Journal of the Korean Institute of Illuminating and Electrical Installation Engineers Vol. 22, No.7, pp. 37~42 July 2008 (22-7-6)

A Real-time Multiview Video Coding System using Fast Disparity Estimation

Kyung-Hoon Bae* · Byung-Kwang Woo**

Abstract

In this paper, a real-time multiview video coding system using fast disparity estimation is proposed. In the multiview encoder, adaptive disparity-motion estimation (DME) for an effective 3-dimensional (3D) processing are proposed. That is, by adaptively predicting the mutual correlation between stereo images in the key-frame using the proposed algorithm, the bandwidth of stereo input images can be compressed to the level of a conventional 2D image and a predicted image also can be effectively reconstructed using a reference image and adaptive disparity vectors. Also, in multiview decoder, intermediate view reconstruction (IVR) using adaptive disparity search algorithm (DSA) for real-time multiview video processing is proposed. The proposed IVR can reduce a processing time of disparity estimation by selecting adaptively disparity search range. Accordingly, the proposed multiview video coding system is able to increase the efficiency of the coding rate and improve the resolution.

Key Words : Disparity estimation, Stereoscopy, Motion estimation, Intermediate-view Reconstruction

1. Introduction

The human eyes observe different images of each view point, and the human brain then synthesizes the acquired stereo images. As a result, people can see a 3-dimensional (3D) object.[1] For several years, various stereoscopic 3D display systems that imitated the human visual system (HVS) have been implemented in a variety of applications. However, stereoscopic display methods are limited in terms of the viewpoint, implying that people cannot sense the 3D effectively when they are not in the viewing region or when they are in an occluded region. Moreover, owing to fatigue of the eyes, applications are limited.[2] Recently, many research works are actively being done on multiview 3D imaging communication technologies[3]. In the conventional multiview 3D imaging system, as the number of views increases, the system complexity in hardware and software also tends to increase at the same time, so that it is difficult to make the practical implementation of the multiview 3D imaging communication system until now. Especially, multimedia-related technologies are now in its rapid developing so that, it is

* Main author : Samsung Thales Co., Ltd.
** Corresponding author : Seoul University of Venture & Information
Tel : +82-31-280-1641, Fax : +82-31-280-1620
E-mail : khbae.bae@samsung.com
Date of submit : 2008. 6. 17
First assessment : 2008. 6. 25, Second : 2008. 7. 15
Completion of assessment : 2008. 7. 24

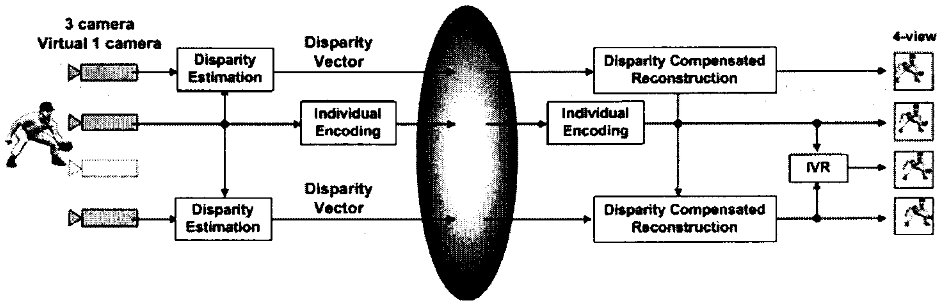


Fig. 1. The proposed multiview video processing

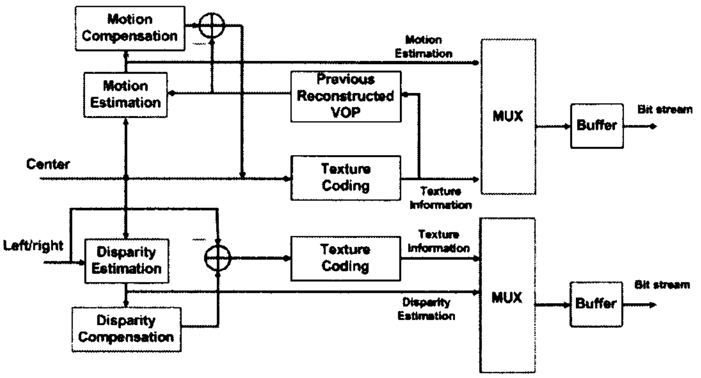


Fig. 2. Multiview video encoder

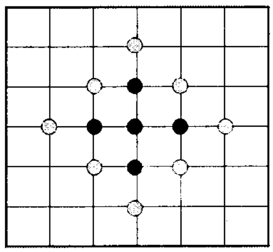


Fig. 3. Diamond search algorithm

obvious that demands on the use of more natural 3D reality media through multiview stereoscopic images[4-5]. However, the images taken from 16~17 cameras are almost impossible to send into one of the media channel. To solve this problem, the technique of taking the intermediate view reconstruction (IVR) is researched a lot[6-7]. But, IVR is impossible to have implementation lively with conventional IVR[8]. Therefore, this paper proposed, in the decoder, IVR using adaptive disparity search range (DSR) for real-time 3D video processing. Also, in this paper, adaptive disparity-motion estimation (DME) is proposed with key-frame coding (KFC) in the encoder. It is possible to develop a fast multiview video coding (MVC) and more efficient compression rate.

2. The Proposed Video Encoding

In the MVC, numerous redundancies exist between stereo images or multiview images. Thus, in this paper, MVC is consist of 4-channel on the MPEG-4 structure. Figure 1 illustrates overall the proposed multiview video processing. Figure 2 illustrates the structure of the proposed multiview video encoder.

2.1 Diamond search algorithm (DSA)

In general, with MPEG-4 video processing, most of the processing time is required for motion estimation. Thus, to reduce this processing time, we use a DSA for fast video processing. Figure 3 illustrates DSA example.

2.2 Adaptive dispairty-motion estimation

In this paper, a cost function for disparity estimation is used to detect the most similar correspondence point between the stereo images. Here, the minimum mean absolute difference (MAD) function is used as a cost function, as expressed by Eq. (1).

$$MAD(k,l) = \frac{1}{N_x N_y}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} | I_L(i,j) - I_R(i+k, j+l) | \qquad (1)$$

In the adaptive DME, feature values extracted from an input stereo image pair are normalized, and the threshold value is optionally set to several steps in the range of 0~1. Correspondingly, an equal number of matching windows with different sizes can be selected. As shown in fig. 4, six steps of the threshold value (0.0, 0.2, 0.4, 0.6, 0.8, 1.0) and six matching window sizes (32×32, 16×16, 8×8, 4×4, 2×2, 1×1) were selected for this study, where n, and Th_n represent an integer in the range of $(1 \le n \le 6)$ and the n-th step's threshold-value, respectively, as explained below.

∂ = *feature value*
WHILE
  IF $(MAD(k,l) > \partial)$
    $matching\_window\_size = 2^{6-\lfloor \frac{\partial}{0.2} \rfloor} \times 2^{6-\lfloor \frac{\partial}{0.2} \rfloor}$
    BREAK
  ELSE
    $\partial = \partial + 0.2$

Fig. 4. Algorithm for adaptive disparity-motion estimation

2.3 Key-frame coding

In this paper, the proposed real-time MVP is used via key-frame coding, as shown in figure 5. That is, disparity estimation is processed between the I-frame of the center images and the P-frame of left and right images, and each view processes an estimation of the motion, excluding the range of the I-frame in center-VOP.

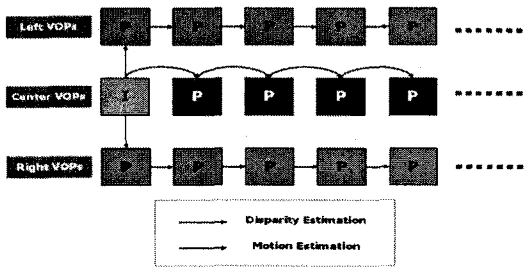


Fig. 5. The proposed key-frame coding

3. The proposed Video Decoding

Fig. 6 illustrates the structure of the proposed multiview video decoder. If the viewpoint is not occluded, the disparity is defined as the distance between the images.

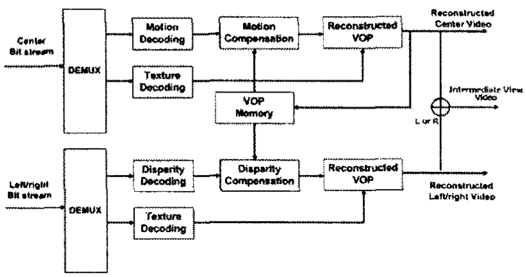


Fig. 6. Multiview video decoding

3.1 Reconstructed Image

Eq. (2) shows the disparity in the horizontal direction and the relationship between the reconstructed (right) image I_r and the reference (left) image I_l.

$$I_r = \begin{bmatrix} i_r \\ j_r \end{bmatrix} = \begin{bmatrix} i_l + DV(i_l, j_l) \\ j_l \end{bmatrix} = I_l + \begin{bmatrix} DV(i_l, j_l) \\ 0 \end{bmatrix} \qquad (2)$$

3.2 Intermediate view reconstruction

By using the IVR technique, more natural 3D image can be acquired through synthesizing the multiview stereo images from the limited given stereo image. Figure 7 shows a general concept of intermediate view synthesis. The position of corresponding point the intermediate view image is defined as a normalized distance α from the left image.

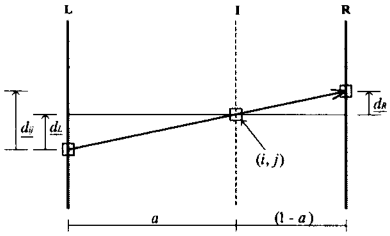


Fig. 7. Corresponding point of intermediate view image from left and right image

The distance from the left to the right image plane is 1, such that $\alpha \in [0,1]$. For example, if $\alpha=0$ is the left image and $\alpha=1$ is the right image and the interval values mean the intermediate view image. Eq. (3) shows the case of interpolation with a weighted average by position α of viewpoint.

$$I_P(i,j) = (1-\alpha)\cdot I_r(i-\hat{d}_{ij}(i,j),j) + \alpha \cdot I_l(i-\hat{d}_{ij}(i,j),j) \quad (3)$$

3.1.1 Adaptive disparity search range

In this paper, adaptive DSR is proposed, which takes the difference between disparity vectors of the blocks and reducing it by comparing to the first disparity vector that is taken from the stereo images. This step is taken by considering the following; the images with less time disparity could also have less search range through the disparity estimation between the blocks of stereo images.

<case 1> Disparity larger than the determined threshold value

If the disparity between stereo images is larger than determined threshold as following eq. (3), it is not necessary to reduce the disparity search range.

$$MAD > Threshold\ Value \quad (3)$$

<case 2> Disparity smaller than the determined threshold value

When the disparity of stereo image value is smaller than the determined threshold, it is necessary to reduce the search range. DSR should be reduced as eq. (4) and DSR should be reduced even more as eq. (5).

$$MAD < Threshold\ Value \quad (4)$$

$$MAD < Threshold\ Value \times 2 \quad (5)$$

When the re-selected absolute value of starting point is smaller than pre-set threshold, its value is smaller than disparity value, but it is bigger than the case 1 so it doesn't reduce the search range, but it reduces the half of case 1.

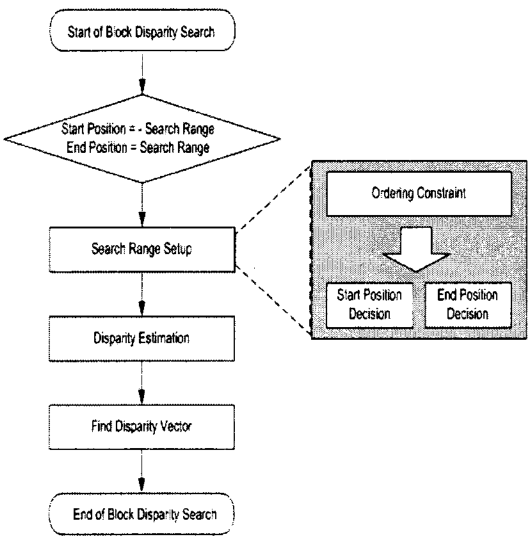
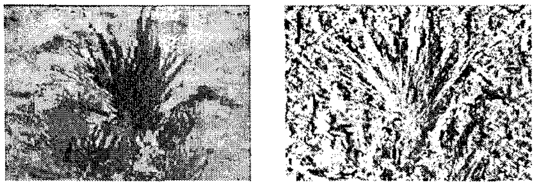


Fig. 8. The flowchart of the proposed algorithm

4. Simulation results and analysis

In the MVC, much redundancy exist between stereo images or multiview images. Therefore, in this paper, MVC using adptive DME is proposed. Figure 9 illustrates a disparity compensated image, and a residual image is illustrated using adaptive disparity estimation between the left and center images. Additionally, Fig. 10 illustrates the reconstructed image 'Pot Plant' using the propsoed algorithm.



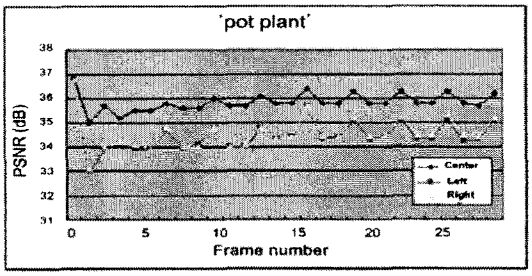
(a) (b)

Fig. 9. 'Pot Plant' (a) Disparity compensated image (b) Residual image

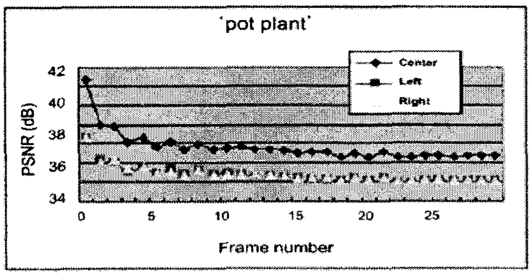
(a) (b)

Fig. 10. Reconstructed image 'Pot Plant' (a) left image (b) right image

Fig. 11 shows PSNR results of 'Pot Plant'. In the conventional method, the average PSNRs of the left, right and center images are 34.47[dB], 34.55[dB] and 35.84[dB], respectively. However, in the proposed algorithm, the average PSNRs of left, right and center images are 36.30[dB], 36.41[dB] and 37.81[dB], respectively. Also, tables 1 shows the PSNRs, bit per pixel and encoding time of 'Pot plane' for 30 frames. in the tables 1, the processing time is 3.962 sec/frame.



(a)



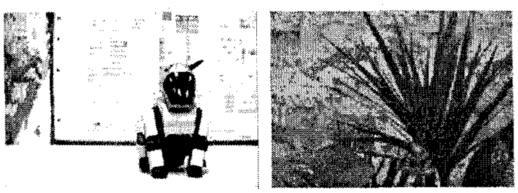
(b)

Fig. 11. Reconstructed image 'Pot Plant' (a) left image (b) right image

Table 1. Result of 'Pot Plant' encoding

pot plant \ Video	center			left			right		
PSNR(dB)	Y	U	V	Y	U	V	Y	U	V
(average)	37.81	40.85	40.95	36.39	39.06	39.16	36.41	39.11	39.13
bpp(bit per pixel)	0.25			0.06			0.06		
Processing Time	4.195 sec. (0.139 sec./frame)								

Figure 12 (a), (b) is synthesized image of 'IVO' and 'Pot Plant' by the proposed IVR.



(a) (b)

Fig. 12. Synthesized intermediate view image 'IVO' and 'Pot Plant'

Table 2 shows the result of the proposed IVR, and the average decoding time is 7.55 sec. and PSNRs are 25.37[dB].

Table 2. The decoding time and PSNRs

30 frame Decoding with IVR	Synthesizing Time (sec)		PSNR [dB]	
	Pot Plant	IVO	Pot Plant	IVO
MAD	13.27	10.98	16.67	20.59
MSE	14.27	12.31	17.46	21.26
Bi-direction	20.27	16.32	22.29	25.14
Proposed Algorithm	8.27	6.82	24.3	26.44

5. Conclusion

In this paper, a real-time multiview video coding system using fast disparity estimation is proposed. In the multiview video encoding, ADME for effective 3-dimensional (3D) video compression is proposed. Also, in this paper, IVR using ADSA for real-time 3D synthesis is proposed. The proposed algorithm can reduce processing time of disparity estimation by selecting adaptive disparity search range. Accordingly, the proposed system is able to increase the efficiency of the coding rate and improve the resolution. Moreover, it can be implemented as a fast mutlivew 3D video system.

References

[1] V. Grinberg et al., "Geometry of binocular imaging", Proc. of SPIE, Vol. 2177, pp. 56-65, 1994.

[2] I. Howard, B. Rogers: Binocular Vision and Stereopsis(Oxford Univ. Press, Oxford, 1995) Chap. 2, p.31.

[3] N. A. Dodgson et al., "Time-multiplexed autostereoscopic camera system", Proc. of SPIE, Vol. 3012, pp. 72-83, 1997.

[4] J. R. Moore et al., "Time-multiplexed color autostereoscopic display", Proc. SPIE, Vol. 2653, pp. 10-19, 1996.

[5] S. T. Barnard, "Dispairty analysis of image", IEEE Trans. Pattern Anal. Mach. Intell. Vol. Pami-2, No. 4, pp. 333-340, 1980.

[6] M. Siegel et al., "Compression of stereo image pairs and streams", Proc. of SPIE, Vol. 2177, pp. 258 -268, 1994.

[7] K. H. Bae et al., "New disparity estimation scheme based on adaptive matching window for intermediate view reconstruction", Optical Engineering, Vol.42, No.6, pp.1778-1786, 2003.

[8] M. Ebroul Izquierdo, "Stereo image analysis for multi-viewpoint telepresence applications", Signal Processing : Image Commnication, Vol.11, pp.231-254, 1998.

Biography

Kyung-Hoon Bae

He received his B.S., M.S. and Ph. D degrees from the department of electronic engineering Kwangwoon University, Seoul Korea, in 2001, 2003 and 2006, respectively. Also he received his MBA degrees from Columbia Southern University, USA, in 2004, and completed his Advanced Project Management (APM) Course from Stanford University, USA, in 2006. He is currently working in Samsung Thales as a senior researcher. His research interests include Visual Communication, 3-D and Robot Vision.

Byung-Kwang Woo

He received his BS degree from the department of electronic engineering Kyungpook University, Daegu, Korea, in 1988, and his MS degree in Business Administration from the graduate school of Kyungpook University in 2004. and he is currently attending Ph.D degree in Seoul university of Venture & Information, Seoul, korea, from 2006. He worked for DTAQ(Defense Agency for Technology and Quality) as QA senior researcher from Mar. 1990 to Mar. 2006. Also he currently working for DAPA(Defense Acquisition Program Administration) as Technology support team member from Apr. 2006. His research interests Visual Communication, 3D Vision.