Data Hiding in H.264 Video for Lossless Reconstruction of Region of Interest

Peter Meuel, Marc Chaumont, William Puech

DATA HIDING IN H.264 VIDEO FOR LOSSLESS RECONSTRUCTION OF REGION OF INTEREST

Peter MEUEL, Marc CHAUMONT, William PUECH

Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier
161 rue Ada, 34090, Montpellier, France
email: {peter.meuel, marc.chaumont, william.puech}@lirmm.fr

ABSTRACT

In this paper, we propose a method to protect faces in video-surveillance scenes. Our method deletes any visible information of faces in a video and uses a data-hiding technique to embed information in the video that allows further reconstruction of the faces if needed. When the entire information needed for reconstruction is embedded, we obtain infinite PSNR for the face regions.

1. INTRODUCTION

Nowadays, video surveillance is growing at an exponential rate. Low costs and high reliability relatively to human agents make video-surveillance the best choice for security monitoring. But two problems appear: privacy and transmission.

The exponential use of video-surveillance in big cities has created a new situation: video-camera webs. Having total access to all the video-camera flows in a city allows tracking of anybody. In order to preserve privacy, a method that forbid immediate people identification by unauthorized person is needed. Of course, authorized persons should be able to identify the people in the video.

Rodrigues *et al.* [1] proposed a method based on partial encryption of the Region of Interest (ROI) for DCT coefficients. The method is fast, effective and reversible but it is based on JPEG image sequence. Dufaux and Ebrahimi [2] proposed a ROI-scrambling method based on an encryption key. Other works on video privacy [3] consider only destructive methods, i.e. methods that do not permit further identification of the characters of a scene. It is obviously non appropriate for video-surveillance as it constraints to produce two accesses: a first one with alterated ROI and a second one with a clear ROI to preserve the possibility of identification.

The second problem is that transmission is often not high enough to allow a high visual quality on the entire image. The optimisation of the available bandwith is a solution to achieve better visual quality. In a video, a ROI is defined by being a part of the image that contains more valuable information than the rest of the image. At the encoding process, more bits will be allocated to the ROI coding to improve its visual quality. Our method makes those extra bits available to the ROI coding by a data hiding method. We will use bits of information of non-interest region to store extra informations about the ROI, after the encoding process.

The proposed method is not a robust watermarking method. It is designed to improve the visual quality of the ROI while minimizing the increase of the size of the marked video. Reference on digital watermarking [4] and state-of-the-art of video watermarking [5] are referenced.

Lee *et al.*[6] recently proposed a similar approach to our method using the H.264 capabilities. After detection of the ROI, their method assigns a lower Quantization Parameter (QP) to the ROI. Paired with a bitrate control mechanism, this method effectively allow optimisation of the bitrate but no protection of the ROI is achieved as any H.264 standard decoder will display the ROI. It can be seen as object coding [7]. Agrafiotis *et al.* [8] worked on a similar problem with the H.264 format. There, the ROI detection is an input of the user. This method uses a two-layer rate controls: GOP-layer rate control and picture-layer rate control to maintain a constant bitrate. As the previous method, the ROI is visible without any constraint. Concerning data embedding in H.264 video, Mobasseri *et al.* [9] recently studied information embedding into the CAVLC bitstream. This method does not affect the visual aspect of the video and does not increase its size but the payload is too small to fulfill the need of our method.

2. THE PROPOSED METHOD

In this section we present the description of the proposed method. Figure 1 presents the global scheme of our method and we explain more deeply each part in the following.

The first step is the detection of the face in the current frame F_n. The detected face is shaped as a rectangular region and defined as the ROI. We then construct the message to embed. This message will consist of the ROI luminance values of the original video. Thoses values are predictive coded and compress with a Huffman coding. We obtain a sequence of bits wich we finally embed in the non-zero quantified DCT coefficients of the coded frame F_n during the encoding process through the HHI JM10 codec [10].

2.1 ROI Detection

We now explain the method used for the ROI detection. Our approach has already been studied in other work [11]. The first step is to categorize every pixels whom fulfill the following condition as *skin pixel* using the following formula:

$$\sqrt{(P_u - \text{ref}_u)^2 + (P_v - ref_v)^2} < d$$

where P_u and P_v are respectively the U (Cb) and the V (Cr) component of the current pixel and d is the threshold determining if the current pixel is marked as *skin pixel*. This detection method is based on the fact that two different skin colors differ mainly by their luminance rather than their chrominance.

This detection method is fast and effective. It allows real-time detection and detects every *skin pixel*. The only draw-
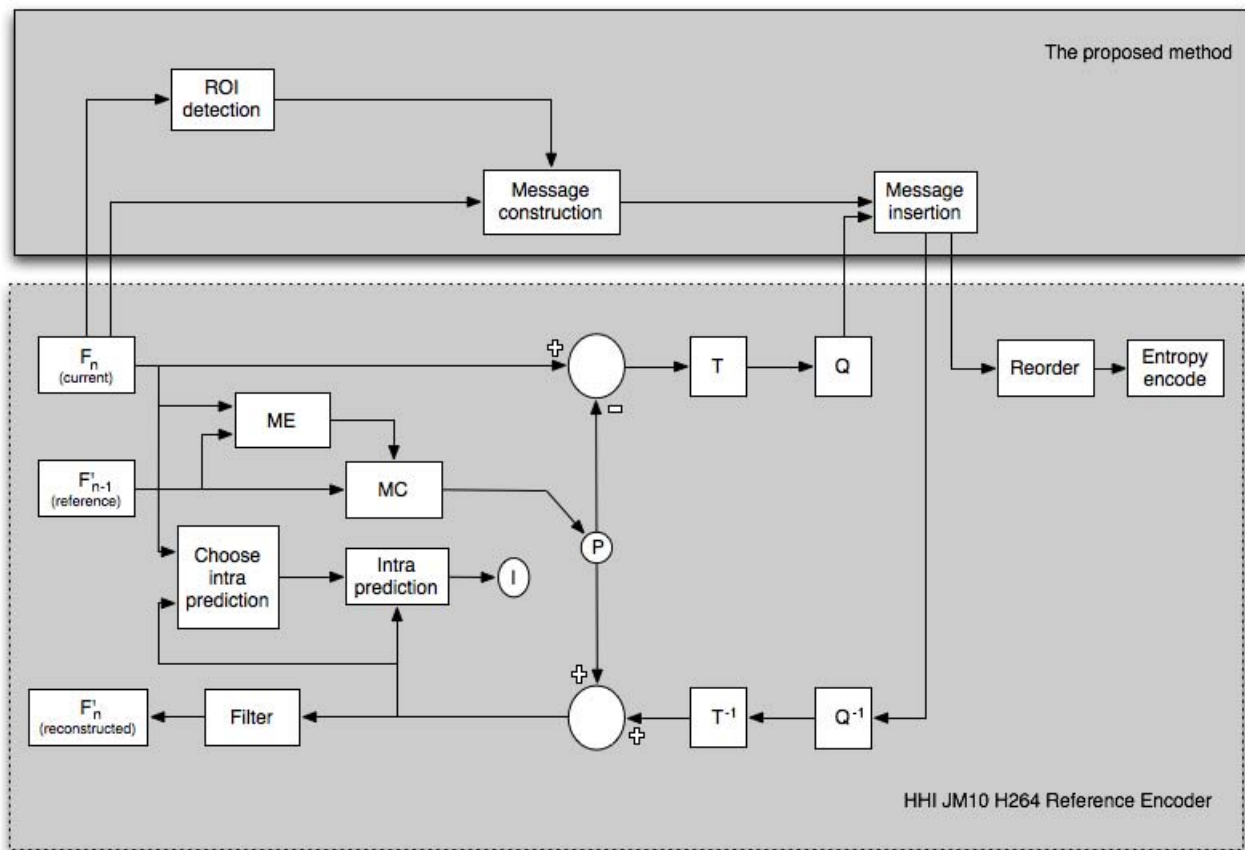
Figure 1: Overview of the proposed method

back is that unwanted pixels are also marked as *skin pixel*. This problem is solved by successives morphological opening and closing operations. When unwanted pixels are no more marked as *skin pixel*, we can format the ROI as a rectangular shape.

As it was implemented, our detection method will found only one ROI. It is only a matter of simplification. As the ROI detection is not the main subject of this paper, we didn't go further. A simple and effective way to detect multiples ROI will be to consider groups of marked pixels superior to a determined size.

2.2 Message construction

To facilitate the reconstruction of the ROI, we put as the four first values of our message the position, the width and the height of the ROI (x,y,w,h above the *luminance values* block in Figure 2). As the goal of the proposed method is to achieve high quality reconstruction in the ROI, we choosed to use the ROI luminance values of the original video instead of the differences of luminance values between the original video and the H.264 encoded one.

Embedding the differences between original values and decoded ones will modify the decoded values. Thus, the embedded differences will not correspond. Construction of the message to embed considering its insertion is hardly achievable as both rely on the other. That is why we choosed to work directly with orignal values.
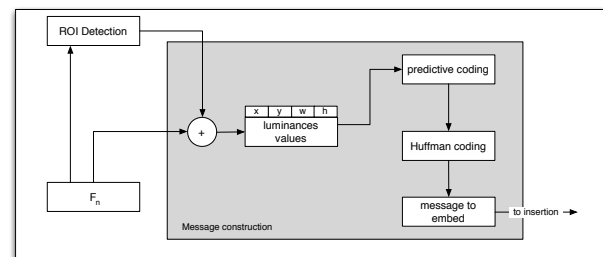

Figure 2: Message construction

As shown in the Figure 2, once the ROI is detected, we encode the ROI luminance value of the original video with a predictive coding. The result is compressed using a Huffman coding. We obtain then the message to embed.

2.3 Message embedding

The embedding method is a substitution of the two Least-Significant Bits (LSB) of the non-zero DCT coefficients of the current frame by the bits of the message. We consider data embedding in the two LSB because after the quantization step of H.264, the number of non-zero DCT coefficients available for insertion is limited. Of course, this method does accentuate distorsion but does not affect our ROI reconstruc-

tion method.

We choose to embed our message into the quantified DCT coefficients even if the visual distortion is more visible than if we embedded into the luminance values prior to H.264 encoding. Embedding into the luminance values also offers a larger payload but it is not possible as the quantization step will suppress many informations in a way we can not control.

In order to increase the protection of the ROI, during the encoding process we specified that the ROI inter prediction must be done in Skip mode, i.e. the decoder predicts the motion vector of the ROI macroblocks, without any indication from the encoder. So an estimation of the ROI based solely on the motion vectors will not be reliable as the Skip mode does not guarantee any fidelity to the original macroblock.

To achieve a maximal protection, we set the ROI macroblock DC coefficients to 1 and all the ROI macroblock AC coefficients to 0. So at decoding, no information can be used to predict the ROI original values. All the information concerning the ROI is the one inserted with our method.

2.4 ROI Reconstruction

The reconstruction is accomplished in three steps. We first extract the embedded data of the current frame. This step is easily performed as it consists in reading the two LSB of each non-zero DCT coefficients. We decode those data previously encoded with a predictive coding. At last, as we have the position, the width and the height of the ROI, we organised the extracted data as a $w \times h$ image and we simply substitute the ROI with that image in the current decoded frame.

In order to achieve infinite ROI PSNR, we desactivate the deblocking filter during the encoding and decoding process. The deblocking filter is a great feature of H.264 as it improves the overall visual quality by smoothing the usual blocking aspect due to the DCT. However, we desactivate the deblocking filter because as we modify the quantified DCT coefficients, it affects the decoded values and the deblocking filter effects. Thus, desactivate the deblocking filter ensures infinite ROI PSNR when the message is totally embedded.

So we clearly see the benefit of insertion into the two LSB of the quantified DCT coefficients: we are able to retrieve 100% of the embedded data. So if we can embed the whole message at encoding, we will extract the exactly same message at decoding and after the ROI reconstruction step, the ROI of the decoded video will be exactly the same as the ROI of the original video and thus, we will obtain an infinite PSNR on the ROI.

3. RESULTS

The main results were obtained on a self-made movie as the reference movies do not fit the typical video-surveillance situation: fixed video-camera recording people moving along the scene. The video was encoded with only the first frame as an I-frame and the other frames as P-frames. We desactivate the Rate-Distorsion optimisation and the deblocking filter and the entropy coding was CAVLC.

The video used for our experimentation is a fixed-camera scene where only one character walks from the middle backside to the right frontside of the picture.

3.1 Visual quality

Figure 3 shows three snapshots of three versions of the same video. The first snapshots are from the standard H.264

video.The second ones are from the H.264 video marked by our method, with the same bitrate. We can already note that the proposed method strongly affect the visual quality. This is a good thing as it does not alter the observation, i.e. we can still monitor what is happening, but does strongly alter the ROI: the face is nearly unrecognizable. The last snapshots are from the same H.264 video with reconstructed ROI. The constrast with the non-reconstructed video is clearly visible.

As shown on the screenshots, the distortions due to the embedding are very strong but these are not a problem as they don't disturb the global visualisation *(ou understanding?)* of the scene.

3.2 PSNR Measure

The Tables 1 and 2 present the numerical results of our method. Each table presents the mean global PSNR and the mean ROI PSNR of the three previous videos: the normal H.264 video, the marked H.264 video without reconstruction of the ROI and the H.264 with reconstruction of the ROI.

Because we modified our DCT coefficients, the PSNR of the whole image for the marked video is lower than the normal H.264 video. But when we focus on the ROI, our method shows its efficiency as we see a real gain on the ROI PSNR. The mean ROI PSNR only considers the frame whom ROI PSNR is not infinite.

H.264 Video	$PSNR_{All}$ (dB)	$PSNR_{ROI}$ (dB)
standard	48.037	53.776
w/o reconstruction	47.668	53.794
with reconstruction	47.714	60.560

Table 1: Bitrate: 3828 kbits/s

At bitrate 3828 kbits/s (Table 1), we obtain more than 80 frames, over 125, with an infinite ROI PSNR, i.e. complete message embedding. Without considering those frames, we obtain a mean ROI PSNR of 60.560 dB. So even with an incomplete embedding, we can improve the ROI with our reconstruction method. At bitrate of 2800 kbits/s (Table 2), we obtain a third of frames with infinite ROI PSNR. The ROI is still improved even with an incomplete message.

H.264 Video	$PSNR_{All}$ (dB)	$PSNR_{ROI}$ (dB)
standard	44.582	49.424
w/o reconstruction	44.097	49.420
with reconstruction	44.097	56.173

Table 2: Bitrate: 2797 kbits/s

3.3 Payload

Figure 6 presents the frame payload for a bitrate of 3828 kbits/s, wich correspond to Table 1. The frame payload is nearly constant for the entire video but we can see that for the last frames, the payload decrease as the message size increase. This corresponds to a size increase of the ROI in the video (see a_3,b_3,c_3 in Figure 3).
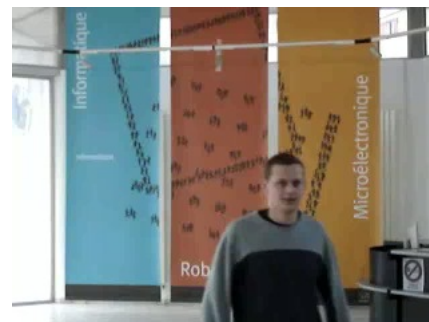
Even with a high bitrate, the payload is not enough for our message. This can be explain by the fact that with a big part of the data of a H.264 video is motion vectors informations. So the part reserved to DCT coefficients is restraint.
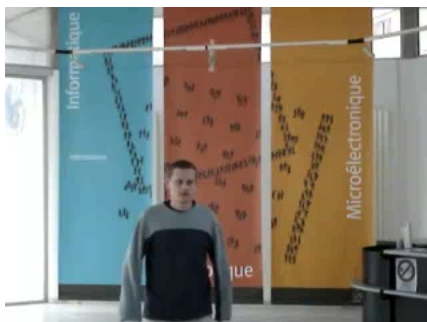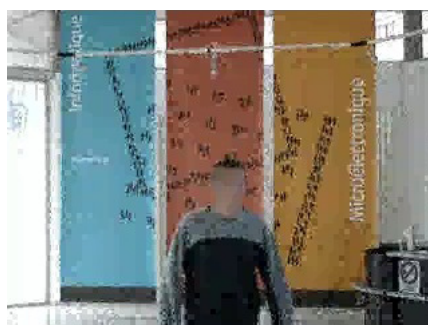
a_1: frame 75 a_2: frame 90 a_3: frame 105
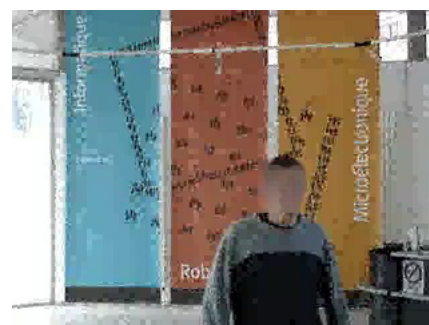
Standard H.264 video

b_1: frame 75 b_2: frame 90 b_3: frame 105

Marked H.264 Video without ROI reconstruction

c_1: frame 75 c_2: frame 90 c_3: frame 105

Marked H.264 Video with ROI reconstruction
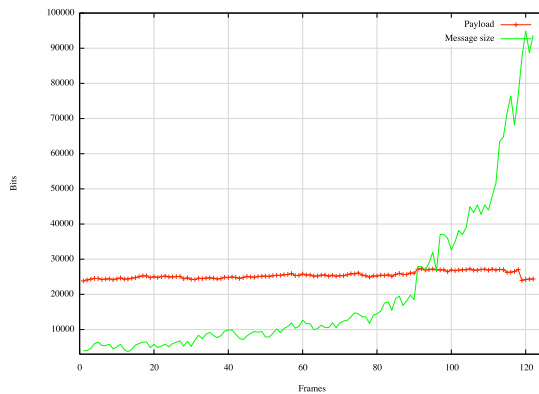
Figure 3: Visual quality comparison

Figure 4: Frame payload for bitrate 3828 kbits/s

When the payload becomes insufficient for a complete embedding of the message, it leads to an incomplete reconstruction of the ROI during the decompression step. Reducing the size of the message is the main solution as the insertion method used is the one offering the most important payload.

4. CONCLUSION

In this paper, we presented a method that automaticaly detects faces as a ROI in a videosurveillance scene, constructs and embeds a message in the video to allow reconstruction of high visual quality in the ROI. We achieved lossless reconstruction (infinite ROI PSNR).

We also modified the HHI JM10 H.264 encoder to minimize the ROI information directly available at the decoding process. The ROI is then secure: direct decoding will not bring any information on the ROI. And by extracting the embedded message with our method, we reconstruct high visual quality in the ROI.

Our next step is to make the embedded message scalable. By applying a DCT to the luminance values prior to predictive encoding and embedding first the low frequencies coefficients, the payload problem will be reduced and we will obtain improved ROI PSNR even on the frames whom can not embed the entire message. We will also take in consideration the H.264 CAVLC coding to propose an insertion method that minimize the size increase.

REFERENCES

[1] J. M. Rodrigues, W. Puech, P. Meuel, J.C. Bajard, and M. Chaumont. Face protection by fast selective encryption in a video. *IET THE CRIME AND SECURITY Conference*, pages 420–425, 2006.

[2] F. Dufaux and T. Ebrahimi. Smart video surveillance system preserving privacy. *Proceedings of SPIE*, 5685:54–63, 2005.

[3] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. *Proceedings of the 2000 ACM Conference on Computer supported Cooperative Work*, pages 1–10, 2000.

[4] J. Bloom I. Cox, M. Miller. *Digital Watermarking*. Morgan Kaufmann Publishers, 2001.

[5] G. Doerr and J.-L. Dugelay. A guide tour of video watermarking. *Signal Processing: Image Communication*, 18:263–282, 2003.

[6] T. C. Thang Y. Ro D. S. Lee, T. Bae. Dynamic fullscalability conversion in scalable video coding. *Multimedia on Mobile Devices 2007 Part of IST/SPIE 19th Annual Symposium, Electronic Imaging*, paper 6507-03, January 2007.

[7] J.S. Shin T. Chen J. Ostermann, E. Jang. Coding of arbitrarily shaped video object in mpeg-4. *International Conference on Image Processing (ICIP'97)*, 1:496, 1997.

[8] D. Agrafiotis, D. R. Bull, N. Canagarajah, and N. Kamnoonwatana. Multiple priority region of interest coding with h.264.

[9] Y. J. Naik Raikar B. G. Mobasseri. Authentication of h.264 streams by direct watermarking of cavlc blocks. *Security, Steganography, and Watermarking of Multimedia Contents IX, Part of IST/SPIE 19th Annual Symposium, Electronic Imaging*, paper 6505-70, February 2007.

[10] S. Wenger J. Lainema B.-M. Jeon Y.-S. Soh T. Stockhammer D. Marpe G. Heising V. George K. Suehring A. M. Tourapis I. Lille-Langoy, R. Sjoberg. Hhi jm10 h.264 codec. *http://iphome.hhi.de/suehring/tml/index.htm*.

[11] D. Chai and K. N. Ngan. Face segmentation using skincolor map in videophone applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9:551–564, 1999.