Disparity-Compensated Stereoscopic Video Coding Using the MAC in MPEG-4

Sukhee Cho, Kugjin Yun, Chunghyun Ahn, and Soo In Lee

ABSTRACT—The MPEG-4 multiple auxiliary component (MAC) is a good mechanism to achieve one-stream stereoscopic video coding. However, there is no syntax or semantics for the residual texture data of the disparity-compensated image in the current MAC. Therefore, we propose a novel disparity-compensated coding method using the MAC for stereoscopic video. We also define a novel MAC semantics in MPEG-4 so as to support the proposed coding algorithm. The major difference between the existing and proposed coding methods using the MAC is the addition of the residual texture coding.

Keywords—Stereoscopic video coding, MPEG-4, multiple auxiliary component (MAC), disparity map.

I. Introduction

Stereoscopic video sequences are comprised of two images, left and right views whose images correspond to the view of left and right eyes. It is well known that there is a high correlation between the two sequences [1].

The multi-view profile in MPEG-2 and temporal scalability (TS) in MPEG-4 use the correlation information between the two views to code stereoscopic video sequences [2], [3]. We call these techniques TS-based coding in this paper.

The TS-based coding usually consists of a base-layer and an enhancement-layer for stereoscopic video; the base-layer is used to code the left-view image while the enhancement layer caters for the right-view image. Left-view images are compressed using motion texture coding algorithms while right-view images are compressed using block-based motion/disparity compensated coding algorithms. However, in the market this coding scheme is scarcely used in practical implementations. The reason is two-fold: First, since the scheme is such a complex structure, the implementation of TS-based coding schemes such as the multi-view profile or TS codec is too costly. Second, this coding scheme outputs two encoded-streams. This two-stream temporal-scalability approach is technically inconvenient because it affects the codec system wide. In particular, the two-stream approach needs to support both multiplexing and de-multiplexing together with frame synchronization among different views from the transport layer. On the other hand, a one-stream approach affects only the video codec part within the whole system chain and is systematically simpler than the two-stream approach in terms of system configuration.

Recently, there have been other kinds of approaches to code a right-view image without using the discrete cosine transform (DCT) for stereoscopic video [5], [6]. These approaches reserve blocking artifacts on the reconstructed images, but have a huge cost of complexity. In particular, Cheung's method in [6] cannot remove the temporal redundancy in the right-view images because the residual texture coding doesn't carry out motion estimation and compensation for the residual texture data.

To improve the performance of stereoscopic video coding, many researchers have also focused on the disparity estimation techniques [7], [8]. Most estimation methods are accurate for simple video sequences which consist of one or two objects with low variation and low motion. However, they don't achieve accurate disparity estimation for the opposite video sequences which consist of many objects with high variation and generally fast motion. Chien and others [7] proposed stereoscopic video coding using mesh-based disparity estimation and compensation without residual texture coding for teleconference video sequences. Their coding method

Manuscript received Aug. 19, 2004; revised Mar. 09, 2005.

This work was supported in part by the Ministry of Information and Communication of Korea under the title "The development of SmarTV technology."

Sukhee Cho (phone: +82 42 860 5357, email: shee@etri.re.kr), Kugjin Yun (email: kjyun@etri.re.kr), Chunghyun Ahn (email: hyun@etri.re.kr), and Soo In Lee (email: silee@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

might have efficient results for only simple video sequences.

The MPEG-4 multiple auxiliary component (MAC) is a good mechanism for generating one-stream stereoscopic video coding. The MAC was added to version 2 of the MPEG-4 Visual part [4] in order to describe the transparency of video objects. Stereoscopic video coding using the MAC compresses the left-view image and then the disparity map for the corresponding right-view image. It can use both motion and disparity compensations, and does not need to multiplex the compressed bit-stream because only one bit-stream is output. However, there is no syntax or semantics for the residual texture data of the disparity compensated image in the current MAC; even the quality of reconstructed images of real stereoscopic video sequences cannot be guaranteed without semantically defined residual texture data.

Therefore, in this paper, we extend the MPEG-4 MAC by defining a novel MAC semantics and propose a stereoscopic coding method using the extended MAC. This paper is organized as follows. Section II introduces the current MAC and extends it. In section III, we propose disparity-compensated coding using the extended MAC. We present experimental results and conclusions in sections IV and V, respectively.

II. Extension of MAC

The MAC was defined for a video object plane (VOP) on a pixel-by-pixel basis and contains data related to video objects such as disparity, depth, and additional texture. Table 1 shows the number and type of auxiliary components that are indicated by *video_object_layer_shape_extension (VOLSE)*. Even though MPEG-4 has MAC syntax and semantics to accommodate the disparity, there are some limitations for real stereoscopic video coding. In particular, syntax and semantics for residual texture data of the disparity compensated image are missing from the current MAC.

Note that in Table 1 the types of all components for the *VOLSE* values between 1101 and 1111 have not been defined yet. Therefore, we have extended the MPEG-4 MAC by defining the novel MAC semantics shown in Table 2. It is possible to use these new MAC semantics to encode the residual texture data of a disparity-compensated image for a right-view image in MPEG-4 [4].

Table 1. Semantic meaning of *VOLSE*.

VOLSE	*AC_type[0]*	*AC_type[1]*	*AC_type[2]*	*count*
0000	ALPHA	NO	NO	1
0001	**DISPARITY**	NO	NO	1
0010	ALPHA	**DISPARITY**	NO	2
0011	**DISPARITY**	**DISPARITY**	NO	2
0100	ALPHA	**DISPARITY**	**DISPARITY**	3
0101	DEPTH	NO	NO	1
0110	ALPHA	DEPTH	NO	2
0111	TEXTURE	NO	NO	1
1000	USER DEFINED	NO	NO	1
1001	USER DEFINED	USER DEFINED	NO	2
1010	USER DEFINED	USER DEFINED	USER DEFINED	3
1011	ALPHA	USER DEFINED	NO	2
1100	ALPHA	USER DEFINED	USER DEFINED	3
1101-1111	t.b.d.	t.b.d.	t.b.d.	t.b.d.

Table 2. Novel semantic meaning of *VOLSE* .

VOLSE	*AC_type[0]*	*AC_type[1]*	*AC_type[2]*	*count*
1101	DISPARITY	Luminance residual texture	Chrominance residual texture	3
1110	DISPARITY	Luminance residual texture	No	2

III. Disparity-Compensated Coding Using the Extended MAC

We propose using the extended MAC with the disparity map and residual texture data for a right-view image to produce a stereoscopic video coder as shown in Fig. 1. The major difference between conventional coding methods and the proposed coding method is the addition of the residual texture coding. Residual texture data represents the difference image between the original right-view image and the disparity-compensated right-view image obtained using the locally reconstructed left-view image and disparity map.

The proposed coding method assigns the disparity map and residual texture data to three components of the MAC: one component for the disparity map, one component for the luminance data of the residual texture, and the remaining one for the chrominance data of the residual texture. Since the chrominance data is half the size of the luminance data in the case of 4:2:0 encoding, the chrominance data will be placed in the first half of the last MAC component. Thus, the coding of the last MAC component should ignore the second half, which is garbage data.

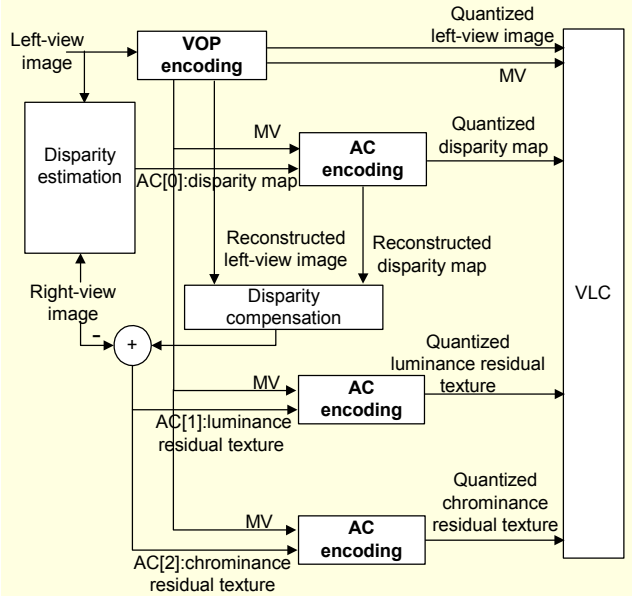

Fig. 1. Block diagram of the proposed stereoscopic video coding.

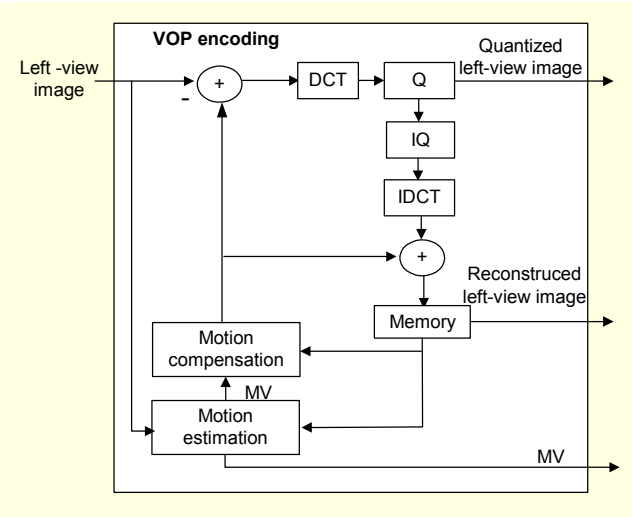

Fig. 2. Procedure of VOP encoding.

The left-view image is compressed by VOP encoding using the motion compensated coding method shown in Fig. 2. The disparity map is generated on a pixel-by-pixel basis using the disparity estimation algorithm [8] and is assigned to a component in the MAC. Then, the disparity map is compressed by the auxiliary component (AC) encoding; this also uses the motion compensated coding algorithm shown in Fig. 3. Since AC encoding does not carry out motion estimation, it compensates for the motion using the motion vector (MV) estimated in VOP encoding.

The proposed coding method for stereoscopic video sequences has the advantage in that multiplexing of the bit-streams is unnecessary. The proposed coding method is also

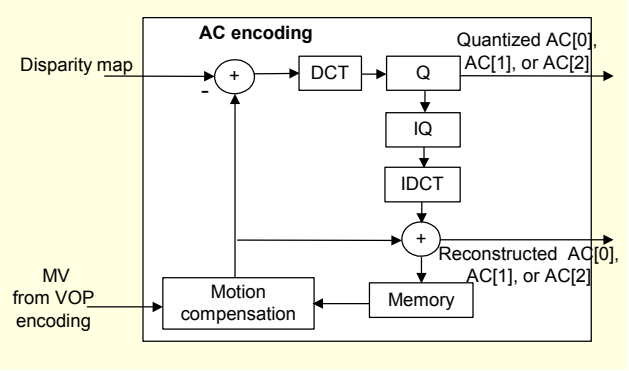

Fig. 3. Procedure of AC encoding.

easy to embody in hardware and/or software and to apply to the current MPEG-4 coding tool.

IV. Experimental Results

We evaluated the performance of the proposed coding method against three conventional coding methods for right-view images. Left-view images are encoded with the same MPEG-4 code for all methods.

The conventional coding using the MAC, *method 1*, compresses left-view images and disparity maps for the corresponding right-view images. The disparity map is assigned to a component in the AC[0] of the MAC and is compressed using the same motion texture coding of MPEG-4. For the proposed coding using the MAC, *method 2*, we evaluated two methods of the extended MAC in Table 2. One is *method 2-1* with a disparity map, luminance, and chrominance residual texture data to the three components, AC[0], AC[1], and AC[2], respectively. *Method 2-2* assigns a disparity map and luminance residual texture data to two components of the MAC, AC[0], and AC[1], respectively. *Method 3* (Independent 2D coding of the two views) independently compresses left- and right-view images using the existing coding technique. Finally, *method 4* (TS-based coding) is a stereoscopic video coding method using temporal scalability.

The stereoscopic video test sequences have left- and right-view images of dimensions 720 × 480 and a Y:Cb:Cr = 4:2:0 format. The test sequence set is composed of two different video sequences: 'Puppy' and 'Soccer2'. The 'Puppy' sequence consists of three objects with low variation and slow motion. In contrast, the 'Soccer2' sequence consists of many objects with high variation and generally fast motion.

Figure 4 shows the average PSNR values of the reconstructed right-view images when left-view images are compressed by QP(I,P,B)=(4,8,12) for the 'Puppy' and 'Soccer2' sequences. The bit-rate of the x-axis is obtained by testing several arbitrary QP values. In Fig. 4, *method 4* and *method 1* have the highest and the lowest PSNR values, respectively, for both sequences.

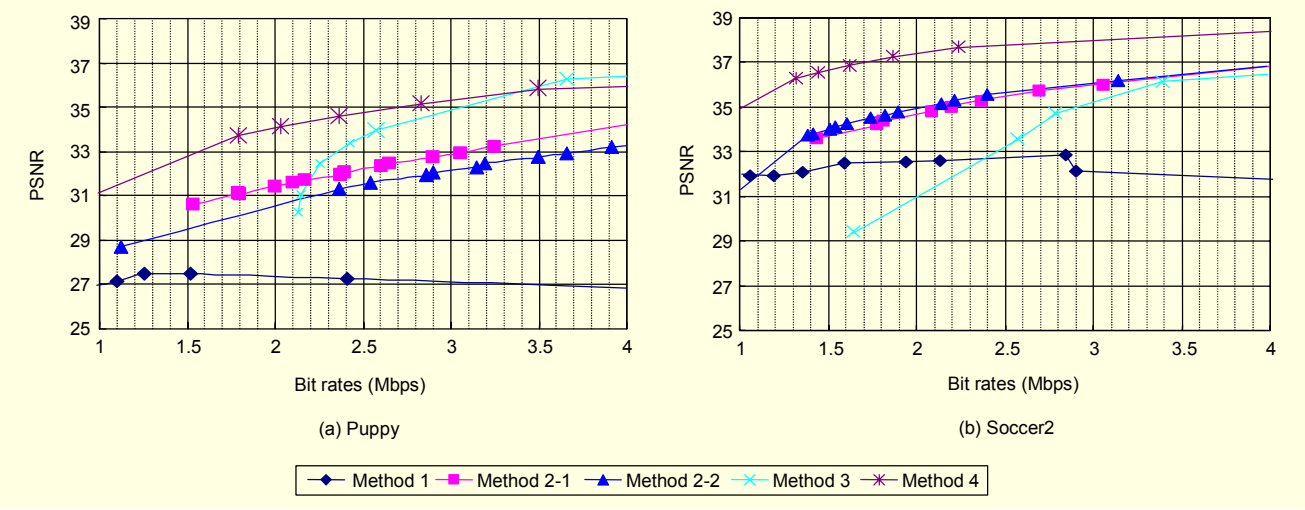

Fig. 4. PSNR values of the reconstructed right-view image when left-view images are compressed by QP(I,P,B)=(4,8,12).

Method 2-1 and *method 2-2* have similar PSNR values for both sequences. Comparing *method 1* and *method 2* using the MAC, *method 2* has higher PSNR values (by about 4 to 7 dB) at all bit-rates than *method 1* for both sequences. *Method 1* has similar PSNR values at all bit-rates (of about 26 and 32 dB for the 'Puppy' and 'Soccer2' sequences, respectively).

To obtain the results of *method 1* and *method 2*, we employed the discrete-cosine-transform-based MPEG-4 coding method for the disparity map; the characteristics of the latter are evidently quite different from those of normal video data. This means that the coding method applied for the disparity map does not match the signal characteristics, and therefore it is far from any form of optimality. We intend a further investigation of more optimal maps.

V. Conclusions

The MPEG-4 MAC is a good mechanism for one-stream stereoscopic video coding. However, there is no defined syntax or semantics for residual texture data of the disparity compensated image in the current MAC. Thus in this paper, we proposed disparity-compensated coding using the extended MAC with a disparity map and residual texture data for the right-view image. We concluded that the TS-based technique is the best method in view of PSNR values for both of our test sequences. However, when considering other coding methods using the MAC, the proposed disparity-compensated coding method with residual-texture gives better results than the conventional coding method. In future work, we plan to investigate better improved coding methods for the disparity map and the residual right-view image together with a subjectively optimal rate allocation among components for the proposed coding method using the MAC.

Acknowledgement

We would like to thank the anonymous reviewers, whose comments improved this paper.

References

[1] Naemura T., Kaneko M., and Harashima H., "Compression and Representation of 3D Image," *IEICE Trans. on Info. & Systems*, vol. E82-D, no.3, March 1999, pp.558-567.

[2] Chen X. and Luthra A., "MPEG-2 Multi-View Profile and Its Application in 3DTV," *Int'l Society for Optical Engineering Proc of SPIE*, vol.3021, San Diego, USA, 1997, pp.212-223.

[3] Naito S. and Matsumoto H., "Advanced Rate Control Technologies for 3D-HDTV Digital Coding Based on MPEG-2 Multi-View Profile," *IEEE Int'l Conf. on Image Proc.*, vol.1, Piscataway, USA, 1999, pp.281-285.

[4] Generic Coding of Audio-Visual Objects – Part 2 : Visual, ISO/IEC 14496-2, 2001.

[5] S.H. Seo and M.R. Azimi-Sadjadi, "A 2-D Filtering Scheme for Stereo Image Compression Using Sequential Orthogonal Subspace Updating," *IEEE Trans. on CASVT*, vol. 11, no. 1, Jan. 2001, pp.52-66.

[6] W.F. Cheung and Y.H. Chan, "A Fast Two-Stage OMT Algorithm for Coding Stereo Image Residuals," *IEEE Int'l Conf. on Image Proc.*, Spain, 2003.

[7] S.Y. Chien, S.H. Yu, L.F. Ding, Y.N. Huang, and L.G. Chen, "Efficient Stereo Video Coding System for Immersive Teleconference with Two-stage Hybrid Disparity Estimation Algorithm," *IEEE Int'l Conf. on Image Proc.*, Spain, 2003.

[8] Y.T. Kim, S. Choi, S. Cho, and K. Sohn, "Efficient Disparity Vector Coding for Multiview Sequences," *Signal Processing: Image Communication*, vol. 19, no. 6, July 2004, pp. 539-553.