A Blind Video Watermarking Scheme Based on 3D Discrete Wavelet Transform

Majid Masoumi and Shervin Amiri

***Abstract*—In this paper a novel digital watermarking method for video based on multi-resolution wavelet decomposition is proposed. The core idea behind our proposed technique is to use the scene change analysis to embed the watermark into DWT coefficients of detected motion scene frames. So every bit of the watermark is spread over the three dimensional wavelet coefficients of LH, HL and HH by using pseudo-random numbers. Extraction process is done without using the original video, namely, blind detection. The resulting watermarking scheme can be used for public watermarking applications, where the original video is not available for watermark extraction. Experimental results show the robustness and the invisibility of the embedded watermark against lots of attacks, containing: frame averaging, frame dropping, frame swapping and lossy compression.**

***Index Terms*—Video watermarking, scene change detection, Discrete Wavelet Transform (DWT), spread spectrum.**

I. Introduction

The rapid proliferation of multimedia over internet demands sophisticated technique for secure and efficient access to information. There is growing need to discourage unauthorized duplication and use of digital data. With the advent of digital video, issues of copyright protection have become more important. In addition to video, watermarking techniques have been proposed to protect images, video, audio, text and other types of data [1].

Watermarking is a concept of embedding a special pattern, namely watermark, into a multimedia document so that a given piece of copyright information is permanently tied to the data. Many algorithms for developing watermarks on images are extended for videos. Applications of video watermarking contain fingerprinting, broadcast monitoring, video authentication and copyright protection. The following aspects are important for designing of video watermarking systems:

Imperceptibility: The watermark embedding should cause as little degradation to the host video as possible.

Robustness: The watermark must be robust to common signal processing manipulations and attempts to remove or impair the watermark.

Security: The embedded information must be secure against tampering.

Capacity: The amount of embedded information that can be hidden in video and be large enough to uniquely identify the owner of the video [2].

Due to the large amount of data and high correlation between frames, unlike in the case of still images, video watermarking poses many problems. It is highly susceptible to pirate attacks such as frame averaging, frame dropping, frame swapping, frame interpolation, statistical analysis and etc. If all the frames are embedded with an identical watermark, it is vulnerable to collusion attack. Applying independent watermark for each frame is also a problem. On the other hand, regions in each video frame with little or no motion remain the same frame after frame. Motionless regions in successive video frames are vulnerable to lossy compression or may be statistically compared or averaged to remove independent watermarks. In order to conquest to these problems we decided to embed the watermark in motion scene frames.

Several watermarking methods have been delivered for video data. In [3] a method is proposed in which video sequence is considered as a three-dimensional signal with two dimensions in space and one dimension in time. Among the delivered techniques in recent years, the ones based on the Discrete Wavelet Transform (DWT) are gaining more popularity due to their excellent spatial localization, frequency spread and multi-resolution characteristics. For these reasons, the proposed method also works in the DWT domain. On the other hand, video watermarking schemes must not use the original video during watermark detection as the video usually is in very large size and it is inconvenient to store it twice. Hence a new video watermarking scheme is proposed to overcome this problem, too.

In this study, a video watermarking scheme which shows the robustness against video attacks and yet enables blind retrieval of the watermark is proposed. The rest of the paper is organized as follows: Section 2 gives information about DWT. Section 3, briefly surveys suggested scene change detection method. Embedding and detection process of the watermark are explained in sections 4 and 5. Results and analysis are discussed in Section 6. Finally section 7 summarizes our delivered video watermarking scheme.

II. Discrete Wavelet Transform

Discrete wavelet transform is a multi-resolution decomposition of a signal. Considering an image, 1 level DWT involves applying a low pass and a high pass filters along the columns and then the rows, respectively [4]. In two dimensional applications, for each level of decomposition, first, DWT is performed in the vertical direction, followed by the DWT in the horizontal direction. After the first level of

Manuscript received April 20, 2012; revised June 19, 2012.
M. Masoumi is with the Electrical Engineering Department, Islamic Azad University Qazvin Branch, Qazvin, Iran (e-mail: masoumiii@ yahoo.com).
S. Amiri is Scientific Member of E.E. Department of Iranian Research Organization for Science and Technology (IROST), Tehran, Iran (e-mail: amiri@irost.org)

decomposition, there are 4 sub-bands: LL1, LH1, HL1 and HH1. For each successive level of decomposition, the LL sub-band of the previous level is used as the input. Each tile component undergoes three levels of decomposition. This results in 10 sub-bands per component. LH1, HL1, and HH1 contain the highest frequency bands present in the image tile, while LL3 contains the lowest frequency band. The three-level DWT decomposition is depicted in Fig.1.

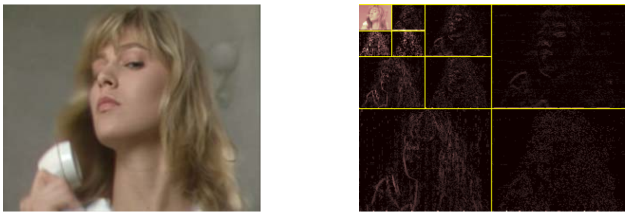

Fig. 1. One frame of the suzie video sequence with three levels of DWT decomposition.

III. Scene Change Detection

Applying independent watermarks to each frame presents a problem, if regions in each video frame remain little or no motion frame after frame. These motionless regions may be statistically compared or averaged to remove the independent watermarks [5]. So we decided to place watermark in motion scene frames. Histogram difference is the method which is used for detecting motion scene. Only the histogram for the red component of the frames is applied. So histograms of the corresponding frames are calculated. Then, the total difference of the whole histogram is calculated, which is given by:

$$D(i,i+1) = \sum_{i=1}^{n} \left| H_i(j) - H_{i+1}(j) \right| \quad (1)$$

Where $H_i(j)$ is the histogram value for the red component j, in the ith frame. If $D(i,i+1) > threshold(T)$ a scene change is detected. By considering a threshold of greater than 3800 adaptive frames for embedding the watermark are achieved. There are 8 frames of video which is detected as scene change frames. Their histograms are shown in Fig. 2.

IV. Watermark Embedding

The process of embedding the watermark is as follows:

- After detecting the motion scene, a three dimensional wavelet in three levels is applied on these frames, it means two dimensional DWT on each frame and one dimensional DWT along temporal axis over frames.
- The three dimensional coefficients of HL, LH and HH are chosen for embedding the watermark. Coefficients of LL (i.e. the low frequency sub-band) are not watermarked, as video energy is concentrated on lower frequency wavelet coefficient. If they are altered, it will affect on perceptual quality.
- A spread-spectrum technique is used to spread the power spectrum of the watermark data, thus, increasing its robustness against attacks. First by creating a random numbers based on Mersenne–Twister algorithm which is proposed by Nishimura and Matsumoto [6] .This method generates numbers with a period of $(2^{19937} - 1)/2$. By using a key, a set of pseudo random numbers $W(i,j,k) \in \{-1,+1\}$ are achieved based on the adaptive size of wavelet coefficients .

According to magnitudes of the 3D-DWT coefficients, watermark image is adaptively spread spectrum and embedded into these coefficients. Following algorithm shows the embedding process:

$$\begin{aligned} & if\ \ I(i,j)=0\ \ then \\ & S'(i,j,k) = S(i,j,k) + \alpha.W(i,j,k) \\ & esle\ \ if \\ & S'(i,j,k) = S(i,j,k) \\ & end\ \ if \end{aligned}$$

Where α is an intensity factor, $S(i,j,k)$ is the 3D-DWT coefficient of frames, $S'(i,j,k)$ is the watermarked 3D-DWT coefficient of frames and $I(i,j)$ is the binary watermark.

After embedding the watermark, the 3D-IDWT is performed on the scene change frames.

V. Watermark Extracting

During extraction process the original video is not needed, namely, blind extraction. The extraction is the inverse process of watermark embedding:

- Performing the 3D-DWT on scene change frames, that is, 2D-DWT on each frame and 1D-DWT along temporal axis over frames.
- Choosing three dimensional coefficients of HL, LH and HH of decomposed coefficients.
- Generating a pseudo-random number based on Mersenne–Twister algorithm using the same key which is used in embedding process.
- Calculating the correlation between extracted coefficients and pseudo-random numbers gives the hidden watermark:

$$\begin{aligned} & if\ \ corr(extracted\ coefficients,\ pseudo\ random\ numbers) > th \\ & watermark\ is\ \text{detectable} \\ & end\ \ if \end{aligned}$$

th (threshold) is achieved by experience .

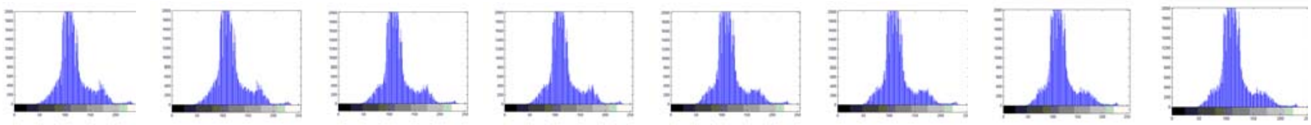

Fig.2. Histograms of scene change frames.

VI. Experimental Results

The image which is considered as the watermark is a binary image of size 36×44. The process is performed on the true color suzie video sequence which contains 300 frames (about 10 seconds) with CIF format. Generally, the accurate measurement of the imperceptibility as perceived by a human observer is a great challenge in image/video processing. The reason is that the amount and visibility of distortions introduced by the watermarking attacks strongly depend on the actual image/video content [7]. To measure the perceptual quality, we calculate the peak of signal-to-noise ratio (PSNR) that is used to estimate the quality of the watermarked frames in comparison with the original ones. The PSNR [8] is defined as follows:

$$PSNR = 20\log_{10}\left(\frac{\max_i}{\sqrt{MSE}}\right) \quad (2)$$

$$MSE = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\|F_{ij} - \hat{F}_{ij}\right\|^2 \quad (3)$$

Where $\max_i = \max\left\{\hat{F}_{ij},\ 1 \le i, j \le m\right\}$ and the MSE is the mean squared error between the cover frame F and the watermarked frame $\hat{F}$.

After extracting and refining the watermark, a similarity measurement between extracted and the referenced watermarks is used for objective judgment of the extraction fidelity and it is defined as:

$$\text{Normalized Correlation}: NC = \frac{\sum_i\sum_j W(i,j)\hat{W}(i,j)}{\sum_i\sum_j [W(i,j)]^2} \quad (4)$$

Which is the cross-correlation normalized by the reference watermark energy to give unity as the peak correlation [9]. In this experiment this measurement is used to evaluate the robustness of proposed scheme.

Approximately the value of PSNR for scene change frames is the same and it varies between $36.52 \le PSNR \le 36.77$. Also, extracted watermark has the NC value of 0.9245. Fig.3 gives a visual comparison between the original and watermark frames, also Fig.4 shows the original and extracted watermark, respectively.

(a) (b)

Fig.3. (a) Original frame; (b) Watermarked frame

To test the robustness of the watermark, different attacks are performed on watermarked video. The embedded watermark is retrieve using the proposed algorithm and the NC value of the recovered watermark is recorded for different attacks. These attacks contain frame dropping, frame averaging, frame swapping and lossy compression.

(a) (b)

Fig.4. (a) Original watermark; (b) Extracted watermark

A. Frame Dropping Attack

There is a little change between frames in a motion scene. Selected frames are removed from the watermarked video and replaced by their corresponding original frames. For this purpose 25% frames of watermarked video are dropped.

B. Frame Averaging Attack

Frame averaging is the average of current frame and its two nearest neighbors to replace the current frame, this process is also perform on 25% frames of the watermarked video. Averaging is defined by:

$$F_k(i,j) = \frac{[F_{k-1}(i,j) + F_k(i,j) + F_{k+1}(i,j)]}{3} \quad ,k=2,3,4,...,n-1 \quad (5)$$

C. Frame Swapping Attack

Frame swapping takes place by defining the $F_k(i,j) = F_{k-1}(i,j)$ and k=1,3,5,...,n-1. This is done on 25% frames of watermarked video, to.

D. Lossy Compression

In this section lots of compression attacks such as MPEG-2, MPEG-4, H.264 and MJPEG are tested. Experimental results show the robustness of the proposed method against various kinds of lossy compression which is shown in Table I.

TABLE I: Performance in Term of Normalized Correlation for Detected Watermark

Attacks	NC	Attacks	NC
No attack	0.9245	MPEG-2	0.8568
Frame dropping	0.9118	MPEG-4	0.7997
Frame averaging	0.9082	H.264	0.9189
Frame swapping	0.8970	MJPEG	0.8942

In order to demonstrate the robustness of the proposed scheme compared with other schemes, our scheme is compared with methods of Hsu [10] and Niu [11]. Table II shows the NC value between the proposed scheme and two other methods when they are exposed to attacks. The rates of frame dropping and averaging are 50%.

TABLE I: COMPARISON OF DETECTED WATERMARK IN TERM OF NORMALIZED CORRELATION WITH OTHER TWO METHODS

Attacks	Hsu's method	Niu's method	Proposed method
Frame dropping	0.614	0.685	0.873
Frame averaging	0.482	0.661	0.911
MPEG-2	0.510	0.812	0.856

At the end, for showing the preference of our scheme, the results of method [12] is compared with our method. Method [12] is a DWT-based watermarking scheme which embeds an identical watermark in all frames .The selected attacks for this comparison are frame dropping and frame averaging.

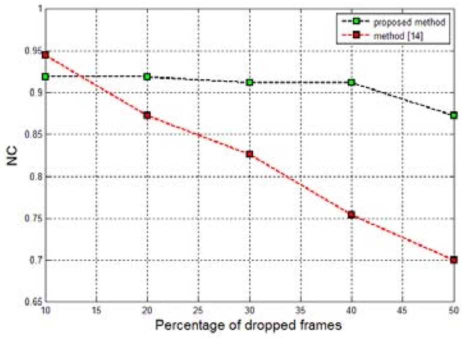

Fig.5. Performance under frame dropping

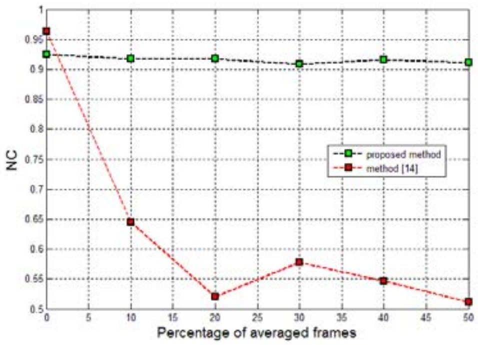

Fig.6. Performance under frame averaging

VII. CONCLUSION

In this paper, a wavelet-based watermarking for video is delivered. The key idea behind the proposed method is using the scene change detection for embedding the watermark. So watermark is embedded into motion scene frames. Experimental results show the invisibility and robustness of our algorithm against various kinds of attacks such as: frame dropping, frame averaging, frame swapping and lossy compression.

REFERENCES

[1] C. I. Podilchuk and E. J. Delp, “Digital watermarking: Algorithms and applications,” *IEEE Signal Process. Mag.*, vol. 18, no. 4, pp. 33–46, July 2001.

[2] N. Deshpande et al, “Review of robust video watermarking algorithms,” *Int. J. Comput. Sci. Info. Secur. (IJCSIS)*, vol. 7, no. 3, March 2010.

[3] F. Hartung and B. Girod, “Watermarking of compressed and uncompressed video,” *Signal Process.* vol. 66, pp. 283–301 May 1998.

[4] G. Davis and A. Nosralinia, “Wavelet-based image coding: an overview,” *IEEE Trans. on Image Proc.*; March 1996.

[5] M. Swanson, B. Zhu, and A. Tewfik, “Multiresolution video watermarking using perceptual models and scene segmentation,” *in Proc. Int. Conf. Image Processing*, vol. 2, Washington, DC, pp. 558–561 Oct. 1997.

[6] M. Matsumoto and T. Nishimura, “Mersenne twister, a 623-dimensionally equilistributed uniform pseudorandom number generator,” *ACM Trans. on Modeling and Comput. Simul.* , vol. 8, no. 1, pp. 3–30, 1998.

[7] S. Winkler, E. D. Gelasca, and T. Ebrahimi, “Toward perceptual metrics for video watermark evaluation,” *Proc. SPIE Appl. Digit. Image Process.* 5203, pp. 371–378, 2003.

[8] A. N. Netravali, and B. G. Haskell, *Digital Pictures: Representation, Compression, and Standards. Plenum*, New York , 1995.

[9] H. Chiou-Tung and W. Ja-Ling, “Digital watermarking for video,” *in Proc. 1997 13th International Conf. on Digital Signal Processing(DSP 97)*, vol. 1, 2-4 July 1997, pp. 217–220.

[10] C. T. Hsu and J. L. Wu, “DCT-based watermarking for video,” *IEEE Trans. Consumer Electronics*, vol. 44, no. 1, pp. 206–216, 1998.

[11] X. Niu, S. Sun, and W. Xiang, “Multiresolution watermarking for video based on gray-level digital watermark,” *IEEE Trans. Consumer Electronics*, vol. 46, no. 2, pp. 375–384, 2000.

[12] M. Ejima and A. Miyazaki, “A wavelet-based watermarking for digital images and video,” *International Conf. on Image Processing* (*ICIP 00)*, Vol. 3, 2000, pp. 678–681.

Majid Masoumi was born in Lahijan, Iran, in 1986. He is a Master student in electrical engineering at Azad University of Qazvin. He received his B.Sc. degree in electrical engineering in 2008. His previously researches include audio watermarking, opto-electronics and photonic crystals. He is currently interested in watermarking, image and video processing, cryptography, optimization, and communication systems. Specially using spread spectrum for making secure the networks against attacks and jammers. He is presently working at Iranian Research Organization for Science and Technology (IROST) as a researcher.

Shervin Amiri was born in Tehran, Iran, in 1966. He received his B.Sc., M.Sc. and Ph.D. from Iran University of Science and Technology (IUST) in communication systems. Now he is a Scientific Member of Electrical Engineering Department in Iranian Research Organization for Science and Technology (IROST). He is supervisor of many PhD and MSc students in the fields of communication system and subsystems.