Robust Video Watermarking Using Image Normalization, Motion Vector and Perceptual Information

Cedillo-Hernández Antonio[1], **Cedillo-Hernández Manuel**[1],
Nakano-Miyatake Mariko[1], **García-Vázquez Mireya S.**[2]
[1]Postgraduate Section of Mechanical and Electrical Engineering School, National Polytechnic Institute of Mexico, Mexico City, Mexico
[2]Research and Development of Digital Technology Center, National Polytechnic Institute of Mexico, Tijuana, Mexico

Abstract – *This paper proposes a video watermarking algorithm robust against geometric distortions, several attacks of signal processing and intentional common attacks for video. Image normalization is used to get geometric invariant feature of video frames. Watermark embedding and detection process are carried out in the Discrete Cosine Transform (DCT) domain. Watermark energy is computed adaptively using perceptual information and motion vectors, to get major Watermark imperceptibility and robustness. Watermark Imperceptibility is evaluated by conventional PSNR and perceptual video quality measurement, taking sufficiently good visual quality. Computer simulation results show the watermark robustness to common signal distortions such as contamination by noise, JPEG compression, geometrical distortions and video common intentional attacks: frame dropping, frame swapping and frame averaging, among others.*

Keywords: Video Watermarking, Motion Vectors, Image Normalization, Block Classification, Perceptual Sensibility

1 Introduction

High speed computer networks, the Internet and the World Wide Web have revolutionized the way in which digital data is distributed. The widespread and easy accesses to multimedia contents and the possibility to make unlimited copy without loss of considerable fidelity arouse the need of digital rights management. Digital watermarking is considered as a technology that can serve this purpose. A large number of watermarking schemes have been proposed to hide copyright marks and other information in digital images, video, audio and other multimedia objects [1].

Usually watermarking algorithms for still images are not efficient when these are used in video sequences, because they are not considered the temporal redundancy of video signal and common attacks to the video signals [2]. In the case of a watermarking system for copyright protection, the embedded watermark should be imperceptible and robust against common attacks such as cropping, contamination by noise, filtering and compression [3]. In addition to the requirement of imperceptibility and robustness techniques, furthermore the video watermarking must satisfy the following requirements: a blind detection, i.e. the detection process does not require the original video signal and the conservation of file size after the insertion of the watermark. Due to the redundancy existing in the video sequences, some intentional attacks, such as frame dropping, frame swapping and frame averaging, that try to destroy the embedded watermark should be considered in design of video watermarking algorithms [3].

In this paper, we propose a video watermarking algorithm based on image normalization, in which a watermark pattern is normalized using the same geometric factors obtained in the image normalization. The proposed algorithm uses three criteria based on Human Visual System (HVS) to insert a robust watermark, preserving their imperceptibility. The first criterion is based on the sensitivity of the HVS to different basic color channels, the second one is based on spatial deficiency of HVS, which is determined using the texture and edge masking proposed by [4], and the last criterion is based on tracing deficiency of the HVS in the regions with high motion speed in video sequences, which is determined using motion vector. The computer simulation results show the watermark imperceptibility and the robustness against common signal processing, geometrical distortions and some intentional attacks to the video sequence. The watermark imperceptibility is measured using the Peak Signal Noise Ration (PSNR) and a quality assessment based on HVS proposed by [5].

The rest of the paper is organized as follows: In Section 2, the proposed system is described in detail and the evaluation results of the proposed system are shown in Section 3. Finally Section 4 provides some conclusions.

2 Proposed system

The proposed video watermarking system consists of several procedures, such as image normalization, block classification using the perceptual information based on HVS, watermark embedding and extraction. In this section, each procedure will be described.

2.1 Image normalization

The normalization procedure of a image f(x,y) consists of the following steps [6]:

1) Center the image $f(x,y)$, through the Affine Transformation as given by:

$$\begin{pmatrix} x_a \\ y_a \end{pmatrix} = A \cdot \begin{pmatrix} x \\ y \end{pmatrix} - d \qquad (1)$$

where matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and the vector $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ with

$$d_1 = \frac{m_{10}}{m_{00}}, d_2 = \frac{m_{01}}{m_{00}} \qquad (2)$$

where m_{10} ,m_{01} and m_{00} are the moments of $f(x,y)$, and d_1, d_2 are the density center of $f(x,y)$ [7]. This step eliminates the translation effect by assigning the center of the normalized image at the density center of the image. The resulting centered image is denoted as $f_1(x,y)$.

2) Apply a shearing transform to $f_1(x,y)$ in the x-direction using the matrix $A_x = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$. The resulting image is denoted by $f_2(x,y)$. This step eliminates shearing effect in the x-direction.

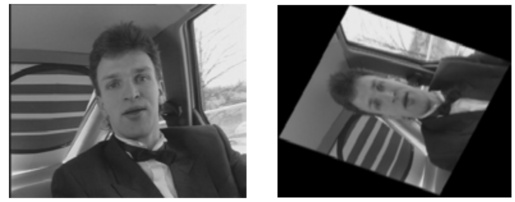
3) Apply a shearing transform to $f_2(x,y)$ in the y-direction using the matrix $A_y = \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix}$. The resulting image is denoted by $f_3(x,y)$. This step eliminates shearing effect in the y-direction.

4) Scale $f_3(x,y)$ in both x and y directions with the matrix $A_s = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$ and the resulting image is denoted by $f_4(x,y)$. This step eliminates the scaling effect by forcing the normalized image to a standard size.

The final image $f_4(x,y)$ is the normalized version. It is important to denote that each step in the normalization procedure is invertible, which allows us to convert the normalized image back to its original version. The determination of transformation parameters β, γ, α and δ, associated with the transforms A_x, A_y, and A_s are shown in detail in [9]. In the figure 1, an example of the original image and the normalized version is shown.

2.2 Perceptual Information

The proposed system uses three criteria, employed previously in [8], to embed an imperceptible and robust watermark in a video signal.



(a) (b)

Figure 1 (a) Original image (b) Normalized version from (a).

These criteria are based on the less sensitivity of the HVS to the blue channel and the detail regions such as texture region, and also it's less ability to track a region with high speed motion. According to the human eye structure, the retina contains two types of photoreceptors, rods and cones. The last one is divided in 3, each sensitive to the three basic colors there are three types of cones that react with the three basic colors: Red, Green and Blue. The number of blue-sensitive cones is 30 times less than the number of cones sensitive to the other two colors [10]. Figure 2 shows the fraction of light absorbed by each of three types of cones, here R, G, B represents the cones sensitive to red, green and blue, respectively. This figure shows that the HVS is less sensitive to blue channel than the other two basic color channels (Red and Green).

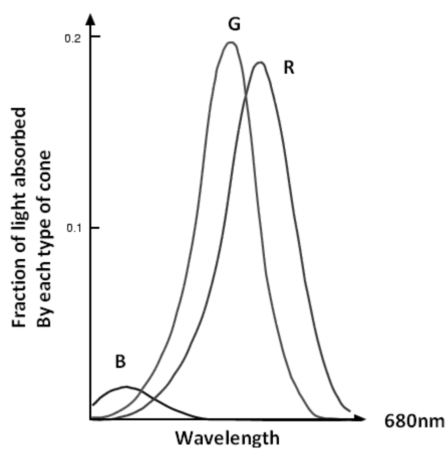


Figure 2 Sensitivity of the three types of cones R (red), G (green) and B (blue). [13]

The proposed algorithm embeds the watermark in the blue channel, taking advantage of the weakness of HVS. The blue channel of each video frame is divided into blocks with 8x8 pixels and then the 2D Discrete Cosine Transform (DCT) is applied in each block. In the DCT domain, each block is classified into three categories: plain block, edge block and texture block according to the algorithm proposed in [4]. Figure 3 shows the result of applying the block classification algorithm to a video frame.

(a) (b)

Figure 3 Block classification (a) video frame and (b) classification of blocks, black blocks are plain, gray blocks are textures and white blocks are edges, respectively.

The last criterion is based on deficiency of HVS to follow regions with high-speed motion. To classify regions of video frame by motion speed, the motion compensation prediction, which is a powerful tool to reduce temporal redundancy in MPEG coding, can be used. The macro-blocks (MB) with larger motion vector are classified as regions of high-speed movement, in which a watermark with greater energy can be embedded without causing degradation in the video signal. The magnitude of the motion vector is calculated using equation (3).

$$Mmv_i = \sqrt{mvh_i^2 + mvv_i^2},\ \ i = 1...N_{MB} \qquad (3)$$

Where mvh^2, mvv^2 are the horizontal and vertical components of motion vector of the ith macro-block and N_{MB} is the total number of macro-blocks. To determine the region with a high-speed motion, we introduce a threshold value *Th_mv* and then each macro-block is classified as high-speed motion block and low-speed motion block (or without movement) as follows:

If Mmv_i < *Th_mv* then block are low-speed

If Mmv_i > *Th_mv* then block are high-speed

Where threshold *Th_mv* is calculated by equation (4)

$$Th\_mv = \frac{1}{N_{MB}} \sum_{i=1}^{N_{MB}} Mmv_i \qquad (4)$$

Figure 4 shows two consecutive video frames (Figure 4 (a) and (b)) and the motion vectors before and after classification (Figure 4 (c) and (d), respectively), the macro-blocks with arrows of some orientation are classified as high-speed blocks and blocks with black point are considered as low-speed blocks (or without movement).

2.3 Adequate watermark energy assignment

Combining the last two criteria: the classification of blocks (8x8) using DCT coefficients and the classification of macro-blocks (16x16) using motion vector, the watermark embedding energy to the video frame is determined experimentally using 10 video sequences, which is shown in Table 1. In Table 1, B_8 is the block of 8x8 pixels in video frames and MB_{MOTION} is the macro-block with high-speed motion, each macro-block contains 4 blocks B_8. This assignation of the watermark embedding energy is applied to the normalized video frames generated in the image normalization process. Figure 5 shows an example of the classification of blocks together with the watermark embedding energy determined using the criteria mentioned above.

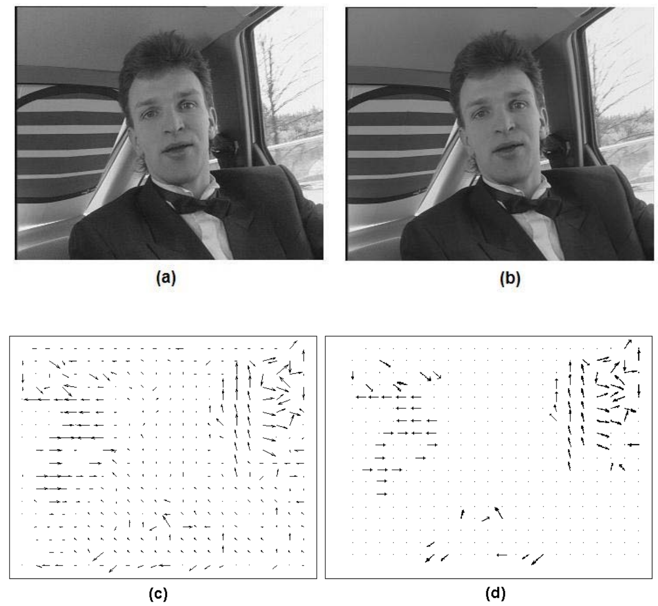


Figure 4 (a), (b)Two consecutive frames of video and (c) and (d) motion vectors before and after classification

Table 1 Watermark embedding energy

	$B_8 \in B_{plain}$	$B_8 \in B_{edge}$	$B_8 \in B_{tex}$
$B_8 \in Frame$ $B_8 \in MB_{MOTION}$	0.9	1.2	1.8

2.4 Watermark embedding process

The watermark generation and embedding process are described as follows:

1) Apply the normalization procedure to the original image to get the normalized image.

2) Divide the normalized image in blocks of 8x8 pixels and get the watermark energy for every block mentioned in 2.3.

The result of this operation is a watermark embedding energy.

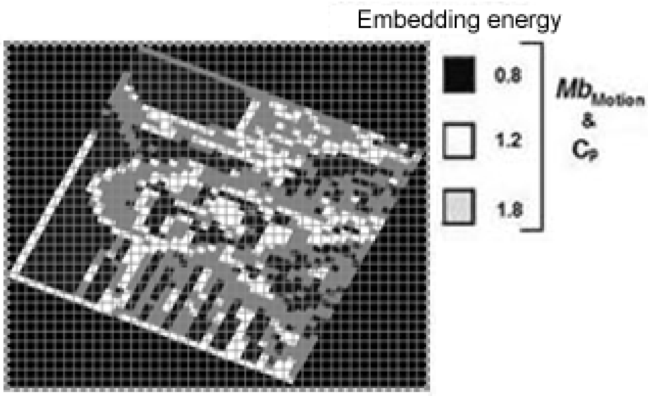


Figure 5 An example of the embedding watermark energy to each block

3) Generate the 2-D pseudo-random pattern R of the same size of original image with any key. Because this pattern is only used as the support of the watermark pattern, any key can be used and it is not necessary to save.

4) Generate a watermark vector $W = [w_1, w_2, ... w_n]$ using a user's secret key, where $w_i = \{1, -1\}$, i=1..n.

5) Create a mask image *M*, which is a binary image, taking 1s within normalized image area and 0s elsewhere, to generate masked pseudo-random pattern *MR* with same scale and rotation factors as these of the normalized image.

6) Watermark vector W is multiplied by a watermark energy vector determined by the previous section.

7) Divide the watermark vector W into N groups of L elements, where L must be one number from 1 to 22 (number of coefficients that compose the middle frequency range in DCT domain), thus, for example, if L = 5 and size of W is 500, the number of groups N = 500 / 5 = 100.

8) The coefficients in middle frequency range of each block are replaced with L elements of each group of watermark sequence. Then apply the IDCT to each watermarked block to get watermarked pattern MR_W .

9) Apply the inverse normalization to the watermarked pattern MR_W to get watermarked pattern *WP* with same size as the cover image.

10) *WP* is embedded into the original image additively with a gain factor α_2. This produces the watermarked image.

$$I_w = I_o + (WP \cdot \alpha_2) \tag{5}$$

where I_o and I_w are original and watermarked image, and *WP* is watermark pattern generated by step 9.

The whole procedure is equivalent to embedding the watermark into the DCT domain of the normalized image. The adequate watermark embedding energy allows embedding strong watermark without causing any visual distortion to watermarked image. The figure 6 shows the watermark generation and embedding procedures.

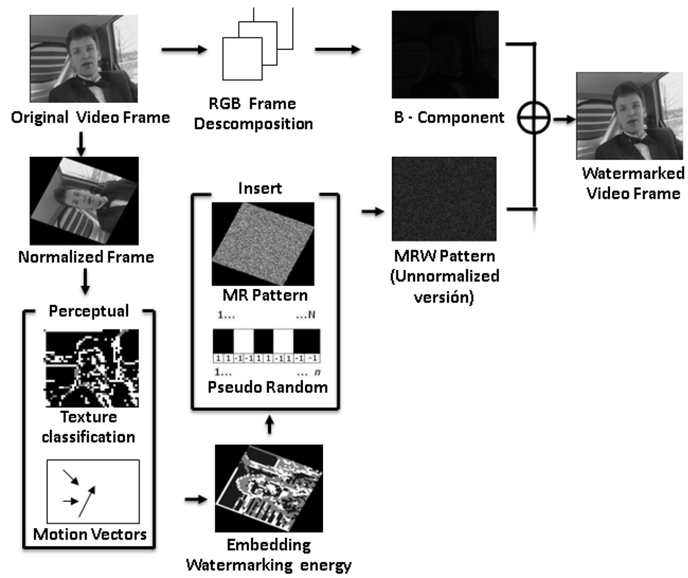


Figure 6 Watermark embedding process

2.5 Watermark Extraction

The process of the watermark detection is as follows:

1) Apply the image normalization procedure to watermarked image to get the normalized watermarked image.

2) Apply the DCT to each block of the normalized watermarked image, the middle range of the DCT coefficients *CW* are extracted. From *CW* the watermark vector $\hat{W}$ is extracted by (6).

$$\begin{aligned} \hat{W} &= [\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_{N-1}, \mathbf{w}_N] \\ \mathbf{w}_k &= sign(CW_k) \end{aligned} \tag{6}$$

where $\mathbf{w}_k$ is the extracted watermark sequence (L bits) from k-th block and *sign* is a sign function.

3 Experimental Results

To evaluate the proposed system, we used 20 video sequences with YUV-CIF format at 30 FPS. All video data have at least 150 frames which are available in [11]. The proposed system is evaluated from the watermark imperceptibility and robustness points of view.

3.1 Watermark imperceptibility

We evaluate watermark imperceptibility of the proposed system using numerical evaluation: Peak Signal Noise Ratio (PSNR) and perceptual objective evaluation proposed in [5]. Figure 7 shows the PSRN value calculated between original and watermarked frames in the proposed method. Perceptual quality assessment proposed in [5] evaluates video quality in perceptual manner, which gives the quality index calculated using image structure and motion vector. The quality index is in range [0.0,1.0], when video sequence under evaluation is numerically identical with its original version, its quality index is 1.0. The quality index of the watermarked video sequence generated by the proposed algorithm is approximately 0.96, which means that degradation of the watermarked video quality is minimum by HVS.

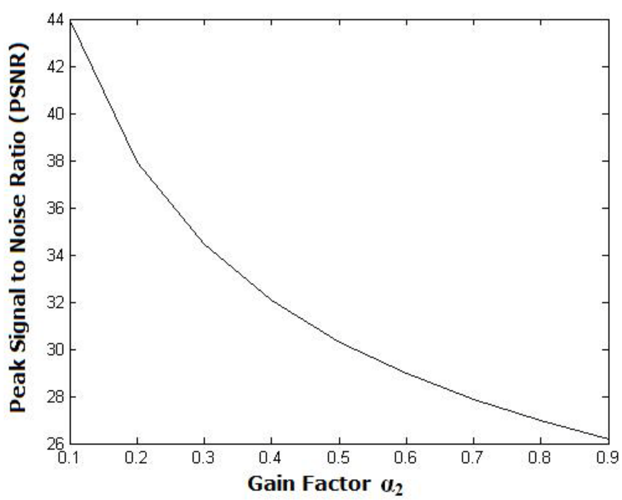
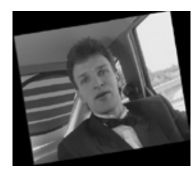
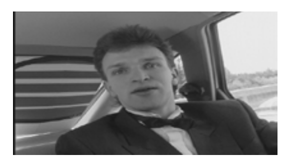
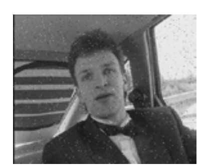
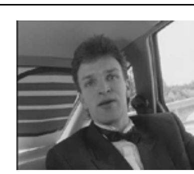


Figure 7 Peak Signal to Noise Ratio of the proposed method.

3.2 Watermark Robustness

To evaluate watermark robustness of the proposed algorithm, the watermarked video sequences are attacked by some common signal and image processing, such as codification rate change, noise contamination by impulsive noise and Gaussian noise, geometrical distortions, frame dropping, frame swapping and frame averaging. The results are shown in Table 2, in which the bit-error-rate (BER) of the extracted watermark bit sequence respect to the embedded one is shown. In all cases the values correspond to the average values using the 20 video sequences with YUV-CIF format. Frame dropping, frame swapping and frame averaging are intentional attacks for video sequence. Frame dropping is drop some frames from video sequence, frame swapping is interchange two frames and frame averaging is generate one frame taking the average of some consecutive frames. Due to that watermark sequence is embedded through temporal video frames in the proposed algorithm; embedded watermark is robust against these types of attacks. Table 3 shows the performance comparison of our proposed system and the recently reported system by Soumik which present an investigation similar to our proposal because it considers the content of video as a sequence of images and the watermark is inserted frame by frame with an invisible watermarking scheme, further extraction of the watermark is blind and is reported as a method extremely appropriate in areas such as copyright and fingerprinting [12].

Table 2. The embedded watermark signal is sufficiently robust to geometrical distortions and common signal processing. .

Attack	Watermark frame attack	BER
Geometric Rotation		0.0960
Aspect Ratio (1:2)		0.070
Gaussian noise		0.097
Impulsive noise		0.098
MPEG Codification (50% quality)		0.091

4 Conclusions

In this paper, we proposed a video watermarking algorithm based on image normalization, in which watermark embedding and detection process are carried out in the

Discrete Cosine Transform (DCT) domain. The watermark extraction has done blindly i.e., neither the watermark nor the original video is needed at the time of the watermark extraction. The proposed algorithm uses three criteria based on Human Visual System (HVS) to embed a robust watermark, preserving their imperceptibility. These criteria are based on the sensitivity of the HVS to different basic color channels, the texture and edge masking classification and the estimation of the motion vectors in video sequences. The computer simulation results show the watermark imperceptibility and the robustness of the scheme against common signal processing, geometrical distortions and some intentional attacks to the video sequence.

Table 3. Robustness comparison between the proposed video watermarking and Soumik method's [12]

Attack	**Method proposed**	**Soumik method's [12]**
Geometric Rotation	Detected	-
Aspect Ratio	Detected	-
Gaussian noise	Detected	-
Impulsive noise	Detected	-
MPEG Codification	Detected	-
Frame Dropping	Detected	Detected
Frame Swapping	Detected	Detected
Frame Averaging	Detected	Detected

5 References

[1] Sourav Bhattacharya, T. Chattopadhyay and Arpan Pal, "A Survey on Different Video Watermarking Techniques and Comparative Analysis with Reference to H.264/AVC". IEEE 10th International Symposium on Consumer Electronics, June 2006, pp.1-6.

[2] M. Swanson, B. Zhu y A. Tewfik, "Multi-resolution scene-based video watermarking using perceptual models", IEEE J. Select areas Communication, vol. 16, no. 4, pp. 540-550, May. 1998

[3] Wolfgang, R.B, C.I. Podilchuk y E.J. Delp, "Perceptual watermarks for digital images and video", Proceeding IEEE, vol. 87, no. 7, pp. 1108-1126, 1999.

[4] H.Y. Tong an d A.N. Venetsanopoulos, " A Perceptual model for JPEG applications based on block classification, texture masking and luminance masking", Int. Conf. on Image Processing (ICIP), vol. 3, pp. 428-432, 1998.

[5] Z. Wang, L. Lu and A. C. Bovik, "Video quality assessment based on structural distortion measurement", Elsevier Signal Processing: Image Communication, vol. 19, pp. 121-132, 2004

[6] M. Cedillo-Hernández, M. Nakano-Miyatake and H. Perez-Meana, "Robust Watermarking Technique based on Image Normalization" (Spanish), Revista Facultad de Ingeniería Universidad de Antioquia, no. 52, pp. 147-160, Marzo 2010.

[7] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", IRE Trans. on Information Theory, vol. 8, pp. 179-187, 1962.

[8] A. Cedillo, M. Nakano, H. Perez and L. Rojas, "Watermarking Technique for MPEG Video using Visual Sensibility and Motion Vector" (Spanish), Revista Información Tecnología, vol. 19, no. 2, pp. 81-92, 2008.

[9] P. Dong, J. B. Brankov, N. P. Galatsanos, Y. Yang and F. Davoine, "Digital Watermarking robust to geometric distortions,", IEEE Trans. on Image Processing, vol. 14, no. 12, pp. 2140-2150, 2005.

[10] K. Sayood, "Introduction to Data Compression", 2nd Edition, Morgan Kaufmann Publishers, 2000.

[11] http://www.antonioch.com/public/?page_id=241

[12] Soumik Das, Pradosh Bandyopadhyay, Shauvik Paul, Chaudhuri Atal and Monalisa Banerjee, "An Invisible Color Watermarking Framework for Uncompressed Video Authentication", International Journal of Computer Applications, vol. 1, no. 11, pp. 22-28, February 2010.

[13]http://science.slashdot.org/story/08/04/08/2213222/What-Font-Color-Is-Best-For-Eyes