Perceptual Watermarking for 3D Stereoscopic Video using Depth Information

Min-Jeong Lee, Ji-Won Lee, Heung-Kyu Lee
 Department of Computer Science
 Korea Advanced Institute of Science and Technology (KAIST)
 Daejeon, Republic of Korea
 Email: mjlee@mmc.kaist.ac.kr

Abstract—Three dimensional (3D) video technologies have made noticeable progress and become common. The 3D videos using depth image based rendering (DIBR) technique are getting the spotlight due to its visual reality, however, they are exposed to illegal distribution. To protect 3D video contents from copyright infringement, a 3D video watermarking scheme can be employed. Since 3D display itself may give rise to unnatural artifacts and arouse visual fatigue to audiences, the imperceptibility is considered as the most significant factor. In this paper, a 3D video watermarking which focuses on perceptual embedding is introduced. The proposed scheme exploits visual characteristics, motion on z -axis and pixels to be hidden by rendering, from depth information. For evaluating the proposed scheme, a stereoscopic display environment is set and the mean opinion score (MOS) is measured as a subjective quality measure for the watermarked 3D videos. Experimental results prove that our proposed scheme has less affect on the perceptibility when watching 3D videos. Also, the proposed method is robust against lossy compression and adding Gaussian noise.

Keywords—3D stereoscopic video, watermarking, copyright protection, perceptual embedding

I. INTRODUCTION

The 3D stereoscopic video technologies have made noticeable progress and become common. The 3D video provides binocular parallax information for inducing the perception of 3D depth to the end-users [1]. The DIBR technique has been commonly used because of its effective data representation and realistic view synthesis [2]. The 3D videos using DIBR techniques are getting the spotlight due to its visual reality, however, they are exposed to illegal distribution. They should be protected from copyright infringement. Like conventional 2D videos, video watermarking techniques can be employed to protect copyright of 3D content. However, 3D display itself may give rise to unnatural artifacts that affect the ability to fuse binocularly and cause visual fatigue [3]. Since the end-users become very sensitive to depth perception and feel visual fatigue when they are watching 3D videos, the imperceptibility is significant when watermarking of 3D video.

There are a few works for 3D content watermarking. Campisi presented an object-oriented watermarking scheme for stereo-images [4]. The objects in the scene were recognized from the depth map and selected as regions to be embedded. Koz *et al.* introduced a watermarking method for free viewpoint television system which allows the user to select freely the viewpoint and generates new views of a dynamic scene from any 3D position [5]. They can detect the watermark from the rendered views both for the known and unknown camera positions. Lin *et al.* proposed a multiple watermarking scheme for DIBR 3D images [6]. Since DIBR generates left-eye and right-eye images by warping the color image according to the depth values for displaying, they embedded three different watermarks into the color image to protect not

Figure 1. DIBR procedure.

only the color image but the left-eye and right-eye images. The watermark embedded by above methods may not be perceptible but they cannot guarantee their imperceptibility when the 3D watermarked video displays.

In this paper, a 3D video watermarking which focuses on imperceptibility is introduced. The proposed scheme exploits visual characteristics from depth information: hidden pixels from view after converting left-eye video and right-eye video and areas with high motion on z -axis. The rest of paper is organized as follows. In Sec. II, we introduce the principle of our perceptual embedding. In Sec. III, we describe the proposed watermark scheme. In Sec. IV, we present experiments to demonstrate the performance of the proposed scheme based on the 3D display environment and Sec. V concludes.

II. PRINCIPLE

The data representation of the DIBR technique consists of monoscopic color video and associated per-pixel depth map. In order to render stereoscopic display, two slightly difference perspective views of a 3D scene are reproduced simultaneously [2] (see Fig. 1). In order for imperceptible watermarking for 3D video, the most of depth characteristics are exploited. In relation to spatial and temporal aspects, two important factors are considered as follows:

1) *Hidden pixels from view after converting left-eye video and right-eye video*: Due to use of depth information, it is obvious that some farther pixels in the color video are covered with some nearer pixels when they are rendered to the left-eye video or right-eye video. Figure 2(a) shows an example of color video frame. The pixels to be hidden are calculated and indicated as green and red pixels. Green (light-gray in monochrome view) pixels represent pixels to be hidden in the left-eye video and red (dark-gray in monochrome view) pixels represent pixels to be hidden in the right-eye video.

Figure 2. Hidden pixels and holes due to left-eye and right-eye image conversion: (a) color video frame, (b) rendered frame of left-eye video, and (c) rendered frame of right-eye video. Green (light-gray in monochrome view) and red (dark-gray in monochrome view) pixels represent pixels to be hidden in the left-eye video and right-eye video, respectively. Sky-blue pixels are holes occurred by sharp depth changes.

By rendering Fig. 2(a) by using its depth map, the left-eye video and right-eye video are generated (see Fig. 2(b) and 2(c)). Due to the fact that the objects which had been occluded in the color video have appeared in the left-eye or right-eye video, holes occur and they are indicated as sky-blue dots. As shown in Fig. 2(b), green pixels are hidden in the left-eye video. It is noticeable that red pixels not hidden in the left-eye video are adjacent to holes. Since holes should be filled by interpolation techniques, red pixels will be used to hole-filling and interpolated with holes. Thus, even though the watermark is strongly embedded into the red pixels of the color video, the hole filling algorithm will be used and it will attenuates the visibility of the embedded watermark. Figure 2(c) shows similar appearance that red pixels are hidden and green pixels are adjacent to holes. Therefore, hidden pixels are considered as high weighted embedding positions.

2) *Areas with high motion on z-axis*: According to [7], the users feel more uncomfortable when they are watching high motion 3D videos, whereas they tend to tolerate some level of channel distortion with less motion 3D videos under packet loss scenarios. However, it is well-known that the depth of an object either much nearer to or farther from where the eyes are fixating can be accurately judged despite the images of the object appearing double. It means that if the depth of an object to which the audiences are paying attention suddenly increases or decreases, they feel the object is blurred. Also, the audiences usually lose their attention to the object which is receding from the scene. Sudden depth increase or decrease on the fixed location can be regarded as high motion on z -axis. Therefore, as a result of compromise of these opposed facts, high motion areas on z -axis are overlooked as the position to be embedded. A simple method of subtracting one frame at t from another frame at $(t-1)$ can obtain information about which parts of the scene have changed in x - and y - axes. In the same manner, we can extract motion on z -axis by subtracting $(t-1)$ th depth map from t th depth map. The higher absolute values represent more change on z -axis.

III. PROPOSED SCHEME

A. Watermark Embedding

The insertion of the watermark must not affect the perceptual quality of videos. This constraint has a bearing on the insertion strength of the watermark, to the extent that it is imperceptible

Figure 3. Embedding procedure.

to the human eye. The watermark pattern is generated and then inserted into the color video frames based on spread spectrum way. Figure 3 shows the embedding procedure. The watermark pattern, which follows a Gaussian distribution with zero mean and unit variance, is generated using a secret key and then consists of 2D random sequence. The 2D pattern is then modulated to contain the bit payload. The modulated pattern is embedded using an additive spread-spectrum method with perceptual scaling. The watermark w is embedded into a color video frame F at time t as follows:

$$F_t^w(x, y) = F_t(x, y) + \alpha_t(x, y) \times w_t(x, y), \quad (1)$$

where the perceptual scaling factor α is calculated by multiplying of

$$\alpha(x, y) = \eta(x, y) \times \tau(x, y) \times \zeta(x, y). \quad (2)$$

η is a local weighting factor of the pixel from human visual system (HVS). We employ an HVS function to be optimized for real-time embedding, which adopts eight compass operators as a local weighting function [8]. The compass operator measures gradients in a selected number of directions and can reduce computational costs by utilizing its separable property. τ refers to the watermark embedding strength according to motion on z -axis. Because human

cannot see the details of fast moving object, fast moving objects have larger value for τ , and slow moving objects have smaller value.

$$\tau(x, y) = s_{min} + \frac{\delta_t(x, y)(s_{max} - s_{min})}{\max(\delta_t) - \min(\delta_t)} \quad (3)$$

where δ_t is calculated by subtracting $(t-1)$ th depth map d_{t-1} from t th depth map d_t . s_{max} and s_{min} are maximum and minimum weights, respectively. ζ refers a map of hidden pixels. It is calculated

$$\zeta(x, y) = \begin{cases} s_{max} & \text{if } \gamma(x, y) > 0 \\ s_{min} & \text{otherwise} \end{cases} \quad (4)$$

where function γ is defined as Table I.

Table I
function γ

```

01:  $\exists x$  and  $y$ ,  $\gamma(x, y) = c_L(x, y) = c_R(x, y) = 0$ ;
02: for  $depth = \min(d) : \max(d)$ 
03:    $index = \text{find}(depth == d)$ ;
04:   for  $\exists index$ 
05:      $x_{left} = x + (t_c \cdot f)/(2 \cdot depth)$ ;
06:      $x_{right} = x - (t_c \cdot f)/(2 \cdot depth)$ ;
07:      $c_L(x_{left}, y)++$ ,  $c_R(x_{right}, y)++$ ;
08:     if  $c_L(x_{left}, y) > 1$  OR  $c_R(x_{right}, y) > 1$ 
09:        $\gamma(x, y)++$ ;
10:   end
11: end
12: end
13: return  $\gamma$ 

```

t_c : camera distance
 f : focal length

B. Watermark Extraction

From the watermarked 3D video, the color video is only used to extract the watermark. The watermark extraction is performed by normalized cross correlation. Since a blind detector is used, an approximation of the embedded watermark is obtained by employing Wiener filtering as denoising filter. Subtracting the denoised frame from the captured frame, we obtain an approximate version of the embedded watermark pattern. If the resulting correlation value exceeds a preset threshold, the hidden messages are correctly extracted. Experimentally, we determine the preset threshold depending on an error probability model which follows Gamma distribution model, since we take the maximum value from cross correlation [9].

IV. EXPERIMENTAL RESULTS

On 3D video sequences *Ballet* and *Breakdancers* [10], the fidelity and detection performance of the proposed scheme are measured. The camera resolution is 1024×768 and the capture rate is 15 fps. We embedded the watermark with 64 bits payload and encoded them as AVI movies. The subjective fidelity and detection performance are tested to measure the proposed scheme.

A. Fidelity

For the subjective analysis, a 23" LG Platron W2363D, a full HD 3D monitor with 120Hz refresh rate and 400cd/m^2 maximum brightness, is used in the experiments to display and view

the watermarked 3D videos. The 3D display is calibrated using a NVIDIA GeForce GTX 460 with 3D Vision active shutter glasses. As the recommendation of [11], we adopted the double stimulus continuous quality scale (DSCQS) method that is used for measuring the quality of systems relative to a reference or stereoscopic image coding. The perceived overall quality and depth are rated on a quality categorical scale from 1 to 5, where 1 represents bad image quality/depth perception and 5 represents excellent image quality/depth perception. Furthermore, the scale is labeled with the adjective terms namely *bad*, *poor*, *fair*, *good*, and *excellent*. Each stimulus set contains one watermarked video sequences and its original, uncompressed version of each scene is used as the reference in the evaluation test. The observers were compelled to determine one of the sequences as the original and score it as *excellent*. Then the observers relatively scored the other one compared to the chosen one. Then the differences of the two ratings are normalized to integer values between 0 (*excellent*) and 100 (*bad*). Ten expert observers participated in this tests. We compared the fidelity of the proposed scheme to the noise visibility function (NVF) [12]. Two testing 3D videos were made by embedding a watermark into a 3D video clip using two different perceptual scaling, the proposed scheme and NVF-based scheme, respectively. The PSNR values of the two test videos were maintained approximately same. At this point, we calculate two PSNR values of left-eye video and right-eye video and take an average of them. Note that the PSNR value of the color video is not used. It is because what the observers watch is the rendered video not the color video.

The MOSs for *Ballet* and *Breakdancers* were obtained as Table II. With the same PSNR, the MOSs of the proposed scheme in both cases are better than those of the NVF-based scheme. It is because the pixels to be hidden are indeed occluded by other pixels and the pixels which are supposed to be hidden in the left-eye video but exposed in the right-eye video and *vice versa* are interpolated and thus refined to be not perceptible. Moreover, it can be interpreted that the observers cannot precisely perceive the objects moving fast on z -axis.

B. Robustness

The robustness of the proposed scheme was measured against Gaussian noise addition and lossy compression. For these attacks, the functions of the Matlab are used and the variance parameter of Gaussian noise is specified as if the intensity of the image is in the range [0,1]. For applying lossy compression, each frame of the watermarked 3D video is encoded by JPEG compression and the quality factor of JPEG is in the range [0, 100]. The proposed scheme is compared to the NVF-based scheme while maintaining

Table II
MOS OF THE PROPOSED SCHEME AND NVF-BASED SCHEME UNDER THE SAME PSNR.

<i>Ballet</i>	NVF	Proposed
PSNR	43 dB	
MOS	30.4	10.0
<i>Breakdancers</i>	NVF	Proposed
PSNR	46 dB	
MOS	21.1	11.8

Figure 4. The bit error rates against attacks: (a) Gaussian noise addition and (b) JPEG compression. The variance parameter of Gaussian noise is specified as if the intensity of the image is in the range [0,1] and the quality factor of JPEG is in the range [0, 100].

same PSNR as shown in Table II. Figure 4 illustrates the bit error rates against two attacks. In both cases, the proposed scheme has better robustness compared to the NVF-based scheme. Therefore, it can be said that our scheme provides higher 3D fidelity and better detection performance.

V. CONCLUSION

In this paper, a 3D video watermarking scheme which focuses on perceptual embedding is proposed. Since 3D display is very sensitive so that small change may give rise to visual fatigue to audiences, the fidelity is one of the most crucial issue of 3D video watermarking. The proposed scheme exploits visual characteristics from depth information of DIBR technique. To minimize the visual fatigue when the audiences are watching 3D contents, the embedded watermark is invisible from the audiences' sight by exploiting hidden pixel and z -axis motion. By doing so, we could obtain both visual comfort and robustness.

ACKNOWLEDGMENTS

This work was partially supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract (UD060048AD).

REFERENCES

- [1] R. Patterson, "Human factors of 3-D displays," *Journal of the Society for Information Display*, vol. 15, no. 11, pp. 861–871, November 2007.
- [2] S. Pastoor C. Fehn, R. Barre, "Interactive 3-DTV Concepts and Key Technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, March 2006.
- [3] D. Hoffman, A. Girshick, K. Akeley, and M. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 1–30, November 2008.
- [4] P. Campisi, "Object-oriented stereo-image digital watermarking," *Journal of Electronic Imaging*, vol. 17, no. 4, pp. 043024–1–5, October–December 2008.
- [5] A. Koz, C. Cigla, and A. Alatan, "Watermarking of free-view video," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1785–1797, July 2010.
- [6] Y. Lin and J. Wu, "A novel blind watermarking scheme for depth-image-based-rendering 3D images," in *Proceedings of the 1st international workshop on 3D video processing*, 2010, pp. 39–44.
- [7] S. Yasakethu, C. Hewage, W. Fernando, and A. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1969–1976, November 2008.
- [8] K. S. Kim, H. Y. Lee, D. H. Im, and H. K. Lee, "Practical, real-time, and robust watermarking on the spatial domain for high-definition video contents," *IEICE Trans. Inf. Syst.*, vol. E91-D, pp. 1359–1368, May 2008.
- [9] H. Y. Lee, H. S. Kim, and H. K. Lee, "Robust image watermarking using local invariant features," *Opt. Eng.*, vol. 45, no. 3, pp. 1–11, March 2006.
- [10] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 600–608, August 2004.
- [11] International Telecommunication Union (ITU) Radio communication Sector, "Methodology for the subjective assessment of the quality of television pictures," in *ITU-R BT.500-12*, Sept. 2009.
- [12] I. G. Karybali and K. Berberidis, "Efficient spatial image watermarking via new perceptual masking and blind detection schemes," *IEEE Trans. Inf. Forensics and Security*, vol. 1, no. 2, pp. 256–274, June 2006.