A Video Watermarking Algorithm Based on the Human Visual System Properties

Ji-Young Moon[1] and Yo-Sung Ho[2]

[1] Samsung Electronics Co., LTD
416, Maetan3-dong, Paldal-gu, Suwon-si, Gyenggi-do, Korea
jiyoung.moon@samsung.com

[2] Kwangju Institute of Science and Technology(K-JIST)
1 Oryoung-dong, Puk-gu, Kwangju, 500-712, Korea
hoyo@kjist.ac.kr

Abstract. In this paper, we propose a new video watermarking algorithm based on the human visual system (HVS) properties to find effective locations in video sequences for robust and imperceptible watermarks. In particular, we define a new HVS-optimized global masking map for hiding watermark signals by combining the frequency masking, the spatial masking, and the motion masking effects of HVS. In this paper, we generate the watermark by the bit-wise exclusive-OR operation of a logo image and a random sequence, and we embed the watermark in the uncompressed video sequence. The amount of inserted watermarks is controlled by the peak-signal-to-noise ratio of the watermarked frame. Experimental results demonstrate that the proposed method is imperceptible and robust against various attacks with a good watermark capacity.

1 Introduction

In recent days, digital video can easily be reproduced and widely distributed over various transmission networks. However, these attractive properties of the digital video lead to serious problems related to copyright protection.

Now that digital watermarking has become an active research area for the digital video copyright protection, several video watermarking algorithms have been proposed. Most of them embed watermark signals in each frame of the video sequence using still-frame image watermarking techniques [1], which would allow inheriting robustness of the two-dimensional approaches. However, they are vulnerable to averaging attacks, where consecutive frames are averaged to remove embedded watermarks. In addition, a simple extension of the image watermarking technique for video watermarking does not consider temporal correlation along the time axis or movement between the consecutive image frames. Other watermarking techniques embed the watermark signal in the DCT domain for the MPEG-2 compression [2]. However, they are not fully exploited in various video codecs. Furthermore, due to the bit rate constraint, only few DCT coefficients of the watermark can be embedded per image block.

A. Yazici and C. Şener (Eds.): ISCIS 2003, LNCS 2869, pp. 699–706, 2003.

In this paper, we propose a new video watermarking method based on the characteristics of the human visual system (HVS). In order to design a general watermarking scheme, we insert the watermark signals in the uncompressed video sequence; therefore, our watermarking method is independent of the coding scheme and the compression ratio.

One of the important goals of watermarking techniques is to embed some useful information in the original content as much as possible, while maintaining invisibility of that information; thus, we consider the HVS characteristics carefully in our design. In this paper, we propose a new HVS-optimized weighting map in order to hide the watermark effectively by combining the frequency masking, the spatial masking, and the motion masking effects of HVS.

2 The Global Masking Map

2.1 Frequency Masking

Our frequency masking operation is based on the Watson's visual model, defined in the DCT domain [3]. After we divide the image I into 8x8 pixel blocks $I[x, y]_k$, where $x, y = 0, 1, \dots, 7$ and k is the block number, each block is transformed into the DCT domain. Then, we use the frequency sensitivity table [4], which contains the smallest magnitude of each DCT coefficient in the block that is perceptible in the absence of any masking noise. Thus, a smaller value in the table indicates that the eye is more sensitive to the frequency component.

In the absence of masking noise in the frequency domain, we can find a visibility threshold value that is the minimum level below which the signal is not perceptible. The visibility threshold is defined by

$$t_{th}[i, j]_k = t[i, j][\frac{\tilde{I}[0, 0]_k}{\tilde{I}_{avg}[0, 0]}]^{\alpha} \tag{1}$$

where $\tilde{I}[0, 0]_k$ is the DC coefficient of the k^{th} block in the original image, $\tilde{I}_{avg}[0, 0]$ is the average of the DC coefficients in the image, and α is a constant. In our experiment, we set $\alpha = 0.649$ empirically. The frequency masking function is also defined by

$$\tilde{F}[i, j]_k = t_{th}[i, j] \cdot max\{1, [\frac{|\tilde{I}[i, j]_k|}{\tilde{I}_{th}[i, j]_k}]^{w}\} \tag{2}$$

where $\tilde{I}[i, j]_k$ is the DCT coefficients of the k^{th} block and w is a constant.

The frequency masking function is shown in Figure 1(a), where we choose $w = 0.7$ for the slope of the line. When $\tilde{I}[i, j]_k$ are larger than $t_{th}[i, j]_k$, the signal threshold values at all frequencies lie along the straight line, which means that the masking effect occurs. Figure 1(c) shows the result of the frequency masking for the *Foreman* sequence.

(a) (b) (c)

Fig. 1. Frequency masking : (a) Frequency masking function (b) Original frame (c) Frequency masking result

2.2 Spatial Masking

The main purpose of the edge map in the proposed watermarking algorithm is to extract connected edges in each image frame. In this paper, we control the contrast of images before the edge detection operation to obtain a good spatial masking map from the following lightness function, proposed by Schreiber [5]:

$$S[x,y] = 1 + 99\frac{log(1 + I[x,y] \cdot a) - log(1 + a)}{log(1 + 100a) - log(1 + a)} \tag{3}$$

where $I[x,y]$ is the luminance value of the original frame. Schreiber indicated that $a = 0.05$ provides a well-adapted luminance scale. After we apply the lightness function, we extract important edges to find the spatial masking effect.

(a) (b) (c)

Fig. 2. Spatial masking and Motion masking : (a) Spatial masking only using the edge detection (b) Spatial masking using the edge detection after controlling contrast of image (c) Motion masking result

Figure 2(a) is the result of the spatial masking using the edge detection operation, and Figure 2(b) is the result of the proposed spatial masking, where we control the contrast of images. In Figure 2, we can observe that Figure 2(b) is better; that is, Figure 2(b) has less spurious edges than Figure 2(a).

2.3 Motion Masking

A video watermarking method can exploit the structural characteristics of the video sequence. After we find displacement parts in the successive image frames, we can apply a suitable filter to extract image contours. Figure 2(c) shows the result of the motion masking, where high values are assigned in the face part because of large motion changes.

2.4 Global Masking Map Modeling

In this paper, we define the global masking map by combining the above frequency, spatial, and motion masking effects together after normalization. In other words, the global masking map G is obtained by

$$G = F + S + M \tag{4}$$

where F is the frequency masking, S is the spatial masking, and M is the motion masking, respectively. Figure 3 shows simulation results with the global masking; Figure 3(a) is obtained by the previous method, and Figure 3(b) is obtained by the proposed method. We can see that Figure 3(b) is more effective than Figure 3(a); that is, we can insert more watermarks effectively using the proposed global masking map more than the previous one.

(a) (b)

Fig. 3. Global masking result: (a) The previous method (b) The proposed method

3 Watermarking System

3.1 Watermarking Embedding System

Figure 4 shows the block diagram of the proposed watermarking scheme, where we perform the exclusive OR operation of a random sequence X_1 and a gray-level logo image X_2 as the watermark W.

1. Divide the test data into non-overlapping 8x8 pixel blocks.
2. Obtain the frequency masking, the spatial masking, and the motion masking factors for each image block, respectively.

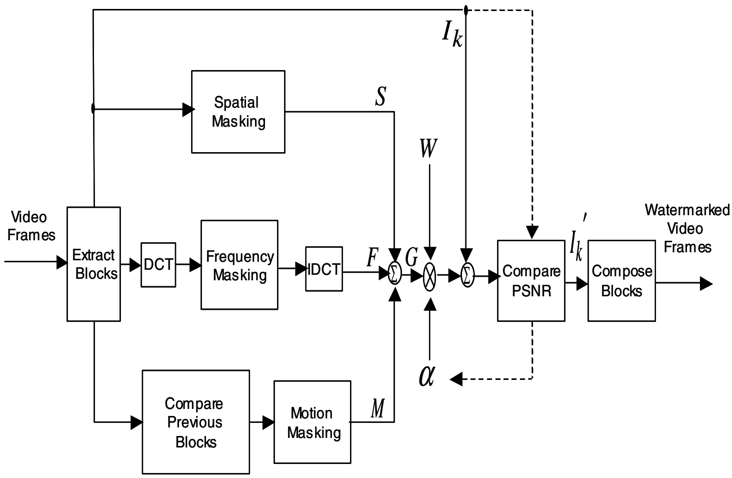


Fig. 4. Proposed watermarking method

3. Generate the global masking map G by combining the frequency masking, spatial masking, and motion masking factors after normalization.
4. Add the watermark weighted by G to the original frame I:

$$I' = I + \alpha \cdot G \cdot W \tag{5}$$

   where the control parameter α is initially set to 1.
5. Find the peak-signal-to-noise ratio(PSNR) of the watermarked frame to control the parameter α: if PSNR is higher than the target, we increase α; otherwise, we decrease α.

3.2 Watermarking Extraction System

1. The original frame I is subtracted from the watermarked frame I':

$$I' - I = G \cdot W^* \tag{6}$$

2. We perform the exclusive OR operation as follows:

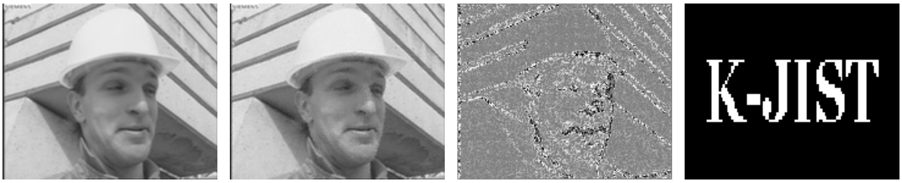
$$G \cdot W^* \bigoplus X_1 = X_2^* \tag{7}$$

   where X_2^* is the extracted logo image.

4 Experimental Results and Analysis

4.1 Invisibility Test

Figure 5 shows the experimental result of the proposed method of the *Foreman* test frame. Figure 5(a) is the original frame, and Figure 5(b) is the watermarked frame. The watermarked frame appears visually identical to the original. Figure 5(c) is the watermark for each frame. It is scaled to gray levels for display, and has the same size as the original frame. Figure 5(d) is the extracted logo.

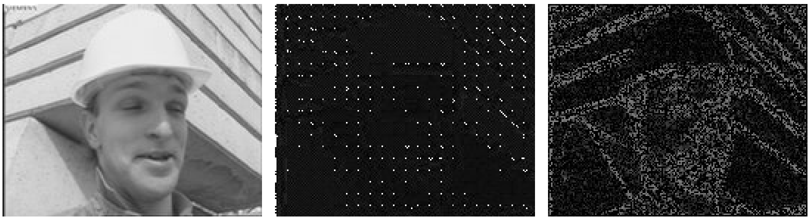


(a) (b) (c) (d)

Fig. 5. Experimental results of the proposed method : (a) the original frame, (b) the watermarked frame, (c) the watermark, and (d) the extracted logo image

4.2 Watermark Capacity Test

Figure 6(a) is the watermarked frame. Figure 6(b) is the extracted watermark from the previous methods using only frequency masking, and Figure 6(c) is the extracted watermark from the proposed method which considered the three masking effects. We can observe that Figure 6(c) is more effective in terms of watermark capacity, compared to Figure 6(b).



(a) (b) (c)

Fig. 6. Watermark Capacity Test : (a) Watermarked frame, (b) Extracted watermark from frequency masking, (c) Extracted watermark from the proposed method

4.3 Robustness Test

Robustness Test on MPEG-2. The test was on three different bit rates: 5 Mbps, 3 Mbps, and 2.5 Mbps. The watermarked frames have been coded and decoded again at different bit rates. Figure 7 shows the extracted logo images at those bit rates. All logo images have been detected only with slight degradation of image quality, but we can see that the performance of the proposed method is not invalidated.

Robustness Test on Re-encoding. We can see that the watermark can always be extracted from the watermarked video frames compressed by MPEG-2 repeatedly. Figure 8 shows the extracted logo images at the different bit rates after MPEG-2 re-encoding.

(a) (b) (c)

Fig. 7. Extracted images after MPEG-2 test : (a) 5 Mbps, (b) 3 Mbps, (c) 2.5 Mbps

(a) (b) (c)

Fig. 8. Extracted images after re-encoding : (a) 5 Mbps, (b) 3 Mbps, (c) 2.5 Mbps

Robustness Test on Frame Cropping. Figure 9(a) is the watermarked frame and Figure 9(b) is the cropped version of Figure 9(a). The missing parts are then replaced by the same parts from the original unwatermarked frame, as shown in Figure 9(c). Figure 9(d) is the extracted logo image after the cropping attack. We can extract the logo image only with small quality degradations from the cropping attack.

(a) (b) (c) (d)

Fig. 9. Experimental results to frame cropping : (a) the watermarked frame, (b) the cropped version of (a), (c) the replaced version, (d) the extracted logo image

5 Conclusions

In this paper, we propose a new masking model for video watermarking based on the characteristics of the human visual system(HVS). In order to design the general watermarking scheme, we embed the watermark signals in the uncompressed video sequence. In this paper, we define an HVS-optimized global masking map for the best trade-off between invisibility and robustness. We generate the global masking map by combining the frequency, the spatial and the motion masking effects. After embedding the watermark signal using the information from the global masking map, we control the amount of watermarks with the control parameters. Experimental results show that the proposed method is imperceptible to human eyes, and also good in terms of watermark capacity. In addition, our method is robust against the various attacks, such as MPEG-2 coding, MPEG-2 re-encoding, and frame cropping. We have observed that the logo images under those attacks are extracted properly only with slight degradation of image quality.

Acknowledgments. This work was supported in part by Kwangju Institute of Science and Technology, in part by the Korea Science and Engineering Foundation (KOSEF) through the Ultra-Fast Fiber-Optic Networks (UFON) Research Center at K-JIST, and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

1. Hsu, C.T.: Digital watermarking for images and videos. Ph. D. dissertation, Commun, Multimedia Lab., National Taiwan Univ. (1997)
2. Hartung, F. and Girod, B.: Digital watermarking of raw and compressed video. Proc. SPIE Digital Compression Technologies and Systems for Video Commun, vol. 2952, Oct. (1996) 205–213
3. Watson, A.B.: DCT quantization matrices optimized for individual images. Human Vision, Visual processing, and Digital Display IV, SPIE-1913 (1993) 202–216
4. Peterson, H.A., Peng, H., Morgan, J.H., and Pennebaker, W.B.: Quantization of color image components in the DCT domain. Human Vision, Visual Processing, and Digital Display. Proc. SPIE (1991) 210–222
5. Schreiber, W.F.: Image processing for quality improvement, Proc. IEEE, vol. 66, Dec. (1978) 1640–1651