Privacy Enabling Technology for Video Surveillance

Frédéric Dufaux [a,b], *Mourad Ouaret* [a,b]*, Yousri Abdeljaoued* [a,b]*,*
Alfonso Navarro [b]*, Fabrice Vergnenègre* [b] *and Touradj Ebrahimi* [a,b]

[a] Institut de Traitement des Signaux
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

[b] Emitall Surveillance SA
Rue du Théâtre 5
CH-1820 Montreux, Switzerland

ABSTRACT

In this paper, we address the problem privacy in video surveillance. We propose an efficient solution based on transform-domain scrambling of regions of interest in a video sequence. More specifically, the sign of selected transform coefficients is flipped during encoding. We address more specifically the case of Motion JPEG 2000. Simulation results show that the technique can be successfully applied to conceal information in regions of interest in the scene while providing with a good level of security. Furthermore, the scrambling is flexible and allows adjusting the amount of distortion introduced. This is achieved with a small impact on coding performance and negligible computational complexity increase. In the proposed video surveillance system, heterogeneous clients can remotely access the system through the Internet or 2G/3G mobile phone network. Thanks to the inherently scalable Motion JPEG 2000 codestream, the server is able to adapt the resolution and bandwidth of the delivered video depending on the usage environment of the client.

Keywords: video surveillance, privacy, scrambling, JPEG 2000, universal multimedia access, heterogeneous clients

1. INTRODUCTION

With the increase of terrorist threats in the recent years and the high level of criminality in urban areas, security remains a major public concern. Video surveillance systems are becoming ubiquitous. They are widely deployed in many strategic places such as airports, banks, public transportation or busy city centers, but also in private houses. However, most current systems are analog and are based on proprietary solutions. It is expected that the next generation of video surveillance systems will be digital and based on standard technologies and IP networking.

While people usually appreciate the sense of increased security brought by video surveillance, they often fear the loss of privacy which comes along. This legitimate concern is often slowing down the deployment of video surveillance. Privacy in video surveillance has been previously addressed in [1][2][3][4][5]. The technique in [1] is based on an object-oriented representation of the scene. The system re-renders a modified video based on the end-user access control authorizations. During re-rendering, areas of the image are blanked out or replaced by computer graphics. The relevant information in the scene is therefore preserved, but privacy-sensitive details are not transmitted. Similarly, in [2] a privacy buffer utilizes privacy filters operating on incoming sensor data to prevent access to sensitive information or transform data to remove private information. These privacy filters are expressed using a privacy grammar. The work in [3] is addressing the threat to privacy brought by face recognition techniques which can automatically identify people in a video surveillance scene. An algorithm is proposed to de-identify faces such that many facial characteristics are preserved but the face cannot be reliably recognized. This is achieved by defining similarity between faces and creating new faces by averaging image components. The techniques in [4][5] are addressing the problem of privacy for Motion

JPEG 2000 video [6]. Both approaches include an analysis module to identify Regions of Interest (ROI), e.g. corresponding to people or faces in the scene. In our earlier work [4], code-blocks corresponding to these ROIs are then scrambled using a wavelet-domain or codestream-domain conditional access control technique. However, the shape of the scrambled region is restricted to match code-block boundaries, which may become a drawback in the case of complex geometry with small arbitrary-shape regions. Conversely, in [5] the data corresponding to the ROIs is downshifted to the lowest quality layer of the codestream. Hence, the ROIs can be decoded to a lower quality by restricting the transmission bandwidth.

In this paper, we present a video surveillance system including a region-based transform-domain scrambling technique to preserve privacy. We consider more specifically the case of Motion JPEG 2000, an extension of JPEG 2000 for the coding of video sequences, which consists of intra-frame coding of each frame using Discrete Wavelet Transform (DWT) JPEG 2000 [6]. However, the method is extensible and can be used with most existing video coding standards, such as Motion JPEG, MPEG-4 or Advanced Video Coding (AVC) as shown in [7]. Video analysis is used to identify regions corresponding to people and assumed to contain privacy-sensitive information. The resulting ROIs are then scrambled. More specifically, Discrete Wavelet Transform (DWT) coefficients corresponding to the ROIs are scrambled by pseudo-randomly inverting their signs. Per consequent, the scene remains understandable, but the people are unidentifiable. The scrambling process depends on a private encryption key which can be in possession of law-enforcement authorities who are therefore the only ones able to unlock and view the whole scene in clear. This scrambling is fully compliant with the emerging Secured JPEG 2000 (JPSEC) [8][9].

Compared to [1][2][5], we believe our approach using scrambling offers a number of advantages and is therefore more appealing. In our proposal, the same protected codestream is transmitted to all clients regardless of their access control credentials. On the one hand, unauthorized clients do not possess the private key required for unscrambling the content. Therefore, they can only view distorted version of the content where private information is not identifiable. On the other hand, authorized clients, e.g. law-enforcement authorities, can unscramble the codestream and recover the complete undistorted scene. In addition, the proposed scrambling technique is very flexible. It can be restricted to arbitrary-shape ROIs, and the amount of distortion introduced can be adjusted from merely fuzzy to completely noisy. Finally, the technique requires a very low computational complexity.

Another important feature of a surveillance system is to be able to remotely access and control the system from heterogeneous clients. In the proposed system, clients such as desktop PCs, laptop PCs, PDAs or mobile phones, can access the system through either the Internet Protocol (IP) (wired or wireless channels) or 2G/3G mobile network. Thanks to the inherently scalable Motion JPEG 2000 codestream, the server is able to adapt the resolution and bandwidth of the delivered video depending on the usage environment and more specifically the terminal capabilities and the network characteristics of the client.

This paper is structured as follow. In Sec. 2, we present the overall architecture of the proposed video surveillance system. Privacy enabling technology is discussed more thoroughly in Sec. 3. The issue of universal multimedia access is addressed in Sec. 4. Finally, we draw some conclusions in Sec. 5.

2. SYSTEM ARCHITECTURE

Hereafter, we describe an efficient, flexible and standard-based architecture for smart video surveillance system. A high level description of the system is illustrated in Figure 1.

The system is composed of several surveillance cameras connected to a surveillance server. The system supports USB or IP cameras. The case of wireless Wi-Fi cameras is especially appealing as it makes it very easy and cost effective to deploy and relocate cameras as the surveillance needs evolve. Analog cameras can also be used providing that the surveillance server is capable of capturing and digitizing the analog video signal.
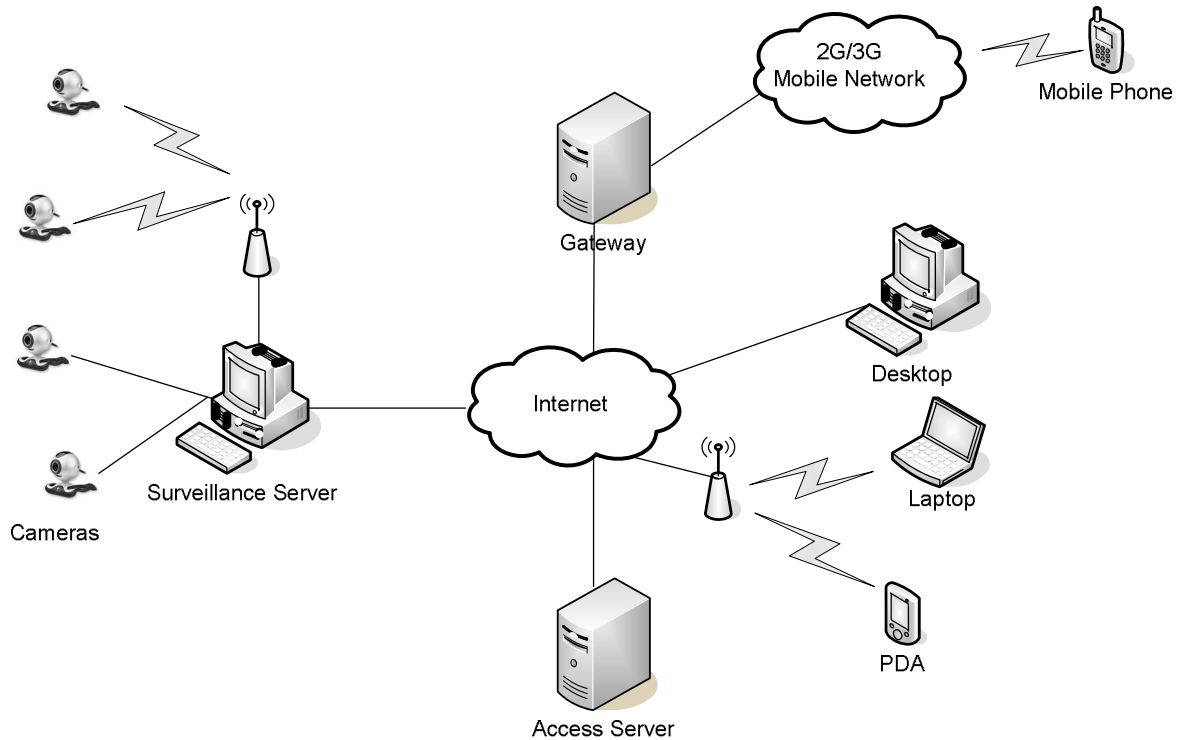
Figure 1 – Video surveillance system architecture.

The video is then processed on a surveillance server. More specifically, the following steps are carried out. First, video analysis identifies ROIs. In case of intrusion detection, the system may take appropriate actions, e.g. by triggering an alarm. The video is then compressed for efficient storage and transmission. For this task Motion JPEG 2000 is used [6], as it is especially well-suited for video surveillance applications [4]. Because of its high coding efficiency, good image quality is obtained. Moreover, Motion JPEG 2000 supports regions of interest coding, enabling previously identified ROIs to be encoded with higher quality. Finally, the resulting codestream is fully scalable, both in resolution and quality. This property is very essential in video surveillance applications when heterogeneous clients have to access the system. Next, scrambling is applied to the ROIs. The scrambling can be adapted to generate from mere fuzziness to complete noise. As a result, people in the scene are not recognizable while the remaining of the scene is visible, successfully addressing the loss of privacy issue. The scrambling is compliant with JPSEC [8][9] which defines an open framework for secure imaging. Finally, Wireless JPEG 2000 (JPWL) [10][11] can be used to achieve the efficient transmission of Motion JPEG 2000 codestream over an error-prone wireless network.

Heterogeneous clients can access the system in order to view live or recorded videos. For instance, policemen or security guards can be equipped with laptops or PDAs while on patrol. Similarly, home owners can receive relevant information on their mobile phone wherever they are. Internet clients access the system with the Internet Protocol (IP) on wired or wireless channels. Alternatively, mobile clients access the system through a gateway by means of 2G or 3G mobile network. Thanks to the inherently scalable Motion JPEG 2000 codestream, the server is able to adapt the resolution and bandwidth of the delivered video depending on the performance and characteristics of each client and its network connection. An access server authenticates the users and controls their authorization level.

The video surveillance system can be controlled and operated remotely from a mobile phone. For instance, commands to start and stop the system can be sent remotely to the server by a mobile phone using Short Message Service (SMS). It is also possible to query the status of the system by sending a SMS and receiving an image from the scene under surveillance by Multimedia Message Service (MMS). Finally, in case of an alarm, the system can send a slideshow of the suspicious event which triggered the alarm on a mobile phone by MMS, hence permitting for quick and efficient alarm

verification. This last functionality is especially appealing in the case of home surveillance where systems are often prone to false alarms.

3. PRIVACY ENABLING TECHNOLOGY

The privacy enabling technology consists of two steps. The scene is first analyzed in order to identify ROIs which are assumed to include privacy-sensitive information. Subsequently, the resulting ROIs are scrambled to make then unidentifiable. We now discuss these two steps in more details.

3.1. Analysis

The goal of the analysis module is to identify ROIs, assumed to contain privacy-sensitive data.

Relying on a human operator monitoring control screens in order to set off an alarm is notoriously inefficient. Therefore, another purpose of the analysis module is to either assist the human operator bringing to his attention abnormal behaviors or events, or even automatically trigger alarms.

It is clear that the automatic segmentation of objects in a video is still an open problem and that the efficiency of the system to preserve privacy depends strongly on the performance of the analysis module. Hereafter, we propose to use change detection or face detection. A similar technique combining face detection and tracking is used in [5]. More sophisticated objects segmentation and tracking techniques such as the one in [12] can also be used.

3.1.1. Change detection

The first step in the video analysis is to detected changes. In the proposed system, we suppose that all cameras remain static. Numerous techniques have been proposed in the literature for change detection. For the sake of low complexity, we choose here a simple frame difference algorithm. For this purpose, the background is captured and stored. Regions corresponding to changes are merely obtained by taking the pixel by pixel difference between the current video frame and the stored background, and by applying a threshold.

In order to handle changes of illumination in the scene, the background is slowly updated. More specifically, a new background is regularly generated by linearly combining a new video frame and the previous background. Furthermore, the threshold can also be adapted based on the level of illumination of the scene and the automatic gain control and white balance in the camera.

Finally, in order to smooth and clean up the resulting segmentation mask, a morphological filter is applied. This step removes small regions and holes in the segmentation mask.

While simple, this approach gives good performances for a very low computational complexity.

3.1.2. Face detection

The detection of the presence of people in the scene is one of the most relevant information a video surveillance system can convey. In our system, we use the face detection technique from the Open Computer Vision Library [13], which is based on [14][15]. In [14], a fast and efficient machine learning technique for object detection is proposed. The detection system works on Haar-like features computed using an integral image representation. A learning algorithm based on AdaBoost is used, leading to extremely efficient classifiers. Finally, these classifiers are combined in a cascade which quickly discards background regions, hence speeding up the algorithm. In [15], the set of Haar-like features is enriched by extending it with a set of rotated features. An improved boosting algorithm is also introduced.

The Open Computer Vision Library provides with a face detection technique trained on frontal faces. This technique not only achieves very high performance, but in addition it is computationally very effective.

3.2. Regions of Interest scrambling

We now address the problem of scrambling arbitrary-shape ROIs in a video sequence.

Earlier works on conditional access control have mostly considered the application of traditional cryptographic techniques to encrypt the codestream resulting from compression [16][17][18]. However, when compared to other types of information (e.g. banking data, confidential documents), video data is characterized by a very high bitrate and a low commercial value [19]. Therefore, conventional cryptographic techniques, which entail a significant complexity increase, are unsuitable in this case. Taking into account the above observations, an efficient video scrambling is proposed in [20] applying bit scrambling to transform coefficients and motion vectors during video encoding. This results in an approach giving a good level of security for a low complexity. The method results in the whole image being completely distorted and thus indecipherable.

In this paper, we address a different problem. Namely, we concentrate on the problem of scrambling ROIs in a video sequence, where the whole scene remains comprehensible but some objects cannot be identified. Whereas we consider its application to video surveillance system preserving privacy, the technique is also applicable to other applications such as to safeguard the anonymity of a source in TV news reporting.

Scrambling is closely linked to the scheme used to encode the video. Hereafter, we consider more explicitly Motion JPEG 2000. Motion JPEG 2000 is an extension of JPEG 2000 for the coding of video sequences which intra-frame encodes each frame using DWT-based JPEG 2000 [6]. However, the approach is extensible to other motion compensated transform-coding techniques such as Motion JPEG, MPEG-4 or AVC as shown in [7].

More specifically, we propose a region-based transform-domain scrambling technique inverting the signs of selected transform coefficients. The amount of distortion introduced by the scrambling can be adjusted, ranging from noise to blur. The technique allows for a good level of security. Finally, this is achieved with a small impact on coding performance and negligible computational complexity increase.

Scrambling can be effectively applied after the DWT and quantization, and before the arithmetic coder, as illustrated in Figure 2 (a). The process is fully reversible. At the decoder side, authorized users have merely to perform the exact inverse operation, as shown in Figure 2 (b).
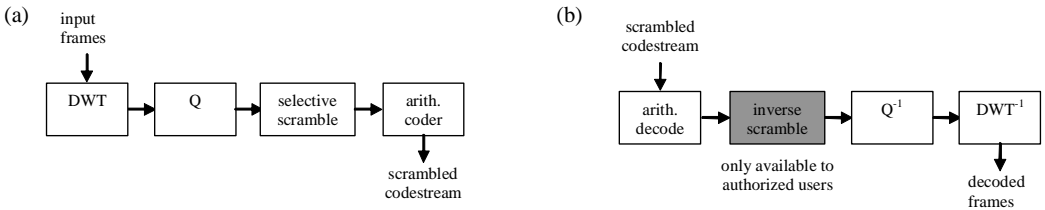
Figure 2 – Transform-domain scrambling in Motion JPEG 2000: (a) encoder and (b) decoder.

The scrambling should have a minimal impact on coding efficiency. As the wavelet coefficients are strongly correlated, scrambling them would reduce coding performance; they are therefore unsuitable for scrambling. However, the signs of wavelet coefficients are typically weakly correlated, and are thus appropriate for scrambling. Furthermore, in general AC coefficients are weakly correlated whereas DC coefficients are strongly correlated. Therefore, AC coefficients are more suitable for scrambling.

Per consequent, in our proposed algorithm quantized wavelet coefficients belonging to the AC subbands and corresponding to the regions of interest are scrambled by randomly flipping their sign, as shown in Figure 3. The amount of scrambling can be adjusted by restricting the scrambling to fewer resolution levels.
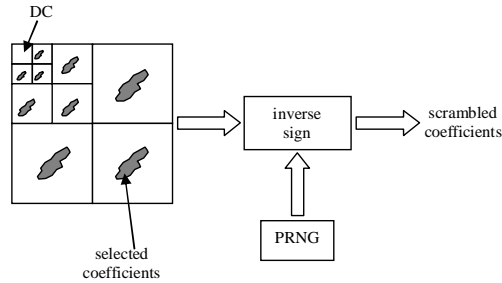
Figure 3 – Wavelet scrambling.

A Pseudo Random Number Generator (PRNG) initialized by a seed value is used to drive the scrambling process. In our implementation, the SHA1PRNG algorithm [21] with a 64-bit seed is used. In order to improve the security of the system, multiple seeds can be used. To communicate the seed values to authorized users, they are encrypted and inserted in the codestream. In our implementation, the RSA algorithm is used for encryption [22]. Note that other PRNG or encryption algorithms could be used as well.

The very same codestream is transmitted to all clients regardless of their access control credentials. On the one hand, unauthorized clients do not possess the private key required for unscrambling the content. Therefore, they can only view distorted version of the content where private information is not identifiable. On the other hand, authorized clients, e.g. law-enforcement authorities, can unscramble the codestream and recover the complete undistorted scene.

With this method, scrambled regions can have arbitrary shapes. The shape of the ROIs has to be available at both the encoder for scrambling and decoder for unscrambling. This could be done by transmitting the shape information as metadata either as part of the Motion JPEG 2000 codestream, or on a separate channel. More efficiently, the shape can be implicitly embedded using the ROI mechanism of JPEG 2000 as proposed in [23]. Furthermore, JPSEC defines an open framework for secure imaging, defining a powerful and flexible syntax [8][9]. Using this JPSEC syntax, the seeds driving the PRNG can be encrypted and embedded in the codestream. In this case, the resulting codestream is fully JPSEC compliant.

3.3. Simulation results

In this section, we present simulation results obtained with the proposed region-based transform-domain scrambling technique in order to evaluate its performance. Results have been obtained using JJ2000 [24].

Three video test sequences in CIF format are used: "Hall Monitor", "Surveillance", and "Road". Each sequence has a ground truth segmentation mask defining ROIs.

3.3.1. Privacy protection

Figure 4 show the capability of the scrambling technique to hide information in ROIs of the video. The strength of the scrambling is adjusted by controlling the number of scrambled coefficients. As can be observed, the scrambling makes it impossible to identify the objects in the scene, e.g. the people for "Hall Monitor" and "Surveillance", or the vehicles for "Road". This technique is therefore suitable to preserve privacy in video surveillance system.
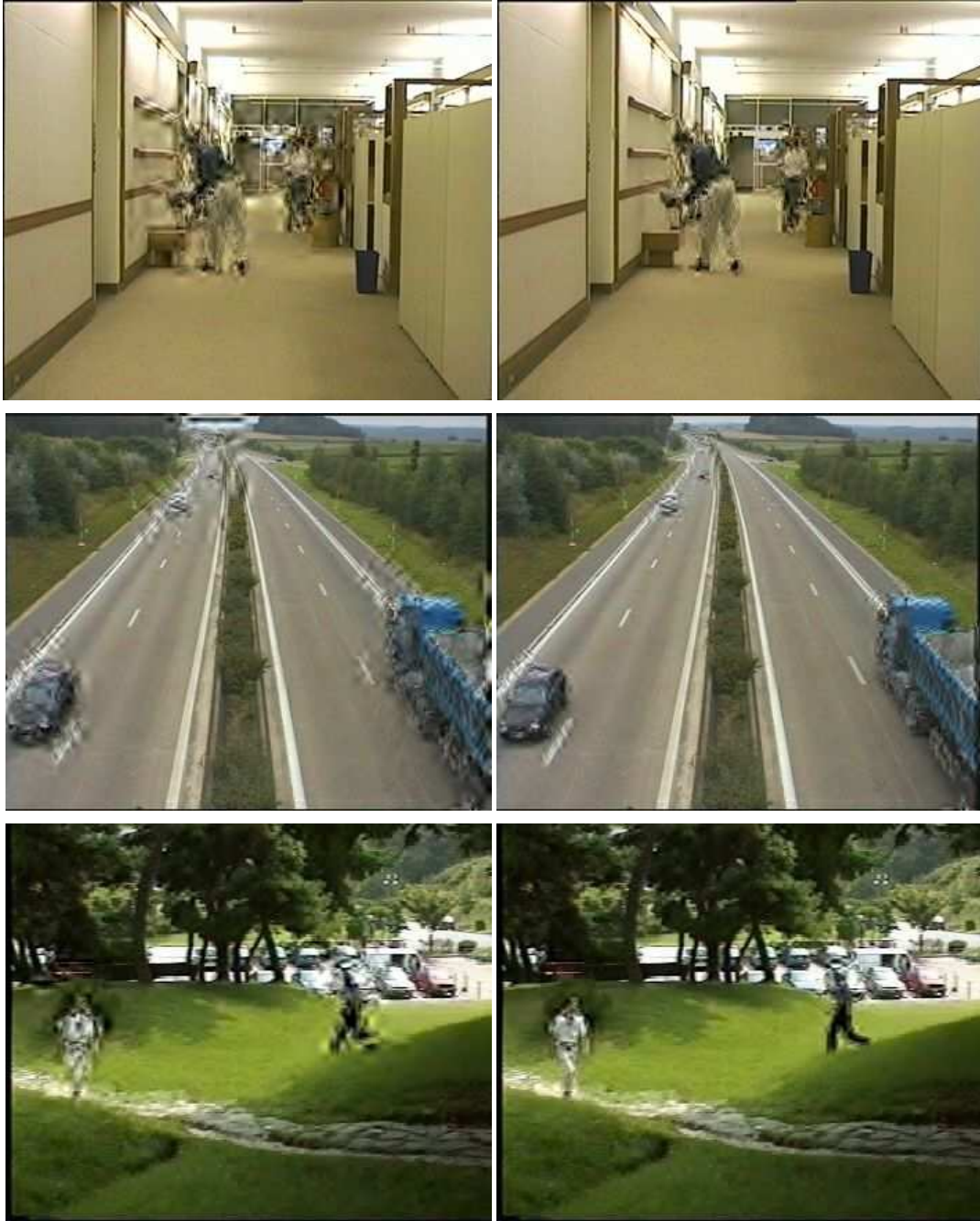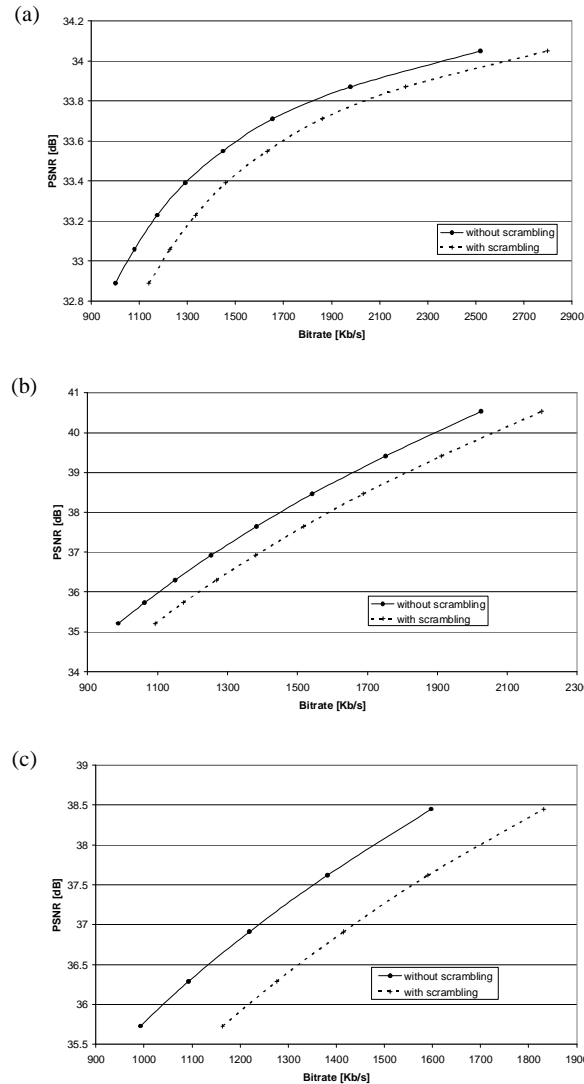
Figure 4 – Heavy or light scrambling: "Hall Monitor", "Surveillance" and "Road".

3.3.2. Coding efficiency

Next, we consider the performance of the scrambling technique in terms of coding efficiency. We compare the two cases when no scrambling is applied and when scrambling and unscrambling is performed. Figure 5 shows the rate-distortion performance.

Figure 5 – Rate distortion coding efficiency comparison between Motion JPEG 2000 without and with scrambling (a) "Hall Monitor", (b) "Surveillance", (c) "Road".

It can be observed that the proposed scrambling has a minimal impact on coding efficiency, resulting in bitrate increases of approximately 10 %.

4. UNIVERSAL MULTIMEDIA ACCESS

Another important feature of a surveillance system is to be able to remotely monitor live or archived video from heterogeneous clients such as desktop PCs and mobile clients such as laptop PCs, PDAs and mobile phones, connecting to the system through the Internet (wired or wireless) or 2G/3G mobile network. These clients should receive alert notifications (e.g. MMS or emails) with slideshow or video corresponding to the triggering event, in order to decide whether to request viewing live video. For instance, policemen or security guards receive video corresponding to suspicious events on their laptops or PDAs while on patrol, allowing for faster intervention and more efficient deployment of human resources.

4.1. Universal Multimedia Access

The property of a multimedia system to be accessible from heterogeneous clients is commonly referred to as Universal Multimedia Access (UMA) [25].

As the Motion JPEG 2000 codestream is scalable, the server is able to adapt the resolution and bandwidth of the delivered video depending on the usage environment and more specifically the terminal capabilities and the network characteristics of the client. For this purpose, we use three XML descriptors from MPEG-21 Digital Item Adaptation (DIA) [26] describing respectively the display capabilities, the network capability and the network condition. Examples of these three descriptors are given in Figure 6, Figure 7 and Figure 8.

The content adaptation is illustrated in Figure 9. The display capabilities and network capability descriptors are static and need to be transmitted to the surveillance server once at the beginning of the session. In contrast, the network condition descriptor is dynamic and is periodically transmitted. Given their very low data rate, all descriptors are simply transmitted in ASCII over TCP/IP. Upon receiving this usage environment information, the surveillance server adapts the Motion JPEG 2000 codestream to be delivered to the client. For instance, the resolution of the video can be easily adjusted by selecting the corresponding resolution levels. Similarly, the bitrate of the codestream can be adapted by dropping quality layers. Therefore, all clients receive a bitstream best adapted to their terminal capabilities and network characteristics. Moreover this adaptation is obtained without transcoding.

```
<DIA>
    <Description xsi:type="UsageEnvironmentType">
        <UsageEnvironment xsi:type="TerminalCapabilitiesType">
            <TerminalCapabilities xsi:type="DisplayCapabilitiesType"
                <Resolution  horizontal="176" vertical="144" activeResolution="1" />
                    bitsPerPixel="24"
                    colorCapable="1"
                    backlightLuminance="1.0000"
                    refreshRate="60.0000"
                    dotPitch="0.2000"
                    activeDisplay="1"
            </TerminalCapabilities>
        </UsageEnvironment>
    </Description>
</DIA>
```

Figure 6 – Display capabilities descriptor.

```
<DIA>
    <Description xsi:type="UsageEnvironmentType">
        <UsageEnvironment xsi:type="NetworkCharacteristicsType">
            <NetworkCharacteristics xsi:type="NetworkCapabilityType"
                maxCapacity="384000"
                minGuaranteed="32000"
                inSequenceDelivery="0"
                errorDelivery="0"
                errorCorrection="0"
            </NetworkCharacteristics>
        </UsageEnvironment>
    </Description>
</DIA>
```
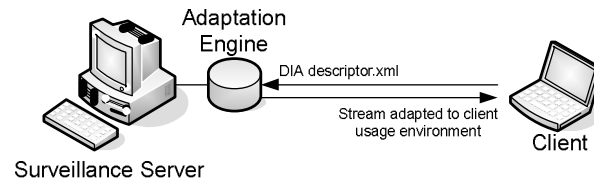
Figure 7 – Network capability descriptor.

```
<DIA>
    <Description xsi:type="UsageEnvironmentType">
        <UsageEnvironment xsi:type="NetworkCharacteristicsType">
            <NetworkCharacteristics xsi:type="NetworkConditionType"
                <AvailableBandwidth minimum="32000" maximum="256000"
                 average="80000" interval="330" />
                <Delay packetTwoWay="330" packetOneWay="165" delayVariation="66" />
                <Error packetLossRate="0.0500" bitErrorRate="3" />
            </NetworkCharacteristics>
        </UsageEnvironment>
    </Description>
</DIA>
```

Figure 8 – Network condition descriptor.

**Figure 9 – Universal multimedia access and content adaptation.**

Finally, it should be pointed out that the scrambling process described in Sec. 3 does not interfere with this content adaptation process. Indeed, the scrambling process does not alter the syntax of the resulting Motion JPEG 2000 codestream. Therefore, a scrambled codestream can be adapted in the same way as any Motion JPEG 2000 codestream. In particular, this guarantees the end-to-end security of the content.

4.2. Mobile phone communication

Commands can be sent to the video surveillance system from a mobile phone by sending SMS. For instance, it includes commands to remotely start, verify the status and stop the system. Conversely, the surveillance system can send information to a mobile phone, either in the form of text (SMS) or images (MMS).

More precisely, a gateway receives commands by SMS and sends information by SMS and/or MMS. In turn, the surveillance server communicates with the gateway by means of web services. This is illustrated in Figure 10.

This functionality is especially appealing for home surveillance applications. For instance, in case of intrusion detection the home owner can receive on his mobile phone a MMS with a slideshow of the incident which set off the alarm. This has two advantages. First, the mobile phone network usually reaches the user wherever he is. Second, the MMS is near-instantaneous. In this way, the alarm can be verified quickly, hence avoiding to a great extend the nuisance of false alarms.
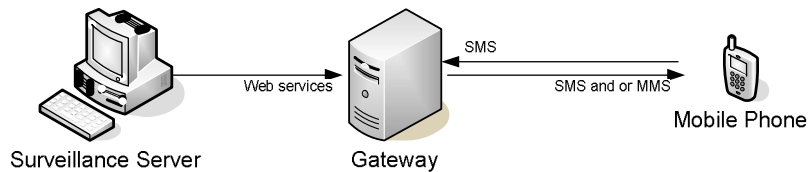
Figure 10 – Communication with mobile phone by means of SMS and MMS.

5. CONCLUSIONS

In this paper, we have described a technique to protect the privacy of people under surveillance in a video surveillance system. An analysis module identifies regions of interest. These regions, assumed to contain privacy-sensitive information, are then scrambled. The scrambling is performed in the transform domain by flipping the sign of transform coefficients during encoding. While the method is generic and can be applied to most existing video coding standards, we discussed in more details the specific case of Motion JPEG 2000 which is especially suited for video surveillance applications.

Simulation results show that the proposed scrambling technique successfully hides information in regions of interest in the scene. Furthermore, the scrambling is flexible and allows adjusting the amount of distortion introduced, from a mere fuzziness to a complete noise. This is achieved with a small impact on coding performance and negligible computation complexity increase. Finally, the method provides with a good security level.

A variety of clients with different capabilities and channel characteristics can access the system to view live or stored video surveillance streams. Thanks to the seamless scalable coding of Motion JPEG 2000, this can be efficiently achieved without transcoding.

REFERENCES

[1] A.W. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, and A. Ekin, "Blinkering Surveillance: Enabling Video Privacy through Computer Vision" IBM Technical Report RC22886, 2003.

[2] D. A. Fidaleo, H.-A. Nguyen, M. Trivedi, "The networked sensor tapestry (NeST): a privacy enhanced software architecture for interactive analysis of data in video-sensor networks", Proc. of the ACM 2nd Int. Workshop on Video Surveillance & Sensor Networks, New York, NY, 2004.

[3] E. Newton, L. Sweeney, and B. Malin, "Preserving Privacy by De-identifying Facial Images", Carnegie Mellon University, Technical Report CMU-CS-03-119, 2003.

[4] F. Dufaux and T. Ebrahimi, "Video Surveillance using JPEG 2000", SPIE Proc. Appl. of Digital Image Proc., Denver, CO, Aug. 2004.

[5] I. Martinez Ponte, X. Desurmont, J. Meessen, and J.-F. Delaigle, "Robust Human Face Hiding Ensuring Privacy" in Proc. of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, April 2005.

[6] D. Taubman and M. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice", Kluwer Academic Publishers, 2002.

[7] F. Dufaux and T. Ebrahimi, "Region-Based Transform-Domain Video Scrambling", in SPIE Proc. Visual Communications and Image Processing 2006, San Jose, CA, Jan. 2006

[8] F. Dufaux, S. Wee, J. Apostolopoulos and T. Ebrahimi, "JPSEC for secure imaging in JPEG 2000", SPIE Proc. Appl. of Digital Image Proc., Denver, CO, Aug. 2004.

[9] "JPEG 2000 Part 8 (JPSEC) Final Draft International Standard", ISO/IEC JTC1/SC29 WG1 N3820, Nov. 2005.

[10] F. Dufaux and D. Nicholson, "JPWL: JPEG 2000 for wireless applications", SPIE Proc. Appl. of Digital Image Proc., Denver, CO, Aug. 2004.

[11] "JPEG 2000 image coding system – Part 11: Wireless JPEG 2000 – Final Draft International Standard", ISO/IEC JTC1/SC29/WG1 WG1N3819, Nov. 2005.

[12] A. Cavallaro, O. Steiger, and T. Ebrahimi "Tracking video objects in cluttered background", in IEEE Trans. on Circuits and Systems for Video Technology, vol. 15, no. 4, April 2005.

[13] http://sourceforge.net/projects/opencvlibrary, Open Computer Vision Library.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in IEEE Proc. CVPR, Hawaii, Dec. 2001.

[15] R. Lienhart, A. Kuranov and V. Pisarevski, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection", MRL Technical Report, Intel Labs, Dec. 2002.

[16] I. Agi and L. Gong, "An empirical study of secure MPEG video transmissions", in Proc. of The Internet Society Symposium on Network And Distributed System Security, Feb. 1996.

[17] T. Maples and G. Spanos, "Performance study of a selective encryption scheme for the security of networked, real-time video", in Proc. 4th Int. Conf. Computer Communications and Networks, Las Vegas, NV, Sept. 1995.

[18] Y. Sadourny and V. Conan, "A proposal for supporting selective encryption in JPSEC", in IEEE Trans. on Consumer Electronics, vol. 49, no. 4, pp 846-849, Nov. 2003.

[19] B. Macq and j. Quisquater, "Cryptology for digital TV broadcasting", Proc. of IEEE, vol. 83, no. 6, pp. 944-957, 1995.

[20] W. Zeng and S. Lei, "Efficient Frequency Domain Video Scrambling for Content Access Control", in Proc. ACM Multimedia, Orlando, FL, Oct. 1999.

[21] http://java.sun.com/j2se/1.4.2/docs/guide/security/CryptoSpec.html, Java Cryptography Architecture API Specification and reference.

[22] R.L. Rivest, A. Shamir, and L.M. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", Communications of the ACM (2) 21, 1978, Page(s): 120-126.

[23] F. Dufaux and T. Ebrahimi, "Smart Video Surveillance System Preserving Privacy", in SPIE Proc. Image and Video Communications and Processing 2005, San Jose, CA, Jan. 2005.

[24] http://jj2000.epfl.ch

[25] A. Vetro, C. Christopoulos, and T. Ebrahimi, "Universal multimedia access", IEEE Signal Processing Magazine, vol. 20, no.2, p. 16, March 2003.

[26] H. Sun, A. Vetro, K. Asai, "Resource adaptation based on MPEG-21 usage environment descriptions", in Proc. of the International Symposium on Circuits and Systems (ISCAS'03), May 2003.