A User Attention Based Visible Video Watermarking Scheme

Chen-Hsiu Huang and Ja-Ling Wu
{chenhsiu,wjl}@cmlab.csie.ntu.edu.tw
Communication and Multimedia Laboratory,
National Taiwan University,
Taipei, Taiwan 106, ROC

Abstract

In this paper, a content dependent visible watermarking scheme for video is proposed. By using the focus detection framework developed in [10], we calculate the unsuitability contrary to the focus region that catches user's attention the most and embed a visible watermark to the region with the highest unsuitability. A scaling coefficient is also calculated according to local properties of the host image to reduce the effect of watermarking. The watermark embedding is performed shot by shot, so that the watermark's movement will not be too annoying.

Keywords: Visible video watermarking, content dependent watermarking, user attention model, focus detection framework.

1. Introduction

Digital watermarking is defined as a process of embedding data, called a watermark, into a multimedia objects such that the embedded watermark can be detected or extracted later to make an assertion about the objects. The multimedia objects can be text, image, audio, video and their compositions.

Digital watermarking techniques can be divided into various categories basing on the applying domain (spatial or frequency domain) and the type of embedding documents (image or video). According to human perceptivity, digital watermarks can be roughly divided into two different types, visible and invisible. Visible watermark is a second transparent pattern or image overlaid with the primary (host) image. The embedded watermark appears visible to a casual viewer undergoes a careful inspection. The invisible watermark is embedded in such a way that alternations made to the pixel value are perceptually not noticeable and can be recovered by using an appropriate decoding mechanism.

Some of the desired characteristics about visible watermarks, as described in [1] and [2], are listed below:

(i) A visible watermark should be obvious in both color and monochrome images.
(ii) The watermark should be spread in a large or important area of the image in order to prevent its deletion by clipping.
(iii) The watermark should be visible yet must not significantly obscure the image details beneath it.
(iv) The watermark must be difficult to remove; removing a watermarked image should be more costly and labor intensive than purchasing original from the owner.
(v) The watermarking process should be applied automatically with little human intervention and labor.

The criterions mentioned above illustrate the general purpose of visible watermarking quite well but there are some challenges when designing visible watermarks:

For example, how can we spread the embedding watermark in a large or important area of the image, without violating the third criterion, i.e. not significantly obscure the image details beneath it? Moreover, if we can achieve these two goals simultaneously, how can this embedding process being labor saving, that is, without too much user intervention so as not to violate the fifth requirement mentioned above?

A few visible image watermarking schemes have been proposed so far [3,4]. The authors in [4] defined a measure equation to modify block DCT coefficients of an image basing on a mathematical model, developed by exploiting the texture sensitivity of the human visual system (HVS). This model based approach ensures that the perceptual quality of host images is better preserved when visible watermarks are embedded. Bo Tao et al. [5] used a regional perceptual classifier to assign a sensitive index to each region for achieving adaptive watermark embedding.

As for video watermarking, Jianhao Meng et al. [6] have proposed a video visible watermarking scheme in the compressed domain. In their work, the embedded watermark adapted to the local video features, such as brightness and complexity, to achieve consistent perceptual visibility. For real-time issues, they also assume that the video content will not changed too dramatically within a Group of Pictures (GOP).

All of the prescribed visible watermarking scheme are trying to do the same thing: modeling the content property from the low-level features (edge, background texture, or intensity) as a clue to adapt the embedded watermarks to the hosting contents. We denote this kind of embedding schemes as content-dependent watermarking. Some of the video watermarking schemes still treat video data as a series of pictures and apply some image watermarking scheme to each frame of the video as an another solution [6], [7] and [8]. This approach may not good enough for video watermarking because the most significant properties of video, temporal correlation and motion information are not taken into account.

We believe that the content dependent watermarking could benefit from some research results in the multimedia content analysis field. In [9], for example, Yu-Fei Ma et al. present a generic framework of video summarization based on the modeling of viewer's attention. By taking advantage of computational attention models, we have more objective references for choosing the locations to insert watermarks.

Recently, Chia-Chiang Ho et al. [10] have proposed a user-attention based focus detection framework for video. In this paper, we based on their work to identify user's attention (i.e. the user's focus region) first, and then embed visible watermarks into regions far away from the identified focus regions in each video shot. Other low-level features, such as texture and intensity, are also considered for providing better perceptual feeling. Since the motion and temporal information in videos are exploited in the user attention based focus detection framework, the proposed visible video watermarking scheme can be regarded as content dependent one.

The rest of paper is organized as follows: Section 2 briefly describes the adopted focus detection framework. Section 3 introduces the proposed watermarking scheme. Section 4 presents the simulation results. Section 5 concludes this write up.

2. The User Attention Model

Attention refers to the ability of a human to focus and concentrate upon some visual or auditory object, by careful observing or listening. Assuming limited processing resources a human has, attention, in some sense, also refers to the allocation of these resources [11]. Here, the resource refers to either neurological or cognitive resources. The former is often referred as bottom-up attention and the later top-down attention. Bottom-up attention [12] models what people are attracted to see. Saliency of early visual features is computed to form a set of feature maps. On the other hand, top-down attention was usually modeled by detecting some meaningful (semantic) objects or video features.

The framework developed in [10], which integrates color and intensity (low level), motion (medium level), and face (high level) saliency maps, to model user attention. For features in different levels, various saliency maps are generated to capture users' focus. Then a priority-based or linear combination rule is applied to fuse them into one integrated saliency map (see Figure 1). Once the final saliency map is generated, the focus point can be detected. Moreover, the focus point detection results will model the observer's attention, successfully.

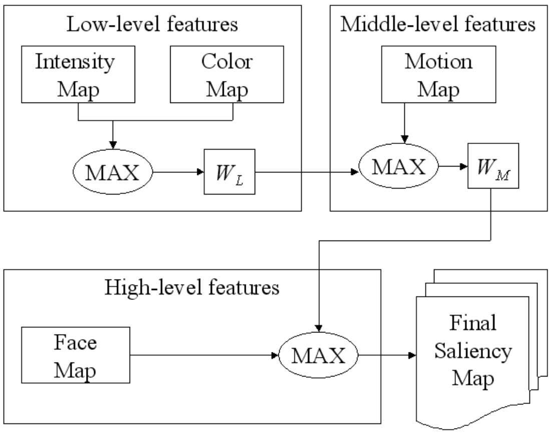
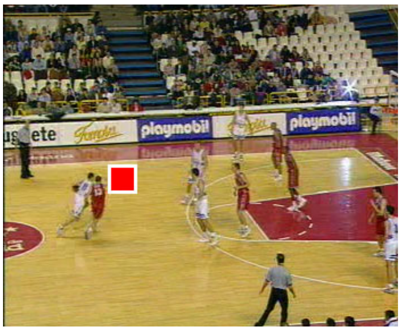


Figure 1. The priority-based fusion scheme for calculating the final saliency map.

(a)



(b)

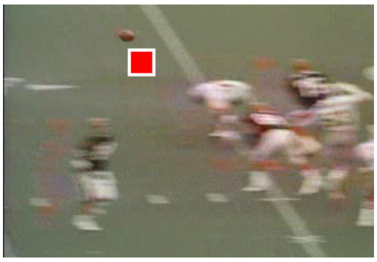


Figure 2. Focus point detection results from (a) a basketball game and (b) a football game video sequences.

Users tend to pay more attention to the focus regions in a video, for example, the major players or characters in sports or movie videos. While users staring at the focus region, they usually pay less attention to other regions, and it is our belief that these non-focus regions are good candidates to embed watermarks.

Figure 2 shows the focus detection results from two sports video sequences. The red square on each frame marks the current focus point while catches user's attention the most. In Figure 2 (a), the player in white cloth was approaching to the three-point line in once offensive and the player in red was trying to defense against the attack; this shot surely catches users' focus the most. The focus detector also identifies the focus point at one offensive long pass of a football game, as shown in Figure 2 (b).

3. Content-dependent Visible Video Watermarking

In this section, we describe the proposed content dependent video watermarking scheme. Firstly, we select the target location for embedding watermarks from the candidate non-focus regions. Secondly, we calculate the watermark's scaling coefficient according to the background attributes; and thirdly, we embed the watermark as a transparent mask onto each frame. At last, we perform the watermarking process on each video shot unit to get the final result.

3.1 Selection of Watermark Location

As we discussed in section 2, the focus point indicates the most attractive region where users eager to see. On the contrary, for those regions far away from the focus point are good candidate positions to embed visible watermarks. Besides, in order to decrease the perceptual distortion after watermarking, the regions of lower intensity and higher texture are the most suitable ones. Thus, we define a weighting function to measure each macroblock's unsuitability $F_{us}(x,y)$:

$$F_{us}(x,y) = \frac{(x-x_f)^2 + (y-y_f)^2}{f_I(x,y) \times f_\rho(x,y)}, \qquad (1)$$

where

$$f_I(x,y) = \frac{1}{mn} \sum_{i=x}^{x+m} \sum_{j=y}^{y+n} \frac{I_{ij}}{255} \qquad (2)$$

and

$$f_\rho(x,y) = \frac{1}{mn} \sum_{i=x}^{x+m} \sum_{j=y}^{y+n} \rho_{ij}\ . \qquad (3)$$

Here mn is the size of embedded watermark in a macroblock, (x,y) denotes the coordinate of each macroblock and (x_f, y_f) represents the macroblock where the focus point resided in. $f_I(x,y)$ and $f_\rho(x,y)$ are functions to measure the strength of local intensity and the degree of texture complexity, respectively. The meterage of intensity is performed on the luminance channel of the YUV colorspace, in which I_{ij} is the average intensity value of the macroblock (i,j). For characterizing textures, we count the percentage of zero quantized DCT coefficients of each macroblock, known as the ρ-value [13]. Both of the luminance intensity, I_{ij}, and ρ-values, ρ_{ij}, can be directly extracted and calculated from the MPEG bitstreams, in the macroblock level.

Each macroblock's unsuitability, in the current video frame, is calculated by equation (1). The one with the highest unsuitability is the selected location for embedding the watermarks.

3.2 Watermark Embedding

A gray scale image of size 144×144 in width and height is showed in Figure 3. It is chosen as the watermark to be added to the original frame in the luminance channel. During embedding, we modify each video frame pixel by pixel in the spatial domain.

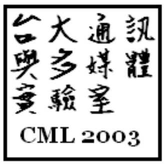


Figure 3. The image used as a watermark

Since our method chooses the non-focus regions to embed watermarks, and therefore, may not raise user's attention when they concentrate very much on video's focus. However, in order not to obscure the beneath image too seriously, we still select a luminance scaling coefficient φ based on the background's intensity and texture to scale the watermark image's strength properly. The scaling coefficient φ is calculated by:

$$\varphi = 0.5 \times (1 + \max(R_I, R_\rho)), \qquad (4)$$

where

$$R_I = (\frac{1}{n} \sum_{i=1}^{n} \frac{I_i}{255}) \qquad (5)$$

and

$$R_\rho = 1 - (\frac{1}{n} \sum_{i=1}^{n} \rho_i)\ . \qquad (6)$$

In eqns. (5) and (6), i denotes each macroblock in the embedding region and n is the total number of macroblocks covered by the watermark image. ρ_i and I_i are the same as the ones described in Section 3.1. R_I and R_ρ represent the strength of luminance intensity and the degree of texture complexity. The selected pixels are modified as follows:

$$\hat{p}_{i'j'} = p_{i'j'} \times \varphi + w_{i''j''} \times (1-\varphi)$$
$$\text{if } w_{i''j''} \neq 255 \qquad (7)$$

where $p_{i'j'}$ and $\hat{p}_{i'j'}$ are respectively the pixels, at frame coordinate (i', j'), before and after modification; $w_{i''j''}$ is the watermark pixel value at the coordinate (i'', j''). Notice that the watermark pixel $w_{i''j''}$ with value 255 is treated as transparent. Figure 4. shows the watermarking results of the two sports videos shown in Figure 2.

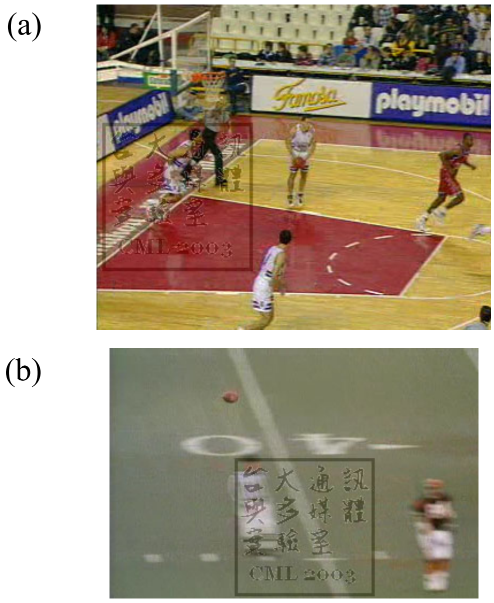


Figure 4. The example frames of two aforecited sport videos. Both of the watermarks appeared in the proper locations, which can declare the copyright and not obscure the video attention focus, at the same time.

3.3 Video Shot Based Watermarking

Since the embedding process is performed on frame level, the selection result may be different from time to time. That implies the embedded watermark may look like another unstable moving object in the video, which is quite annoying. For resolving this shortage, we change our policy to calculate the watermarking location only once in a video shot. The watermark appears in the same location during each shot and only moves when video shot-change occurs.

The shot-change detection is realized by using the measurement of minimal square error of pixel values between two adjacent video frames. More advanced shot detection algorithms can be applied to get better performance and some of them have been discussed in [14].

4. Results and Discussion

The watermark used for embedding is approximately quarter of the video frame size, which is large enough for preventing from deletion by clipping and achieving the goal of copyright assertion.. The watermark is prominent in the non-focus regions and can easily be identified by users.

Comparing with [6], our watermark may be moved to any regions that out of the users' attention, so that the content owner's rights claim is preserved without spoiling the video's presentation. Besides, the selection of location to embed watermark is performed only once in each video shot and will not annoy the original video due to frequently movements. Visible video watermarking scheme proposed in [6] generate a randomized location shifting for watermark masks to enhance the security issue, and hence increase the difficulty for attackers to remove the watermark. In our approach, watermark differs in each shot both in its location and strength. The appearance of watermark cannot be predicted in advance due to its content-dependent nature, and therefore, can hardly be removed by attackers, either.

At last, the proposed watermarking can be done fully automatic, i.e. it is labor saving and without user's intervention. This goal is not easily to be achieved especially for video watermarking where keeping good perceptual quality is a must.

5. Conclusion

In our opinion, borrowing techniques developed in the content analysis field, such as modeling of user's attention, is beneficial to digital watermarking. Currently, many research efforts have been devoted to multimedia content analysis, trying to break the barriers between the low-level features and high-level semantics. And we believe that by exploiting more content features with higher semantic meanings, approaching a better digital watermarking scenario could be possible.

6. Reference

[1] Yeung M.M., et al., "Digital Watermarking for High- Quality Imaging", *Proc. of IEEE First Workshop on Multimedia Signal Processing*, Princeton, NJ, pp. 357-362, June 1997.
[2] F. Mintzer, G. Braudaway and M. Yeung, "Effective and Ineffective Digital Watermarks", *Proc. of IEEE International Conference on Image Processing ICIP-97*, Vol.3, pp. 9-12, 1997
[3] G. W. Braudaway, et. al., "Protecting Publicly Available Images with a Visible Image Watermark", *Proc. SPIE Conf. Optical Security and Counterfiet Deterrence Technique*, Vol. SPIE- 2659, pp.126-132,
[4] Saraju P. Mohanty, K. R. Ramakrishnan and Mohan S. Kankanhalli, "A DCT Domain Visible Watermarking Technique for Images", *IEEE International Conference on Multimedia and Expo (II)*, pp. 1029-1032, 2000
[5] Bo Tao and Bradley Dickinson, "Adaptive Watermarking in the DCT Domain", *International Conf.on Accoustics, Speech, and Signal Processing, ICASSP '97*
[6] Jianhao Meng and Shih-Fu Chang, "Embedding Visible Video Watermarks in the Compressed Domain," *Proc. of ICIP International Conference on linage Processing*, Oct. 1998
[7] Swanson, Mitchell D., Zhu et. al., "Object-based Transparent Video Watermarking", *Electronic Proceedings of the IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing*, June, 1997
[8] Chiou-Ting Hsu and Ja-Ling Wu, "DCT-Based Watermarking for Video," *IEEE Trans. Consumer Electronics*, vol. 44, no. 1, Feb. 1998, pp. 206-216
[9] Yu-Fei Ma, Lie Lu, Hong-Jiang Zhang and Mingjing Li, "A User Attention Model for Video Summarization," *ACM Multimedia, Dec. 2002*
[10] Chia-Chiang Ho, Wen-Huang Cheng, Ting-Jian Pan, and Ja-Ling Wu, "A User-Attention Based Focus Detection Framework and Its Applications", *the 4th IEEE Pacific-Rim Conference on Multimedia, 2003*
[11] Chia-Chiang Ho, "A Study of Effective Techniques for User-Centric Video Streaming", Ph.D. dissertation, National Taiwan University, Taipei, Taiwan, June, 2003
[12] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194-203, 2001.
[13] Zhihai He; Yong Kwan Kim; Mitra, S.K., "Low-delay rate control for DCT video coding via rho-domain source modeling," *Circuits and Systems for Video Technology, IEEE Transactions on , Volume: 11 Issue: 8 , Aug. 2001 Page(s): 928 -940*
[14] R Lienhart, "Comparison of automatic shot boundary detection algorithms". *In SPIE Conf. on Storage and Retrieval for Image & Video Databases VII*, volume 3656, pages 290--301, 1999