Watermarking-Based Content Authentication of Motion-JPEG Sequences

Weilin Huang, Anthony T. S. Ho and Vinod Pankajakshan
Department of Computing, Faculty of Engineering and Physical Sciences,
University of Surrey, UK
Email: {wh00006, a.ho, v.pankajakshan}@surrey.ac.uk

Keywords: Content authentication, Digital watermarking Motion-JPEG.

Abstract

Content authentication has become an important issue for surveillance video. This paper presents a watermarking system based on Discrete Cosine Transform (DCT) for Motion-JPEG video authentication. To protect the integrity of the video object, a content-based watermark is embedded into the frames of the video. Robust watermark and semi-fragile watermark are used for temporal authentication and spatial authentication respectively. For analysis and evaluation, the system trades-off the robustness and the visual quality of the watermarked frames. The results show that the authentication rate almost reaches 100% in temporal authentication and is more than 98% in spatial authentication when the PSNR of watermarked frames is 36dB. It can robust to 55% JPEG compression and additive white Gaussian noise with variance of 0.002.

1 Introduction

Surveillance systems have attracted much attention in recent years and are widely used in commercial, law enforcement and military applications. For example, IP or Network cameras are already prevalent in banks, stores, and parking lots. They are also used to measure traffic flow, detect accidents on highways, and even as the evidence of crimes in court. For all these applications, the security of video data becomes an important issue. The video data must be protected from illegitimate distortions. Thus, technical solutions are needed to address the problems associated with data integrity and authentication of origin.

Digital watermarking is a technique which can be used for protecting the integrity of video data by embedding a signal permanently into video data and extracting it later for the purpose of data verification. Content authentication is one of the important applications of video watermarking. According to different applications, it could be divided into two categories: *exact authentication* and *selective authentication*. Exact authentication uses the *fragile* watermarking scheme, which can detect any changes to the bits of the watermarked data. In contrast, selective authentication is used to detect some significant changes which will lead to inauthentic content. The *semi-fragile* watermark scheme is proposed to address the issue of selective authentication. It seeks to verify that the content of the multimedia has not been modified by any predefined set of illegitimate distortions, while allowing modification by legitimate distortions [1]. It is able to resist some degree of legitimate distortions to the watermarked data (e.g. Gaussian noise and JPEG compression) but unable to survive under illegitimate attack (e.g. Cutting and Pasting).

The task of video authentication is to detect whether the video has been tampered with or not. The tamper methods of video can be divided into temporal tampering and spatial tampering. Temporal tampering occurs on a time sequence. It includes adding, dropping and reordering frames. Spatial tampering focuses on the modification of the image, such as replacing, removing and adding the content of the image. There are different levels of tamper authentication. Authentication for temporal tampering can range from video level, scene level, shot level to frame level; spatial tampering can be detected from video level, scene level, shot level, picture level, block level to pixel level [2].

Since most digital videos are used in compressed standard-compliant format, numerous applications of digital watermark in compressed-format videos have been proposed. For example, Hartung *et al*. [3] embedded spread-spectrum watermark into compressed video. Langelaar *et al.* [4] proposed the *differential energy watermark* algorithm, in which enforcing energy difference between the DCT coefficient blocks embeds the watermark bits. A semi-fragile watermarking scheme for the authentication of images as well as MPEG-1/2 videos is proposed in [5]. Most of the previous works, for example, [2]-[7], were focusing on highly compressed video watermarking (e.g. MPEG-2 and H.264/ACV). However, few of them presented their works in low-compressed video like the Motion-JPEG. Motion-JPEG is a low-compressed video encoding technique which is widely used in video surveillance systems, [8]-[9]. Each frame of the Motion-JPEG video is compressed by JPEG standard separately, and frames have all of the information they need stored in them. So the algorithm for encoding and decoding the Motion-JPEG video is simple, which makes the video editing easier. The visual quality of M-JPEG video higher than those of highly compressed video, but it has the disadvantages of high bandwidth requirements for transmission and low resilience to error due to missing frames.

In this paper, we present a DCT-based watermarking system for M-JPEG video authentication. In our authentication system, the tampering is detected in frame level in time sequence and in block level in spatial authentication. Two M-JPEG traffic enforcement surveillance videos are used as example videos to test and analyze the authentication system. The rest of the paper is organized as follows: Section 2 presents the details of the proposed authentication system. The experimental results and evaluation of the system are presented in Section 3. Finally, Section 4 concludes the paper.

2 Proposed authentication system

For surveillance video authentication, the system needs to detect both the spatial and temporal tampering on video data. Semi-fragile watermarking, which is mentioned above, is often used for spatial authentication. For temporal authentication, the watermark is needed to survive under malicious manipulations on frames. So robust watermarking can be adapted for temporal authentication. Embedding a semi-fragile and a robust watermark simultaneously on video frames may increase the perceptual distortion of the watermarked sequence. In the proposed authentication system, the semi-fragile watermark is derived from the robust watermark. So, only one watermark is embedded in each frame of the video sequence. The watermark embedding and detection processes are explained in the following Subsections.

2.1 Watermark embedding process

In M-JPEG, each frame of the video is independly coded using JPEG standard. The watermark embedding is done by modifying selected mid-frequency DCT coefficients in the 8×8 blocks of video frames. The mid-frequency coefficients are chosen as a compromise between the visual quality of the watermarked video and robustness against unintentional content modifications like the frame-wise JPEG compression. The block diagram of the watermark embedding process is shown in Figure 1. The first step in the watermark embedding process is the watermark generation. For each frame of the video, a *frame watermark*, W_f is generated as:

$$W_f = \overbrace{10\cdots11}^{S(20bits)}\overbrace{10\cdots11}^{T(20bits)} \quad (1)$$

where S is the sequence number of the frame and T is the total number of frames in the sequence.

Each bit of the *frame watermark* is embedded in multiple 8×8 blocks of the frame. The 8×8 block in which a particular bit of W_f is embedded is randomly chosen by using a secret key K_1. In each of the 8×8 block, a pair of mid-frequency DCT coefficients, chosen randomly using a secret key K_2, is used for watermark embedding. The purpose of using the secret keys K_1 and K_2 is to increase the security of the proposed scheme. For simplicity, the same set of secret keys is used for watermark embedding in all the frames of the sequence.

Assuming that the video frames are in the RGB format, the watermark is embedded in the R, G and B planes of each frame. The locations of the DCT coefficients modified by the watermark embedding are the same in all the three planes of a frame. So, corresponding to each 8×8 blocks in a frame, six DCT coefficients (two each in the R, G and B planes) are modified by the watermark embedding process. The watermark embedded in the entire frame is the robust watermark, used for detecting temporal tampering. The watermark bits added to each 8×8 block of the frame constitutes the semi-fragile watermark of the block, which is used for detecting any spatial tampering to that block.

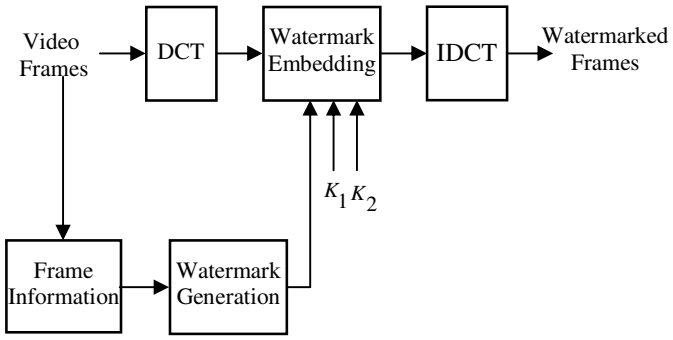

Figure 1: Watermarking embedding process

The algorithm for watermark embedding in each block is given as:

$$X' = \begin{cases} X & \left((W_b = 1 \,\&\, X \geq \lambda) \| (W_b = 0 \,\&\&\, X \leq -\lambda)\right) \\ \lambda & (W_b = 1 \,\&\&\, X < \lambda) \\ -\lambda & (W_b = 0 \,\&\, X > -\lambda) \end{cases} \quad (2)$$

where $\lambda > 0$ is the threshold, X and X' are the DCT coefficients before and after embedding and W_b is the watermark bit. The threshold λ controls the watermark embedding strength, which is used as a trade-off between the robustness of the system and the visual quality of the watermarked video. Finally, the 8×8 IDCT is applied to each block to obtain the watermarked frame.

2.2 Watermark extration and authentication process

The block diagram of the watermark extraction and authentication process is shown in Figure 2. The watermark extraction does not require the host frames or the original watermark. First, the semi-fragile watermark from each block is extracted. The watermarked DCT coefficients corresponding to an 8×8 block are obtained by using the secret key K_2. From each watermarked DCT coefficient, the watermark bit W_b' is extracted as:

$$W_b' = \begin{cases} 1 & (X > \lambda_1) \\ -1 & (X < -\lambda_1) \\ 0 & (|X| \leq \lambda_1) \end{cases} \quad (3)$$

where $\lambda_1 \geq 0$ is a threshold used to increase the robustness of the authentication process. Then the semi-fragile watermark W_s' is extracted as :

$$W'_s = \begin{cases} 1 & \left(\sum W_b' \geq \lambda_2\right) \\ -1 & \left(\sum W_b' \leq -\lambda_2\right) \\ 0 & \left(\left|\sum W_b'\right| < \lambda_2\right) \end{cases} \quad (4)$$

where the summation is over all the six DCT coefficient locations in which the watermark is embedded. If the embedded watermark bit is 1 and the block is untampered, then $\sum W_b' = 6$. Similarly, if the embedded watermark bits is 0 and the block is untampered, then $\sum W_b' = -6$. If the block is tampered, then we will have $-6 < \sum W_b' < 6$. So, on comparing the value of $\sum W_b'$ against a threshold λ_2, where $\lambda_2 \leq 6$, we can detect the tampered blocks. i.e, if $W_s' = 1$ or $W_s' = -1$, we classify the block as untampered and if $W_s' = 0$, the block is detected as tampered. The threshold λ_2 is important in controlling the accuracy of authentication. Using a larger value of λ_2 increases the probability of detection of tampered blocks, but with an increase in the probability of false detection in untampered blocks.

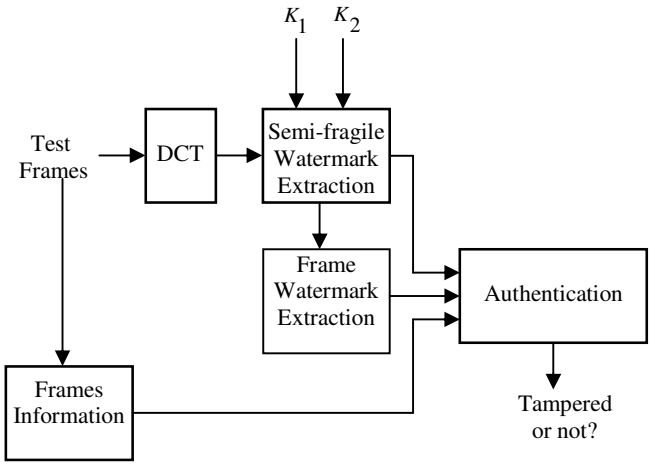

Figure 2: Watermark extraction and authentication process

The final step in the authentication process is the extraction of the frame watermark W_f'. The two components of W_f': the extracted sequence number S' and the extracted total number of frames T', are computed from the semi-fragile watermark W_s'. The i^{th} bit of S' is obtained by:

$$S'(i) = \begin{cases} 1 & \left(\sum W_s' > 0\right) \\ 0 & \left(\sum W_s' < 0\right) \end{cases} \quad (5)$$

where the summation is over all the 8×8 blocks in which the i^{th} bit of S is embedded, which are identified by using the secret key K_1. Since the value of T embedded in all the frames are the same, we use a slightly different method to extract T'. The i^{th} bit of T' is obtained by:

$$T'(i) = \begin{cases} 1 & \left(\sum\sum W_s' > 0\right) \\ 0 & \left(\sum\sum W_s' < 0\right) \end{cases} \quad (6)$$

where the first summation is over the blocks in which the i^{th} bit of T is embedded and the second summation is over the corresponding blocks in different frames. Note that since $W_s' = 0$ for tampered blocks, those blocks will not contribute to computing of S' or T'. Temporal tampering such as reordering, insertion or deletion of frames can be detected by comparing S' and T' with the frame information of the test M-JPEG video.

3 Experimental results

In order to evaluate the performance of the proposed authentication scheme, detailed experiments have been carried out. Two traffic enforcement videos, captured with an IP camera, are used in the experiments. Each video consists of 100 frames of size 240×320 pixels and is M-JPEG encoded at 15 frames per second with Quality Factor (QF) 75. The values of $\lambda_1 = \lambda/2$ and $\lambda_2 = 4$ in Equations (3) and (4), determined empirically, are used in the experiments.

The visual quality of the watermarked sequence and the authentication rates are used as the performance measures. The PSNR of the watermarked sequence is used as the visual quality metric. The spatial authentication rates are defined as:

$$R_{ST} = \frac{N_C}{N_T} \times 100 \quad (7)$$

and

$$R_{SU} = \frac{N_U - N_E}{N_U} \times 100 \quad (8)$$

where N_C is the number of blocks correctly detected as tampered, N_T is the total number of tampered blocks, N_E is the number of blocks wrongly detected as tampered and N_U is the total number of untampered blocks. The temporal authentication rate is defined as:

$$R_T = \frac{\sum_{i=1}^{N} C_i}{N} \times 100 \quad (9)$$

where N is the total number of frames and

$$C_i = \begin{cases} 1 & (S'_i = S_i) \\ 0 & (S'_i \neq S_i) \end{cases} \quad (10)$$

Where S_i and $S_i^{'}$ are the original and the extracted sequence number of the i^{th} test frame.

As mentioned in Subsection 2.1, the embedding strength of the watermark λ is an important parameter in controlling the authentication performance of the system. For evaluating the dependency of λ on the authentication performance, the test sequences are watermarked with different value of λ and each watermarked frame is JPEG compressed with QF=75. The watermarked frames are then subjected to spatial tampering by 'cut and paste' attack in which randomly chosen blocks in each watermarked frame are replaced with blocks from different frames in the same video. The authentication rates for 20% tampering are plotted against the PSNR of the watermarked frames in Figure 3. Note that the lower values of PSNR correspond to higher λ, and vice-versa.

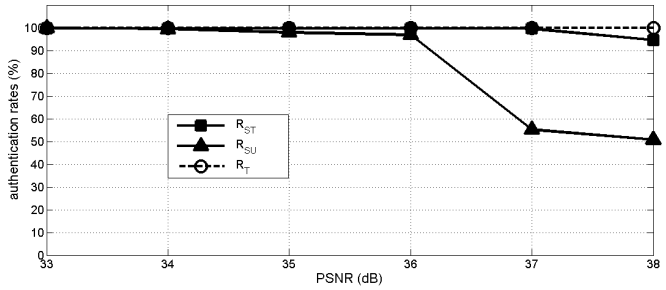

Figure 3: Authentication rates for different watermark embedding strengths and 20% spatial tampering.

It is observed that the temporal authentication rate is not affected by the decrease in embedding strength λ. The spatial authentication rate R_{ST} slightly decreases with decrease in the embedding strength. But the number of blocks wrongly detected as tampered increases considerably when the PSNR of the watermarked sequence is above 37 dB, as indicated by the low values of R_{SU}. Sample frames from the videos shown in Figures 4 and 5 demonstrate the spatial authentication performance of the proposed system for two different values of λ. The areas shown by the black blocks are the blocks those are detected by the authentication scheme as being tampered. The figures show that when the PSNR of the watermarked sequence is 36 dB, the proposed scheme detects the tampered blocks with good accuracy. But the false detection rate is considerably increased when the value of λ is decreased to maintain the PSNR at 37 dB.

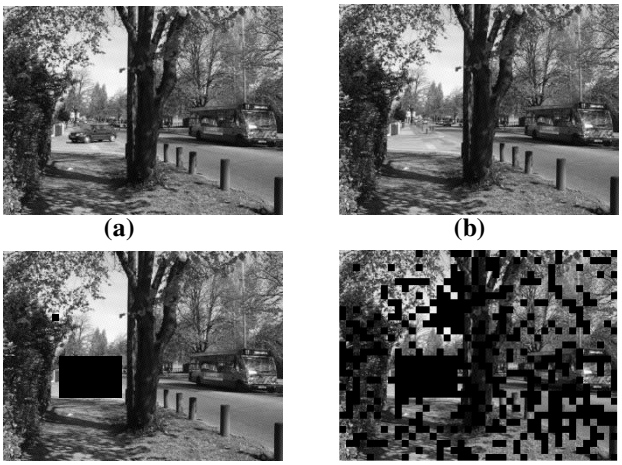

Figure 4: Spatial authentication results for the example frames Video I with 5% spatial tampering: (a) Original frame (b) Tampered frame (c) Authenticated frame (PSNR=36 dB); (d) Authenticated frame (PSNR=37 dB).

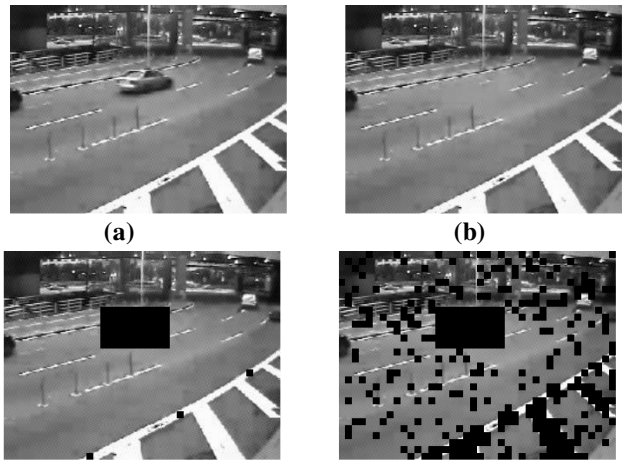

Figure 5: Spatial authentication results for the example frames Video II with 5% spatial tampering: (a) Original frame (b) Tampered frame (c) Authenticated frame (PSNR=36 dB); (d) Authenticated frame (PSNR=37 dB).

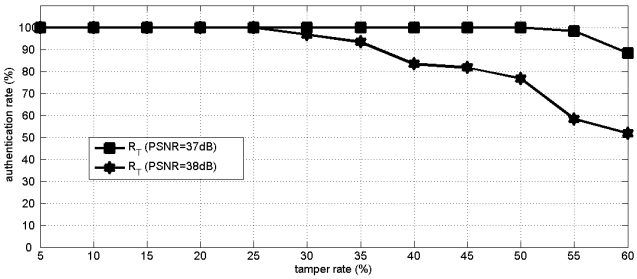

Figure 6: Temporal authentication rates for different percentages of spatial tamper.

The next experiment evaluates the temporal authentication rate when the sequences are subjected to different percentages of spatial tampering. The sequences are watermarked using two different λ such that the PSNR of the watermarked sequences are 37 dB and 38 dB. The watermarked sequences are then subjected to different spatial tampering ranging from 5% to 60% of the blocks in each frame and the temporal authentication rates are shown in Figure 6. It is observed that with the higher embedding strength (PSNR =37 dB), the temporal authentication rate is nearly 100 up to a spatial tampering of 55 %. On the other hand, with lower embedding strength (PSNR =38 dB), the authentication rate considerably decreases above 30% spatial tamper. As a compromise between the visual quality of the watermarked sequence and the authentication rates, the embedding strength λ in the subsequent experiments is chosen such that the PSNR of the watermarked sequence is 34 dB.

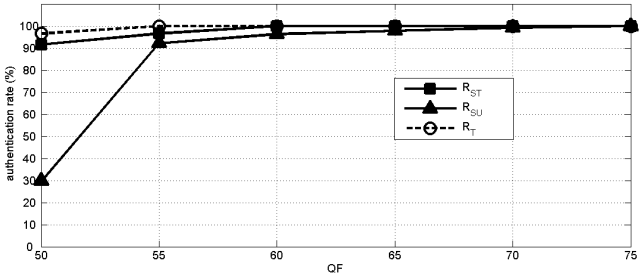

Figure 7: Authentication rates in presence of JPEG compression with different quality factors.

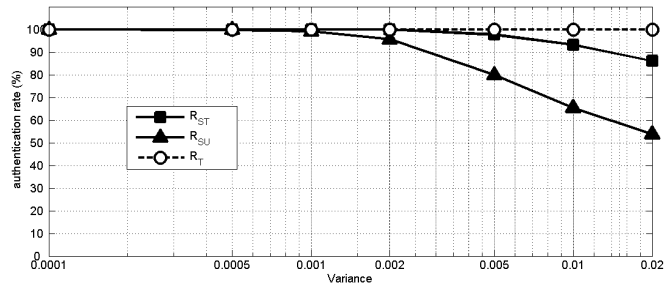

Figure 8: Authentication rates in presence of AWGN with different variances.

The performance of the proposed authentication scheme in presence of legitimate content modifications such as frame-wise JPEG compression and additive white Gaussian noise are also evaluated. First, the authentication performance in presence of JPEG compression is evaluated. The watermarked sequences are frame-wise JPEG compressed with varying quality factors from 50 to 75. The JPEG compressed frames are then subjected to 20% spatial tampering and the authentication rates are shown in Figure 7. The authentication rates are above 90% for QF above 55. When the QF is below 55, the number of blocks wrongly detected as tampered increases, as indicated by the sharp decrease in R_{SU} . Finally, the effect additive white Gaussian noise (AWGN) on the authentication performance is evaluated. White Gaussian noise with different variances are added to the watermarked sequences and the resulting sequences are then subjected to 20% spatial tampering. From the authentication results plotted in Figure 8, it can be observed that the scheme is robust up to a noise variance of .002 and then the spatial authentication rates decreases. Note that for adding white Gaussian noise we used the IMNOISE function of MATLAB which normalizes the pixel values in the range of [0,255] to [0,1] before adding the noise. The PSNRs of the noise added sequences with respect to the corresponding watermarked sequences for variances .0001, .002 and .02 are approximately 40 dB, 27 dB and 17 dB, respectively.

4 Conclusion and future work

This paper proposed a watermarking-based authentication system for M-JPEG sequences. Robust watermark is used to detect temporal tampering and semi-fragile watermark is used for detecting spatial tampering. The watermark embedding process modifies selected mid-frequency DCT coefficients in the 8×8 blocks of the frames. The watermark detection is *blind* in the sense that neither the original videos nor the original watermark is required. The authentication performance of the proposed system is evaluated against a number of legitimate and illegitimate content modifications. The experimental results show that the system can detect spatial and temporal tampering with a good accuracy.

As a future work, the recovery capability for tampered video could be added to the authentication system. There are some algorithms for image recovery, such as LSB self-correcting [10], irregular sampling [11]. It would be interesting to investigate how such recovery techniques can be developed for video authentication.

Acknowledgements

The authors would like to thank Dr Helen Treharne and Mr. Chris Culnane, Department of Computing, University of Surrey.

References

[1] Ho, A.T.S., Zhu, X., Guan, Y.L., "Image Content Authentication Using Pinned Sine Transform," EURASIP Journal on Applied Signal processing (JASP), Special Issue on Multimedia Security and Rights Management, No. 14, Vol 2004, pp2174-2184, October 2004.

[2] Peng,Yin, Yu, H.H., "Semi-fragile watermarking system for MPEG video authentication" *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on,* Volume 4, 13-17 May 2002.

[3] G. Langelaar and R. Lagendijk, “Optimal differential energy watermarking of DCT encoded images and video,” *IEEE Trans. Signal Process*, vol. 10, no. 1, pp. 148–158, Jan. 2001.

[4] F. Hartung and B. Girod, “Watermarking of uncompressed and compressed video,” *Signal Process*, vol. 66, no. 3, pp. 283–301, May 1998.

[5] S. Thiemert, H. Sahbi, M. Steinebach. “Using Entropy for Image and Video Authentication Watermarks,” in *IS and T/SPIE Conference on Security, Steganography and Watermarking of Multimedia Contents (SPIE)*, San Jose, California, to appear, January 2006.

[6] Zhang, J., Ho, A.T.S., Qiu, G., Marziliano, P., "Robust Video Watermarking of H.264/AVC," *IEEE Transactions on Circuits and Systems (TCAS)*, Part II, Vol. 54, Issue 2, p205-209, February 2007.

[7] A. Giannoula, N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, “Watermark detection for noisy interpolated images,” *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 53, no. 5, pp. 359–363, May 2006.

[8] D. Nicholson and J. Meessen, "Technologies for multimedia and video surveillance convergence", in *SPIE Proc. Image and Video Communications and Processing 2005*, San Jose, CA, January 2005.

[9] J. Meessen, C. Parisot, X. Desurmont, and J.-F. Delaigle, "Scene Analysis for Reducing Motion JPEG 2000 Video Surveillance Delivery Bandwidth and Complexity", in *Proc. of IEEE International Conference on Image Processing (ICIP'05)*, Genova, Italy, September 2005.

[10] J. Fridrich and M. Goljan, “Images with self-correcting capabilities,” in *Proc. IEEE International Conference on Image Processing (ICIP ’99)*, vol. 3, pp. 792–796, Kobe, Japan,October 1999.

[11] Zhu, X., Ho, A.T.S., Marziliano, P., "A New Semi-fragile Image Watermarking With Robust Tampering Restoration Using Irregular Sampling," *Elsevier Signal Processing: Image Communication,* Vol. 22, Issue 5, p515-528, June 2007