

Scrambling for Anonymous Visual Communications

Frederic Dufaux and Touradj Ebrahimi

Emitall S.A.

Rue du Théâtre 5, 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 present a system for anonymous visual communications. Target application is an anonymous video chat. The system is identifying faces in the video sequence by means of face detection or skin detection. The corresponding regions are subsequently scrambled. We investigate several approaches for scrambling, either in the image-domain or in the transform-domain. Experiment results show the effectiveness of the proposed system.

Keywords: scrambling, anonymity, video chat, region of interest, media security, MPEG-4, JPEG 2000

1. INTRODUCTION

In this paper, we propose a system to provide anonymous visual communications. The target application is an anonymous Internet video chat. Chat is one of the very popular activities on the Internet. Besides its ease and convenience to communicate, part of its appeal resides in the anonymity it provides. Thanks to technological advances, many chats, such as Yahoo Messenger [1] and MSN Messenger [2], now offer the possibility of a video link in order to enhance communications. This provides with a desirable sense of human contact.

With the proposed system, we can advantageously combine the functionality of visual communications with the preservation of anonymity. More specifically, we describe a technique to scramble regions identified as corresponding to faces in a video sequence. In this way, a user can transmit video but decide to hide his face until being more acquainted with the other party.

Note that the proposed system can also be exploited in other applications. For instance, it can be used in video telephony, whenever one of the parties wishes temporarily not to be seen. The same approach can also be used in TV news or documentaries to preserve the anonymity of a source. The proposed technique also shares some similarities with the video surveillance system preserving privacy presented in [3], which is scrambling regions corresponding to people so that they cannot be recognized.

More specifically, the proposed system first performs video analysis in order to locate the position of human faces in the scene. This step can be achieved for instance using face detection [4][5], skin detection [6][7] or change detection. Once regions of interest are identified, they are scrambled. The focus of this paper is to explore and compare various approaches to perform video scrambling. We address the problem of scrambling regions of arbitrary shape, and we consider scrambling in the image-domain prior to coding, in the transform-domain during coding, or in the codestream-

domain after coding. We deal more particularly with scrambling in conjunction with two video coding schemes: MPEG-4 [8] and Motion JPEG 2000 [9][10].

This paper is structured as follow. In Sec. 2, we discuss in more details the proposed system, with an emphasis on the scrambling aspects. In order to evaluate the performance of the method, experimental results are reported in Sec. 3. Finally, we draw conclusions in Sec. 4.

2. PROPOSED SCRAMBLING SYSTEM

In this section, we discuss in more details the proposed system to identify and scramble regions of interest corresponding to faces in the video, in order to provide anonymous visual communications. The video processing steps taking place are depicted in Figure 1.

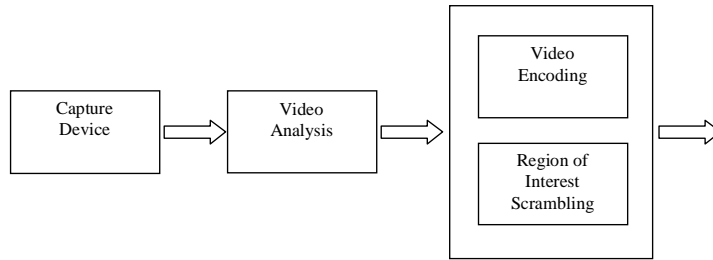


Figure 1 – Processing steps.

The video content is first analyzed. This step will detect the location of faces in the scene. In the proposed system, we use two techniques: face detection [4][5] and skin detection [6][7].

Once the regions of interest are identified, scrambling and encoding is performed. This is the focus of this paper. 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. In this paper, we consider more specifically two well-known video coding schemes: MPEG-4 and Motion JPEG 2000. MPEG-4 is based on a motion compensated block-based DCT [8]. 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 [9][10]. 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. Video Analysis

To identify faces in a video sequence is a well-known and very challenging problem. In this paper, we merely use some of the state-of-the-art techniques for face detection and skin detection.

Many face detection techniques have been proposed in the literature. For instance, a neural network technique is proposed in [4]. In [5], a machine learning algorithm based on AdaBoost, Haar-like features and a cascade of classifiers is introduced. Typically, these techniques achieve high performance for frontal faces. Furthermore, they are computationally very effective.

As far as skin detection is concerned, a learning-based system trained on a large corpus of data is proposed in [6]. Skin detectors in various color spaces are discussed in [7]. These approaches lead to good performance with a very low computational complexity.

Finally, in order to smooth and clean up the resulting segmentation mask, a morphological filter is applied [11]. More specifically, we perform an *opening* (i.e. erosion followed by dilation), then a *closing* (i.e. dilation followed by erosion). This step removes small regions and holes in the segmentation mask.

2.2. Image-Domain Scrambling

We now address the problem of scrambling a region of interest in the video sequence. The first approach is to perform scrambling in the original image prior to encoding, as illustrated in Figure 2.

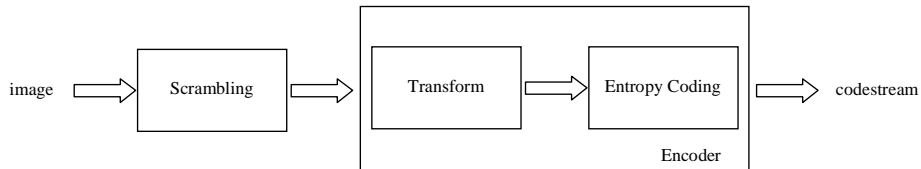


Figure 2 – Image-domain scrambling.

This can be achieved by randomly flipping the most significant bit plane of the pixels belonging to the region to be scrambled using a pseudo-random number generator (PRNG), as shown in Figure 3.

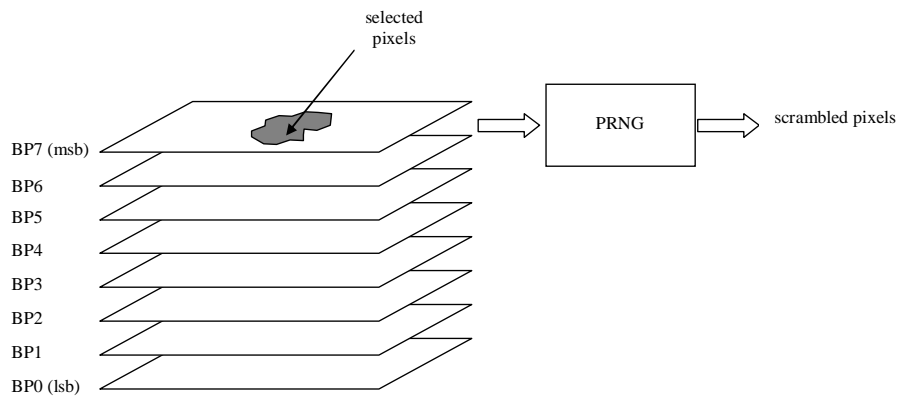


Figure 3 – Bit plane scrambling: most significant bit plane is scrambled.

This approach has the advantage of being very simple and independent from the encoding scheme subsequently used. However, it has the disadvantage of introducing noise in the image prior to coding, possibly leading to lower coding performance.

Note that the same effect could be achieved to some extent by masking the pixels corresponding to the region of interest (e.g. replacing them by a given 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, which may be a desirable feature in some applications. In addition, the masking approach provides an all-or-nothing solution without flexibility to control the amount of distortion introduced.

2.3. Transform-Domain Scrambling

A second approach is to apply scrambling during encoding, as shown in Figure 4. Scrambling is taking place after the DCT or wavelet transform and before entropy coding. More specifically, we propose to randomly flip the sign of transform coefficients corresponding to the region 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.

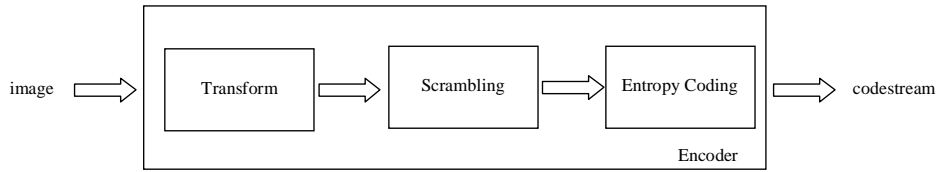


Figure 4 – Transform-domain scrambling.

2.3.1. MPEG-4

In the case of MPEG-4, each frame is subdivided in 16x16 MacroBlocks (MB). Each MB is composed of four 8x8 luminance blocks and two 8x8 chrominance blocks. The DCT is performed on these 8x8 blocks, resulting in 64 DCT coefficients: one DC and 63 AC coefficients.

We first identify all the blocks corresponding to the region to be scrambled. For these blocks, all 63 AC coefficients are scrambled as illustrated in Figure 5. A PRNG is then used to randomly inverse their sign. Note that it would be possible to scramble fewer AC coefficients in order to obtain a lighter scrambling.

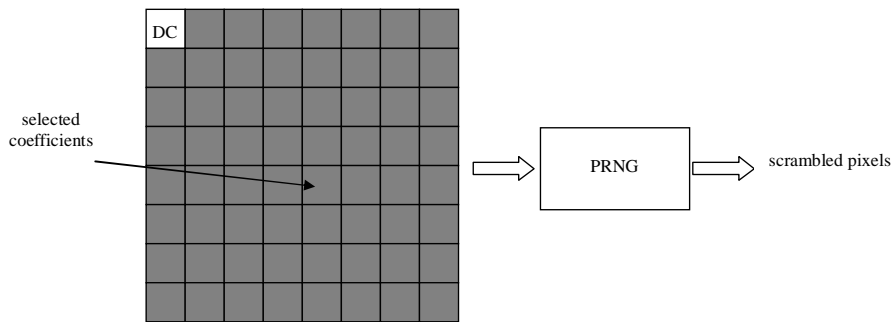


Figure 5 – 8x8 DCT block scrambling: all 63 AC coefficients are scrambled.

Straightforwardly, in this case, the shape of the scrambled region is restricted to match the 8x8 DCT blocks boundaries.

It can also be pointed out that the same technique could be used for other DCT-based schemes, such as MPEG-4 Advanced Video Coding (AVC) / H.264 or Motion JPEG.

2.3.2. Motion JPEG 2000

The technique is similar in the case of Motion JPEG 2000. Wavelet coefficients belonging to the AC subbands and corresponding to the region to be scrambled have their sign randomly flipped, as shown in Figure 6 for an image decomposed with 3 resolution levels. Scrambling coefficients in all AC subbands, i.e. levels 1, 2 and 3, results in a strong scrambling. Subsequently, as previously a PRNG is used to randomly inverse the sign of the corresponding coefficients. The amount of scrambling can also be decreased by restricting the scrambling to fewer resolution levels.

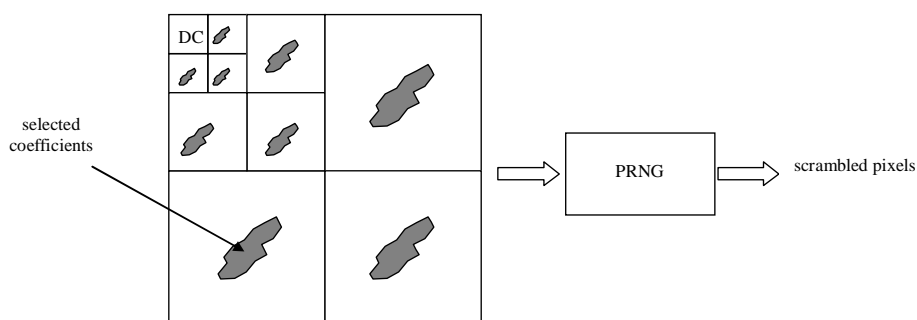


Figure 6 – Wavelet scrambling: coefficients in subbands from levels 1, 2 and 3 are scrambled.

Unlike the MPEG-4 case, with Motion JPEG 2000 the scrambled region can have an arbitrary shape.

2.4. Codestream-Domain Scrambling

In the third approach, scrambling is applied after encoding, as illustrated in Figure 7. More specifically, the compressed codestream is directly scrambled. Again, this can be efficiently done by pseudo-randomly flipping bits in the stream.

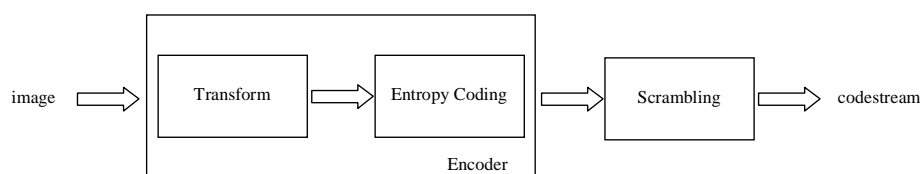


Figure 7 – Codestream-domain scrambling.

One of the drawbacks of this approach is that the codestream has to be parsed in order to identify which parts correspond to the region to be scrambled. Furthermore, another severe drawback is that the scrambled codestream may crash the decoder. Finally, it may be difficult to adjust the strength of the scrambling. For these reasons, we will not investigate further this approach.

2.5. Pseudo-Random Number Generator (PRNG)

The above scrambling techniques rely on a PRNG driven by a seed value. In our implementation, the SHA1PRNG algorithm [12] with a 64-bit seed is used. In order to improve the security of the system, the seed can be frequently changed. Note that other PRNG could be used as well.

3. RESULTS

In this section, we present results showing the effectiveness of the proposed technique to provide anonymous video communications. We also compare image-domain and transform-domain scrambling in conjunction with MPEG-4 and Motion JPEG 2000.

3.1. Scrambling for Anonymity

In this experiment, the Miss America video sequence in CIF format is used. We consider two segmentation masks. The first one is obtained by the neural networks face detection from [4], and the second one is generated by the machine learning skin detection from [6].

Figure 8 compares image-domain and transform-domain scrambling in the case of MPEG-4 encoded at 1 Mb/s. Figure 9 shows similar results for Motion JPEG 2000 at 2 Mb/s. Whereas the scrambled region is strongly masked in the case of image-domain scrambling, transform-domain scrambling is achieving sufficient concealment for the target application.

Note that this video sequence has very low frequency content, which is not favorable for transform-domain scrambling. Nevertheless, both approaches are effective to preserve anonymity.

3.2. Coding Efficiency

We now study the performance of the image-domain and transform-domain scrambling techniques in terms of coding efficiency. More specifically, we compare the quality of the background (i.e. the unscrambled region).

From Figure 8 and Figure 9, it can be observed that the visual quality of the background is significantly degraded in the case of image-domain scrambling when compared to transform-domain scrambling. This observation can be easily explained by the fact that for image-domain scrambling many bits are wasted encoding the noise in the foreground.

We now consider results expressed in terms of PSNR calculated on the background. For this study, we use three additional test sequences: Hall Monitor, Road and Surveillance, along with ground-truth background/foreground segmentation masks. As a reference, we also consider the case when no scrambling is performed. Results are given in Table 1 and Table 2 for MPEG-4 and Motion JPEG 2000 respectively. They show that image-domain scrambling leads to very significant background PSNR decrease. Conversely, transform-domain scrambling achieves a quality of the background very close to the case without scrambling.

Sequence	no scrambling	image-domain scrambling	transform-domain scrambling
Miss America (skin detection mask)	42.36	39.11	41.41
Miss America (face detection mask)	42.07	38.54	41.41
Hall Monitor	40.00	37.20	39.52
Road	38.38	34.35	37.37
Surveillance	40.99	36.18	39.90

Table 1 – Background PSNR: image-domain versus transform-domain scrambling for MPEG-4 at 1 Mb/s.

Sequence	no scrambling	image-domain scrambling	transform-domain scrambling
Miss America (skin detection mask)	41.77	33.06	40.58
Miss America (face detection mask)	41.43	36.91	40.22
Hall Monitor	32.75	31.00	32.33
Road	35.64	30.60	34.95
Surveillance	34.96	32.31	34.67

Table 2 – Background PSNR: image-domain versus transform-domain scrambling for Motion JPEG 2000 at 2 Mb/s.

4. CONCLUSIONS

In this paper, we presented a system to identify and scramble regions corresponding to faces in a video sequence. We considered image-domain and transform-domain scrambling approaches. In the first case, the most significant bit plane is randomly inverted prior to coding. In the second case, the method is flipping the sign of transform coefficients during encoding. Simulation results show that both approaches can successfully be applied to hide information in a region of interest in the scene. Furthermore, for transform-domain scrambling this is achieved without decreasing coding performance. We have shown that the proposed system can be used for anonymous visual communications, and in particular for anonymous Internet video chat.

ACKNOWLEDGEMENT

EPFL's contribution to this work was partially supported by the European Network of Excellence VISNET <http://www.visnet-noe.org> (IST Contract 506946) funded under the European Commission IST 6th Framework Program.

REFERENCES

- [1] <http://messenger.yahoo.com/>
- [2] <http://messenger.msn.com/>
- [3] 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.
- [4] H.A. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection", IEEE Trans. On PAMI, vol. 20, no. 1, pp. 23-38, 1998.
- [5] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", in IEEE Proc. CVPR, Hawaii, Dec. 2001.
- [6] M. Jones and J. Rehg, "Statistical Color Models with Applications to Skin Detection", TR-98-11, CRL, Compaq Computer Corp., Dec. 1998.
- [7] A. Albiol, L. Torres, Ed. Delp. "Optimum Color Spaces for Skin Detection", in IEEE Proc. Inter. Conf. on Image Proc., Thessaloniki, Greece, October 2001.
- [8] T. Ebrahimi and F. Pereira, "The MPEG-4 Book", Prentice Hall, 2002.
- [9] 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.
- [10] D. Taubman and M. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice", Kluwer Academic Publishers, 2002.
- [11] P. Salembier and J. Serra. Flat zones filtering, connected operators and filters by reconstruction. IEEE Transactions on Image Processing, vol. 3, no. 8, pp. 1153-1160, August 1995.
- [12] <http://java.sun.com/j2se/1.4.2/docs/guide/security/CryptoSpec.html>, Java Cryptography Architecture API Specification and reference.

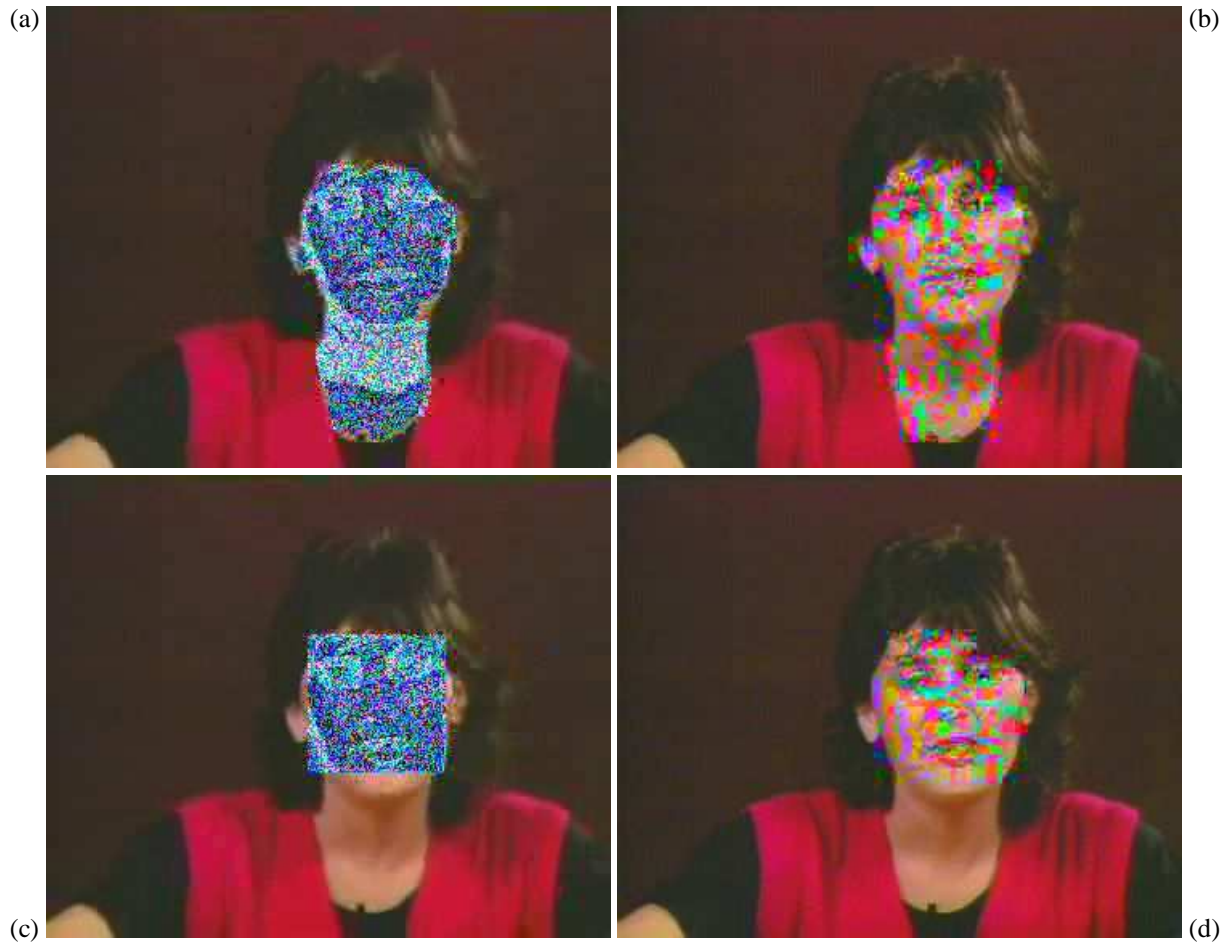


Figure 8 – Image-domain versus transform-domain scrambling for MPEG-4 at 1 Mb/s
(a) skin detection mask and image-domain scrambling
(b) skin detection mask and transform-domain scrambling
(c) face detection mask and image-domain scrambling
(d) face detection mask and transform-domain scrambling.



Figure 9 – Image-domain versus transform-domain scrambling for Motion JPEG 2000 at 2 Mb/s
(a) skin detection mask and image-domain scrambling
(b) skin detection mask and transform-domain scrambling
(c) face detection mask and image-domain scrambling
(d) face detection mask and transform-domain scrambling.