Scrambling for Video Surveillance with Privacy

Frédéric Dufaux and Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
frederic.dufaux@epfl.ch, touradj.ebrahimi@epfl.ch

Emitall Surveillance SA, CH-1820 Montreux, Switzerland
frederic.dufaux@emitall.com, touradj.ebrahimi@emitall.com

Abstract

In this paper, we address the problem of scrambling regions of interest in a video sequence for the purpose of preserving privacy in video surveillance. We propose an efficient solution based on transform-domain scrambling. More specifically, the sign of selected transform coefficients is pseudo-randomly flipped during encoding. We address more specifically the two cases of MPEG-4 and 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. Finally, this is achieved with a small impact on coding performance and negligible computational complexity increase.

1. Introduction

Video surveillance systems are becoming ubiquitous. They are widely deployed in many strategic places such as airports, banks, public transportation or busy city centers.

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][6]. 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. 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. In [4], a cryptographic technique is introduced to obscure faces, hence preserving the privacy of people under surveillance. The process is invertible for authorized personnel in possession of the necessary encryption keys.

The techniques in [5][6] are addressing the problem of privacy for Motion JPEG 2000 video [7]. In both approaches, Regions of Interest (ROIs) are identified by an analysis module, e.g. corresponding to people of faces in the scene. In [5], 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. In [6], the code-blocks corresponding to the ROIs are 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 propose a region-based transform-domain scrambling technique to preserve privacy in video surveillance systems. Similarly to [5][6], video analysis is used to identify regions corresponding to people and assumed to contain privacy-sensitive information. The resulting ROIs are then scrambled. Unlike [5][6], the approach can be applied to any transform-coding techniques such as the Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT). More specifically, transform 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 or a trusted third party who

are therefore the only ones able to unlock and view the whole scene in clear.

Compared to [1][2][3][4][5][6], we believe our approach using scrambling offers a number of advantages and is therefore more appealing. In our proposal, the system outputs a single protected code-stream. This very same code-stream 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 code-stream and recover the complete undistorted scene. The method can be used with most existing video coding standards, such as DCT-based Motion JPEG, MPEG-4 or Advanced Video Coding (AVC), or DWT-based Motion JPEG 2000. 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 has a small impact on coding performance and requires a very low computational complexity.

2. System Architecture

The architecture of the proposed video surveillance system is illustrated in Figure 1.

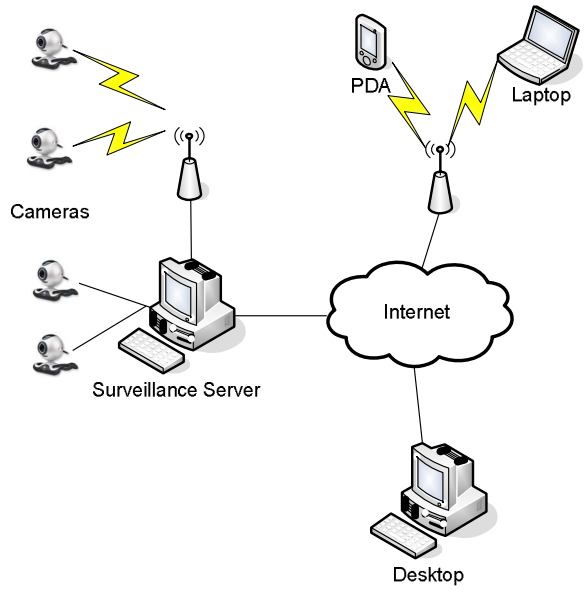


Figure 1 – Video surveillance system architecture.

The system is composed of several simple wired or wireless surveillance cameras connected to a video surveillance server. Video processing is then carried out on the server. First, a video analysis module identifies ROIs. The video is then compressed for efficient storage and transmission. Simultaneously, scrambling is applied to the ROIs. As a result, people in the scene are not recognizable while the remaining of the scene remains clear, successfully addressing the loss of privacy issue. Finally, heterogeneous clients can access the system through the Internet in order to view live or recorded videos.

While this paper is not addressing the issue of video analysis, 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. The use of face detection [8][9] is proposed in [4][5], change detection in [5], and a combination of face detection and tracking in [6]. Skin detection [10], more sophisticated objects segmentation and tracking [11], or a combination of all the above analysis techniques can also be used. In the remaining, we will assume that the ROIs containing privacy-sensitive data are known, and we will focus on the issue of scrambling these regions.

3. Video 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 [12][13][14]. 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 [15]. 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 [16] 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.

The following features are important for an efficient solution. The scrambling should not entail lower coding performance or significant complexity increase. It should cope with arbitrary-shape regions. Finally, it should be flexible, allowing for the adjustment of the amount of

distortion introduced.

3.1. Transform-domain Scrambling

Scrambling is closely linked to the scheme used to encode the video. Most video coding schemes are based on transform-coding. Namely, frames are transformed using an energy compaction transform such as the DCT or DWT. The resulting coefficients are then entropy coded using techniques such as Huffman or arithmetic coding. Hereafter, we consider more explicitly two video coding schemes: MPEG-4 [17] and Motion JPEG 2000 [7]. However, the approach is straightforwardly extensible to all transform-coding techniques based on DCT or DWT.

One approach is to perform scrambling in the original image prior to encoding. This approach has the advantage of being very simple and independent from the encoding scheme subsequently used. However, it has the disadvantage of significantly altering the statistics of the video signal, hence making the ensuing compression less efficient.

Another approach is to apply scrambling after encoding. More specifically, the compressed codestream is directly scrambled. The main drawback of this approach is that it is very difficult to guarantee that the scrambled codestream will not crash a standard decoder.

Taking into account the above remarks, we propose to perform scrambling in the transform-domain. More specifically, we introduce 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.

3.2. MPEG-4

We now consider the scrambling of ROIs in MPEG-4 encoded video. MPEG-4 is based on a motion compensated block-based DCT [17]. Each frame is subdivided in 16x16 MacroBlocks (MB). In turn, each MB is composed of four 8x8 luminance blocks and two 8x8 chrominance blocks. The DCT is performed on each of these 8x8 blocks, resulting in 64 DCT coefficients: one DC and 63 AC coefficients. As both the encoder and decoder contain the motion compensation loop, attention has to be paid for the scrambling process not to introduce a drift between these two loops.

Taking into account the above remark, scrambling can be effectively applied on the quantized DCT coefficients, and outside of the motion compensation loop, as illustrated in Figure 2 (a). At the decoder side, authorized users perform unscrambling of the coefficients resulting from entropy decoding, prior to the motion compensation loop, as depicted in Figure 2 (b). Straightforwardly, as the scrambling is kept out of the motion compensation loop, this allows for a fully reversible process for authorized users.

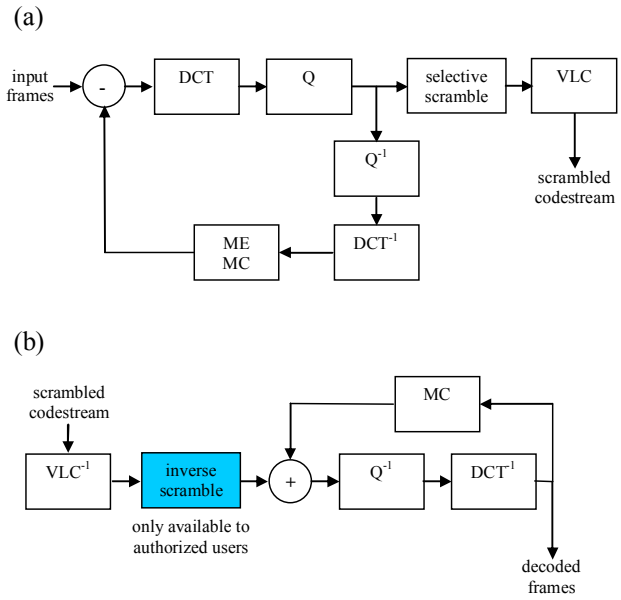


Figure 2 – Transform-domain scrambling in MPEG-4: (a) encoder and (b) decoder.

From Figure 2 (b), it can be deduced that an unauthorized decoder, i.e. which is not capable of unscrambling, will use a different motion compensation loop than an authorized decoder. As a result, an unauthorized decoder may experience a drift, resulting in artifacts in the scrambled sequence, as depicted in Figure 3 (a). This undesirable effect can be avoided by modifying the MB type decision during encoding. More precisely, unscrambled MBs in the current frame, co-located with a scrambled MB in the reference frame, are always INTRA coded. This modification of the MB type decision prevents the drift in motion compensation loop and consequently removes the artifacts in the scrambled sequence, as shown in Figure 3 (b).

(a)

(b)

Figure 3 – Scrambled video sequence
(a) Verification Model MB type decision resulting in drift,
(b) Modified MB type decision removing drift.

The scrambling should have a minimal impact on coding efficiency. In general DC coefficients are strongly correlated and are therefore unsuitable for scrambling. Furthermore, whereas the amplitude of AC coefficients is correlated, their signs are not. Per consequent, in our proposed algorithm, all 63 quantized AC coefficients of the blocks corresponding to the ROIs are scrambled by pseudo-randomly flipping their sign, as shown in Figure 4. Straightforwardly, the shape of the scrambled region is restricted to match the 8x8 DCT blocks boundaries. The amount of scrambling can be adjusted by restricting the scrambling to fewer AC coefficients.

A Pseudo Random Number Generator (PRNG) initialized by a seed value is used to drive the scrambling process. In our implementation, the SHA1PRNG algorithm [18] 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 code-stream. In our implementation, the RSA algorithm is used for encryption [19]. Note that other PRNG or encryption algorithms could be used as well.

In order to unscramble the codestream, authorized decoders need to know the shape of the ROIs. The latter has therefore to be transmitted as metadata either in private data in the MPEG-4 codestream, or on a separate channel.

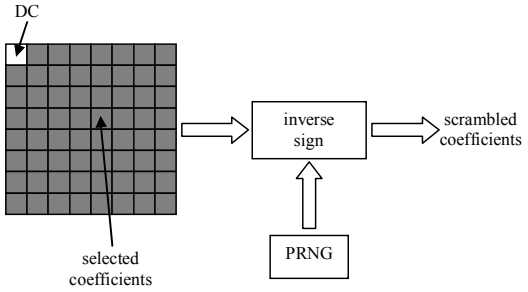


Figure 4 – 8x8 DCT block scrambling.

Straightforwardly, as the scrambling is merely flipping signs of selected coefficients, the technique requires negligible computational complexity.

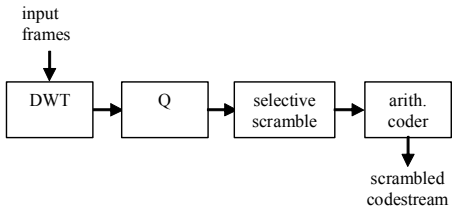
Finally, it should be pointed out that the same technique can be extended to other DCT-based schemes, such as Motion JPEG or AVC.

3.3. Motion JPEG 2000

In this section, we address the scrambling of ROIs for Motion JPEG 2000 encoded video sequences. Motion JPEG 2000 is an extension of JPEG 2000 for the coding of video sequences. It consists of the intra-frame coding of each frame using wavelet-based JPEG 2000 [7].

Scrambling can be effectively applied after the DWT and quantization, and before the arithmetic coder, as illustrated in Figure 5 (a). The process is fully reversible. At the decoder, authorized users have merely to perform the exact inverse operation, as shown in Figure 5 (b).

(a)



(b)

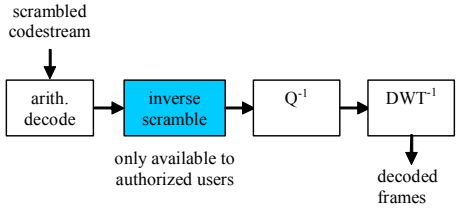


Figure 5 – Transform-domain scrambling in Motion JPEG 2000:
(a) encoder and (b) decoder.

The scrambling technique itself follows a similar approach as proposed for MPEG-4 in Sec. 3.2. Quantized wavelet coefficients belonging to the AC subbands and corresponding to the ROIs are scrambled by pseudo-randomly flipping their sign, as shown in Figure 6. The amount of scrambling can be adjusted by restricting the scrambling to fewer resolution levels.

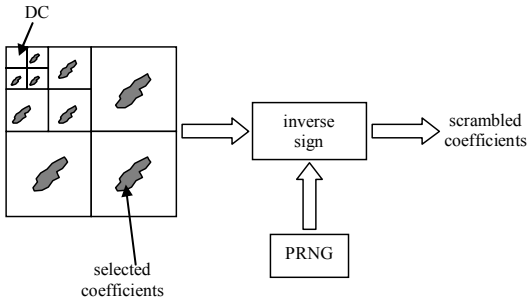


Figure 6 – Wavelet scrambling.

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 [7].

Furthermore, an extension of the baseline JPEG 2000, Secured JPEG 2000 (JPSEC) is of special interest [20][21]. JPSEC defines an open framework for secure imaging, defining a powerful and flexible syntax. 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.

4. Simulation Results

In this section, we present simulation results obtained with the proposed region-based transform-domain scrambling technique in MPEG-4 and Motion JPEG 2000 in order to evaluate its performance. Two video test sequences in CIF format are used: Hall Monitor and Surveillance. Each sequence has a ground truth segmentation mask defining ROIs.

Results for MPEG-4 have been obtained with the MoMuSys Verification Model [22], whereas results for Motion JPEG 2000 have been obtained using JJ2000 [23].

4.1. Scrambling for Privacy

We first consider the capability of the scrambling technique to hide information in ROIs of the video. Figure 7 and Figure 8 show scrambling results for MPEG-4 and Motion JPEG 2000 respectively. In both cases, the strength of the scrambling is adjusted by controlling the number of scrambled coefficients.

Figure 7 – Scrambling with varying strengths for MPEG-4: Surveillance.

Figure 8 – Scrambling with varying strengths for Motion JPEG 2000: Hall Monitor.

As can be observed, the scrambling makes it impossible to identify the people in the scene. This technique is

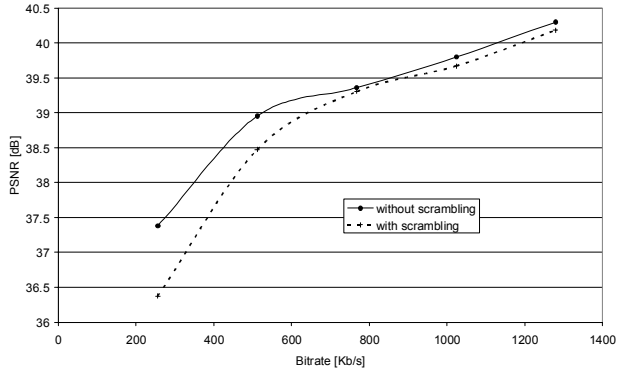
therefore suitable to preserve privacy in video surveillance system.

4.2. 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 9 and Figure 10 show the rate-distortion performance for MPEG-4 and Motion JPEG 2000 respectively.

It can be observed that the proposed scrambling has a minimal impact on coding efficiency, resulting in bitrate increases of less than 10 % for MPEG-4, and approximately 10 % for Motion JPEG 2000.

(a)



(b)

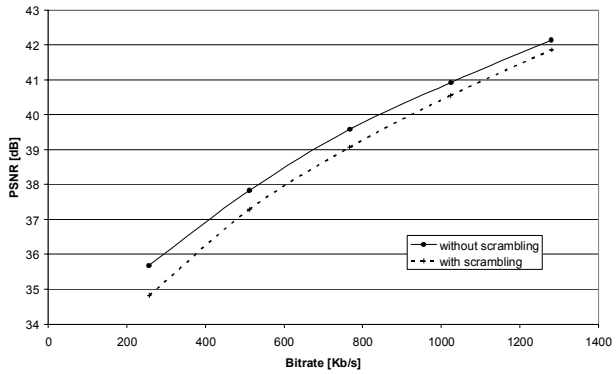
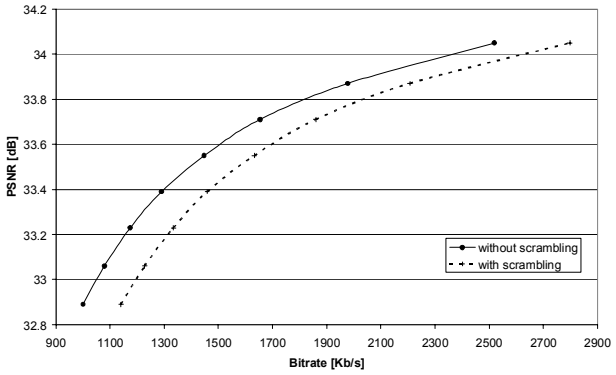


Figure 9 – Rate distortion coding efficiency comparison between MPEG-4 without and with scrambling
(a) Hall Monitor, (b) Surveillance.

(a)



(b)

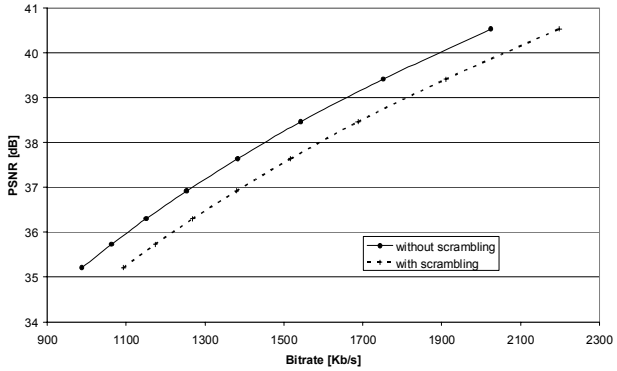


Figure 10 – Rate distortion coding efficiency comparison between Motion JPEG 2000 without and with scrambling (a) Hall Monitor, (b) Surveillance.

4.3. Security

We now consider the security of the proposed scrambling technique. Let us consider a brute-force attack where the attacker tries all combinations reversing the signs of all non-zero AC coefficients. If we consider the luminance component of a CIF frame, i.e. 352x288 pixels, both MPEG-4 and Motion JPEG 2000 (3 levels of decomposition as shown in Figure 6) results in 99'792 AC coefficients. We further suppose that the attacker knows the ROIs which cover 5% of the image, henceforth restricting the number of corresponding AC coefficients to 4'990. Finally, assuming that only 5 % of those are non-zero, an attacker would have to try reversing the signs of 250 coefficients, representing therefore 2^{250} combinations for each frame. Therefore, the method provides with a good level of security.

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 pseudo-randomly flipping the sign of transform coefficients during encoding. While the method is generic and can be applied to all existing video coding standards, we discussed in more details the two specific cases of MPEG-4 and Motion JPEG 2000. Simulation results show that the proposed scrambling can be successfully applied to hide information in regions of interest in the scene. Furthermore, it 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

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] T.E. Boult, "PICO: Privacy through Invertible Cryptographic Obscuration", IEEE/NFS Workshop on Computer Vision for Interactive and Intelligent Environments, Nov. 2005.

[5] F. Dufaux, and T. Ebrahimi, "Video Surveillance using JPEG 2000", in SPIE Proc. Applications of Digital Image Processing XXVII, Denver, CO, Aug. 2004.

[6] 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.

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

[8] http://sourceforge.net/projects/opencvlibrary

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

[10] M. Jones and J. Rehg, "Statistical Color Models with Applications to Skin Detection", TR-98-11, CRL, Compaq Computer Corp., Dec. 1998.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] T. Ebrahimi and F. Pereira, "The MPEG-4 Book", Prentice Hall, 2002.

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

[19] 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.

[20] JPEG 2000 Part 8 (JPSEC) FDIS, ISO/IEC JTC1/SC29 WG1 N3820, November 2005.

[21] J. Apostolopoulos, S. Wee, F. Dufaux, T. Ebrahimi, Q. Sun and Z. Zhang, "The Emerging JPEG 2000 Security (JPSEC) Standard", in IEEE Proc. Int. Symp. on Circuits and Systems (ISCAS), Island of Kos, Greece, May 2006.

[22] ISO/IEC JTC1/SC29/WG11 WG11N5550, "ISO/IEC 14496-7/DAM1 Optimized reference software for coding of audio-visual objects", March 2003.

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