H.264/AVC VIDEO SCRAMBLING FOR PRIVACY PROTECTION

Frederic Dufaux and Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Emitall Surveillance SA, CH-1820 Montreux, Switzerland

ABSTRACT

In this paper, we address the problem of privacy in video surveillance systems. More specifically, we consider the case of H.264/AVC which is the state-of-the-art in video coding. We assume that Regions of Interest (ROI), containing privacy-sensitive information, have been identified. The content of these regions are then concealed using scrambling. More specifically, we introduce two region-based scrambling techniques. The first one pseudo-randomly flips the sign of transform coefficients during encoding. The second one is performing a pseudo-random permutation of transform coefficients in a block. The Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC is exploited to discriminate between the ROI which are scrambled and the background which remains clear. Experimental results show that both techniques are able to effectively hide private information in ROI, while the scene remains comprehensible. Furthermore, the loss in coding efficiency stays small, whereas the required additional computational complexity is negligible.

1. INTRODUCTION

The issue of personal privacy is increasingly becoming prominent with the widespread use of large video surveillance systems. While the deployment of video surveillance systems is justified by the perception of insecurity due to terrorist threats and high criminality rate, the rightful fear of privacy invasion is turning into a significant concern. In this paper, we attempt to reconcile on the one hand the need for video surveillance and on the other hand the concern of privacy protection.

Previous works addressing the topic of privacy protection have previously reported. The system in [1] is based on an object-based representation of the scene. Basically, an altered rendering of the video is produced where some objects are masked out depending on the user authorizations, preventing the transmission of privacy-sensitive objects. The issue of automatic face recognition techniques to identify people is discussed in [2]. A technique is proposed to de-identify faces such that they can no longer be reliably recognized but many facial characteristics are preserved. Privacy filters, expressed using a privacy grammar, are applied on incoming video sensor data in [3], preventing access to privacy-sensitive information.

In [4], wavelet-domain and codestream-domain conditional access control techniques are proposed for JPEG 2000 to scramble code-blocks corresponding to Regions of Interest (ROI) containing for instance people or faces. In [5], it is extended to a region-based transform-domain scrambling method applicable to Motion JPEG 2000 or MPEG-4. More specifically, AC transform coefficients corresponding to ROI are scrambled by pseudo-randomly inverting their signs, concealing any privacy-sensitive data. Similarly, encryption is used to conceal faces in [6]. A secret encryption key is required in order to invert the process, thus guaranteeing privacy protection.

Finally, [7] and [8] propose MPEG-7 cameras which feature an embedded processor to perform video analysis. The camera does not output an actual video stream, but rather an MPEG-7 compliant descriptor data stream sufficient for video monitoring and surveillance.

In this paper, we propose an extension of our earlier work in [5] to the state-of-the-art H-264/AVC standard [9]. The approach scrambles ROI while leaving the background intact, with the resulting scrambled stream still complying with the standard syntax. We present two scrambling approaches. The first one pseudo-randomly inverts the sign of AC transform coefficients of blocks belonging to ROI similarly to [5]. The second one applies a pseudo-random permutation of the AC transform coefficients in blocks corresponding to ROI. The pseudo-random process is driven by one or more seed values which are encrypted and transmitted to the decoder either as private data or on a separate channel. Decoders in possession of the secret encryption key can reverse the scrambling process and recover the truthful scene. Conversely, other decoders obtain a video sequences where ROI have severe noise, concealing privacy-sensitive information. To discriminate between scrambled and unscrambled regions, we exploit the Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC to define two slice groups composed of MacroBlocks (MB) corresponding to the foreground and background respectively.

The proposed approach to privacy protection provides with a number of advantages. The same scrambled stream is transmitted to all clients independently from their access rights. The scrambling is confined to ROI, whereas the background remains unaltered. Finally, it has a small impact in terms of coding efficiency, and requires a low computational complexity.

This paper is structured as follow. In Sec. 2, we discuss the requirements and specificities of the problem addressed. We introduce two scrambling schemes for H.264/AVC in Sec. 3. To evaluate performance, experimental results are presented and discussed in Sec. 4. Finally, we draw some conclusions in Sec. 5.

2. VIDEO SCRAMBLING FOR PRIVACY

In this section, we discuss the use of scrambling for privacy protection. We assume that the video surveillance system is performing video analysis in order to identify ROI containing privacy-sensitive information.

The following requirements are identified for an effective solution:

- After scrambling, the scene should remain comprehensible, but privacy-sensitive objects cannot be identified
- The process should be fully reversible for those in possession of a secret key
- The scrambled content should be secured
- The syntax of the scrambled stream should remain standard compliant
- The scrambling should not entail lower coding performance
- The scrambling should not result in a significant complexity increase

While earlier research works has been reported on scrambling techniques for H.264/AVC [10][11][12], they have addressed the problem of conditional access control in a traditional way, proposing techniques where the whole frame is completely distorted and it is impossible to make out the scene. The problem we address in this paper has very different requirements which demands different approaches.

3. H.264/AVC SCRAMBLING

In this section, we consider the problem of concealing ROI data in a video sequence compressed using the H.264/AVC video coding standard [9] and more precisely the baseline profile.

H.264/AVC is based on a motion compensated block-based DCT-like transform [9]. Among many design innovations, H.264 supports spatial intra prediction on top of the traditional temporal motion compensated inter prediction. It also features a 4x4 transform which allows for a better representation the video signals thanks to localized adaptation.

We propose a transform-domain scrambling approach. More specifically, the scrambling can effectively be applied on the quantized transform coefficients, and outside of the prediction loop, as illustrated in Figure 1 (a). Straightforwardly, a benefit of this approach is that it guarantees to preserves the standard compliant syntax of the compressed stream. Note that in the encoder, unscrambled data is used in the prediction loop.

At the decoder side, authorized users perform unscrambling of the coefficients prior to the prediction loop, as depicted in Figure 1 (b). Note that these users will use unscrambled data in the prediction loop, identical to the one used in the encoder. Therefore, in this design the scrambling process can be made fully reversible. Conversely, unauthorized users are still able to correctly decode the video stream, except for the scrambled coefficients. However, in this case scrambled data is used in the prediction loop, hence introducing a drift.

One of the challenges is to be able to correctly decode both foreground and background for unauthorized decoders due to the drift in the intra/inter prediction loop. Note that this issue has not been addressed in previous conditional access control techniques, as it only happen when applying scrambling on given regions.

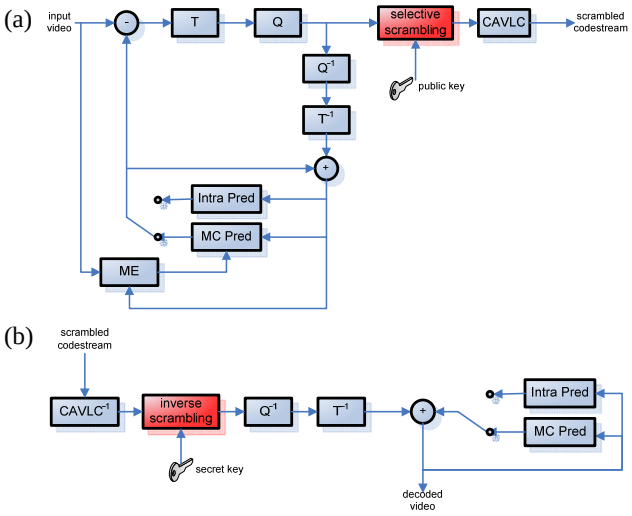

Figure 1: Transform-domain scrambling in H.264/AVC: (a) encoder and (b) decoder.

For this purpose, we propose to exploit the Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC essentially developed for robustness to transmission errors. More specifically, we use the Slice Group Map Type 6 which specifies an explicit MacroBlocks (MB) assignment of slice group. We define two slice groups composed of MB corresponding to the foreground and background respectively. By definition, MB in I slice can be decoded using intra prediction only from decoded samples within the same slice. Similarly, P slice MB can be decoded using intra prediction from decoded samples within the same slice or inter prediction from previously-decoded reference frames.

In other words, by using FMO we ensure that background MB will not use scrambled foreground MB for spatial intra prediction. Nevertheless, the problem of temporal inter prediction remains. This can be solved by merely modifying the mode selection in the encoder to force some MB to be coded in intra mode to prevent the use of scrambled data for inter prediction.

An added benefit of FMO is that the shape of the scrambled region is conveyed to the decoder which needs this information for unscrambling. Obviously, in this case the foreground/background segmentation mask is restricted to match the 16x16 MB boundaries.

3.1. Scrambling Process

The scrambling process is based on a Pseudo Random Number Generator (PRNG) initialized by a seed value. Multiple seeds can be used to strengthen the security. The seed values are then encrypted, preferably using asymmetric encryption, and transmitted to the decoder either as private data or on a separate channel. Authorized users, in possession of the secret encryption key, can recover the seed values and hence reproduce the same pseudo-random sequence to descramble the coefficients.

3.1.1. Random sign inversion

As previously stated, the scrambling process should not have a negative impact on coding efficiency. A natural choice is therefore to apply scrambling to the AC coefficients. Furthermore, whereas the amplitude of AC coefficients is correlated, their signs are not.

Per consequent, we propose to scramble the quantized AC coefficients of each 4x4 block of the MB in the foreground slice group by pseudo-randomly flipping their sign. Defining the vector of quantized AC coefficients *qACcoeff*[*i*] with *i*=0..15, the scrambling consists in performing the following operation for each *i*

$$qACcoeff[i] = \begin{cases} -qACcoeff[i] & if\ random\_bit = 1 \\ +qACcoeff[i] & otherwise \end{cases}$$

Straightforwardly, this technique requires negligible computational complexity.

3.1.2. Random permutation

Alternatively, we propose a second scrambling method applying a random permutation to rearrange the order of AC coefficients in 4x4 blocks corresponding to MB in the foreground slice. The random permutation can be expressed as follow

$$\begin{pmatrix} 0 & 1 & \dots & 14 & 15 \\ x_0 & x_1 & \dots & x_{14} & x_{15} \end{pmatrix}$$

We use the Knuth shuffle to generate a permutation of *n* items with uniform random distribution. Starting from the identity permutation, we scan through each position *i* from 0 to 14, and swap the element currently at position *i* with the element at an arbitrarily chosen positions from *i* through 15.

4. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed scrambling techniques. The Hall Monitor and Road video test sequences in CIF format are used, with ground truth segmentation masks. Experiments have been performed with the JM13.2 reference software [13].

4.1. Privacy Protection

First, we assess the capability of the proposed scrambling to hide information in ROI. Figure 2 shows results for the random sign inversion and the random permutation methods. It can be observed that both approaches are efficient at concealing the content of the ROI so that people can no longer be identified. It can also be noted that despite the scrambling the scene can still be correctly comprehended.

4.2. Coding Efficiency

In this section, we evaluate the impact of the two proposed scrambling methods on coding efficiency when compared to regular H.264/AVC. The rate-distortion performances obtained with the random sign inversion and random permutation are given in Figure 3.

Figure 2: Scrambling for "Hall Monitor":
(a) random sign inversion, (b) random permutation.

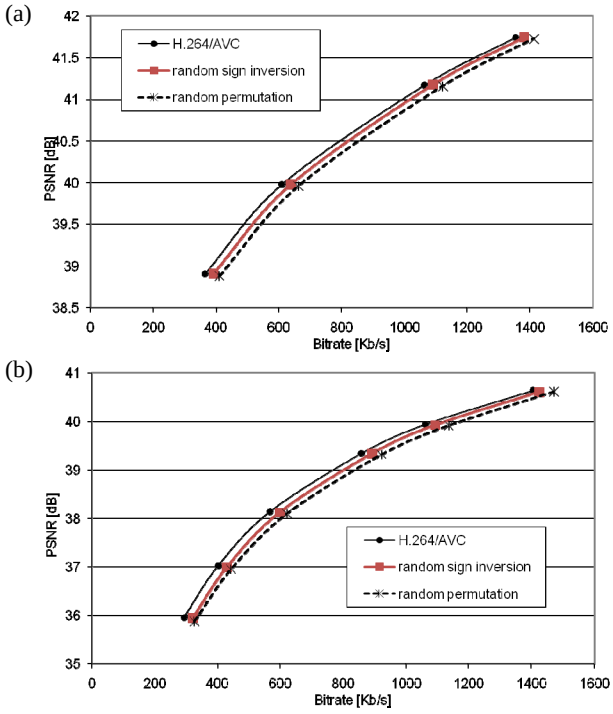

Figure 3: Rate distortion coding efficiency comparison:
(a) Hall Monitor, (b) Road.

It can be observed that the proposed scrambling has minimal impact on coding efficiency. For the random sign inversion scrambling, the bit rate increase is of the order of 1% at high rates and 8% at low rates. The random permutation

scrambling produces a slightly larger penalty with a rate increase of 4% at the high end and 11% at the low end.

The Slice Group Map Type 6 signaling for FMO requires 1 bit per MB, hence resulting in an overhead of approximately 12 Kb/s for CIF accounting for a significant percentage of the penalty at low rates. The remaining of the bit rate loss is due to the alteration of coefficient statistics as well as the modification of MB mode decision. The random permutation scrambling results in a slightly larger penalty due to the coefficient re-ordering which is consuming more bits than the simple sign inversion.

4.3. Security

4.3.1. Brute-force attack

We now consider the security of the two proposed scrambling technique against a brute-force attack which exhaustively tries all combinations reversing the scrambling.

For the random sign inversion scheme, the luminance component of a CIF frame, i.e. 352x288 pixels, generates 101376 coefficients. Further assuming that the ROI covers 5% of the image and is known to the attacker, the number of corresponding coefficients is 5068. Finally, supposing that only 5 % of those are non-zero, an attacker has to try reversing the signs of 253 coefficients, representing 2^{253} combinations for each frame.

For the random permutation approach, a set of n elements has $n!$ possible permutations. In other words, for each 4x4 scrambled block, there are 16! permutations. A CIF frame has 6336 4x4 blocks. Assuming that 5% are corresponding to the ROI, 316 blocks are scrambled. It results in $(16!)^{316}$ possibilities, offering therefore an even stronger protection than the sign inversion approach.

4.3.2. Error concealment attack

Protected multimedia content can be vulnerable to error concealment attacks which attempt to conceal scrambled/encrypted data. In this section, we assess the security of our approaches against a simple error concealment attack where scrambled AC coefficients are simply set to 0 at the decoder. We observe in Figure 4 that this attack is inefficient.

Figure 4: Attack by setting to 0 scrambled AC coefficients.

Furthermore, more sophisticated error concealment is unlikely to produce better results, as by definition the scrambled foreground objects have different characteristics when compared to the background and therefore cannot be extrapolated from the latter.

5. CONCLUSIONS

In this paper, we have described a technique to address the issue of privacy in video surveillance. More specifically, we have introduced two scrambling techniques for H.264/AVC. Simulation results show that the proposed schemes are successful at concealing privacy-sensitive information while leaving the scene comprehensible. The resulting scrambled streams are still standard compliant. Simulations results show that this is obtained with a small impact on coding efficiency and provides good security.

ACKNOWLEDGEMENT

This work was partially supported by the European Network of Excellence VISNET II (http://www.visnet-noe.org) funded under the European Commission IST 6th Framework Program.

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

[3] 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.

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

[5] F. Dufaux and T. Ebrahimi, “Scrambling for Video Surveillance with Privacy”, in IEEE Proc. Workshop on Privacy Research In Vision, New York, NY, June 2006.

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

[7] F. Dufaux and T. Ebrahimi, “Recent Advances in MPEG-7 Cameras”, in SPIE Proc. Applications of Digital Image Processing XXIX, San Diego, CA, August 2006

[8] http://www.eptascape.com/products/eptacam.htm

[9] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003.

[10] J. Ahn, H.J. Shim, B. Jeon and I. Choi, “Digital Video Scrambling Method Using Intra Prediction Mode”, in Pacific Rim Conference on Multimedia, Tokyo, Japan, 2004.

[11] H.J. Lee, “Low complexity controllable scrambler / descrambler for H.264/AVC in compressed domain”, in Proceedings of the 14th annual ACM international conference, Santa Barbara, CA, Oct. 2006.

[12] J. Wang, Y. Fan, T. Ikenaga and S. Goto, , “A partial scramble scheme for H.264 video”, in Int. Conf. ASIC’2007, Oct. 2007.

[13] http://iphome.hhi.de/suehring/tml/