Abstract

We present a system prototype for self-determination and privacy enhancement in video surveilled areas by integrating computer vision and cryptographic techniques into networked building automation systems. This paper describes research work that has been done within the first half of the collaborative blue-c-II project and is conducted by an interdisciplinary team of researchers. Persons in a video stream control their visibility on a per-viewer base and can choose to allow either the real view or an obscured image to be seen. The parts of the video stream that show a person are protected by an AES cipher and can be sent over untrusted networks. Experimental results are presented by the example of a meeting room scenario. The paper concludes with remarks on the usability and encountered problems.

Surveillance, Cryptography, Computer Vision, Building Automation

1 Introduction

Today video surveillance systems are being deployed worldwide, covering public places, corporate buildings and private homes with CCTV and networked cameras. In our work we focus on technical solutions to preserve the privacy of individuals within buildings despite the omnipresence of cameras. Our scenarios for which we built a system prototype are placed in typical office buildings.

To enhance privacy in such an environment our approach is to use techniques provided by computer vision and cryptography and supported by facilities provided by a building automation system, to give surveilled persons the power to control their video information. This approach is embedded in a larger research project where the capabilities of conventional and 3D video cameras are explored [10]. Within this context our sub-project is deploying results from research in the field of computer-vision based tracking to a prototype system. In our collaborative project the prototype is examined from Architects and Computer Engineers’ points of view.

Furthermore, the actual use of our rather hidden and invisible technology by a monitored person is also of great importance to us. Additional demands are ease of use and a tight integration with existing controls of a networked building automation system. Fig. 1(a) presents a use case diagram of the involved actors and the system. An observed person can allow or deny access to the parts of a video stream that contain images of her or him. A viewer can see the original video stream where no person is visible. For every person present in the video stream the viewer sees either the original clear image, if access has been allowed (Fig. 1(b)) or an obscured image, if access has been denied (Fig. 1(c)).

1.1 Related Work

In this paper different techniques from computer vision, cryptography, distributed computing and building automation are presented. All four disciplines are by themselves

Privacy in Video Surveilled Areas

Torsten Spindler
ETH Zurich, Prof. Ludger Hovestadt
Institute of Building Technology, CAAD
Wolfgang Pauli Str. 15
8093 Zurich, Switzerland
spindler@arch.ethz.ch

Daniel Roth
ETH Zurich, Prof. Luc van Gool
Computer Vision Laboratory
Sternwartstrasse 7
8092 Zurich, Switzerland
droth@vision.ee.ethz.ch

Christoph Wartmann
ETH Zurich, Prof. Ludger Hovestadt
Institute of Building Technology, CAAD
Wolfgang Pauli Str. 15
8093 Zurich, Switzerland
wartmann@arch.ethz.ch

Prof. Andreas Steffen
University of Applied Sciences Rapperswil
Institute for Internet Technologies and Applications
Oberseestr. 10
8640 Rapperswil, Switzerland
andreas.steffen@hsr.ch
research fields with a rich tradition. As far as we know, the combination of all four is novel. However, combination between just two of them were presented in previous work. Senior et al. [20] reviewed privacy in video surveillance techniques for future systems. Furthermore they presented a privacy console and PrivacyCam, both implementing a subset of the proposed methods allowing some vision analysis, transformation and encryption of the video. However, the access control was fixed to a privacy console and only based on users privileges given by the system owner, for example a security company. Our system in contrast allows individual persons to grant or deny access to other persons or groups with help of a variety of devices. Additionally, our system focuses on being used within a building and being integrated with a building automation system to allow the ease of use for surveilled individuals. Dufaux et al. [3] presented a system using computer vision detectors to scramble and encrypt image regions into JPEG 2000 image format. The amount of image scrambling can be controlled and the system has been tested with different detectors such as skin color and face detectors. However, the system does not use tracking methods and thus is not able to apply an individual privacy status when multiple people are present in a scene. A startup company Emitall [7] now uses this technology to encrypt the videos of surveillance cameras, unless the need arises for law enforcement agencies to view the clear image, which then can decrypt the stream. Another startup company Eptascape [8] is trying to bring their privacy enhancing technologies to the market. The first product of this start-up company is an MPEG-7 encoder and real-time tracking hardware box, which can be directly attached to CCTV cameras. The proprietary tracking algorithm seems to be base segmentation and therefore similar to our tracker. However, the example videos do not show changing lighting conditions and the tracker seems not be able to handle occlusion, a prerequisite for tracking multiple people individually.

The outline of this paper is as follows. Section 2 gives an overview of our interdisciplinary approach. The system architecture will be explained in Section 3, while the deployed computer vision methods are examined more thoroughly in Section 4. Section 5 presents the user interface. Results are discussed in Section 6 and Section 7 concludes the paper with remarks on problems encountered, including an outlook.

2 Overview

The Privacy in Video Surveilled Areas (PiViSA) project combines computer vision and cryptographic technologies. It is meant to be integrated with a networked building automation system. The projects environment is the building, more specifically an office building, and its scope is a room. Our goal is to enhance privacy for a person that is surveilled by networked video cameras.

With privacy we mean the ability of a person or a group to be unobserved by others. In a conventional building the privacy of a room is determined by the ability to see or listen to what is happening in the room. Based on our informal definition of privacy, walls, windows and doors influence the privacy of a room in a building. The privacy of a room can be controlled by opening and closing doors or curtains. When a room is observed, the observed persons can either be aware of the observation or unaware. This concept can be found in different architectural settings. The most famous example is most likely Jeremy Bentham’s Panopticon [16], though it never has been built. Invented as prison it enables guards to oversee inmates without being seen. A number of buildings provide more limited observation systems. For example, a city house by Gaudi allowed the house owner to spy on the women room. For similar purposes the famous Opera Garnier in Paris contains portholes towards the dancers preparation area. We believe that in a modern building, with use of automation infrastructure and video surveillance equipment, a complete loss of privacy can happen. In this context our system is meant to be used as an electronic successor to a curtain. While the easiest way to achieve this result would be to stop all surveillance of a room when privacy is wanted, our approach allows for fine grained control. Unlike a curtain, that blocks visibility and reduces audibility, our system enables an observed person to control who can see them in taken videos.

To achieve this goal, the video stream taken by a camera
is available in different versions. Similar to the approach presented by Senior et al. [20] a public version is available where privacy relevant information is removed or obscured. If an observed person grants access to an observer, an additional clear video is sent protected by a cipher to the observer’s display, where it is merged with the publicly available obscured version. The roles of video controlled spaces, building automation and cryptography for our project are explained before details of the system are presented.

2.1 Video Controlled Spaces

Over the last decade a growing number of public spaces and commercial buildings have installed surveillance cameras. Access to recordings of these cameras is usually strictly limited to security staff members and alike. However, this does not necessarily prevent misuse. The question of who monitors the surveillant is still as important as ever. Recently three guards were sentenced for abusing their video surveillance equipment for spying on a woman [18]. Despite the possibility of abuse, more and more video surveillance system are being deployed as the advantages seem to outweigh disadvantages. For example, residents of the London neighborhood Shoreditch have access to local surveillance cameras on their home television with the help of a set top box [25]. While on a different scale, it is also noteworthy that in the UK a system to monitor license plates of cars nationwide is being installed [12]. The possibility to profile not only general traffic patterns, but also the use of individual cars is quite intriguing. Especially the
countrywide scope of this project is of interest. We believe that it is more than likely that such serious effort to automatically monitor the traffic in a whole country will lead to similar projects within buildings. In Variations on a Theme Park [21] the role of controlled spaces, foremost in form of malls, has been examined. The ability to track and identify customers in such a closed environment might offer new possibilities. If an economic benefit is to be expected or a customers desire can be satisfied it will be implemented eventually.

2.2 Building Automation

We look at the issue of privacy in a video surveilled area against the background of an intelligent building, a smart home or, probably better termed, a computer integrated building as first introduced by Frank Duffy [4]. Such a building utilizes computers to improve the level of comfort by providing personalized services and can also provide information for facility management systems. Most of current building automation systems use field bus systems that can be seen as a closed and proprietary network technology. However, the transition to more versatile IP-network based systems is on the way. It is noteworthy that the older field bus systems are already retrofitted with IP-network gateways to offer their services via http and other protocols. New systems are often using TCP/IP networks as a core technology.

In addition to the adoption of network technologies the tasks are expanded beyond traditional automation. Next to the control of heating, ventilation and air conditioning the integration of services like IP based telephony, television and video on demand is possible. The common communication and control channel for such a system is the already available network infrastructure. At the chair for Computer Aided Architectural Design (CAAD) at the Swiss Federal Institute of Technology research was undertaken how to add multimedia services for an IP based building automation system. The chosen system was RaumComputer [17], which uses the Java based OSGi service platform [15]. The multimedia services were added and a generic interface for hand held devices was developed [24] [23]. With a single interface it is possible to control all services a room provides, from light control to video services. The technical devices in the room, ranging from speaker, microphone, light to TV were abstracted as services and are no longer used as single devices. Rather the room itself becomes a container of available services and provides a unified user interface. Within this context surveillance cameras and video conferencing systems were used to extend visibility between rooms. A personalized user interface for controlling the rooms service were provided after an RFID based authentication. We expect that such a building automation system
will know if a room is currently occupied and by whom.

As a next step the building automation system could provide valuable information for a computer vision based tracking system. The accuracy of the tracking is influenced by changes in the observed environment. Some of these changes are known by a computer integrated building and could be relayed as events to the tracking systems. Examples for such events are opening and closing of curtains and doors, changing illumination, giving presentations. The tracking system could react to these changes and either ignore certain parts of the scene or adopt to the new environment. We think that this communication could increase the robustness of a visual tracker in the mentioned situations.

2.3 Cryptography

Cryptography is used to protect information against spying and tampering. In our system we define two types of data that need protection. The clear regions of a video stream, where a person or object is visible, are one type. The other type is the management and administrative data. For these two types we have use for two different cryptographic components. Primary concerns for the protection of management and administrative data are secrecy and tamper resistance. Primary concern for the protection of the video streams is speed, because typical surveillance cameras produce between a few hundred kilobit to several megabit of video data per second.

For the first task asymmetric cryptographic technologies can be used. For example the SSL/TLS protocol suite allows to establish a secure channel between nodes on a network. With use of passwords, certificates or challenge-response mechanisms nodes can authenticate and information can then be sent securely between nodes. Also, information that is necessary to authenticate and authorize may need to be protected. If possible, facilities provided by a building automation system should be used to store and access these.

For the second task a symmetric cipher in stream mode can be used [9]. Secret keys for the cipher can be transported securely by the above described secure channel between nodes. Important characteristics for the cipher are speed, security and hardware availability.

3 System Architecture

With the focus on a building and rooms therein, we have chosen a distributed approach. Nodes connected by a network are the building blocks of our system. Based on our experience with adding media streaming to building automation systems, we adhere to the idea that every node provides a specialized service, that is loosely coupled with the rest of the system. While this enhances the chance to use parts of the systems in different contexts, there is a price to pay in terms of performance.

Fig. 2 gives an overview of the different nodes involved in sending a video stream. The outer ring shows the different nodes, with either clear or obscured images which are sent over the network. Nodes can be specialized hardware, such as cameras, or reside on a host PC alone or together.

Fig. 3 presents a typical data flow from camera to display. Starting from the camera node the clear video stream is delivered to the nodes that track, obscure and encrypt. The track node analyzes the video and is described in Section 4. The result of the tracking process is sent to nodes responsible to encrypt and obscure the stream. The result sent by the track node contains information on regions which define the location of objects and persons found and internal identifiers for these. While the track node is following an object it will use the same identifier for a region in every tracking result. However when a person leaves the observed area and enters again, the person will get a different identifier. The task to relate those identifiers to identities of persons is given to an identify node. This node uses information about identities from the configuration node and the relative or absolute positions of tracked regions as well as the images of the video streams at these position. Once the person is identified and their identity is related to the tracker’s identifier, the encrypting node protects the clear image by a cipher. The key to be used in this cipher is issued by the key manager node. Here session keys are created when needed. The session key is distributed by the key manager node to participating encrypt nodes and decrypt nodes. An encrypt or decrypt node participates when it is associated with a merge node that is assigned to a display node, which is finally controlled by an authorized observer.
The merge node takes the clear image information sent by the decrypt node and replaces the part of the video stream that has been obscured by the obscure node. Lastly the display node shows the final video stream. Depending on the number of persons and objects tracked and permissions for these, the displayed video will contain a combination of clear and obscured parts.

### 3.1 Network Protocol

In our project the network communication uses the UDP protocol to transfer large amounts of data. The video streams themselves contribute to most of the network traffic. A video stream sent by the obscuring node is named *obscured stream* and one sent by a encrypting node is named *protected stream*. Where possible we use UDP multicast to reduce the load on the network. In Fig. 3 the camera can use multicast to reach the obscure, track and encrypt nodes. All our protocols, except the ones for video streaming, are based on XML messages. While XML is not a lightweight approach it is considered to be best suitable for integration into existing frameworks. Here is an example for the message sent by the tracking node to the encrypt and obscure nodes:

```xml
<obscure>
  <frame number="804"
    timestamp="1203932276">
    <objectlist>
      <object id="1">
        <box xc="446" yc="242"
          w="68" h="78"/>
      </object>
      <object id="0">
        <box xc="254" yc="262"
          w="132" h="172"/>
      </object>
    </objectlist>
  </frame>
</obscure>
```

The messages refers to an internal frame count (804) and uses a timestamp within the MPEG stream of the camera for synchronization with other components. Within this frame, two objects have been detected and their coordinates are given as attributes of the box tag.

The obscure node takes the information sent by the tracker and applies two filters on the stream. First, a block filter is applied that averages color information and removes the texture from the image. Second, all colors are reversed by an invert filter. The later is used to make an obscured region noticeable when applied to an area with the same color and not much texture. This is mainly useful for illustration and debugging purposes. Fig. 4 shows an image as sent out by the obscure node.

### 3.2 Scalability

For better scalability we group our system in two parts: One part produces protected video streams, the other consumes it. The producer acts like a streaming server and the consumer like a streaming client. The grouping of the nodes is shown in Fig. 5. Camera, tracking, encrypting and obscuring nodes define the producer while decrypting, merging and displaying nodes define the consumer. Due to the use of multicasts to send the streams, new consumers can be added with no additional cost in network traffic, when the streams are already present in the network segment. If a new consumer has permission to view a protected stream sent by an encrypt node, only this will be sent in addition to existing...
protected streams and the obscured stream. A producer will only send its streams when a consumer has a demand for it. The number of producers that can send parallel streams is limited by the available network bandwidth. Also the number of protected streams can change over time. Eventually a situation may occur, where permission is available for a protected stream, but the network cannot take the additional traffic. The number of nodes tasked with encrypting and decrypting is flexible and it is possible to balance the load between them. The identifier sent by the tracking node can, for example, be used for allocating a specific node to a region. The tracking node is one of the computationally most demanding processes in our setup. Therefore, it would be desirable to have multiple tracking nodes working on only a subpart of the camera image. However, the effort to keep a consistent view on the surveilled scene may outweigh the performance gained and we do not use multiple trackers in our prototype. Another reason to increase the number of tracking nodes could be the demand for very specialized tracking methods, where multiple complementary tracking methods could improve the overall reliability.

### 3.3 Key Management

The key manager node controls the protected parts of the video stream. For every observed and identified person a list of persons with authorization to see a clear image of the observed is maintained. Additionally it is tasked with authentication. Authentication can be either implemented within the key manager or be delegated by the key manager to an external service.

To transfer the clear video securely between the producer and consumer (Fig. 5) the AES-256 cipher in counter mode is used. More specifically, it is used in the encrypt and decrypt nodes of all participating producers and consumers. A ciphertext \( C \) is created by \( C_i := P_i \oplus K_i \), with \( P \) being the payload and \( K \) the key [9]. For every ciphertext that is created the key \( K \) is composed of a random key and an initialization vector. The initialization vector is composed of a nonce\(^1\) and a counter. The secrecy of the protected information is depending on the key’s secrecy. Therefore, the random key and nonce part of the initialization has to be transfered securely between key manager and nodes tasked with encrypting or decrypting.

We do not assume that in our distributed system all nodes share the same knowledge on the system’s state. Therefore the session key is changed periodically by the key manager. When either an encrypt or decrypt node fails to receive the key change message, the decrypt node cannot decode the incoming stream and the obscured region is shown. When all participating encrypt and decrypt nodes fail to receive the key change message, they will continue to work until the counter repeats. Then the encrypt node is obliged to stop using the session key, as the same key would be used twice and thus possibly infringing security.

For playback of a recorded video it is necessary to store the session keys, assuming that the obscured and encrypted streams and regions are saved and not the original video. The key manager can use files or databases to store the keys and is responsible for protecting them. When the key manager is providing authentication, the keys may be protected by a secret of the person, for example a password or private key.

The key manager is also responsible for providing authentication. Authentication is needed for assigning a person to a display node and for observed persons. As authentication is also needed by the building automation system it is considered advantageous, when the key manager interfaces with this facility. The key manager may utilize a service like LDAP or RADIUS to delegate authentication.

### 3.4 Security

As our system follows an open and distributed approach there are several areas that need to be protected. Within the consumer and producer (Fig. 5) the video is sent unprotected. The channel between key manager and encrypt and decrypt nodes needs protection when session keys are transferred. The authentication process and transmission of authorization information are vulnerable as well. Further the obscured stream may be altered.

To open a secure channel between key manager and encrypting and decrypting node we use the SSL protocol [14]. With the use of host keys we identify participating nodes. The https protocol is often supported by streaming video cameras. This protocol uses SSL to transport information securely and we use it therefore in the producer, if possible. As access via https will be unicast, a multicast reflector can be used to spread the stream. Within the consumer SSL and host keys can be used as well. To protect against attacks from the person using the system, a client-side certificate or pass-phrase can be used. Communication between the system and the client can be done over SSL again.

Additionally there are a number of security problems the system does not attempt to address: First, a person with access permission to a clear video can record or stream the video without protection. Second, cameras involved might be accessed directly, circumventing the complete system. Moreover, control over the key management node will result in control over all protected streams.

### 3.5 Prototype

A prototype was developed and deployed to gain insight in the technical feasibility of our approach. First, a reduced

---

\(^1\)Number used Once, see [9]
producer was implemented consisting of a manual identification, camera, track and obscure node. A prototype interface was implemented that allowed to switch between the clear and obscured stream. A first usability test with colleagues showed high interest for an even improved ease of use of the system and a wish to also manually control the size of the obscured area. Second, the cryptographic system made of key management, configuration, encrypt and decrypt node was put together. Also a user interface to manually add an obscured area to the video was implemented. Finally, the merging node was implemented and replaced the semi-automatic switching between obscured and protected streams. Simplications were made in respect to the number of allowed regions to track and the absence of an automatic identification. Authentication is provided by the building automation’s bar code reader, which is used to read an employees or students identification card. A unique identifier is embedded there for use with the school’s libraries. Currently RFID is evaluated if it is suited as replacement for the bar code reader.

4 Visual Tracking

Automatic detection and tracking of objects from video cameras is a very active research topic in the computer vision community. Applications such as visual surveillance, intelligent living environment and human behavior analysis rely directly on new visual tracking algorithms. Within the context of “privacy in video surveilled areas” a real-time visual tracking method [19] will be presented, showing the capabilities of vision based trackers.

The visual tracking presented in this section is most closely related to the following publications [13, 2]. Mittal and Davis [13] developed a multi-camera system which also uses Bayesian classification. However, the 3D segmentation approach owes its robustness to the use of multiple cameras, that stand in the way of real-time implementation. Capellades et al. [2] implemented an appearance based tracker which applies Bayesian classification only on pre-segmented foreground pixels, while we let all the models – object models, background model and model explaining newly appeared objects – compete for all pixels, thus maintaining a consistent probabilistic approach throughout the whole algorithm.

4.1 Bayesian Per-Pixel Classification

The proposed method performs a per-pixel classification to assign every pixel to one of the different objects that have been identified, including a background. The classification is based on the probability that a given pixel belongs to one of the objects given its specific color and position. The object probabilities are determined on the basis of two components. On the one hand, the appearance of the different objects is learned and updated, which yields indications of how compatible observed pixel colors are with these models. On the other hand, a motion model makes predictions of where to expect the different objects, based on their previous position. The approach is akin to similar Bayesian filtering approaches, but has been slimmed down to strike a good balance between robustness and speed.

Fig. 6 sketches the tracking framework with its probability images. Different characteristics for every object such as its appearance and motion are incorporated in specialized models and updated over time. The segmentation image assigns every pixel individually to the object with the highest multiplied probability of its appearance and motion model.

![Figure 6. Tracking framework](image)

4.2 Tracking Models

Color Model All our appearance models use Gaussian mixtures in RGB color space. Stauffer and Grimson [22] proposed this popular choice for modeling scene backgrounds with time-adaptive per-pixel mixtures of Gaussians (TAPMOGs). However, we modified this approach to fit into our multi-model approach.

Appearance Background Model In contrast to Stauffer’s algorithm which combines foreground and background in one color model per pixel we use individual models for whole foreground objects. Therefore we model the background only with a single Gaussian per pixel. The background model is initialized at system startup.

Detecting new Objects Pixels with a significant different color than the background model have a low probability. If it is lower than a certain threshold the pixel is
assigned to a generic ‘new object model’. If whole regions fall on this model, a new foreground object is detected and it’s motion and appearance models are initialized.

**Appearance Foreground Model** For the appearance models we use a so called ‘sliced object model’, as it divides the person into a fixed number of horizontal slices of equal height. For each slice, color models with multiple Gaussians are generated representing the most important colors in that region of a person. (Fig. 7(b) and 7(c))

**Motion model** The size and movement of each foreground object is predicted individually. We use a linear Kalman filter for the 2D position in the image and a weighted average filter for the object size (Fig. 7(a)).

![Motion model](image)

(a) Motion model  
(b) 7 slices  
(c) probability image

Figure 7. Different specialized models

5 Interface

The primary task of the user interface is to control a person’s appearance on a viewer’s display in our privacy enhancing system. The observed person has to be capable to allow or deny access to their clear view for every viewer or whole groups.

Requirements on this interface for us are

- Ease of use
- Ease of integration in other interfaces
- Deployable on fixed terminals as well as on hand held devices

To fulfill these requirements we took a look at related areas first. We found that internet based chatting applications deal with similar problems: Services like ICQ or Jabber use the concept of friends or buddies to manage a potentially very high number of users very easily with the help of lists.

The design and use of our interface prototype adapts several techniques found in instant messaging clients to the special needs of our privacy application. The prototype of the user interface is programmed in Flash [11] which is well known to work with most browsers and on many computer platforms. This allows us to use a variety of end user devices like fixed workplaces and laptops, touch-screen displays and very compact and mobile internet tablets like the Nokia 770. It would even be possible to deploy the interface on a flash capable cellphone.

In our system prototype the user interface queries the key manager for a list of persons. The protocol used for information exchange between key manager and user interface is XML based. Information can be relayed between the two with a http proxy. This assures that the key manager can work from a private network. After the interface received the list of pre-registered persons, a login screen looking familiar to many computer users is shown (Fig. 8).

![Login Screen of User Interface Prototype](image)

Figure 8. Login Screen of User Interface Prototype

After identification with a valid password the list of potential viewers is displayed. In Fig. 9 the list of viewers that will see the obscured image are called ‘user’. Those viewers that are allowed to see the clear image are found below ‘superuser’. The terms ‘user’ and ‘superuser’ are configurable and the ones chosen in the example give a hint on the observed persons favorite operating system.

To give or revoke viewers access to the clear video images an intuitive drag and drop mechanism is used. The prototype interface follows standards set by online chatting systems. These are known to work for a large number of people.

6 Results

This section demonstrates several of theses features using different sequences and datasets recorded with our sys-
Figure 9. Privacy Control Screen of User Interface Prototype

our system prototype. A more detailed tracking evaluation with public video databases is available in [19]. Our first scene contains two people entering a meeting room for a discussion. Fig. 10 shows the camera view with the two persons and a bounding-box around them showing the successful tracking by the algorithm. Due to our multi model tracking framework the persons are detected during the whole meeting without fading to the background.

Figure 10. Multi-person tracking in a conference room

The advantages of the presented tracking method for the proposed scenario are:

- Discrete and different models for foreground objects and the background allow the tracking of moving and motionless objects.
- Adaptive color models for foreground and background adapt individually to changing lighting conditions.
- Multiple people can be tracked individually and occlusions between two objects can be handled for objects with different color-models.
- The actual implementation runs in real-time on a Linux workstation at QVGA (320x240) resolution.

Occlusion can be handled by the tracking framework as shown in Fig. 11. However, sufficiently discriminable color models of the objects are a prerequisite for correct operation. During the occlusion phase, appearance and motion models are no more updated. This results in frozen color models and a fixed dimension of the objects during the whole occlusion phase. While partial occlusion, the object positions are taken from non-overlapping object parts. During complete occlusion the velocity of the object is assumed constant until some object parts reappear.

6.1 Limitations of visual tracking

Current tracking methods are not able to fully model all possible events happening in front of a camera, nor are they able to have an in-depth understanding of the scene due to limited computational resources. For our segmentation based approach all changes in the image of a reasonable size and not related to persons are a source of detection or tracking errors. Fig. 12 shows such problems in the form of a projection screen, where the tracker cannot distinguish between real objects or projected ones. Furthermore, changes in the environment like the door which was left open in Fig. 12 can lead to wrong objects if new images significantly differ from the initially learned background model. The Adaptation of the background is designed to handle slow changes in the image like varying lighting conditions.

In most security applications, it is not sufficient to use just a single technology to provide an accurate security standard. Therefore, we do not propose that computer vision based techniques alone can provide a foolproof system able to provide privacy. Instead a combination of multiple techniques should be used. Information from the building automation system about open doors, projector screens or lights turning on or off are possible solutions to these problems.

6.2 Synchronization

A synchronization mechanism had to be deployed as the different nodes process their information with differ-
Figure 11. Occlusion handling between two persons. The lower image shows the per-pixel classification.

Figure 12. Large change in the background can lead to wrong objects.

Figure 13. Synchronization problem

Due to the need for synchronizing and thus waiting for information to arrive the prototype has a high latency that differs on the systems configuration between 1 and 10 seconds. Several nodes add to the latency. Beginning with the camera some latency is added due to the use of MPEG-2. Next the tracking node has to analyze and understand the scene. The merging node has to wait for the decrypting node to deliver the unobscured information and it has to buffer the video stream to send a continuous stream to the displaying node.

To reduce the latency within the camera an open camera

ent speed. Influenced by the utilized MPEG-2 video cameras (Axis-230 [1]) a timestamp embedded within the initial camera video-stream is used. Due to the difference in frame rates between the camera which delivers 25 frames per second (fps) and the tracking node which delivers 10-15 fps, a delay on the obscured video is very noticeable. This excludes the use of our system for a live video conference setting.

Fig. 13 illustrates another problem found. The face of the person sitting to the right is not obscured. Even a sitting person will move slightly and can move out of the region detected by the tracking node between updates of the frames. Therefore the obscure node adds a fixed margin of 16 pixels to each side of the tracked region. This value has been derived from observations of the running system and is adjustable.

When information arrives at a merging node it also needs to be synchronized, as the obscured stream will usually arrive ahead of the regions sent by decrypting nodes. Again the same timestamp is used.

6.3 Latency

Due to the need for synchronizing and thus waiting for information to arrive the prototype has a high latency that differs on the systems configuration between 1 and 10 seconds. Several nodes add to the latency. Beginning with the camera some latency is added due to the use of MPEG-2. Next the tracking node has to analyze and understand the scene. The merging node has to wait for the decrypting node to deliver the unobscured information and it has to buffer the video stream to send a continuous stream to the displaying node.
system like the Elphel 333 [5] might be used. The camera is open to modification and comes with a programmable FPGA. The FPGA source code that runs on the camera is available as Open Source [6]. The task of obscuring and encrypting parts of the video can be done on the camera itself. The latency of the merging and displaying node could further be reduced when they would be combined into a single node.

6.4 Multiple identities

When persons are standing very close to each other it might not able to separate and identify them individually. Instead the tracker node will only notice a single region and send a single identity. Therefore, the system needs to support groups of identities. In such a case the group of people will only be visible to a viewer holds the permission to view all persons in the group. However, the whole systems will act trivial in very crowded situation, when many people cover the whole camera image and either the tracker combines them to a single big group or the massive occlusions lead to wrong tracking results. But given the scope of the project to a room within an office building we cannot imagine that this happens very often.

6.5 Interface Usability

The interface prototype was presented to colleagues at the chair for CAAD. Overall the interface was found to be intuitively usable. The comments made during this usability test were:

1. the ability to control the margin to the tracked region by the observed person
2. the ability to obscure the complete video
3. the ability to define groups of persons
4. color coding to identify persons currently viewing
5. ability to control a display in the room to view the taken video

The first four items are already implemented or planned to be while item 5 should be solved by closer integration with the interface of the building automation system.

7 Conclusion

The presented work focused on an automatic distributed system that provides reliable, scalable and efficient protection of a persons privacy while being surveilled by video cameras in a building. Using computer vision, cryptographic technologies and an integrated interface we can achieve this partially.

A prototype was successfully built on top of a building automation system. We added specialized processing nodes to build a scalable and distributed setup containing nodes for camera, tracking, key management, encryption, decryption and display. An easy to use graphical user interface allows everybody in a building to control the system from a variety of different electronic devices.

The functionality of the presented system prototype also showed today's technological limitations. What has been achieved is a system where a person can give and revoke access to others. Only authorized persons can view the clear video while others see only an obscured version. If a viewer has the permission to see a subset off all persons in a surveilled scene, the resulting video contains mixed clear and obscured images of persons. However, the people have to be correctly identified manually and a reliable tracking cannot always be guaranteed.

Due to the limitations in identifying and today's visual tracking methods further components will have to be added. A barcode scanner is one possible solution for the identification problem, as it is not feasible to robustly identify people from CCTV cameras automatically. Even with a barcode on an ID card, the observed person has still to be willing to provide its identity truthfully. It is very possible that a person might be tempted to take on a false identity. Deployment of an authentication method that makes this harder will affect the ease of use and acceptance of the system. For example entering a pass phrase or identifying by biometric methods (iris scan, fingerprint) is thought to be an awkward procedure in an everyday situation in an office.

Even when the above mentioned limitations are solved some problems are permanent. What happens to video information that can be viewed by an authorized person who distributes the video further? Currently there is no way to control a digital video distribution when the system is open. Only with very restrictive digital rights management (DRM) methods is it possible to limit a distribution. But many methods used in DRM are not applicable here. Most prominently a loss in quality of the video will still be an acceptable option for abuse of the video. Also people with access to the personal keys will be able to see all videos.

8 Acknowledgments

The authors gratefully acknowledge support by the ETH Zürich project blue-c-II and Swiss SNF NCCR project IM2.

References


[12] Britain will be first country to monitor every car journey. news.independent.co.uk/uk/transport/article334686.ece.


[18] The Register: CCTV Peeping Toms jailed. www.theregister.co.uk/2006/01/13/cctv_men_jailed/.


