Video Watermarking on Overlay Layer

Ji-Won Lee, Tae-Woo Oh, Min-Jeong Lee, and Heung-Kyu Lee
Department of Computer Science
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
{jwlee, twoh, mjlee, hklee}@mmc.kaist.ac.kr

Hae-Yeoun Lee
Department of Computer Software Engineering
Kumoh National Institute of Technology
Gumi, Gyeongbuk, Republic of Korea
haeyeoun.lee@kumoh.ac.kr

***Abstract*—There exist various video watermarking schemes to protect copyrighted video content. Since most of these schemes exploit the perceptual properties of an original video content, they ineludibly had high computing complexity to embed watermark. However, some video applications such as settop box (STB) don't have enough computing power to insert the watermark into the high definition (HD) content using existing schemes. In this study, we propose an overlay video watermarking scheme that embeds watermark into the overlay layer of the multimedia device. Since the proposed scheme exploits only the traits of human visual system (HVS) without any information about original video data, it achieves both low computational complexity and good imperceptibility. Further, the experimental results prove that the proposed scheme has good fidelity and enough robustness.**

***Keywords*-copyright protection; overlay layer; overlay watermarking; video watermarking**

I. Introduction

Copyright protection of the multimedia content is one of the most sensitive issues in the age of information. Content providers try to protect their intellectual property, but it is difficult due to the advance of networks and the availability of low-cost multimedia devices. We can easily find out the illegal copies of the copyrighted multimedia content in the peer to peer (P2P) or torrent sites. Many illegal copies were recorded in the cinema or from the internet protocol television (IPTV) using the camcorder. Practically, the portion of illegal copies from the IPTV contents is getting larger. Therefore, it is urgent to find an appropriate solution to resolve this problem.

To prevent the distribution of illegal contents, many practical video watermarking techniques are introduced [1], [2], [3], [4], [5]. The watermark embedding domain of each technique is different: Some techniques embed the watermark signal into the spatial domain, and some other techniques use the discrete cosine transform (DCT) domain. However, since every technique has to have the robustness against the various attacks, they embed the watermark signal as strong as possible unless the embedded watermark is invisible. To do this, these schemes make the watermark pattern considering the perceptual properties of an original content. In this process, the original content undergoes the inevitable quality degradation by modifying the original data. Further, it is required the massive computing power when the scheme computes the perceptual properties of a content and embeds the watermark signal. If the resolution of a target video content is small and the device has the high computing power, real-time embedding may be possible. However, the portion of HD video contents is getting larger and the most multimedia devices such as STB have not enough computing power to embed the watermark into the HD content in real-time yet. Therefore, it is required to develop a video watermarking scheme that satisfies both low computational complexity for high quality videos and robustness against the various attacks. In this situation, the overlay layer can be exploited as an embedding domain. In this study, we propose an invisible overlay video watermarking scheme by exploiting the overlay layer and the HVS.

The remainder of this paper is organized as follow. In Sec. II, the overlay layer where the watermark information will be embedded is introduced. In Sec. III, the embedding and detection processes of the proposed scheme are discussed. Experimental results to prove the effectiveness of the scheme are presented in Sec. IV, and Sec. V concludes.

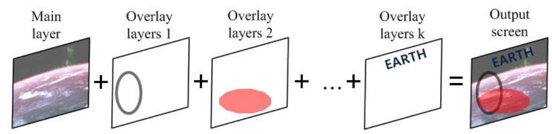


Figure 1. Relationship between the output screen and the layers

II. Overlay layer

The output screen of general multimedia devices consists of several layers. These layers can be categorized into two types: the main layer and the overlay layer. The main layer is compulsory and used for displaying the video content. On the other hand, the overlay layer is optional and used for displaying the status of the content or device. Also, there can exist more than one overlay layer in one output screen. Fig. 1 shows the relationship between the output screen and the layers. Since every layer is independent and each pixel in the layer has own opacity level, the pixel color of the output screen is determined by blending the corresponding pixel colors and opacity levels of every layer.

978-0-7695-4517-2/11 $26.00 © 2011 IEEE
DOI 10.1109/IIHMSP.2011.52

There are a few alpha blending formulas to determine the color of output screen. Among the alpha blending formulars, Eq. 1 is the most widely used formula [6], [7].

$$L^{k}(x,y) = L^{k-1}(x,y)\cdot\frac{255-\alpha^{k}(x,y)}{255} + L_{o}^{k}(x,y)\cdot\frac{\alpha^{k}(x,y)}{255} \quad (1)$$

where $1 \leq k \leq Num_o$. Here, Num_o is the number of overlay layers. $\alpha^k(x,y)$ is the opacity level of k-th overlay layer and the range of $\alpha^k(\cdot)$ is from 0 to 255. $L^0(x,y)$, $L^k(x,y)$, and $L_o^k(x,y)$ are the pixel value of the main layer, the k-th temporary blended layer, and the k-th overlay layer, respectively. When k is equal to Num_o, $L^k(x,y)$ represents the pixel value of corresponding output screen.

In this study, we assume that the multimedia device employs Eq. 1 to compute the output screen.

III. WATERMARK DESIGN

A. Embedding process

The goals of the proposed embedding scheme are achieving good robustness, high imperceptibility, and low computational complexity. Since the watermark signal is generated without helping the perceptual properties of an original content and simply added to the overlay layer, the proposed scheme achieves low computational complexity. However, it is difficult to achieve both good robustness and high imperceptibility together. To achieve them, we generated the embedding pattern using the traits of human vision. The proposed embedding process is illustrated in Fig. 2.

The spread spectrum (SS) method can embed the noise-like signal into the host channel without knowledge of that channel. Since we do not use the data of an original content, the SS method is the best embedding method for us. To make the pattern to be embedded, the watermark information is used as keys to generate the rectangular binary pattern whose size is $m \times n$ using the pseudo-random number generator. Since there are many research results about the luminance stimuli of HVS claiming that the human vision cannot perceive small luminance difference, we select the luminance channel as the host channel. To express the luminance in the pattern, we used the white and black color and the opacity level. When the pattern element is

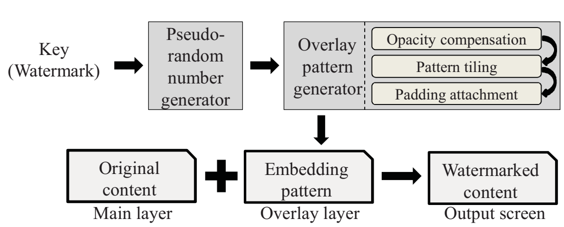


Figure 2. Overlay watermark embedding process

1, the corresponding color is white (whose luminance value is 255). On the contrary, the corresponding color is black (whose luminance value is 0) when the element of the pattern is 0.

The generated pattern is then post-processed to be appropriated for the invisible and robust pattern. First, the adequate opacity level of the pattern is calculated in the opacity compensation step. If the opacity level of the pattern is too low, the embedded watermark is not robust at all. On the other hand, if the opacity level is too high, the pattern may visible. Therefore, the selection of adequate opacity level is the most important step in the embedding process.

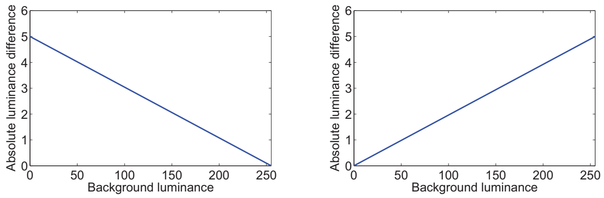
According to Weber's law and the Chou's study [8], the visibility threshold of human vision depends on the background luminance value. Also, Chou *et al.* proposed the formula of visibility threshold. However, since the luminance expressiveness of digital devices is advanced significantly and the quality of video contents is improved, the confidence of visibility threshold formula in [8] is weak for current environments. Therefore, we re-tested and adjusted the coefficients of Chou's visibility threshold formula that can express the difference of maximum imperceptible luminance of human eyes. To get the adequate coefficients, twenty observers were participated in the test. The adjusted visibility threshold formula is as follows.

$$f(bg(x,y)) = \begin{cases} T_0\cdot(1-(bg(x,y)/120)^{1/2})+1.5 & \text{for } bg(x,y)\leq 120 \\ \gamma\cdot(bg(x,y)-120)+1.5 & \text{for } bg(x,y) > 120 \end{cases} \quad (2)$$

Here, $f(\cdot)$ is the visibility threshold due to background luminance, and $bg(x,y)$ is the background luminance of pixel position (x,y). T_0 and γ denote the visibility threshold when background grey level is 0, and the slope of the line that models the function at higher background luminance, respectively. In here, we empirically decide that 4 and $1/64$ is appropriate for the value of T_0 and γ, respectively.

To select the adequate opacity level of the pattern, we have to find out the relationship between the opacity level and the luminance value first. If the background area is overlaid with the maximum luminance valued area, i.e. white colored area, with the opacity level α_w, the luminance difference between the overlaid and overlaying area decreases linearly when the background luminance value increases. The maximum luminance difference is the same as α_w when the background luminance value is 0. On the other hand, if the minimum luminance valued area, i.e. black colored area, with the opacity level α_b overlays on the background area, the aspect is reversed. Fig. 3 shows the absolute luminance difference between overlaid and overlaying area.

To get both robust and imperceptible watermark pattern, the luminance difference caused by overlaying the pattern should have the maximum value under the condition of keep-



(a) Overlay the maximum luminance valued area (255) with certain fixed opacity level

(b) Overlay the minimum luminance valued area (0) with certain fixed opacity level

Figure 3. Absolute luminance difference between overlaid and overlaying area when $\alpha_w = \alpha_b = 5$

ing lower than the maximum visibility threshold induced in Eq. 2. Thus, we set α_w, α_b to 2, 3 in the embedding process. The relationship among α_w, α_b, and the visibility threshold is shown in Fig. 4.

In the pattern tiling step, we tiled the pattern vertically v times and horizontally h times to have more robust pattern. Then, the vanished padding is attached to make the pattern more imperceptible. The pattern is theoretically invisible, however, the boundaries could be visible due to mach band effect [8]. Human eyes perceive darker or brighter next to the boundary between dark and bright area. To alleviate it, the proposed scheme attaches the padding that vanished gradually next to the boundaries of the pattern. Each pixel opacity level $\alpha(x, y)$ in the padding is calculated as

$$\alpha(x, y) = \frac{D_{pl}((x, y), B_{pad})}{D_{ll}(B_{emb}, B_{pad})} \times \alpha_w \tag{3}$$

where B_{emb} and B_{pad} are the boundary lines of the pattern and padding, respectively. $D_{pl}(P, l_a)$ is the point-line distance between a point P and a line l_a. $D_{ll}(l_b, l_c)$ is the line-line distance between two lines l_b and l_c. The pixel luminance in the padding is randomly chosen between 0 and 255. Fig. 5 depicts the padding of the pattern.

After padding attachment step, the pattern is added to region of interest (ROI) of the overlay layer and the overlay layer is merged to the main layer.

B. Detection process

Since the watermark information is embedded in the video clip as a sort of noise, we can estimate the watermark

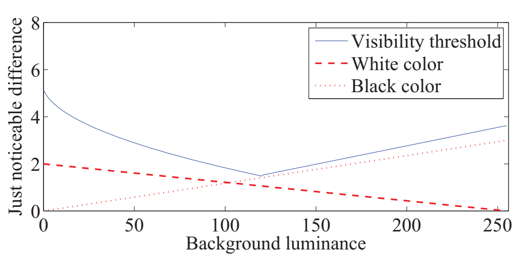


Figure 4. Maximum opacity level that satisfies the imperceptibility

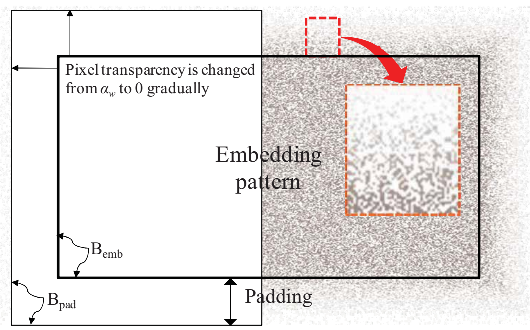


Figure 5. Padding of the pattern (Left half: concept description, Right half: real watermark pattern)

pattern of each frame by calculating difference between the watermarked ROI and the denoising filtered ROI. In this study, we use an adaptive Wiener filter as a denoising filter.

The embedded pattern and the estimated pattern after accumulation is denoted by P_{emb} and P_{est}, respectively. After estimating P_{est}, normalized cross correlation (NCC) between P_{emb} and P_{est} is calculated to detect the embedded watermark information. More mathematically we have

$$P_{det} = \begin{cases} P_{emb} & \text{if } NCC(P_{est}, P_{emb}) > \tau \\ Undetermined & \text{otherwise} \end{cases} \tag{4}$$

where $NCC(\cdot, \cdot)$ is the maximum element of NCC. τ is the detection threshold defined by

$$\tau = \mu_{ncc} + \kappa_{ncc}\sigma_{ncc} \tag{5}$$

Here, μ_{ncc} and σ_{ncc} are the average and standard deviation of NCC, respectively. κ_{ncc} is a constant that controls a false positive error.

Then, we can determine the embedded watermark information using the calculated P_{det}.

IV. Experimental results

This section shows experimental results to measure the fidelity and robustness of the proposed scheme. We carried out experiments under two different environments: computer simulated environment and STB ported environment which is one of the most representative multimedia device. Since objective fidelity cannot be measured in the STB directly, we simulated the proposed scheme in personal computer. We used four 5-minute length HD video clips from various genres. The size of pattern $m \times n$ was 512×256 and the tiling size v and h were 2 and 2, respectively. Thus, the final pattern size was 1024×512 and it was embedded in the middle of the HD frame.

A. Simulation test

In order to show the fidelity of the proposed embedding scheme, we simulated the overlay layer using Eq. 1 and measured the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [9]. The high values of PSNR and

Table I
THE BIT ERROR RATE UNDER VARIOUS DISTORTIONS (UNIT: %)

Distortion type	Test video clips			
	Clip 1	Clip 2	Clip 3	Clip 4
No attack	0	0	0	0
Scaling from 1920×1080 to 960×540	2.46	2.56	2.57	2.54
Format conversion from MPEG2 to MPEG4	1.50	1.57	1.49	1.51
Frame rate conversion from 30 fps to 24 fps	2.44	2.53	2.51	2.50
Compound attack (apply above distortions)	4.43	4.42	4.43	4.45

SSIM indicate that the watermarked video clips are similar with original one. The average PSNR and SSIM values were 48.5376 dB and 0.9952, respectively. Therefore, human eyes may not feel uncomfortable and unsatisfactory caused by the embedded watermark signal.

To measure the robustness, the watermark detection was performed on the watermarked video clips which were distorted by various attacks. The results of the bit error rate (BER) are described in Table I. The calculated BER proves that the proposed scheme has the robustness even if watermarked video clips were distorted in various ways.

B. STB ported test

We ported the proposed scheme into two STBs (Dasan networks H925K and Celrun TD9000) and recorded the video contents using the DVI recorder. As shown in Table II, the proposed scheme has robustness when the watermarked content undergoes the HDMI-VGA / VGA-HDMI attack.

To measure the imperceptibility of the scheme, we performed the subjective fidelity test similar with [10]. Five expert observers, adapt at our scheme and able to detect visual artifact of watermarked video content, were participated in the experiment in which each trial consisted of two presentation of the same video, once with and once without watermark. After six trials, the observers reported which of the two videos contained the watermark. In this experiment, any observer could not exactly distinguish between the watermarked one and the original one.

V. CONCLUSION

There are many types of video applications these days. Among them, some applications have not enough computing power that can process high complexity jobs. In this study, we proposed an overlay video watermarking scheme which has both low computing complexity and good robustness. We exploited the traits of human vision to make the pattern and simply added it into the overlay layer of multimedia devices without accessing and modifying the data of an original content. The experimental results show the proposed scheme has high imperceptibility and enough robustness to be applied in low computing powered video applications dealing with the HD content.

Table II
RESULTS OF BER UNDER HDMI-VGA / VGA-HDMI ATTACK FOR STB PORTED SCHEME (UNIT: %)

Test video clips	STB model	
	H925K	TD9000
Clip 1	7.22	7.08
Clip 2	7.06	7.04
Clip 3	6.13	6.11
Clip 4	7.19	7.10

ACKNOWLEDGMENT

This research project was supported by Ministry of Culture, Sports and Tourism(MCST) and from Korea Copyright Commission in 2011.

REFERENCES

[1] P. Callau, R. Darazi, and B. Marq, " Exhibition QIM-based watermarking for digital cinema," in *Proc. SPIE Electronic Imaging*, pp. 72540U-1–72540-U-11, San Jose, CA, USA, January 2009.

[2] M. J. Lee, K. S. Kim, T. W. Oh, H. Y. Lee, and H. K. Lee, "Improved watermark synchronization based on local autocorrelation function," *Journal of Electronic Imaging*, 18(2):1–11, April 2009.

[3] D. Choi, H. Do, H. Choi, and T. Kim, "A blind MPEG-2 video watermarking robust to camcorder recording," *Signal processing*, 90(4):1327–1332, April 2010.

[4] S. Wohlgemuth, I. Echizen, N. Sonehara, and G. Muller, "On privacy-compliant disclosure of personal data to third parties using digital watermarking," *Journal of Information Hiding and Multimedia Signal Processing*, 2(2):270–281, 2011.

[5] F. C. Chang, H. C. Huang, and H. M. Hang, "Layered access control schemes on watermarked scalable media," *Journal of VLSI Signal Processing System for Signal, Image, and Video Technology*, 49(3):443–455, 2007.

[6] T. Porter and T. Duff, "Compositing digital images," in *Proc. SIGGRAPH.*, vol.18, pp.253–259, New York, NY, USA, July 1984.

[7] J. Blinn and F. James,*Jim Blinn's Corner: Dirty Pixels*, Morgan Kaufmann, 1998.

[8] C. J. Nielsen, "Effect of scenario and experience on lnterpretation of mach bands," *Journal of Endodontics*, 27(11):687–691, 2001.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on image Processing*, 13(4):600–612, 2007.

[10] J. Lubin, J. Bloom, and H. Cheng, "Robust, content-dependent, high=fidelity watermark for tracking in digital cinema," in *Proc. SPIE Electronic Imaging*, pp. 536–545, Santa Clara, CA, USA, January 2003.