IEICE TRANS. INF. & SYST., VOL.E96–D, NO.9 SEPTEMBER 2013

LETTER

Frame Synchronization for Depth-Based 3D Video Using Edge Coherence

Youngsoo PARK[†a)], ***Student Member***, **Taewon KIM**[††b)], ***and*** **Namho HUR**[†c)], ***Nonmembers***

SUMMARY A method of frame synchronization between the color video and depth-map video for depth based 3D video using edge coherence is proposed. We find a synchronized pair of frames using edge coherence by computing the maximum number of overlapped edge pixels between the color video and depth-map video in regions of temporal frame difference. The experimental results show that the proposed method can be used for synchronization of depth-based 3D video and that it is robust against Gaussian noise with σ = less than 30 and video compression by H.264/AVC with QP = less than 44.

key words: *synchronization, 3D, depth-map, edge coherence*

1. Introduction

A stereoscopic 3D image consists of two images: a left-eye viewed image and a right-eye viewed image, based on binocular disparity, which is one of the characteristics of the human visual system (HVS) [1], [2]. However, stereoscopic 3D video requires double the data space for storage and transmission than conventional 2D video [3]. Therefore, a solution to reduce the data size is necessary, and depth-based 3D imaging is one alternative approach [4]. A depth-based 3D image consists of a mono-viewed color image and its gray-scaled depth-map image [4]. It can be used for synthesizing a stereoscopic 3D image using the depth image based rendering (DIBR) technique; it is also possible to reduce the data size to less than 2/3 compared to that of a classical stereoscopic 3D video [5]. In spite of these merits, depth-based 3D imaging has many problems that must be solved before it can be widely used [6]. In addition to the previously addressed problems [6], synchronization between a color video and depth-map video is another problem. For stereoscopic 3D video, synchronization errors between a left-eye viewed video and right-eye viewed video occur from the use of cameras without generator locking (genlock) or from different recording speeds of the two cameras [3]. Likewise, synchronization errors between a color video and depth video can also occur during the acquiring process. In depth-map video generated from color videos of two views by the stereo matching algorithm, mistakes in the editting process can cause syncronization errors [7]. In addition, syncronization errors can also occur during video storing, encoding/decoding and transmitting for 3D cinema and 3DTV broadcasting service. If the color video and depth-map video of a depth-based 3D video have been desynchronized, it is difficult to convert into a stereoscopic 3D video, and the viewer cannot perceive the proper 3D effects from the video because of the incorrect depth information. Therefore, we propose a method of frame synchronization between the color video and depth-map video for 3D content using edge coherence. The remainder of this letter is organized as follows: In Sect. 2, we explain the synchronization processes, including the proposed method. Section 3 provides information and a discussion on the experimental environments and results. Finally, some concluding remarks are given in Sect. 4.

Manuscript received March 12, 2013.
Manuscript revised May 9, 2013.

[†]The authors are with University of Science and Technology (UST), Korea.
[††]The author is with ETRI, Korea.
a) E-mail: nextstep@ust.ac.kr
b) E-mail: kimm@etri.re.kr
c) E-mail: namho@etri.re.kr

DOI: 10.1587/transinf.E96.D.2166

2. Proposed Method

The color video and depth-map video of depth-based 3D video are visualizations of the same scene. However, these videos have different characteristics caused by their different acquisition systems and methods. For synchronization between a color video and depth-map video, it is necessary to determine the common characteristics of the two video types, which will be used for measuring the correlation according to the delayed or advanced frames. In this letter, we propose a method of measuring the edge coherence by computing the overlapped edge pixels between color video and depth-map video in regions with a temporal frame difference. To reduce the effects of noise, we include the bilateral filtering of videos in the proposed synchronization method.

2.1 ROI Masking with Temporal Frame Difference

We assume that the difference in the edge coherence mainly occurrs in regions of temporal frame difference, which subtract a previous frame from the current frame. Thus, we set a region of interest (ROI) with regions of temporal frame difference to reduce errors and the processing time. We obtain the regions of temporal frame difference from depth-map video because it has a simpler texture than that of color video. ΔD_i, a region of temporal frame difference from a depth-map video, is calculated as follows:

$$\Delta D_i(x, y) = D_i(x, y) - D_{i-1}(x, y) \quad (1)$$

where D_i is the ith frame of the depth-map video, and x and

y are the horizontal and vertical pixel coordinates, respectively. B_i, the binary-map of ΔD_i, is defined as follows:

$$B_i(x,y) = \begin{cases} 1 & |\Delta D_i| \geq T_b \\ 0 & |\Delta D_i| < T_b \end{cases} \quad (2)$$

where T_b is the threshold value. In this letter, we use Otsu's method for an automatic determination of the threshold [8]. To apply the bilateral filter and edge detection, it is necessary to expand the ROI. The expanding ROI mask, M_i, is calculated as

$$M_i = B_i \oplus K \quad (3)$$

where $\oplus$ is the dilation operator. K is the $N \times N$ kernel, all values of which are 1 [9]. We then apply the ROI mask to color video and depth-map video. The masked frames, C'_i and D'_i, are computed as follows:

$$C'_i(x,y) = C_i(x,y) \cdot M_i(x,y) \quad (4)$$

$$D'_i(x,y) = D_i(x,y) \cdot M_i(x,y) \quad (5)$$

where C_i is the ith frame of a gray-scaled color video.

2.2 Bilateral Filtering

The bilateral filter is a non-linear filter proposed by Tomasi and Manduchi to smooth images while preserving the edges [10]. We adopt this to detect edges while reducing the effects of noise. The bilateral filtered image of C'_i, $\hat{C}_i$, is computed as follows:

$$\hat{C}_i(x,y) = \frac{1}{k(x,y)} \sum_{(l,m)\in\Omega} w_s(x,y;l,m) \cdot w_r(x,y;l,m) \cdot C'_i(l,m) \quad (6)$$

where w_s is the domain weight, which indicates the geometric closeness; w_r is the range weight, which is the photometric similarity; and $k(\cdot)$ is the normalizing factor. (x,y) is the central pixel and (l,m) is its neighboring pixel of $N \times N$ filter window, Ω [11]. Similarly, $\hat{D}_i$, the bilateral filtered image of D'_i, can also be computed.

2.3 Detecting Overlapped Edge

To detect an edge, we adopt the Sobel edge detector [9]. The edge binary maps of $\hat{C}_i$ and $\hat{D}_i$, EC_i and ED_i, are calculated as follows:

$$EC_i = \begin{cases} 1 & \hat{C}_i * S \geq T_e \\ 0 & \hat{C}_i * S < T_e \end{cases} \quad (7)$$

$$ED_i = \begin{cases} 1 & \hat{D}_i * S \geq T_e \\ 0 & \hat{D}_i * S < T_e \end{cases} \quad (8)$$

where S is the Sobel edge operator and $*$ is the convolution operator. The threshold of edge detection, T_e, is determined automatically depending on the magnitude of the gradient [9]. The set of overlapped edge pixels between EC_i and ED_i, $H_{i,i}$, is computed as follows:

$$H_{i,i} = \{(x,y) | EC_i(x,y) \cdot ED_i(x,y) = 1\} \quad (9)$$

2.4 Synchronization Using Edge Coherence

For synchronization, we find a pair of frames that has a maximum number of overlapped edge pixels. The calculation used to find a pair of synchronized frames between color video and depth-map video is performed as follows:

$$j = \arg\max_t [n(H_{i,i+t})] \quad (10)$$

where the operator, $n(\cdot)$, is the number of elements in the set; and t is the delayed or advanced frames. From (10), we find that a pair of synchronized frames is (C_i, D_{i+j}).

3. Experiments

3.1 Test Conditions

We used the four depth-based 3D video sequences shown in Fig. 1, *Ballet* (1024×768) and *Break Dancers* (1024×768) provided by Microsoft Research [12]; and *Cafe* (1920×1080) and *Book Arrival* (1024×768) provided by MPEG. In Fig. 1, where the left image is the color video and the right image is the depth-map video that is generated by the stereo matching algorithm from color videos of two views. The sequences of *Ballet* and *Break Dancers* consist of dynamic scenes with fast motions. On the other hand, the sequences of *Cafe* and *Book Arrival* consist of scenes with slow motions. The maximum number of delayed and advanced frames is set to 15, which is the same length as the group of pictures (GOP) of the Advanced Television Systems Committee (ATSC) standard [13]. The parameters of the bilateral filter are set to $N = 5$, the standard deviation for $w_s = 5$, and the standard deviation for $w_r = 20$.

3.2 Test on Original Video

To verify the proposed method, we synchronized the test sequences manually and placed them in order of frame num-

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

Fig. 1 The test sequences, where the left image is the color video and the right image is the depth-map video: (a) *Ballet*; (b) *Break Dancers*; (c) *Cafe*; and (d) *Book Arrival*.

(a) (b)

Fig. 2 A comparison of a portion of the resulting images using the proposed method for *Ballet*; the yellow regions are the ROI, and the blue lines indicate the overlapped edge between the color video and depth-map video: (a) synchronous video ($t = 0$); and (b) non-synchronous video ($t = 1$).

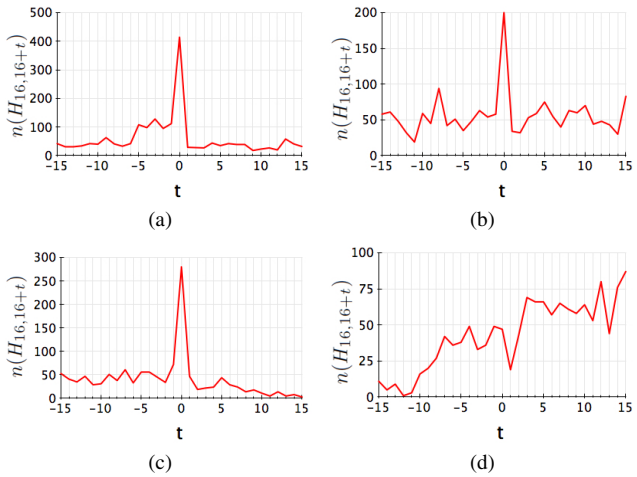

Fig. 3 Graph results for the original video: (a) *Ballet*; (b) *Break Dancers*; (c) *Cafe*; and (d) *Book Arrival*.

ber. We then found the frame of the depth-map video corresponding to the 16th frame of the color video for synchronization using the proposed method. Figure 2 illustrates a comparison of the edge coherence with the images resulting from the proposed method. The left image shows synchronous video ($t = 0$), and the right image is non-synchronous video ($t = 1$). The number of overlapped edge pixels is greater in the left image than in the right image. The graphs in Fig. 3 show the test results for the original video. The number of overlapped edge pixels reaches a peak at $t = 0$ for the graphs of *Ballet*, *Break Dancers*, and *Cafe*. In contrast, the graph of *Book Arrival* is such that a peak is not located at $t = 0$. In other words, the proposed method can be used for the synchronization of *Ballet*, *Break Dancers*, and *Cafe*, but cannot be applied to *Book Arrival*. This is caused by the level of accuracy of the depth-map video. We assume that the depth-map video is correct, and that the synchronized color video and depth-map video have overlapped edge pixels in the boundary. Thus, in the exceptional case in which the accuracy of the depth-map video is very low, it is difficult to obtain a correct result for finding a synchronized pair of frames by the proposed method. Table 1 shows the estimated accuracy of the depth-map video with the proportion of overlapped edge pixels between the color video and depth-map video to the edge pixels in the depth-map video. It is clear from the table that the depth-map video of *Book Arrival* has a lower accuracy than the other sequences.

Table 1 Estimated accuracy of the depth-map video with the proportion of overlapped edge pixels between the color video and depth-map video to the edge pixels in the depth-map video.

Test sequences	Estimated accuracy of the depth-map video
Ballet	0.90
Break Dancers	0.76
Cafe	0.71
Book Arrival	0.34

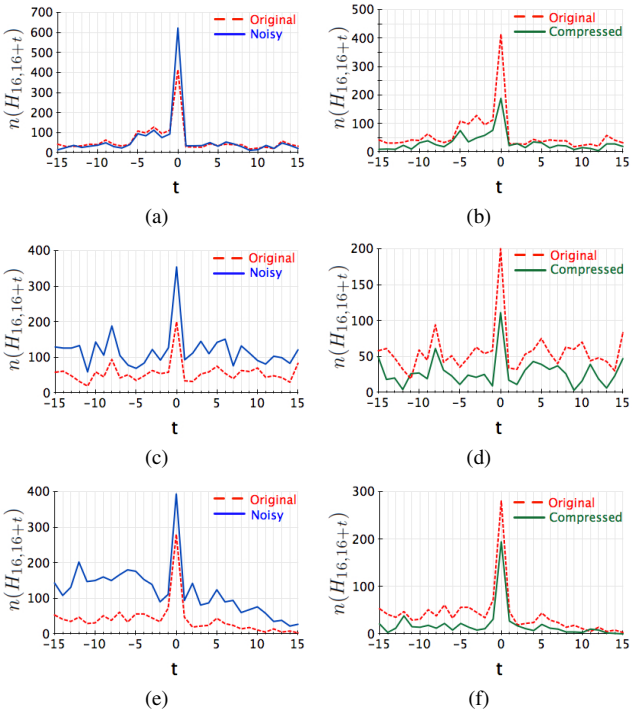

Fig. 4 Graph results of robustness test sequences (PSNR (dB)): noisy video (Gaussian noise, $\sigma = 30$) of (a) *Ballet* (C, 18.68; D, 18.83), (c) *Break Dancers* (C, 19.22; D, 18.70), and (e) *Cafe* (C, 19.10; D, 20.50); and compressed video by H.264/AVC ($QP = 44$) of (b) *Ballet* (C, 33.99; D, 34.20), (d) *Break Dancers* (C, 33.30; D, 34.88), and (f) *Cafe* (C, 34.29; D, 37.80).

3.3 Robustness Test

To verify the robustness against noise and video compression, we added Gaussian noise with $\sigma = 10$, 20, and 30 to the three sequences but omitted for *Book Arrival* because the proposed method could not find a pair of synchronized frames in the test on the original video for the sequence of *Book Arrival*, so we consider that the robustness test against noise and video compression for *Book Arrival* is meaningless. We also compressed the videos using H.264/AVC reference software, JM 18.4, with $QP = 32$, 38, and 44. The graphs provided in Figs. 4 (a), 4 (c), and 4 (e) show the results for robustness against noise. The blue lines indicate the noisiest video ($\sigma = 30$), and the red dotted lines indicate the original video. In this case, the proposed method can find a synchronized pair of frames for all sequnces. The graphs provided in Figs. 4 (b), 4 (d), and 4 (f) show the results for robustness against video compression. The green lines indi-

cate the most compressed video by H.264/AVC ($QP = 44$), and the red dotted lines indicate the original video. In this case, the proposed method can also find a synchronized pair of frames for all sequences.

4. Conclusion

In this letter, we propose a method of frame synchronization between color video and depth-map video for 3D content using edge coherence. The experimental results show that the proposed method can be used for synchronization of depth-based 3D video, and that it has robustness against Gaussian noise with less than $\sigma = 30$ and video compression by H.264/AVC with less than $QP = 44$.

Acknowledgments

This research was supported by the KCC (Korea Communications Commission), Korea, under the ETRI R&D support program supervised by the KCA (Korea Communications Agency) (KCA-2012-11921-02001).

References

[1] M.J. Tovee, An Introduction to the visual system, Cambridge Univ. Press, 2008.
[2] H. Harashima, 3D image and human science, Ohmsha, 2010.
[3] Y. Choi et al., Stereoscopic 3D production workbook, Kocca, 2010.
[4] C. Fehn, "3D-TV approach using depth-based-image-rendering (DIBR)," Proc. VIIP '03, pp.482–487, 2003.
[5] A. Smolic et al., "Coding algorithms for 3DTV — A survey," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.11, pp.1606–1621, 2007.
[6] L. Zhang and W.J. Tam, "Stereoscopic image generation based on depth images for 3DTV," IEEE Trans. Broadcast., vol.51, no.2, pp.191–199, 2005.
[7] B. Mendiburu et al., 3DTV and 3D cinema: Tools and processes for creative stereoscopy, Focal Press, 2011.
[8] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern., vol.9, no.1, pp.62–66, 1975.
[9] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 3rd ed., Peason Education, 2010.
[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," Proc. ICCV '98, pp.839–846, 1998.
[11] L. Ying-Hui, G. Kun, and N. Guo-Qiang, "An improved trilateral filter for Gaussian and impulse noise removal," Proc. ICIMA '10, pp.385–388, 2010.
[12] C.L. Zitnic et al., "High-quality video view interpolation using a layered representation," ACM SIGGRAPH '03, pp.600–608, 2003.
[13] G.A. Davidson, "ATSC video and audio coding," Proc. IEEE, pp.60–76, 2006.