation for watched individuals. This report presents suggestions of how to adress these problems. We introduce four new concepts for surveillance in an urban environment, and suggestions on how to realize these concepts. We believe that these concepts are beneficial, possibly necessary, for the development of future successful surveillance systems. Finally, we belive that the kind of surveillance systems suggested are realizable within a near future.

Keywords

Surveillance, Visualization, Multiple Sensors, Integrity Preserving Surveillance, Human-Aware Surveillance

Further bibliographic information	**Language** English
ISSN 1650-1942	**Pages** 29 p.
	Price acc. to pricelist

Utgivare FOI - Totalförsvarets Forskningsinstitut - Sensorteknik Box 1165 581 11 Linköping	**Rapportnummer, ISRN** FOI-R--1668--SE	**Klassificering** Användarrapport
	Forskningsområde 4. Ledning, informationsteknik och sensorer	
	Månad, år Juni 2005	**Projektnummer** I3475
	Delområde 42 Spaningssensorer	
	Delområde 2	
Författare/redaktör Hedvig Sidenbladh Jörgen Ahlberg Lena Klasén	**Projektledare** Jörgen Ahlberg	
	Godkänd av Mattias Severin	
	Uppdragsgivare/kundbeteckning FOI	
	Tekniskt och/eller vetenskapligt ansvarig Tomas Chevalier	

Rapportens titel (i översättning)

Nya system för urban övervakning

Sammanfattning (högst 200 ord)

Det växande antalet övervakningssensorer i samhället idag leder till ett antal nya utmaningar, såsom problem med översikt över stora datamängder och intrång i den personliga integriteten hos övervakade individer. Denna rapport presenterar förslag på hur dessa problem kan överbryggas. Vi föreslår fyra nya koncept för övervakning i urban miljö, vilka vi tror är gynnsamma, kanske också nödvändiga för utvecklingen av framtida framgångsrika övervakningssystem. Slutligen tror vi att de föreslagna övervakningskoncepten kan bli realitet inom en nära framtid.

Nyckelord

Övervakning, Visualisering, Multipla sensorer, Integritetsskyddad övervakning, Människomedveten övervakning

Övriga bibliografiska uppgifter	**Språk** Engelska
ISSN 1650-1942	**Antal sidor:** 29 s.
Distribution enligt missiv	**Pris: Enligt prislista**

Contents

Chapter 1

Introduction

Visual surveillance systems are increasingly common in our society today. You can hardly take a walk in the center of a modern city without being recorded by several surveillance cameras, even less so inside shops. During military operations, surveillance systems are useful for detection of trespassing, tactical decision support, training, and documentation.

Currently, there is a growing interest in these issues at FOI. Research is conducted with applications such as anti-terrorist operations, urban crisis and airport security in mind. This report is not directed towards a specific application, but discusses concepts for future surveillance in general.

The rising numbers of surveillance sensors introduce problems, both on how to get an overview of the surveillance data, and how to preserve the personal integrity of the people being watched by the sensors. This report present suggestions on how to address these problems. We introduce four new concepts for surveillance in an urban environment, and suggestions on how to realize these concepts using technology developed at FOI.

1.1 Problem 1: Overview of all the surveillance data

The traditional surveillance system in urban areas consists of a set of CCTV cameras acquiring images that are recorded and monitored at a surveillance central. In a surveillance central, a set of TV screens show the images from one or more cameras per screen. The problem with this approach is that each camera records micro events, and these micro events are hard to relate to other micro events recorded by other cameras. Thus, it is difficult to put the micro events in a correct spatial and temporal context, and also to get an overview of the entire situation, i.e., situation awareness.

3D presentation. One solution is to use a 3D model of the area, and to project the images from all cameras as texture on the 3D structure. In Chapter 2 we describe this concept in more detail.

Multiple heterogeneous sensors. Using the 3D model, we can make full use of the capabilities of diverse sensor systems, and fuse sensor data from heterogeneous sensors by projecting them into the model. The concept of using multiple heterogeneous sensors is more closely described in Chapter 3.

Human-aware systems. We propose to use computer vision techniques that can detect and classify human motion and behavior, enabling functionality like warning for mobs, fights, accidents, or other abnormal behaviors. Detection of human behavior can for example be used as an "alarm clock" to steer the attention of a human operator of a camera surveillance system. The concept of human-aware systems is presented in Chapter 5.

1.2 Problem 2: Personal integrity in a world full of cameras

There is an inherent conflict between the demands for security of the public, and the demands to preserve the personal integrity of individuals in the public. In other words, people in general (understandably) do not like to be watched, especially if it is not clear who has access to the recorded surveillance data (Senior et al., 2003).

Integrity preserving surveillance. One solution to this problem is to cover parts of the surveillance images in order to conceal peoples' identities, but not their actions or activity. We call this *integrity preserving surveillance* (IPS) since the technique strives to ensure the security of the watched subjects while not intruding on their personal integrity. In Chapter 4, IPS is discussed in detail.

Human-aware systems. The concept of IPS requires techniques for locating humans in images and recognizing human activities in video. This links IPS to Human-aware systems in an intimate way.

Chapter 2

3D Presentation

One of the greatest obstacles in a surveillance system with a large number of sensors is overview. A human operator of a surveillance system is showered with a large number of images of micro events that are difficult to relate in space and, sometimes, in time.

A suggestion is to use a 3D model of the area. In this virtual environment, the cameras from the real environment are represented by projectors, that project the camera views as texture onto the 3D model, see Figure 2.1. This approach has several advantages:

1. The context in which each camera is placed is visualized and becomes obvious.
2. The spatial relation between different cameras become obvious.
3. Imagery from several cameras can be studied simultaneously, and an overview of the entire area is easily acquired.

We propose to exploit this approach to create a framework for surveillance of urban areas. Even if the idea is not completely new, it is not widely used, and it improves the general situation awareness tremendously. In addition, there is a great need for methods in many applications, e.g. military operations, law enforcement and anti-terrorism.

The concept is built around a 3D model of the area to be surveyed. In this 3D model, all available sensor data can be visualized in such a way that their context and mutual relations are immediately visible, see Figure 2.1.

We have developed a research platform for visualization of the surveyed area. The platform is a visualization tool called SceneServer, built at FOI on open source software. SceneServer visualizes 3D models and projects textures from input video, and is controlled using either a GUI or by commands over a network.

The concept of projecting sensor data on a 3D model of the environment was demonstrated in May 2004 in Norrköping (Ahlberg, 2004; Ahlberg and Klasén, 2004).

2.1 Building the 3D model

The actual key to an operational system is that the 3D model can be automatically generated. If this is the case, the surveillance system can be deployed quickly in a previously unknown area.

In order to construct high fidelity 3D environment models, detailed and reliable information of the environment is needed. In an indoor environment, a CAD model can often be built directly from blueprints of the building. However, in outdoor environments and in cases where blueprints of a building is not available, this is not a feasible approach. Furthermore, this approach does not allow for updating the model for areas undergoing changes e.g. from explosions, earth quakes etc.

An alternative is to scan the environment and reconstruct 3D structure from sensor data. Modern laser scanner systems for 3D sensing, usually combined with passive high resolution electro-optical image sensors, (visual and/or infrared) provide an excellent source of data for obtaining the information needed. To utilize the data provided by the sensors and obtain an automatic process from sensor data to environment models, new and efficient methods for data processing and environment model construction need to be developed. This is the topic and long term goal for our work on automatic methods for rapid construction of high-fidelity natural environment models at FOI.

Generally, research and development on methods for processing laser data is a growing and active area today. Results have been reported for many problem areas of which several are of interest for high-fidelity natural environment modelling, e.g. ground surface modelling, 3D reconstruction of buildings, tree identification and forest mapping, etc.

For the construction of 3D environment models the raw data from the laser scanner and camera system must be processed in order to produce a number of specific data sets, e.g., digital elevations models (DEMs), orthophoto mosaics, 3D object models and various feature data in terms of points, vectors, and polygons. These data sets can then be put together to form the desired environment models using some COTS software package.

For the processing of sensor data we have developed several methods, including a new method for ground surface modelling and ground point classification based on active contours and a method for subsequent classification of the non-ground points into e.g. buildings, vegetation, roads, (lamp)posts etc.

For vegetation we have developed a novel method for identification of single trees and estimation of tree position, height and crown diameter. This method has recently been extended to discriminate between tree species.

For buildings we have developed a new method for 3D building reconstruction. This method is very general and allows for complicated building structures like doom shaped roofs and curved walls. The evaluation is going on and no results have been published yet.

A sample result is shown in Figure 2.2.

2.2 Technical requirements

Technology required to realize this concept is:

- A method for automatic building of 3D models covering the surveyed area. In an indoor environment, these can be build from blueprints. In an outdoor environment, or an environment that changes over time due to explosions etc., data from an airborn laser range scanner can be used to build the model.

- Techniques for keeping track of sensor positions and configurations in the model, and for projecting sensor data onto the model.

- Techniques for presenting the 3D model with the projected sensor data.

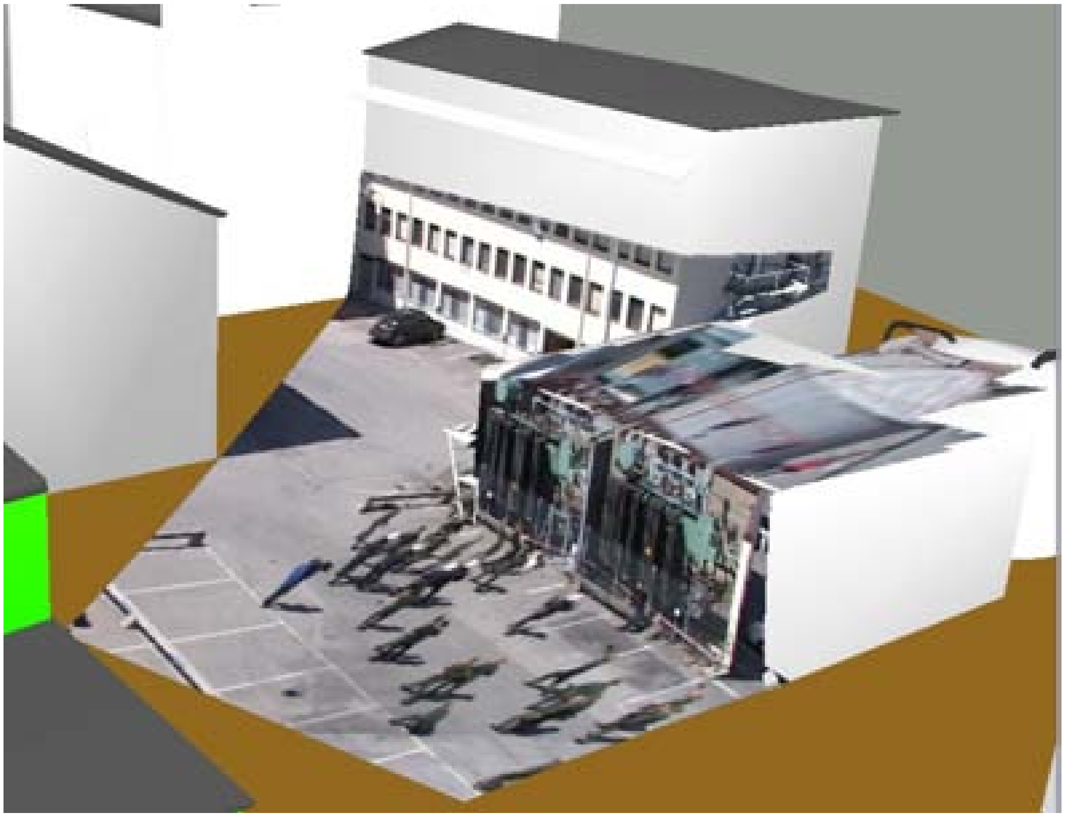


Figure 2.1: Projecting video onto a 3D model.

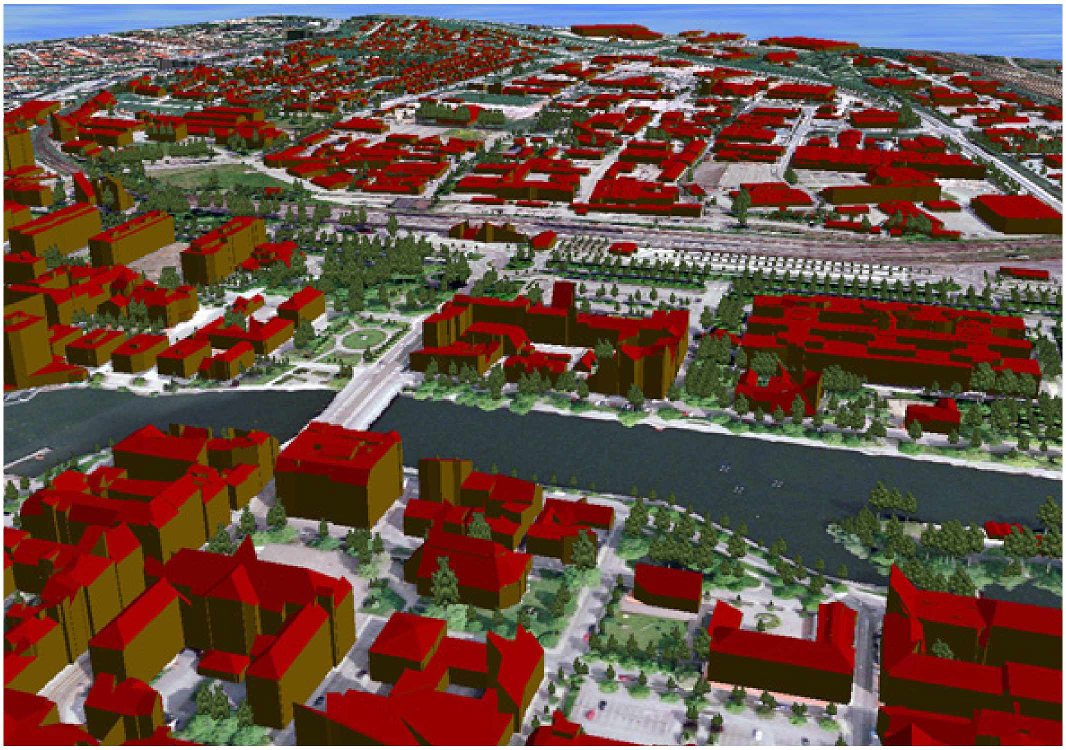


Figure 2.2: Automatically generated 3D model of Norrköping.

Chapter 3

Multiple Heterogeneous Sensors

Different types of sensors give different type of information. Sensors can be active (such as laser radar), passive (such as video), of long or short range, of low or high resolution. Using the 3D model, we can make full use of the capabilities of diverse sensor systems, and fuse sensor data from many different types of sensors by projecting them into the model.

The key issue with the multiple heterogeneous sensors concept is to make use of the benefit brought by new capabilities by new and cooperating sensor systems. Besides conventional acoustic, seismic, electro-optical and infrared sensors, this can e.g. include range gated imaging, full 3D imaging laser radar sensors, multispectral imaging, mm-wave imaging or the use of low frequency radars in urban environment. Assume, for example, that we have a sensor that can localize gunfire. The position of the shooter can then immediately be marked in the 3D model, which gives several interesting possibilities:

- If the shooter is within the field of view of a camera, he is pointed out by marking the location of the shot in the 3D model (see Figure 3.1). The shooter can then be tracked forwards and backwards in time, searching for pictures suitable for identification. The information can also be used to warn others in the area.
- Regardless if the shooter is within the field of view of a camera or not, the shooter's field of view can be marked in the 3D model. The marked area is a risk area that should be avoided and warned for.
- The same functionality can be used in a deployment scenario, aiding the placement of sensors, snipers and people.

Other sensor examples are passage detection sensors, sensors that track or classify vehicles, sensors that detect suspicious events or behavior.

The placement of surveillance sensors is a non-trivial task, determined by laws and by the functionality and limitations of the sensors (Bergstrom, 2004; Nastell, 2002). Well-placed sensors is a necessary condition for a well-functioning surveillance system. Although important, this issue is not discussed here.

3.1 Sensor types

The most common type of surveillance sensor is CCTV cameras. However, a number of different sensors provide useful data, which can be visualized (see Chapter 2).

3.1.1 Cameras

CCTV cameras are the most commonly used sensors in surveillance systems. Image and video give rich information about the world, but are difficult to interpret automatically. Therefore, it is most common that the images are interpreted by a human operator of the surveillance system.

In Chapter 5, automatic interpretation of human activities in surveillance video is discussed. Other interesting problems include automatic analysis of traffic flow, detection of anomalies in industrial production, and recognition of car number plates.

Camera calibration and ego-motion. As stated in the Chapter 2, the 3D model of the environment is obtained before the sensor data is acquired and projected into the model. However, with knowledge of the intrinsic and extrinsic parameters of cameras watching the scene, the 3D model can be updated with information from the cameras. The camera parameters can thereafter be re-calibrated with information from the updated 3D model, so that the two models can improve in an iterative manner.

Camera calibration is also interesting for reconstruction of the 3D positions of people and other moving objects in the scene. From 2D image coordinates in multiple camera views, the 3D positions can be triangulated, given that enough is known about the camera parameters.

3.1.2 Acoustic sensors

A network of acoustic sensor nodes can be used to locate gunshots, acoustic hotspot areas and track sound sources. For example, technology used in military applications for tracking ground vehicles in terrain can modified to fit in with an urban scenario. The output of the sensor network is synchronized with all other information in the system and areas of interest is automatically displayed in the 3D model with a classification tag to indicate the type of event. Figure 3.1 illustrates how the acoustic sensor nodes interact to pinpoint a gunshot location. Figure 3.2 shows the real time acoustic hotspot that represents the movements of a crowd.

3.1.3 Imaging radar system

Researchers at FOI have developed an imaging radar system, capable of delivering through-the-wall measurements of a person. Figure 3.3 shows the radar images when measuring a person through three different inner wall types at 94 GHz.

3.1.4 Passage detection sensors

Passage detection sensors can be used for determining when people and/or vehicles enter a surveyed area and other sensors should be activated. Examples are:

- **Fiber optic perimeter sensors** that react on pressure, i.e., when someone walks on the sensor (that consequently should be placed slightly below the ground's surface). Several types of such sensors exist – a fiber-optic pressure-sensitive cable has been used at FOI.
- **Laser beam sensors** that react when someone breaks an (invisible) laser beam.
- **Seismic sensors**, e.g., geophones, that register vibrations in the ground.

Several other types of passage detectors are commercially available.

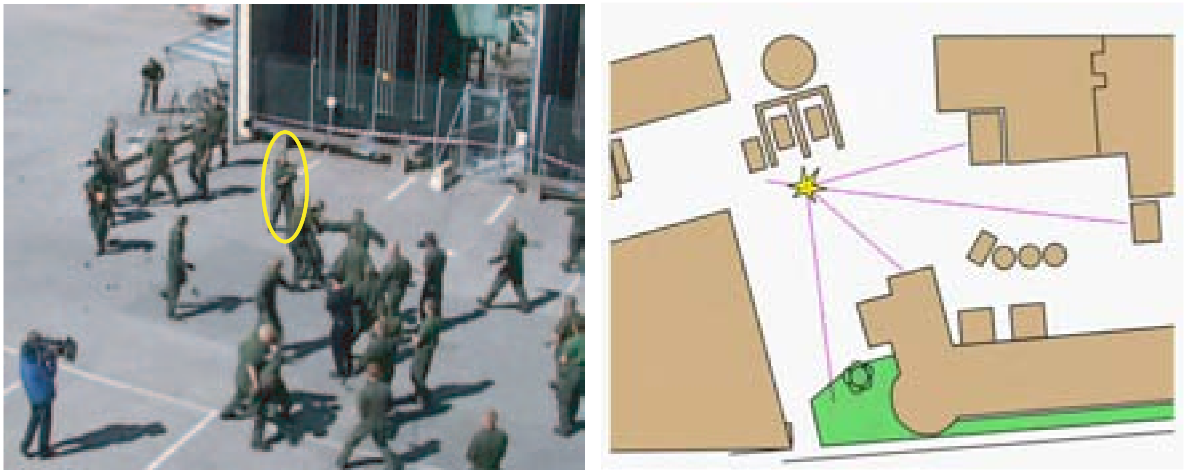


Figure 3.1: Localization of gunfire using acoustic sensors.

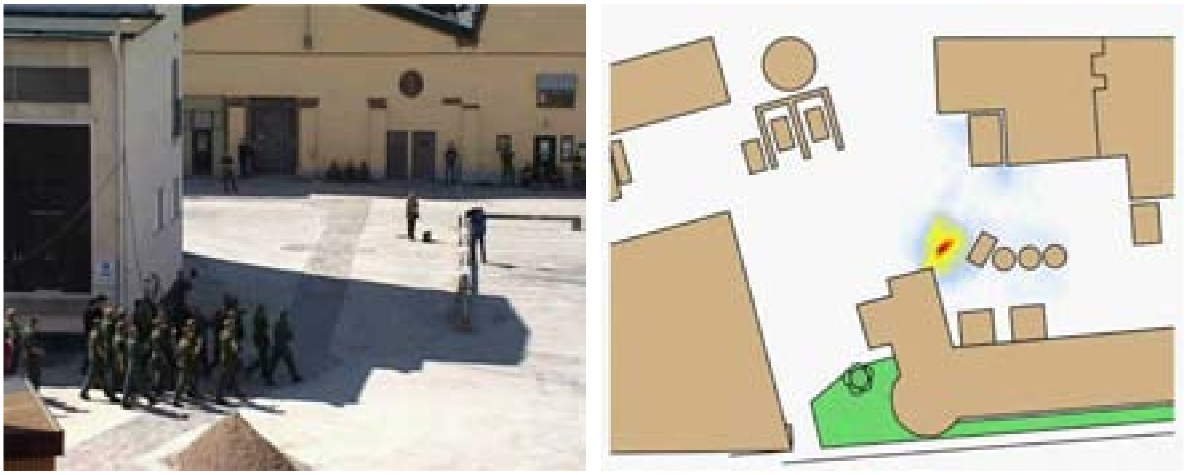


Figure 3.2: Localization of people using acoustic sensors.

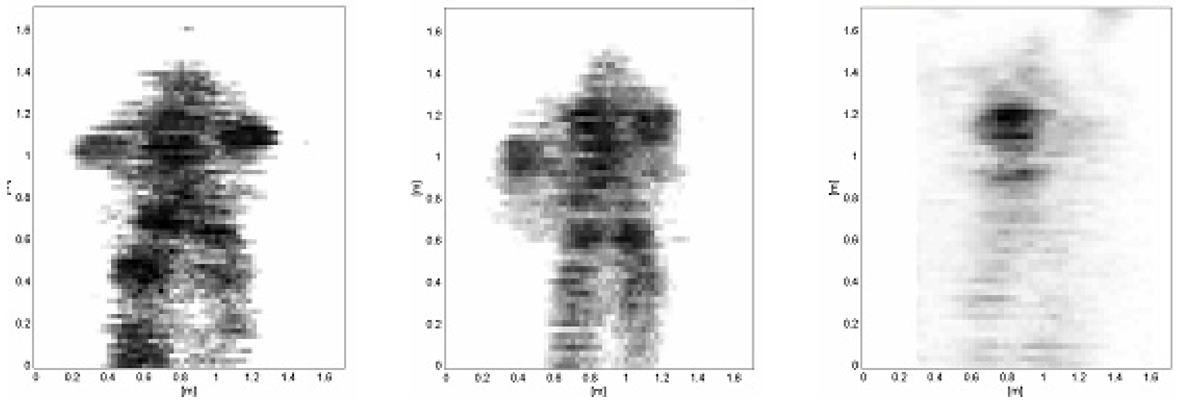


Figure 3.3: Imaging of a person behind a wall by measurements carried out at FOI with an in-house developed imaging radar system. Radar images measured through three different inner wall types at 94 GHz are shown. Left: A 12.5 mm thick plasterboard. Middle: Two 12.5 mm thick plasterboards separated by an 45 mm air slit. Right: A 12.5 mm thick chipboard.

3.2 Technical requirements

Technology required to realize this concept is:

- Signal processing methods to extract information from the raw data obtained from each sensor. For some sensor types, e.g., passage detectors, this is basically a solved issue. For others, e.g., cameras, this requires a large research and development effort over several years. Extraction of information from images is further discussed in Chapter 5.
- Methods to fuse the sensor data. Here, we suggest projecting all sensor data onto the 3D model discussed in the previous chapter.

Chapter 4

Integrity Preserving Surveillance

We use the term *integrity preserving surveillance* (IPS) to denote various technologies enabling surveillance that does not reveal people's identities. The implication for IPS is that people generally do not like to be watched and/or identified, and, furthermore, the use of surveillance cameras is often restricted by law. A similar concept has also been developed at IBM (Senior et al., 2003), which implies a growing interest in this type of issues.

The two scenarios below explain the potentials. IPS systems put high demands on functionalities like robust classification and tracking of people and vehicles.

Military scenario. *In a peace keeping operation we want to deploy a surveillance system in certain areas in a city. The problem is that we know that this is unpopular among the city's inhabitants, and the solution is an IPS system. The system maps, as described above, the videos on a 3D-model of the areas, but replaces people and vehicles with blobs or symbols. The non-manipulated videos are encrypted and stored at an institution that the local population have trust in. The manipulated videos can even be publicly displayed, for example on a web server. The semantic data used for image manipulation is also used for behavior analysis and warning.*

Commercial scenario. *A shop-keeper wants to know how bypassers respond to different arrangements and items in his shop window. A surveillance camera looking out through the window would only be a partial solution, and might also be illegal or require special permissions. An IPS camera, on the other hand, would not reveal any identities. In fact, it would not even reveal any images, but only the wanted statistics: "14 people passed, 5 of them stayed to look, 3 of them looked at item X."*

One morning, when the shop-keeper arrives, his window is smashed. The police comes and unlocks the camera, enabling it to show the stored and encrypted images. These images do still not reveal any identities, since people are covered with blobs or replaced by drawn stick figures in the same pose. This is enough to point out the blobs or stick figure that commit crimes, and these specific persons are, after the suitable legal decisions are made by the suitable authorities, unlocked and their images shown.

4.1 Technical requirements

Technology required to realize this concept is:

- Methods for locating and tracking moving objects in the video image. The requirements on robustness is very high in this application, much higher than in, e.g., gaming applications where a certain failure ratio is accepted by the user. Commercial video tracking software is available for this task, although the robustness of these methods is not tested for the high robustness requirements of our application.

- Methods for identifying human activity, direction of view, etc. These methods will be used to label the masked objects in the image, and to perform counts like ”14 people passed, 5 of them stayed to look, 3 of them looked at item X”.

4.2 Legal issues

There is today a large legal apparatus around surveillance cameras (Bergstrom, 2004). This body of rules and regulations connected to the placement and handling of surveillance cameras would be even larger with the introduction of IPS systems.

A certifying authority would be needed, to issue certificates for IPS systems that reach a certain level of robustness and security to hackers and viruses. Great care must be put down to make the systems trustworthy for the public, otherwise they fail their purpose.

Chapter 5

Human-Aware Systems

Most environments that are interesting to survey contain humans. Currently, automatic analysis of humans in sensor data is limited to passage detectors and simple infrared motion detectors. More complex analysis, like interpretation of human behavior from video, is performed by human operators.

With the recent rapid development in computing power, image processing and computer vision algorithms are now applicable in an entirely different way than a few years ago, especially those for looking at humans in images and video. The benefits of automating analysis of human behavior are mainly robustness. If the video surveillance data is scanned by a human, a certain error ratio is to be expected due to the human factor, i.e., fatigue and information overload. By automating parts of the process, the human operator can concentrate on interpretation based on the refined information from the human-aware system.

5.1 Detection of human motion

A basic capability of a human-aware system is to be able to detect and locate humans and other moving objects in the video images. This could either be used in a stand-alone manner in the same way a threspassing sensor is used, or for initializing tracking or recognition systems.

A method for detection of human motion in video, based on the optical flow pattern, has been developed at FOI (Sidenbladh, 2004). The method uses a support vector machine (Cristianini and Shawe-Taylor, 2000) to distinguish between human and non-human motion patterns. Figure 5.1 shows the performance of the method in initial tests. The detector misses small people and people standing still. This is due to the choice of signal; the optical flow method can not detect objects that are smaller than a certain number of pixels. Furthermore, humans standing still do not give rise to any optical flow and are thus missed by the detector.

Presently, alternatives to the optical flow, e.g., temporal and spatial filter responses at different scales, are investigated. In Section 5.3, possible extensions of this detector is discussed.

5.2 Tracking of multiple humans in surveillance video

This capability is especially relevant for the concept of IPS described in Chapter 4. For the purpose of masking out individuals or groups of people from a surveillance video sequence, in

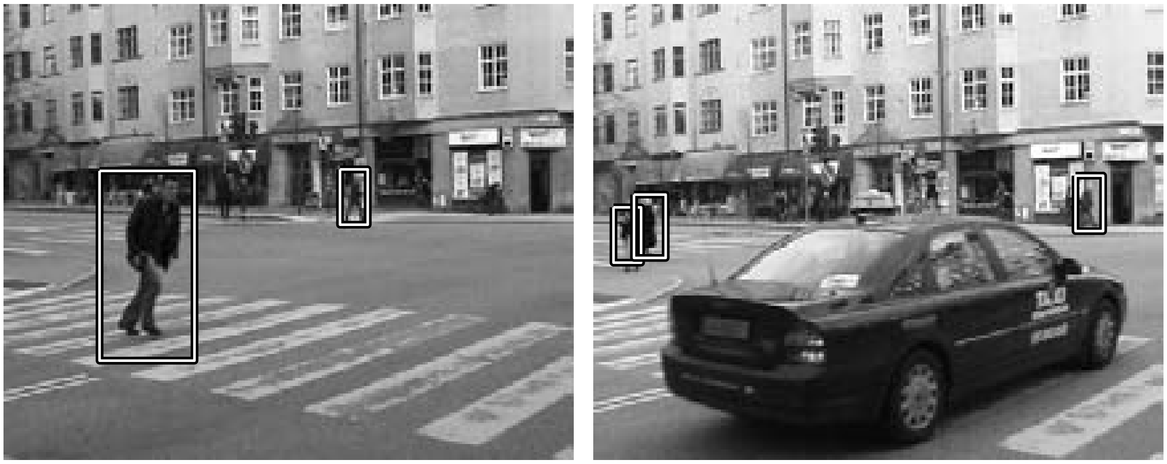


Figure 5.1: Detection of human motion based on optical flow.

order to reveal their activities to a human observer but not their identity, we need any of the following technologies.

The most basic thing to do is foreground-background separation (Figure 5.2). The foreground areas can then possibly be classified according to shape and motion as human or non-human. The foreground areas classified as human is masked out, i.e., the pixel values replaced with a uniform color. In this way, a human observer can see the silhouettes of the humans in the image, and interpret their activities from this information.

A development of this method is to separate the foreground into different individuals (Figure 5.3). There are many methods for this in the literature. A presentation where each individual in the image is masked out with a separate color would greatly enhance the human understanding of the activity in the scene. Further developments could be to automatically recognize what activities are taking place in the scene (e.g., the presence of violence) from the silhouettes (Section 5.3).

The moving humans and cars do not get far in the short time between two video frames (0.04 seconds in European PAL video). Thus, two consecutive video frames look pretty much the same. This information can be used by applying some kind of tracking method, for example a Kalman filter. The basic idea of a filter is to give suggestions on where in the image to look for motion, based on where motion occured in the last video frame.

The problem of tracking multiple moving blobs in the image is more difficult than tracking a single blob. The reason is the risk of confusion when two blobs are crossing each other, and when new blobs enter the scene. There are however mathematical methods to deal with these problems in a robust manner.

The challenge in the IPS application is, as stated above, that a low failure rate is extremely important for the IPS system to be acceptable (see also Section 4.2).

5.3 Recognition of human activity in surveillance video

The next step of a human-aware system is to recognize what activities the detected or tracked humans in the image are involved in. For example, it could be valuable to recognize different events taking place in a scene, like the presence of violence, people falling, people standing

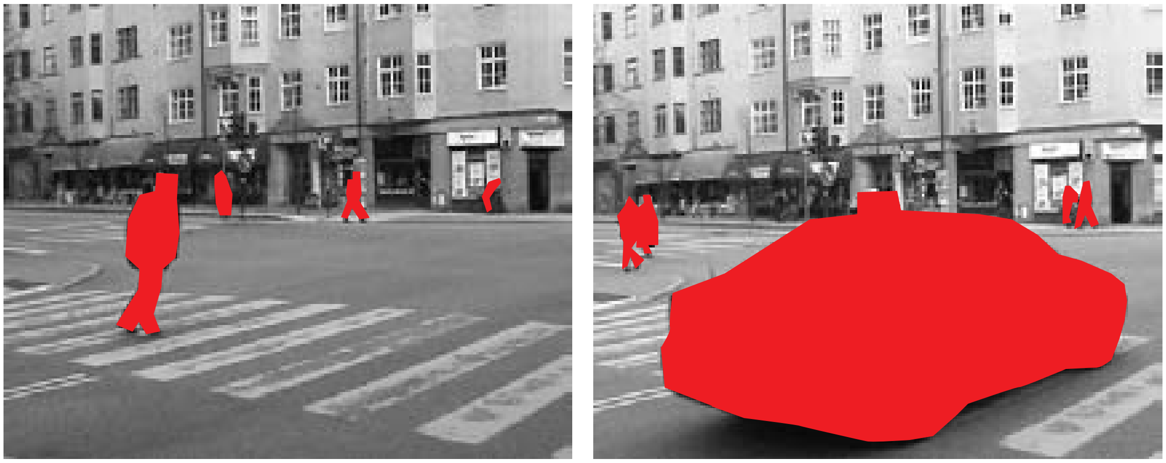


Figure 5.2: Foreground-background separation. *Note: These images are hand drawn and do not originate from an implemented method.*

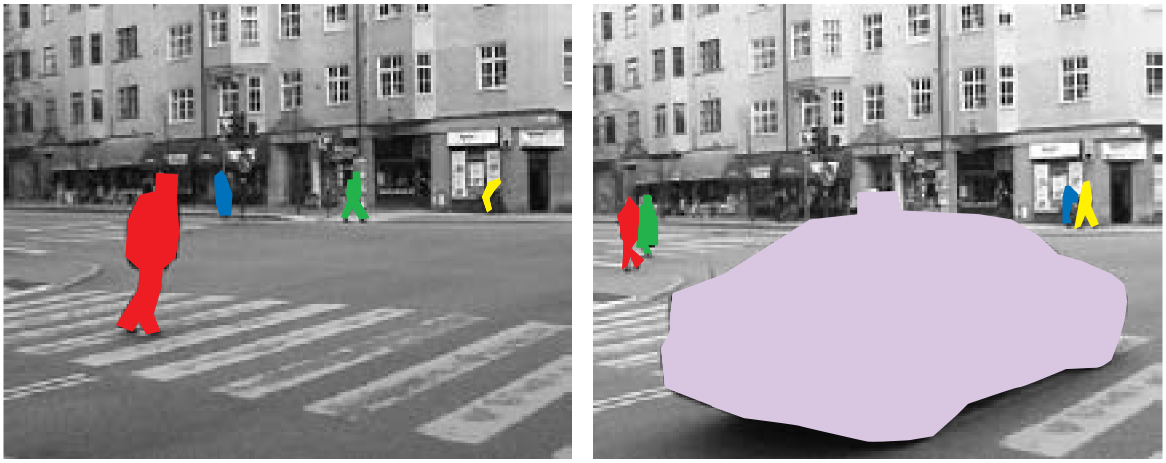


Figure 5.3: Separating foreground into distinct objects. *Note: These images are hand drawn and do not originate from an implemented method.*

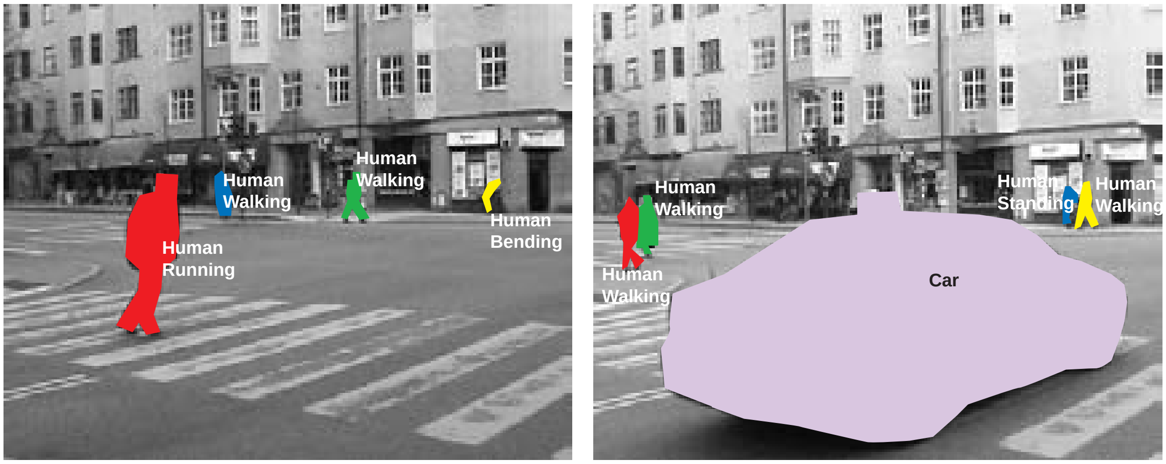


Figure 5.4: Activity recognition from shape. *Note: These images are hand drawn and do not originate from an implemented method.*

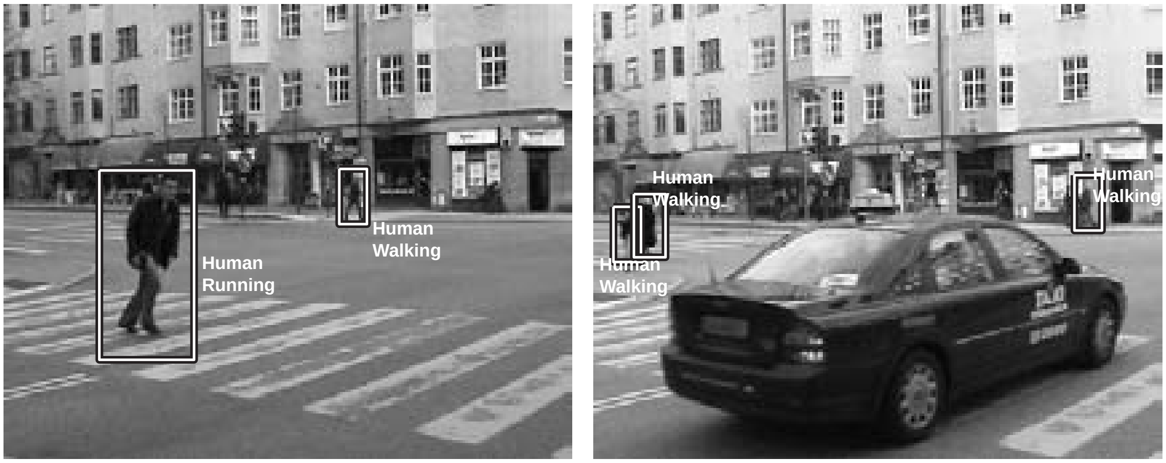


Figure 5.5: Activity recognition from motion. *Note: These images are hand drawn and do not originate from an implemented method.*

around, people behaving nervously, or people running.

One possible application of this is to detect anomalous human activities for surveillance purposes. While automatic surveillance in a calm environment works well when all sightings of people are reported, an environment with many people but only few notable events is better surveyed by a system that only reports certain events. An activity recognition method would be useful for automatically detecting such notable events. Examples of crowded environments with many surveillance cameras in need for an event detection system include airports, underground stations, banks and stores. Notable events in an underground station could be someone falling in the escalator, someone leaving a bag, or the presence of violence. Of course, the evaluation of what events are notable is context dependent; notable events in a bank could be someone running – this would not be a notable event in the underground.

Another application is classification of human motion patterns with respect to age, gender or identity. Humans can identify people known to them over long distances, just by their gait. This implies that the motion pattern is characteristic for each person, and even more so, includes enough information to distinguish young humans from old, or women from men based on the image motion pattern.

Recognizing human activities is of course considerably more difficult than just detecting and tracking humans, and a stable activity recognition system will require many years of research and method development.

5.3.1 Recognition based on motion

An interesting extension of the method described in Section 5.1 is to distinguish between different types of human motion patterns. The output of this method could, for example, look something like the image in Figure 5.5.

5.3.2 Recognition based on shape

The downside of looking at motion only is of course that the system is blind to non-moving activities, like sitting. Another cue, which gives in many respects complementary information, is shape, i.e. the silhouette of the human. The output from a recognition system based on shape could possibly look like the image in Figure 5.4.

5.3.3 Multiple cues

Of course, the shape and the motion cue can be combined, so that the recognition takes both into account when determining type of activity. Such a method is more robust than any of the single-cue methods, provided that the information from the two cues are combined in a correct manner.

The downside of using two cues is that the method becomes computationally heavier – the computational cost is in fact the largest problem in many computer vision algorithms, since images are very large signals in terms of bytes required to store them.

We can here see an analogy to the multiple heterogeneous sensors concept described in Chapter 3 – exactly the same argument applies to the benefits and downsides of using several different sensors in a surveillance system.

Chapter 6

Applications and Impact

This chapter describes various applications of the proposed concepts. Each application is associated with a realistic scenario and provides solutions to real problems for military and police forces, surveillance personnel or airport security personnel.

6.1 Tactical decision support

For soldiers moving through a hostile urban area, the 3D presentation system can be used as a tactical decision support system. In our vision, each squad leader would have the 3D model displayed, for example, inside his APC. The positions of his own soldiers as well as detected threats and risk areas should be visualized. Furthermore, selected information might be viewed by the individual soldiers directly on helmet mounted displays.

Note that this is applicable to both high and low intensity conflicts, as well as in stabilization and reconstruction (S&R) or peace keeping operations (PKO) where civilian riots, terrorists or mobs might be the enemy. The key factor is that the system must be able to work in real-time. This puts high demands on several components of the system.

6.2 Documentation

In MOUT and PKO, the need for documentation is large and growing. There are similar problems as in surveillance for tactical decision support. However, the real-time aspect is not present.

6.3 Warning

In a strictly military scenario, warning systems are essentially trespassing detectors. In a more mixed urban environment, for example in a PKO, the situation is different. It is pointless to warn every time someone enters a city street. However, intelligent systems that can warn for certain or anomalous behaviors would be much more useful.

In the underground, a system that warns when a person has fallen in the escalators or on the tracks would be extremely useful – it is difficult for a human operator to notice these events among the large amounts of data collected from surveillance cameras.

6.4 Data mining

In many surveillance systems, large volumes of data is collected and saved on disc for a certain period of time for future reference. Automatic detection of events, such as the presence of violence, could be used to seach the collected data in a more efficient manner than would be the case with manual search.

6.5 Course of events

A similar problem arises when a situation, for example a riot or an act of crime or terrorism, has already occurred, and available video (or still image) material is to be analyzed. The available material can be a mix of CCTV recordings, police recordings, and recordings from bypassers (with the advent of cellular phones equipped with video cameras, this is very likely). Forensic analysis of such material typically starts with placing all the imagery in a common time frame in order to facilitate the reconstruction of the situation. This step involves time-consuming manual work, and when the work is done, the following analysis is still difficult.

The 3D presentation concept provides a solution by inserting the available video material in a common space and time frame. Reconstruction of complex events becomes much easier. However, high demands are put on the system's ability to correctly position all sensors on the model. Thus, ego-motion and camera calibration techniques are essential (see Section 3.1.1).

Chapter 7

Conclusions

In this report, we present four different concepts for urban surveillance in the future, which are presently discussed and investigated at FOI. The concepts are in different stages of realization, with Human-Aware Surveillance as the one needing the largest research effort to become reality.

We believe that these concepts are beneficial, possibly necessary, for the development of future successful surveillance systems. Therefore, FOI will strive to pursue these research tracks further, in cooperation with companies and authorities with an interest in surveillance and surveillance systems.

The ultimate goal or our research are surveillance systems which

- give *good situation awareness*, e.g. by using a 3D model of the environment,
- use *multiple heterogeneous sensors* to robustly obtain all possible information of the environment,
- preserve the *personal integrity* of surveyed individuals,
- *automatically interpret image data* do determine what human activities are taking place, to relieve human operators of surveillance systems of tedious work.

Such systems are not as far off in the future as one might think!

Bibliography

Ahlberg, J. (2004). Gatustrider i Norrköping: Gemensam slutrapport för projekten I30342 Övervakningssystem för kris i bebyggelse och I30344 Analys av övervakningsdata *(in Swedish)*, *FOI Memo 1180*, Swedish Defence Research Agency, Linköping, Sweden.

Ahlberg, J. and Klasén, L. (2004). Surveillance systems for urban crisis management, *SSBA Annual Symposium on Image Analysis*.

Bergstrom, P. (2004). Rekommendationer vid användande av kameraövervakningssystem *(in Swedish)*, *SKL Rapport 2004:12*, National Laboratory of Forensic Science, Linköping, Sweden.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK.

Nastell, P. (2002). Teknisk och personell bevakning – realiserbarhetsstudie *(in Swedish)*, Svenska Kraftnät, Stockholm, Sweden.

Senior, A., Pankanti, S., Hampapur, A., Brown, L. and Ekin, A. (2003). Blinkering surveillance: Enabling video privacy through computer vision, *Technical Report RC22886*, IBM Research Division, Yorktown Heights, NY, USA.

Sidenbladh, H. (2004). Detecting human motion with support vector machines, *IAPR International Conference on Image Processing*, Vol. 2, pp. 188–191.