International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:10, 2008

Shadow Detection for Increased Accuracy of Privacy Enhancing Methods in Video Surveillance Edge Devices

F. Matusek, G. Pujolle, and R. Reda

***Abstract*—**Shadow detection is still considered as one of the potential challenges for intelligent automated video surveillance systems. A pre requisite for reliable and accurate detection and tracking is the correct shadow detection and classification. In such a landscape of conditions, privacy issues add more and more complexity and require reliable shadow detection.

In this work the intertwining between security, accuracy, reliability and privacy is analyzed and, accordingly, a novel architecture for Privacy Enhancing Video Surveillance (PEVS) is introduced. Shadow detection and masking are dealt with through the combination of two different approaches simultaneously. This results in a unique privacy enhancement, without affecting security. Subsequently, the methodology was employed successfully in a large-scale wireless video surveillance system; privacy relevant information was stored and encrypted on the unit, without transferring it over an un-trusted network.

***Keywords*—**Video Surveillance, Intelligent Video Surveillance, Physical Security, WSSU, Privacy, Shadow Detection.

I. Introduction

COMPUTER vision in video surveillance systems usually involves tracking objects through scenes to detect unwanted behavior. Such intelligent video surveillance systems started to appear recently in several commercial applications. Further concepts were developed to provide a video surveillance architecture,which supports large-scale video surveillance systems [1]. One of the big challenges is to develop a reliable method, which is able to distinguish clearly between a detected object and its own shadow. A further challenge is that this algorithm must be able to deal with the problem of interference between different objects and their shadows and between different shadows. An even more complex problem with shadows emerges if the shadow of an object is cast on another object, like two cars driving parallel to each other. In this case those two objects are "connected" through the shadow and are detected as one single object [2]. The solution to these challenges is presented in this work: an intelligent reliable shadow detection algorithm.

F. Matusek is with KiwiSecurity, Austria (www.kiwi-security.com).
G. Pujolle is with Université Pierre et Marie Curie, Labortoire d'informatique de Paris 6 (www.lip6.fr).
R. Reda is with Innovation Communication Technologies, Austria / Germany (www.ictmc.com).

Accordingly, the video surveillance system is able to recognize a shadow as a shadow and not as part of the object. Shadows can be classified into different categories, depending on how they were generated [3]. First, a distinction can be made depending on movement. Shadows, which are cast by static objects, such as buildings or parking cars, are called static shadows, while shadows, which are generated by moving objects, are called dynamic or moving shadows. Second, shadows can be distinguished by the surface they appear on. The part of the object, which is not illuminated by the light sources, is called self-shadow. Cast shadows are shadows, which are cast on a surface by an object which blocks light from the light source. In this work moving cast shadows are relevant since they present a problem to algorithms relying on motion information. Contrary to usual approaches to shadow detection, we view moving cast shadow detection in the context of video surveillance not only as a problem of object tracking, as described above and as it is usually considered [4], but as a problem of masking privacy data, such as faces, in video images. Even though shadow detection plays a crucial role in providing satisfying results, little has been done in this area. This work describes the challenge, privacy enhancing methods and solutions in an embedded video surveillance system, the WSSU [5].

Section II presents the general analysis and motivation behind masking and privacy enhanced video surveillance systems. Section III discusses the specifics of the shadow detection method used in the WSSU. Finally, Section IV presents the results obtained with this method.

II. Masking of Privacy Data in Video Surveillance Systems

As more and more video surveillance systems are installed, concerns about privacy of citizens and personal information are raised, followed by a set of new privacy protecting laws in different countries, especially in the EU.

In recent years, computer vision methods have proven to be effective to enhance privacy of persons in video surveillance systems. Effectively, information in an image that makes a person identifiable, i.e. a face, is cut out and saved separately in an encrypted way by the system before security personnel view it. Only if a criminal act takes place, a defined person can access this critical data. This way, the privacy of innocent

Fig. 1 Privacy enhanced video surveillance systems can mask the faces of persons

While preserving privacy, as much relevant information as possible should be sent to the Video Surveillance Head Quarters (VS HQ)[5]. However, concerning shadow detection, the system must be able to choose the critical balance between accuracy and privacy.

Fig. 2 shows a novel process / architecture for video surveillance systems including masking of privacy information, the Privacy Enhancing Video Surveillance architecture (PEVS). The input is formed by an image and meta-data. This meta-data is motion information with removed shadows. In the PEVS a behavior detection module uses pre-defined rules from the behavior library to detect unwanted behavior. It also uses the Feature Matching Engine (FME) to classify objects using the object library. The system constantly updates itself in order to adapt to new situations. The output is a masked image and the corresponding meta-data of the masked region.

Fig. 3 shows a typical masked person and cart with no shadow detection, demonstrating the shadows effect. Shadows are classified as part of the person. All masked pixels including the cart reach approximately 14.000 pixels. Roughly half of the pixels are shadow pixels, which do not need to be masked in order to preserve privacy. With a shadow detection algorithm these pixels would not be masked and more scene information could be kept.

In privacy enhancing video surveillance systems the raw video image has to be transferred over insecure communication channels, thus exposing private information to anyone who can intercept the transmission.

The current system employs smart cameras, which are able to cut out the privacy information already on the camera site, thus not sending any private information over the network. In addition, privacy information is saved locally on the processing unit of the smart camera in an encrypted way [5]. Furthermore the Wireless Self-contained Surveillance Unit (WSSU) was used. The WSSU (which is an advanced smart camera) is able to operate independently of external power supply or wired connections. Fig. 5 shows how the WSSU is integrated in video surveillance architecture. Depending on the storage technique chosen, the privacy information is either stored centrally or in each WSSU individually.

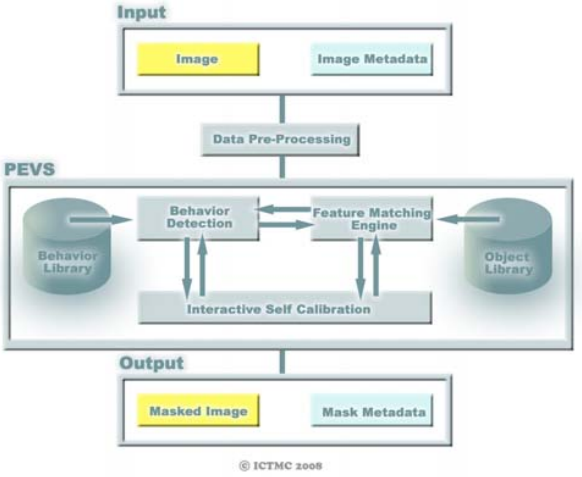


Fig. 2 The PEVS architecture. The input consists of the image of interest and meta-data (motion segmented regions) and is analyzed to output a masked image with accompanying meta-data

Fig. 3 A masked person and cart. Roughly half of the masked pixels (excluding the cart) are shadow pixels

These storage strategies are further elaborated in [6]. In this example it is assumed that it is stored in the WSSUs themselves. The WSSUs send alarms and masked frames to the video surveillance headquarters (VS HQ), which in turn sends them to storage, user interface clients and mobile clients such as VSLCs (Video Surveillance Local Control). VSLCs are a new concept, introduced in [7], for a mobile video surveillance client to be used in high security applications. They are physically as well as on a software level highly secure, being virtually indestructible. They can be used to access video surveillance information as well as databases such as national security and border control databases.

If a VSLC or the user interface requests private information with the correct authentication from the VS HQ, it forwards this information to the WSSUs, which send the private (encrypted) information. The VS HQ sends images and meta-data, such as alarms and object information, to the data fusion server, which tries to find coherences between alarms and decides which alarms are worth keeping. In the storage server only masked frames are stored.

III. Shadow Detection in the WSSU

Since in the WSSU embedded resources are limited shadow detection algorithms need to be implemented efficiently. Currently, the WSSU is processing video streams using an Analog Devices Blackfin BF561 dual-core processor, featuring 600MHz and 128KB L2 cache per core and 64MB

SDRAM. Currently, the processing of the background model, i.e. separating foreground and background to produce a binary image to be used as input to shadow detection, takes 40% of the available processing power. The other 60% consists of tracking, management processes and 10% of unused processing power. So when using shadow detection, it has to use 10% or less of the processor. This limits the possibilities for shadow detection significantly. Rather than using more complex shadow detection methods, several established methods were combined in order to create a highly accurate and efficient shadow detection method. More specifically, a mixture between a method used by Horprasert [4] and Nadimi [8] was used. This method is further discussed in section III B).

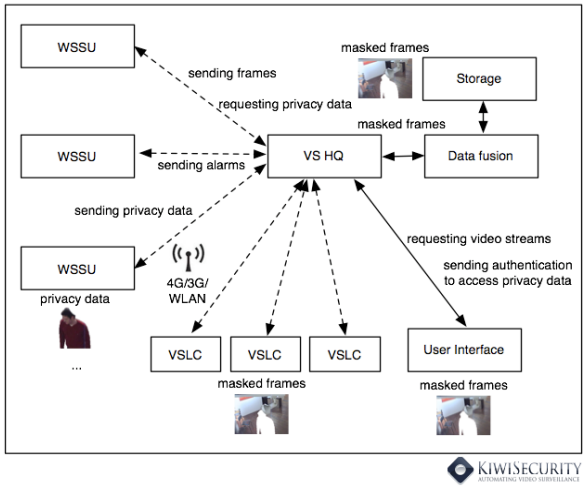


Fig. 4 WSSUs in a video surveillance system

A. *Video Analysis Processing Queue*

Fig. 5 shows the video analysis process in the WSSU. Once the CMOS sensor captures the image, it is sent to the background subtraction module, which separates foreground from background. After that, shadow detection is performed on the separated image. Following that detected image regions are clustered using a fast Mean Shift implementation [9] and faces are detected using the Viola & Jones face detector [10]. Next, tracking on the still unmasked image regions is performed. Depending on the user's choice, either only the faces or the whole body of persons is now masked. This is currently done by either setting the value of pixels to white or by inverting the color value. The original, unmasked, image regions are encrypted and stored locally in the internal storage of the WSSU. Finally, a reasoning step after the tracking is "behavior analysis" which applies pre-defined rules to the collected data and sends produced alarms, together with the masked frames to the VS HQ. In the following the shadow detection method is discussed and results are presented.

B. *Shadow Detection and Removal*

For the implemented shadow detection method a novel combination of shadow detection methods introduced by Horprasert et al. [4] and Nadimi et al. [8] is used. The complete algorithm is elaborated and every stage of the shadow detection process is explained. Fig. 6 shows the shadow detection pipeline. Each step in this pipeline was applied carefully in the system, resulting in a reliable, privacy oriented system architecture.

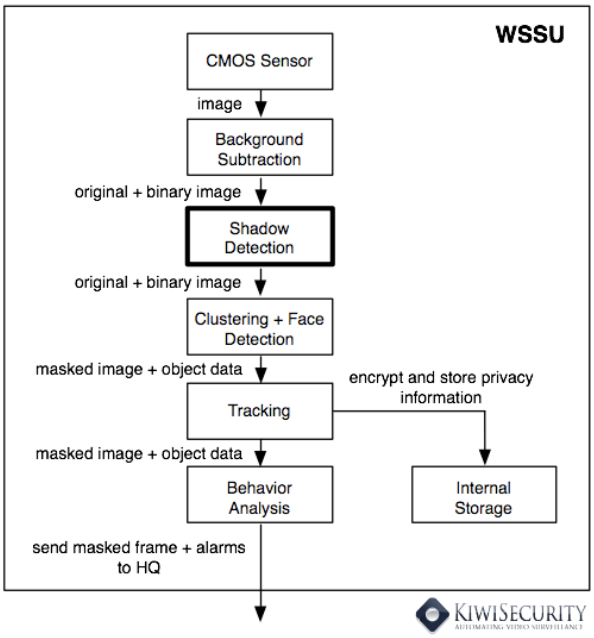


Fig. 5 Video analysis process in the WSSU

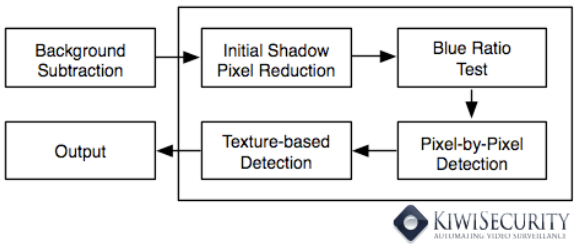


Fig. 6 Shadow detection pipeline

1) Background Subtraction using Mixture of Gaussians (MoG)

In this first stage motion in the image is detected. The output of this stage is a binary mask with moving pixels marked, as is shown in Fig. 3. The result of this stage can be used to reduce shadow detection to just the areas that are detected as motion and that therefore could cause problems later on.

The motion detection implementation is based on the method "Mixture of Gaussians", proposed by Stauffer et al. [11]. This method was chosen because of its ability to model background motion, thus being robust against repeating background motion such as waving trees or flags. This way, false positives during motion detection can be reduced and a good model of the foreground to perform shadow detection on is produced. A drawback of this method is its speed. Since it

models each pixel with at least three distributions this algorithm is slower than simple background subtraction. However, accuracy is an important factor and since background subtraction provides the basis for all following processing steps, the data has to be as good as possible. This is why processing power was sacrificed in exchange for more accurate results. This method for motion detection uses not only one single Gaussian to represent the background, but a mixture of Gaussians. In this approach multiple Gaussians can represent the background, based on their persistence, variance and a threshold T. T is a measure of how much of the data should be accounted as background.

2) Initial Shadow Pixel Reduction

In this stage pixels, which are candidates for shadow pixels are removed. It is assumed that pixels on a detected surface cannot be shadows if they have a higher intensity than the actual background. So if a pixel has a higher intensity than the background it is either a highlight, which is also detected as moving foreground, or an actual foreground object but it is not a shadow. This pixel is left out from the object mask and does not have to be checked again.

3) Blue Ratio Test

Similar to the previous stage, the blue ratio test reduces the number of potential shadow pixels by using a condition on the pixels. It exploits the observation that shadow pixels, which fall on neutral or grey surfaces, such as asphalt roads, tend to be more bluish [8].

4) Pixel-by-pixel Shadow Detection (PbP)

In order to improve the results of the detection, two different detection methods are used simultaneously: Pixel-by-pixel and Texture-based detection. PbP shadow detection applies the shadow detection rules to each pixel separately while texture-based (TB) shadow detection checks bigger texture patches (usually between 7x7 to 9x9 windows). This is an attempt to reduce the number of foreground pixels, which are classified as shadows because the foreground object has a similar color like the background (e.g. a grey car on a grey road). However, the TB method is only feasible if the background is textured. Asphalt roads, for example, do not contain enough texture information to support reliable TB shadow detection.

The results of these two methods weight each pixel according to the result. If a pixel is classified by either of the methods as shadow pixel, it gets a confident rating of 0.5. If both classify it as shadow, it accumulates to 1. Only pixels that have a rating of 1 get removed in the end of the process.

In the PbP shadow detection option, a pixel is classified as shadow if the pixel has similar chromaticity but lower brightness than the background. Since at this stage we already performed background subtraction our goal is to delete shadow pixels that are incorrectly detected as foreground motion. So it is assumed that all pixels, which are no shadow pixels but were detected as foreground pixels are actual pixels of foreground objects. So the first classification, foreground or background, is already done in the background subtraction stage. That leaves the shadow detection. A pixel is classified as a shadow pixel if it has a similar chromaticity but a significantly lower intensity that the background. Because of color variation in the images of the sequence, chromaticity values can vary in a small range. The same goes with the intensity values. Therefore, thresholds have to be set. If the change of the intensity is over a certain threshold and the change of the chromaticity under a certain threshold, a pixel is classified as a shadow pixel:

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

where:

px	Current pixel
c	Current chromaticity value
i	Current intensity value
cpre	Mean of past chromaticity of non-shadow pixels
ipre	Mean of past intensity values of non-shadow pixels
Tc	Chromaticity threshold
Ti	Intensity threshold

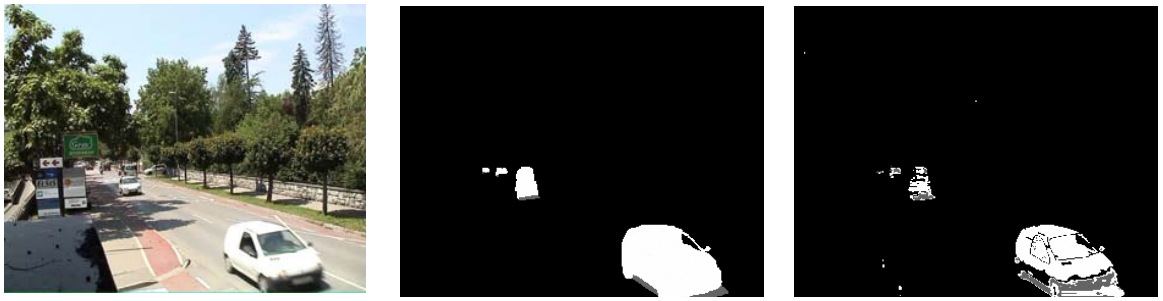
If the first condition is met, a shadow pixel is found and it is marked as such. Otherwise the pixel keeps its classification.

5) Texture-based Shadow Detection (TB)

The difference between a TB shadow detection method and PbP detection is that in TB methods a number of pixels are treated as a patch. This way, noise resulting from camera sensors, bad image compression or wrong motion detection can be avoided. Also, if shadow detection is performed, it is likely that shadows on foreground objects that have a similar color than the background, are also detected as cast shadows and therefore deleted. If a texture patch is used, then the pattern of the background and the foreground object has to match in order get false shadow detection.

One main disadvantage of this method is that is does not work properly on background surfaces that are not highly textured. Examples are asphalt streets or pavements. This method works well on surfaces such as grass or brick walls, which have many distinctive features.

The classification rule is similar as in the pixel-based approach, with the difference that new thresholds are used.



(a) (b) (c)

Fig. 4 Shadow detection results from scene 1. (a) shows the original input frame, (b) ground-truth and (c) shadow detection results. White pixels represent foreground, grey pixels shadows and black pixels background

(a) (b) (c)

Fig. 5 Shadow detection results from scene 2. (a) shows the original input frame, (b) ground-truth and (c) shadow detection results. White pixels represent foreground, grey pixels shadows and black pixels background

$$p_x = \begin{cases} 1 & if\ (c_{pre} - c) < T_{tx\_c}\ \ and\ \ (i_{pre} - i) > T_{tx\_i} \\ 0 & otherwise \end{cases} \quad (2)$$

where:

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

This rule is checked against every pixel value in the current texture patch, which is usually a 7x7 window that is moved over the image. If all pixels in the patch meet the condition above the texture-patch is declared to be a shadow patch. Otherwise it is left as it is. This results in detection results that have the block form of the patch window.

6) Output

The output of the algorithm is a shadow pixel mask that can, in further steps, be easily be removed. This leaves just the foreground objects of the scene. This is essential in the video analysis process in the WSSU, shown in Fig. 5. After this stage, clustering and face detection can be performed.

IV. Experimental Results

In Fig. 4 an example frame from the results of pixel-wise shadow detection in a street (scene 1) is shown. From left to right the original frame (a), ground truth (b) and the actual shadow detection result (c) are shown. White pixel represent detected foreground, grey pixel detected shadows and black pixel background. Due to high image resolution of the input (720x576 pixels), apart from four spots in the picture, no noise is produced. Shadows under the car in front are detected, however also parts of the front of the car is detected. This is caused by the dark area at this part of the car that has the same color value as detected shadows. Shadows under the car further back are also detected. Shadow detection rate η and shadow discrimination rate ξ are shown in Table II.

In Fig. 5 an example frame of the results of the shadow detection in scene 2 is presented. The image resolution of the input frames in this scene was lower (320x240 pixels). Compression artifacts were the reason that noise was produced. As in scene 1, the original frame (a), ground truth (b) and the shadow detection result (c) are shown. White pixels represent foreground, grey pixel shadows and black pixel background. While most parts of the cast shadows are detected, so are also parts of the objects. Specifically, dark cars pose a problem due to their color value, which is similar to color values of shadows. In Table I shadow detection rate η and shadow discrimination rate ξ from scene 2 are shown. Also, results from comparable shadow detection algorithms (SNP, statistical nonparametric, and SP, statistical parametric), which yielded the highest shadow detection and shadow

discrimination rate on this scene in the evaluation by Prati et al., are given. The shadow detection rate without post-processing is slightly lower than with the SNP algorithm (81.07% compared to 81.59%), however the shadow discrimination rate is higher (65.66% compared to 63.76%). As in scene 1 the shadow detection rate is higher with post-processing while the shadow discrimination rate decreases.

TABLE I
SHADOW DETECTION RATE H AND SHADOW DISCRIMINATION RATE Ξ FOR THE TESTED SEQUENCES. RESULTS WITHOUT POST-PROCESSING (W/O PP), WITH POST-PROCESSING (W/ PP) AND FROM COMPARABLE ALGORITHMS (SNP, SP) ARE GIVEN

	Street/Scene 1		Highway/Scene 2	
	η%	ξ%	η%	ξ%
w/o pp	65.73%	83.72%	81.07%	65.66%
w/ pp	66.43%	78.40%	85.06%	60.84%
SNP	n/a	n/a	81.59%	63.76%
SP	n/a	n/a	59.59%	84.70%

ACKNOWLEDGMENT

This work was supported by the Austrian Federal Ministry for Transport, Innovation and Technology www.bmvit.gv.at, the Austrian Research Promotion Agency "ÖsterreichischeForschungsförderungs-gesellschaft" www.ffg.at and the Academic Business Incubator INiTS, www.inits.at. This work will be partially included in a PhD thesis at the Pierre & Marie Curie University (UPMC).

REFERENCES

[1] F. Matusek, S. Sutor, K. Kraus, F. Kruse and R. Reda, NIVSS: A Nearly Indestructible Video Surveillance System, The Third International Conference on Internet Monitoring and Protection, Bucharest, July 2008.
[2] A. Prati, I. Mikic, M.M. Trivedi and R. Cucchiara. Detecting Moving Shadows: Algorithms and Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25: 918 – 923, July 2003.
[3] A. Prati, I. Mikic, R. Cucchiara and M.M. Trivedi. Comparative Evaluation of Moving Shadow Detection Algorithms. *Proceedings of 3rd Workshop on Empirical Evaluation in Computer Vision (in conjunction with CVPR 2001)*, December 2001.
[4] T. Horprasert, D. Harwood and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. *Proceedings of IEEE ICCV'99 FRAME-RATE Workshop*, 1999.
[5] S. Sutor, F. Matusek, and R. Reda, "WSSU: Wireless Self-Contained Surveillance Unit; An Ad-Hoc Video Surveillance System", in Press, presented at The Fourth Advanced International Conference on Telecommunications, Athens, June, 2008.
[6] F. Matusek and R. Reda, "Efficient and Secure Storage of Privacy Enhanced Video Surveillance Data in Intelligent Video Surveillance Systems", to be presented at the International Symposium on Computer and Information Sciences, October 2008.
[7] F. Matusek and R Reda, VSLC: Video Surveillance Network Control, Mobile Video Surveillance Local Control Engineering and Applications to be presented at IEEE, IFIP Wireless Days Conference 2008.
[8] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1079 – 1087, August 2004.
[9] Y. Cheng, "Mean Shift, Mode Seeking and Clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, pp. 790-799, 1995.
[10] P. Viola, and Michael Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features," cvpr, p. 511, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01) - Volume 1, 2001.
[11] C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:747 – 757, August 2000.

Florian Matusek graduated and received his BSc and MScat the Vienna University of Technology in computer science. During his study he co-founded the company KiwiSecurity, http://www.kiwi-security.com, which is developing an automatedvideo surveillance system using video analytic methods, KiwiVision. KiwiSecurity has won a number of awards and was elected among the most innovative and promising start-ups of Austria. KiwiSecurity is providing products and services for large infrastructure operators and public institutions with high security requirements. Florian is currently acting as Managing Director for Operations. Florian also co-founded the European Security and Trust Experts Alliance "ESTEAlliance", www.estealliance.com, which provides security expertise by some of the leading security experts of Europe. Florian is acting as TPC in a number of security related conferences. Further, Florian is currently writing his PhD thesis at the Pierre & Marie Curie University (UPMC) in the area of wireless security and its application in an automatic video surveillance system.