Region-Based Transform-Domain Video Scrambling

Frederic Dufaux and Touradj Ebrahimi

Emitall S.A.
Rue du Théâtre 5
CH-1820 Montreux, Switzerland
Frederic.Dufaux@emitall.com, Touradj.Ebrahimi@emitall.com

Institut de Traitement des Signaux
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Frederic.Dufaux@epfl.ch, Touradj.Ebrahimi@epfl.ch

ABSTRACT

In this paper, we address the problem of scrambling regions of interest in a video sequence. We target applications such as video surveillance preserving privacy, anonymous video communications, or TV news safeguarding the anonymity of a source. We propose an efficient solution based on transform-domain scrambling. More specifically, the sign of selected transform coefficients is flipped during encoding. We address the two cases of Motion JPEG 2000 and MPEG-4. Simulation results show that it 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.

Keywords: scrambling, privacy, region of interest, media security, MPEG-4, JPEG 2000

1. INTRODUCTION

Security is a major concern in many digital video applications. Indeed, the ease to manipulate, copy and distribute digital content at negligible cost raises the issues of confidentiality, authentication, data integrity and conditional access control.

In this paper, we focus on conditional access control. Earlier works have mostly considered the application of traditional cryptographic techniques to encrypt the codestream resulting from compression [1][2][3]. 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 [4]. 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 [5] 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 regions of interest in a video sequence, where the whole scene remains comprehensible but some objects cannot be identified. Target applications include video surveillance system preserving privacy [6], anonymous Internet video chat or video telephony [7], or TV news protecting the anonymity of an informant. We consider two video coding schemes: Motion JPEG 2000 [8][9] and MPEG-4 [10].

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.

This paper is structured as follow. In Sec. 2, we first give arguments for performing scrambling in the transform-domain. We propose two region-based transform-domain scrambling techniques for Motion JPEG 2000 and MPEG-4 in Sec. 3 and Sec. 4 respectively. Experimental results showing the effectiveness of the approach are presented in Sec. 5. Finally, we draw conclusions in Sec. 6.

2. VIDEO SCRAMBLING

We now address the problem of scrambling regions of interest in video sequences. 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 Discrete Cosine Transform (DCT) or wavelet transform. The resulting coefficients are then entropy coded using techniques such as Huffman or arithmetic coding. Basically, scrambling can be applied at three different stages: in the image-domain prior to coding, in the transform-domain during coding, or in the codestream-domain after coding. We more thoroughly discuss these approaches hereafter.

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.

2.1. Image-Domain Scrambling

The first approach is to perform scrambling in the original image prior to encoding, as illustrated in Figure 1. This can be achieved for instance by randomly flipping the bits in one or more bit planes of the pixels belonging to the regions of interest.

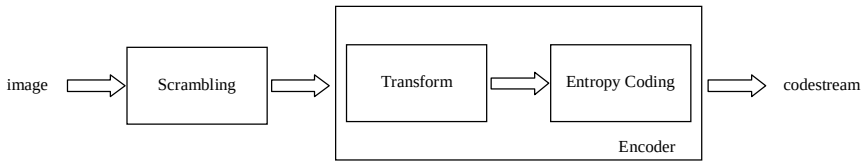


Figure 1 – Image-domain scrambling.

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.

Note that the same effect could be achieved to some extend by masking the pixels corresponding to the regions of interest (e.g. replacing them by a solid color), or by applying a low-pass filter (e.g. making the region sufficiently blurred). However, these two approaches have the drawback to preclude the possibility to ever unscramble the video. In addition, the masking approach provides an all-or-nothing solution without flexibility to control the amount of distortion introduced.

2.2. Transform-Domain Scrambling

A second approach is to apply scrambling during encoding, as shown in Figure 2. Scrambling is taking place after the DCT or wavelet transform and before entropy coding. More specifically, this can be done by randomly flipping the sign of transform coefficients corresponding to the regions to be scrambled. Besides its simplicity, this approach does not adversely affect the subsequent entropy coding. Furthermore, thanks to the frequency analysis property of the transform, the strength of the scrambling can be controlled by restricting the scrambling to some frequencies.

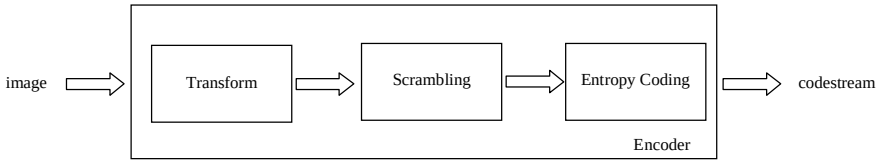


Figure 2 – Transform-domain scrambling.

Another benefit of this approach is that it preserves the syntax of the codestream, e.g. maintaining standard compliance. This enables content adaptation or transcoding at mid-network nodes or proxies, as is often required in a video delivery system.

2.3. Codestream-Domain Scrambling

In the third approach, scrambling is applied after encoding, as illustrated in Figure 3. More specifically, the compressed codestream is directly scrambled. Again, this can be efficiently done by randomly flipping bits in the stream.

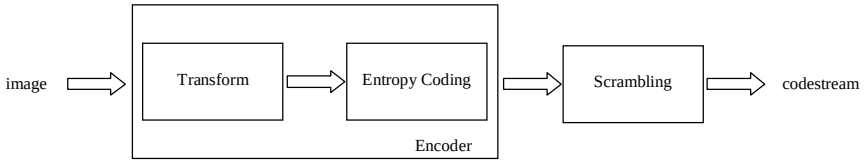


Figure 3 – Codestream-domain scrambling.

One of the drawbacks of this approach is that the codestream has to be parsed in order to identify which parts correspond to the regions to be scrambled, hence entailing a larger computational complexity. Furthermore, another severe drawback is that it may be difficult to guarantee that the scrambled codestream will not crash a decoder. Finally, the strength of the scrambling cannot be easily adjusted.

3. REGION-BASED TRANSFORM-DOMAIN SCRAMBLING FOR MOTION JPEG 2000

In this section, we address the scrambling of regions of interest 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 [8][9].

Given the arguments discussed in Sec. 2, we consider the transform-domain approach. Scrambling can be effectively applied after the Discrete Wavelet Transform (DWT) and quantization, and before the arithmetic coder, as illustrated in Figure 4 (a). The process is fully reversible. At the decoder side, authorized users have merely to perform the exact inverse operation, as shown in Figure 4 (b).

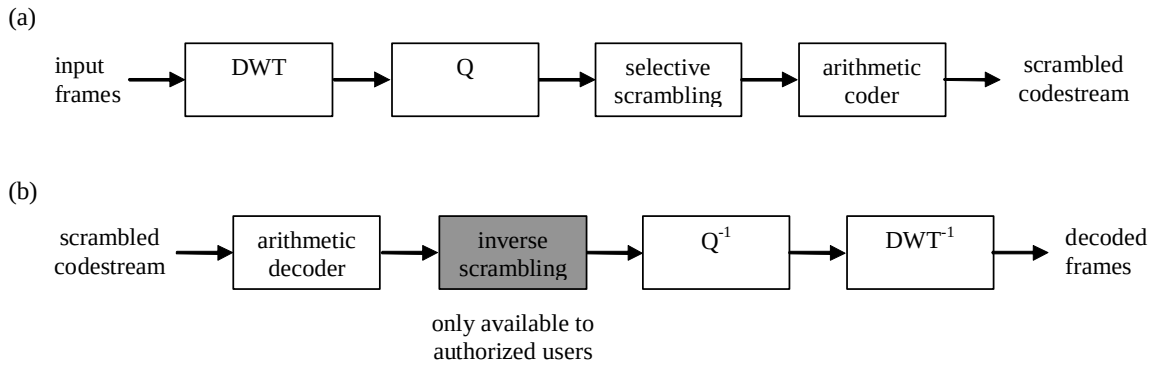


Figure 4 – Transform-domain scrambling in Motion JPEG 2000: (a) encoder and (b) decoder.

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 5. A Pseudo Random Number Generator (PRNG) is used to drive the scrambling process. The amount of scrambling can be adjusted by restricting the scrambling to fewer resolution levels.

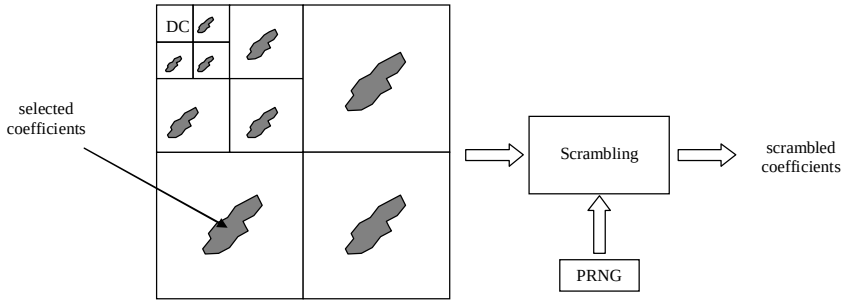


Figure 5 – Wavelet scrambling: coefficients in subbands from levels 1, 2 and 3 are scrambled for an image decomposed with 3 resolution levels.

The proposed scrambling technique relies on a PRNG driven by a seed value. In our implementation, the SHA1PRNG algorithm [11] with a 64-bit seed is used. Note that other PRNG could be used as well. In order to improve the security of the system, the seed can be frequently changed.

With this method, scrambled regions can have arbitrary shapes. The shape of the regions of interest 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, as proposed in [6], the shape can be implicitly embedded using the Region of Interest (ROI) mechanism of JPEG 2000.

Furthermore, an extension of the baseline JPEG 2000, Secured JPEG 2000 (JPSEC) [12], is of special interest. JPSEC defines an open framework for secure imaging, defining a powerful and flexible syntax. Using this JPSEC syntax, the seeds driving the PRNG and the scrambling process can be encrypted and embedded in the codestream. In this case, the resulting codestream is fully JPSEC compliant.

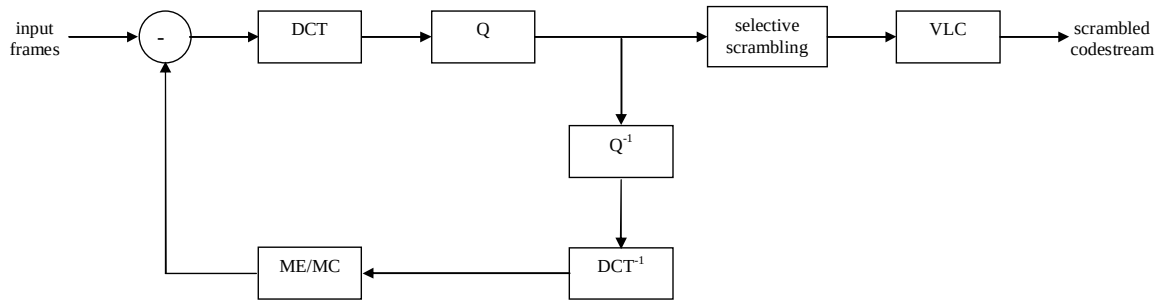
Straightforwardly, as the scrambling is merely flipping signs of selected wavelet coefficients, the technique requires negligible computational complexity.

4. REGION-BASED TRANSFORM-DOMAIN SCRAMBLING FOR MPEG-4

We now consider the scrambling of regions of interest in MPEG-4 encoded video. MPEG-4 is based on a motion compensated block-based Discrete Cosine Transform (DCT) [10]. As both DCT and DWT can be seen as special cases of subband decompositions, the same scrambling approach as for Motion JPEG 2000 can be used. However, in contrast with Motion JPEG 2000 which is based on intra-frame coding, MPEG-4 uses inter-frame coding. 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 remarks, scrambling can be effectively applied on the quantized DCT coefficients, and outside of the motion compensation loop, as illustrated in Figure 6 (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 6 (b). Straightforwardly, as the scrambling is kept out of the motion compensation loop, this allows for a fully reversible process for authorized users.

(a)



(b)

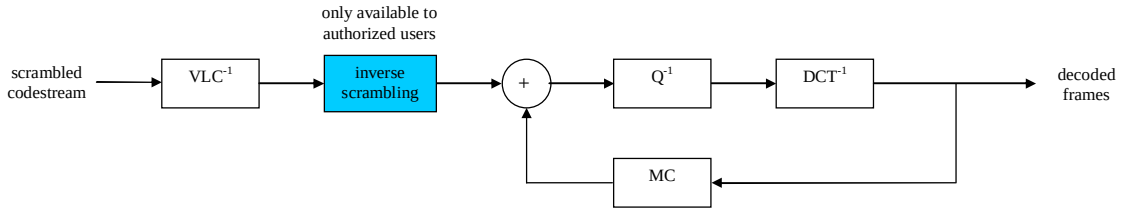


Figure 6 – Transform-domain scrambling in MPEG-4: (a) encoder and (b) decoder.

From Figure 6 (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 will experience a drift, resulting in artifacts in the scrambled sequence, as depicted in Figure 7 (a). This undesirable effect can be removed by modifying the MacroBlock (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 7 (b).

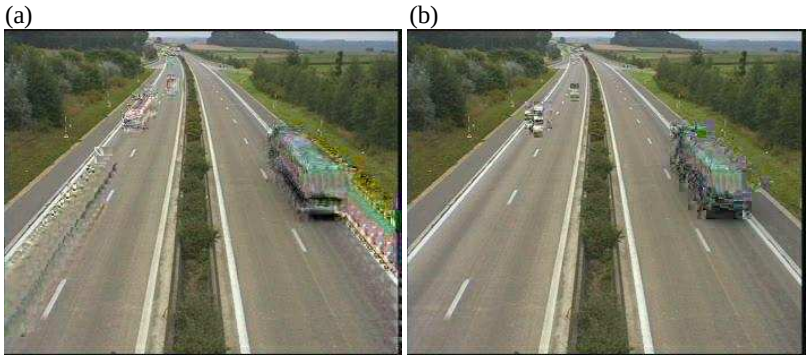


Figure 7 – Scrambled video sequence
(a) Verification Model MB type decision resulting in drift, (b) Modified MB type decision removing drift.

The scrambling technique itself follows a similar approach as proposed for Motion JPEG 2000 in Sec. 3. With MPEG-4, each frame is subdivided in 16x16 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.

We first identify all the blocks corresponding to the regions to be scrambled. For these blocks, all 63 AC coefficients are scrambled by randomly reversing their sign, as illustrated in Figure 8. As before, PRNG is used to drive the scrambling process. Note that it is possible to scramble fewer AC coefficients in order to obtain a lighter scrambling.

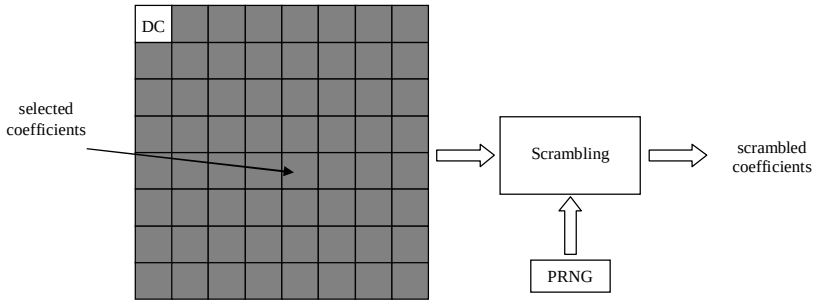


Figure 8 – 8x8 DCT block scrambling: all 63 AC coefficients are scrambled.

Straightforwardly, in this case, the shape of the scrambled region is restricted to match the 8x8 DCT blocks boundaries. In order to unscramble the codestream, authorized decoders need to know the shape of the regions of interest. The latter has therefore to be transmitted as metadata either in private data in the MPEG-4 codestream, or on a separate channel. In parallel, the encrypted seeds can be transmitted in a similar way.

Finally, it should be pointed out that the same technique could be used for other DCT-based schemes, such as Advanced Video Coding (AVC) / H.264 or Motion JPEG.

5. EXPERIMENTAL RESULTS

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

Results for Motion JPEG 2000 have been obtained using JJ2000 [13], whereas results for MPEG-4 have been obtained with the MoMuSys Verification Model [14].

5.1. Scrambling

We first consider the capability of the scrambling technique to hide information in regions of interest of the video. Figure 9 and Figure 10 show scrambling results for Motion JPEG 2000 and MPEG-4 respectively. In both cases, 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 vehicles in Figure 9 or the people in Figure 10.

This technique is therefore suitable to preserve privacy in video surveillance system [6], to guarantee the anonymity of participants in Internet video chat or video telephony [7], or to safeguard the anonymity of a source in TV news reporting.

Figure 9 – Scrambling with varying strengths for Motion JPEG 2000: Road.

Figure 10 – Scrambling with varying strengths for MPEG-4: Surveillance.

5.2. 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 11 and Figure 12 show the rate-distortion performance for Motion JPEG 2000 and MPEG-4 respectively.

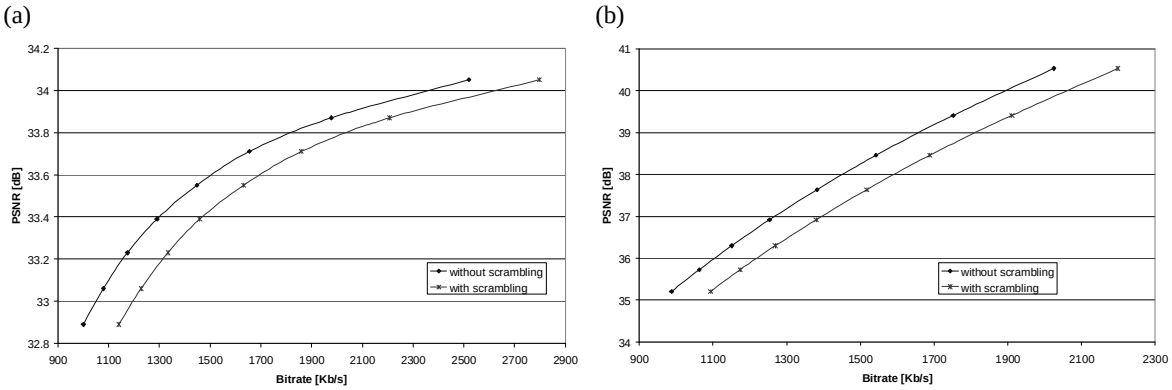


Figure 11 – Rate distortion coding efficiency comparison between Motion JPEG 2000 without and with scrambling (a) Hall Monitor, (b) Surveillance.

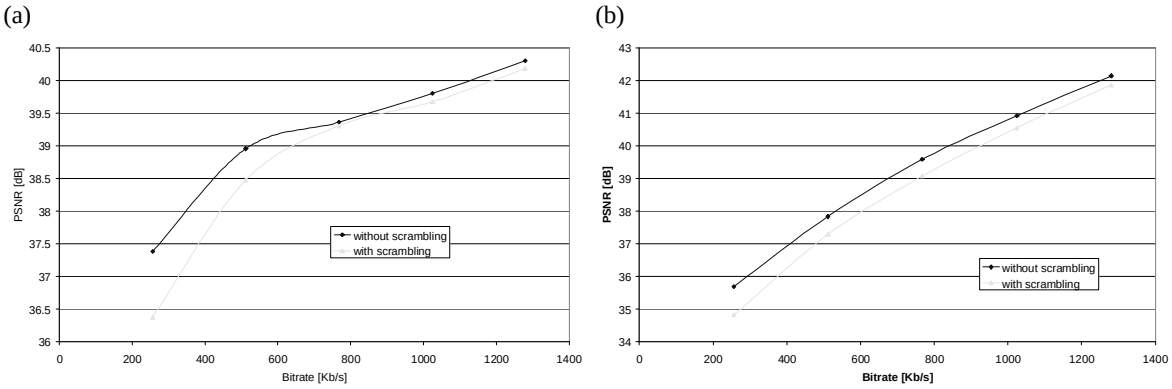


Figure 12 – Rate distortion coding efficiency comparison between MPEG-4 without and with scrambling (a) Hall Monitor, (b) Surveillance.

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

5.3. Security Analysis

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 Motion JPEG 2000 (3 levels of decomposition as shown in Figure 5) and MPEG-4 results in 99'792 AC coefficients. We further suppose that the attacker knows the regions of interest 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.

6. SUMMARY

In this paper, we presented an efficient region-based transform-domain technique to scramble video. Basically, the method is flipping the sign of transform coefficients during encoding. We proposed solutions for the two specific cases of Motion JPEG 2000 and MPEG-4. Simulation results show that it can be successfully applied to hide information in regions of interest in the scene. Furthermore, the scrambling is flexible and allows adjusting the amount of distortion introduced, from noise to mere blur. This is achieved with a small impact on coding performance and negligible computation complexity increase. Finally, the method provides with a good security level.

This proposed scrambling technique is suitable for applications such as video surveillance preserving privacy, anonymous Internet video chat or video telephony, or TV news safeguarding the anonymity of a source.

REFERENCES

[1] 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.
[2] 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.
[3] 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.
[4] B. Macq and j. Quisquater, "Cryptology for digital TV broadcasting", Proc. of IEEE, vol. 83, no. 6, pp. 944-957, 1995.
[5] W. Zeng and S. Lei, "Efficient Frequency Domain Video Scrambling for Content Access Control", in Proc. ACM Multimedia, Orlando, FL, Oct. 1999.
[6] 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.
[7] F. Dufaux, and T. Ebrahimi, "Scrambling for Anonymous Video Chat", in SPIE Proc. Applications of Digital Image Processing XXVIII, San Diego, CA, Aug. 2005.
[8] A. Skodras, C. Christopoulos and T. Ebrahimi "The JPEG 2000 still image compression standard", IEEE Signal Processing Magazine , vol. 18, no. 5, pp. 36 -58, Sept. 2001.
[9] D. Taubman and M. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice", Kluwer Academic Publishers, 2002.
[10] T. Ebrahimi and F. Pereira, "The MPEG-4 Book", Prentice Hall, 2002.
[11] http://java.sun.com/j2se/1.4.2/docs/guide/security/CryptoSpec.html, Java Cryptography Architecture API Specification and reference.
[12] JPEG 2000 Part 8 (JPSEC) FCD, ISO/IEC JTC1/SC29 WG1 N3480, November 2004.
[13] http://jj2000.epfl.ch
[14] ISO/IEC JTC1/SC29/WG11 WG11N5550, "ISO/IEC 14496-7/DAM1 Optimized reference software for coding of audio-visual objects", March 2003.