MPEG-4 BASED STEREOSCOPIC VIDEO SEQUENCES ENCODER

Wenxian Yang[1] *and Ngan King Ngi*[1, 2]

[1]Nanyang Technological University, Singapore
[2]Chinese University of Hong Kong, Hong Kong

ABSTRACT

In this paper we propose an object-based MPEG-4 compatible stereoscopic video sequence encoder. We aim at efficient stereoscopic video compression for video-conferencing and 3D telepresence systems. The stereoscopic video sequence includes one main view and one auxiliary view. The main view is encoded using MPEG-4 encoder and the auxiliary view is encoded by joint motion and disparity compensation. After the input sequences are balanced to compensate for lighting conditions and camera differences, the joint disparity and motion regularization is performed on the VOP basis. The output of the encoder contains two bitstreams, a main bitstream, which can be decoded by a standard MPEG-4 decoder, and an auxiliary bitstream. Simulation results show that the joint estimation and regularization of disparity and motion fields provide more accurate vector fields and efficient compression for the auxiliary stream. The proposed system achieves high image quality at lower bitrate than existing stereoscopic video encoders.

1. INTRODUCTION

Stereoscopic or multiview video can provide more vivid and accurate information about the scene structure than the monocular video as they can provide the depth information. When each image of a stereo pair is viewed by its respective eye, the stereo image will provide the viewers a 3D feeling.

Stereoscopic and multiview video systems are widely used in application areas of entertainment, manufacturing industry, remote operations, telemedicine, telerobotics, 3D visual communications, and virtual reality, etc. For transmission and storage of stereoscopic and multiview video data, compression is important as the required bandwidth linearly increases with the number of camera channels. Stereoscopic sequences can be compressed much more efficiently than the independent compression of its two image streams by exploiting, in addition to the intra and inter frame redundancy, the high inter channel correlations.

The first real-time transmission of stereoscopic video over ATM network took place in Europe as a result of the RACE DISTIMA project [1]. DISTIMA has developed a system for capturing, coding, transmitting and presenting digital stereoscopic image sequences. Later, another project called PANORAMA was developed to enhance the visual information exchange in telecommunications with 3-D telepresence [2]. The two major achievements of these systems are the auto-stereoscopic display and the multi-viewpoint capability. Stathis [3] developed an object-based stereo sequence encoder, which can be used for any type of scene and can provide better picture quality than MPEG-2 at the same bitrate. Luo [4] and Puri [5] also proposed MPEG-2 compatible stereoscopic video encoders, which adopt disparity compensation to remove the inter-channel redundancy.

There are two main purposes for existing stereoscopic video encoders. One is to interpolate intermediate images at the receiver so as to provide viewpoint adaptation [1][2] and the other is to provide the viewers the 3D feeling [5][6]. For the former case, high qualities are required for both image sequences and the disparity fields, thus the computational complexity for obtaining the disparity field and the bitrate are usually very high. For the latter case, the human visual system (HVS) states that it is not necessary to code both views with the same quality. The quality of the auxiliary views can be significantly lower, for example, 3 dB or more [5] compared to the main view, without incurring significant visual distortions. Further support for unequal distribution of quality between the two views is provided by the leading eye theory, according to which the overall subjective impression in stereoscopic 3D viewing depends mainly on the quality perceived by the leading eye of viewers. Thus, significant benefits can accrue from unequal distribution of bits between the two views.

In order to develop a stereoscopic or multiview video sequences CODEC, which provides good quality at low bitrate, there are three main issues we consider important. Firstly, compatibility with existing monoscopic video coding standards should be maintained. The MPEG-2 multiview profile (MVP) is a straightforward way under such purpose. While the main application area of the MPEG-2 MVP is in stereoscopic TV, it is expected that multiview aspects of MPEG-4 will play a major role in

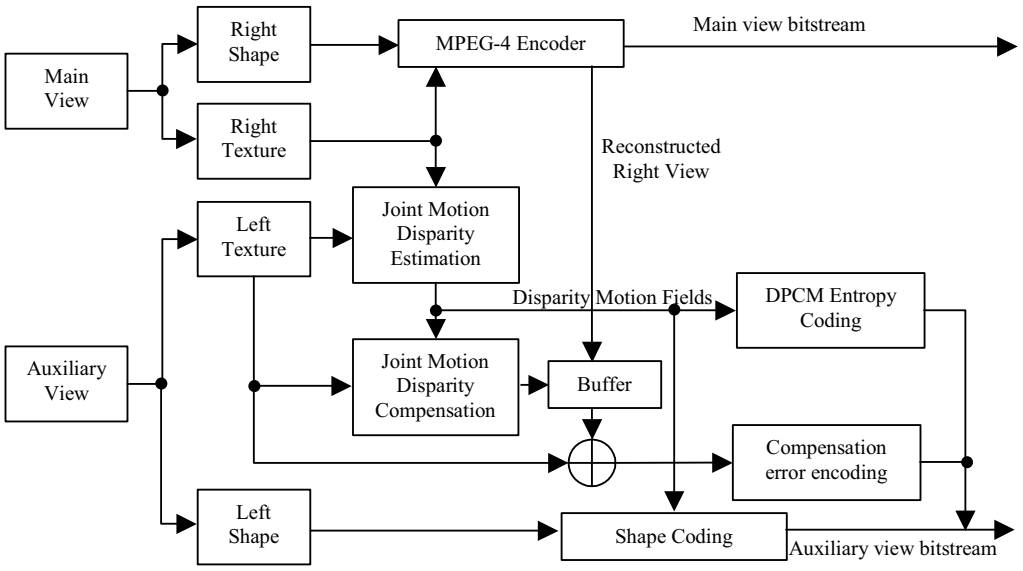

Figure 1: Block diagram of the proposed stereoscopic video encoder

interactive applications, e.g., navigation through virtual 3-D worlds with embedded natural video objects [4]. In our proposed system, one video sequence, namely, the main view is encoded using MPEG-4. Secondly, the computational complexity of the disparity and motion estimation algorithm should be low and the relationship between the disparity and motion fields should be exploited. In our work, a joint disparity and motion estimation/regularization scheme is proposed, which provides more accurate and smoother vector fields at relatively low complexity. Thirdly, the stereoscopic video encoder should be easily extendable to multiview video encoder. Since the frames of the auxiliary view may become reference frames in a multiview sequence encoder, good image quality is also required for the auxiliary view. The additional bandwidth needed for the auxiliary view can be adjusted depending on the demand of the users.

In this paper, we propose a stereoscopic video encoder subject to these three considerations. The structure of the proposed system and the implementation details are described in Section 2. We will discuss the simulation results in Section 3 and draw the conclusions in Section 4.

2. STEREOSCOPIC VIDEO CODING

Image balancing is performed as a preprocessing step. The purpose of image balancing is to eliminate the potential signal difference between the stereo images, which is due to light conditions and camera differences. The balancing procedure is defined in [7]. The auxiliary video sequence is input to the encoder after balancing with the main video sequence.

For every two images pairs, joint regularization of disparity and motion fields is performed. The joint regularization procedure is performed iteratively under the stereo consistency constraint:

$$\vec{M}_l + \vec{D}_t - \vec{M}_r - \vec{D}_{t+1} \approx \vec{0} \qquad \textbf{(1)}$$

A detailed description of the regularization procedure is given in [8].

As depicted in Figure 1, the main view is encoded using a MPEG-4 encoder and the auxiliary view is encoded using joint disparity and motion compensation. Disparity and motion estimation will be VOP-based. Figure 2 shows a simple GOP structure that is typically used in MPEG. Since the theoretically optimal GOP structure can be defined only under certain application requirements and source data characteristics, in our proposed encoder, user can define the GOP structure by setting the M and N parameters. Here, N is intra distance, which is the length of GOP, and M is the prediction distance. As shown in Figure 2, we introduce new picture types P_D and J for the auxiliary view, where P_D VOPs are predicted by disparity fields and J VOPs are predicted jointly by disparity and motion fields. The joint regularization is performed using the original image data,

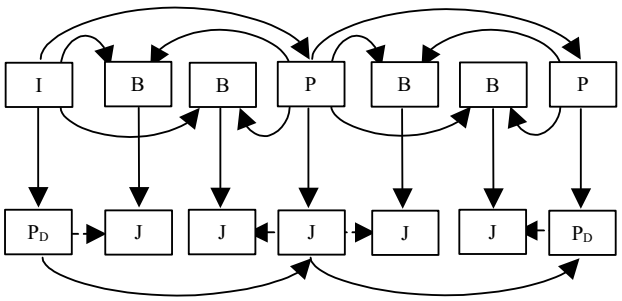

Figure 2: GOP structure

Table 1: VOP types and prediction types

VOP Type		MB Prediction type			
Main view	I	Intra			
	P	Intra	Inter	Inter4V	
	B	Direct	Interpolate	Forward	Backward
Auxi. view	P_D	Intra	Inter	Inter4V	
	J	Interpolate	Motion	Disparity	

Table 2: Test sequences

Sequence name	Frame size	Frame rate	Tested pairs	Mask	Remarks
MAN	512x512	30	25	Yes	Head and shoulder
Train & Tunnel	720x576	25	25	No	Not head and shoulder

thus refinement of the fields within a very small search range and half-pixel search is required during the encoding process using reconstructed reference frames so that the results are reproducible at the decoder.

Different MB prediction modes of texture data for different VOP types are defined in Table 1. The prediction mode which gets the minimum SAD (Sum of Absolute Difference) value is selected for the current macroblock. For P_D VOPs, shape is also coded by disparity compensation and for J VOPs, shape is coded either by disparity or by motion, the selection is VOP-based but not MB-based.

Disparity fields are encoded by DPCM and Huffman coding, which is similar to the coding of motion fields used in MPEG-4. Residual for P, B, P_D, B and J VOPs after disparity/motion compensation is encoded by DCT coding as defined in MPEG-4.

3. SIMULATION RESULTS AND ANALYSIS

To evaluate the performance of the proposed stereoscopic video encoder, two stereo sequences were used, *MAN* and *Train & Tunnel*. Description of the sequences is given in Table 2.

Without loss of generality, we set the intra distance N=9 frames and the prediction distance M=3 frames. For *MAN*, luminance and shape information are encoded, while for *Train & Tunnel*, only luminance information is encoded. The coding results of the proposed system are given in Table 3 and the results of separately encoding the two sequences by MPEG-4 are given in Table 4. The overall values are over the 25 frames tested, i.e., three GOPs except the last two B/J frames. Since rate control scheme is not implemented, we give the results at different QP values of the main view and auxiliary view. In our results, the image quality of the auxiliary view can be as good as the main view with a significant reduction in bitrate.

Comparison of Table 3 and Table 4 shows that the image quality obtained for the auxiliary view is very close to that of the main view, with a lower bitrate. There is a slight difference between the coding results of the main view data in our proposed system and that in MPEG-4, although the main view data is encoded as a MPEG-4 bit stream. The reason is that in our proposed system, motion fields for the main view are obtained from joint regularization but not by full search used in MPEG-4. When the image quality of the main view is fixed, we can still adjust the image quality and bitrate of the auxiliary view by modify the quantization parameter. The results are shown in Table 5. From these results we can see that

Table 3: Selected coding results of the proposed encoder

Seq. name	QP		I		P / P_D		B / J		Overall		
			PSNR	bpp	PSNR	bpp	PSNR	bpp	PSNR	bits per VOP	bpp
MAN	8	Main	38.22	0.38	37.35	0.24	37.96	0.11	37.84	20607	0.17
		Auxi.	-	-	37.08	0.26	37.66	0.14	37.59	18075	0.15
	16	Main	34.58	0.23	34.67	0.20	35.38	0.09	35.12	15772	0.13
		Auxi.	-	-	34.89	0.18	34.98	0.09	34.97	12241	0.10
Train & Tunnel	8	Main	34.57	0.76	33.73	0.34	33.84	0.25	33.90	137414	0.33
		Auxi.	-	-	33.35	0.38	33.59	0.24	33.56	106602	0.26
	16	Main	30.71	0.38	30.34	0.18	30.39	0.12	30.42	68268	0.16
		Auxi.	-	-	31.04	0.25	30.52	0.11	30.58	53136	0.13

Table 4: Selected coding results of MPEG-4

Seq. name	QP		I		P		B		Overall		
			PSNR	bpp	PSNR	bpp	PSNR	bpp	PSNR	bits per VOP	bpp
MAN	8	Right	38.22	0.38	36.97	0.26	35.46	0.09	37.62	21875	0.18
		Left	38.74	0.36	37.56	0.25	36.08	0.09	38.25	20147	0.17
	16	Right	34.58	0.23	34.00	0.20	35.07	0.10	34.75	16491	0.14
		Left	35.01	0.22	34.53	0.19	35.63	0.09	35.29	15748	0.13
Train & Tunnel	8	Right	34.57	0.76	33.64	0.34	33.85	0.22	33.89	130304	0.31
		Left	34.77	0.74	33.84	0.33	34.08	0.21	33.95	125930	0.31
	16	Right	30.71	0.38	30.21	0.17	30.38	0.10	30.38	63140	0.15
		Left	30.86	0.37	30.39	0.17	30.56	0.10	30.55	62095	0.15

*Here, "Right" and "Left" correspond to "Main" and "Auxi." views in Table 3, respectively.

Table 5: PSNR and bpp of the proposed encoder

Seq. name		PSNR	bpp		PSNR	bpp
MAN	Main	37.84	0.17	Main	35.12	0.13
	Auxi. view	34.97	0.10	Auxi. view	32.93	0.08
		35.95	0.11		33.90	0.09
		36.52	0.12		34.32	0.10
		36.83	0.13		34.60	0.10
		37.59	0.15		34.97	0.10
Train & Tunnel	Main	33.90	0.33	Main	30.42	0.16
	Auxi. view	30.26	0.09	Auxi. view	28.52	0.07
		30.89	0.10		28.67	0.08
		31.93	0.13		29.72	0.09
		32.47	0.17		30.14	0.12
		33.56	0.26		30.58	0.13

our proposed system is also efficient for providing the stereoscopic view, when the PSNR of the auxiliary view is 2 to 3 dB lower than that of the main view, the bitrate can be less than 30% of that of the main view, subject to the characteristics of the images.

4. CONCLUSIONS

A stereoscopic video encoder has been proposed in this paper to provide good image quality of both views at low bitrate. The system is compatible with a conventional coding standard as the main view is encoded as a MPEG-4 bitstream, and the quality of the auxiliary view can be adjusted to cater for different user requirements.

The system can be easily extended to a multiview video encoder which encodes three or more sequences simultaneously, and also be extended to handle multiple video objects, given the original shape masks of the sequences.

5. REFERENCES

[1] Dimitrios Tzovaras, Nikos Grammalidis, and Michael G.Strintzis, "Object-based Coding of Stereo Image Sequences using Joint 3-D Motion/Disparity Compensation", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 312-327, April 1997.

[2] Jens-Rainer Ohm, Karsten Gruneberg, Emile Hendriks, M.Ebroul Izquierdo, Dimitris Kalivas, Michael Karl, Dionysis Papadimatos, and Andre Redert, "A Realtime Hardware System for Stereoscopic Videoconferencing with Viewpoint Adaptation", *Signal Processing: Image Communication,* vol. 14, pp. 147-171, 1998.

[3] Stathis Panis, Manfred Ziegler, and John P.Cosmas, "The Use of Stereo and Motion in a Generic Object-based Coder", *Signal Processing: Image Communication*, vol. 9, pp. 221-238, 1997.

[4] Luo Yan, Zhang Zhaoyang, and An Ping, "Stereo Video Coding Based on Frame Estimation and Interpolation", *IEEE Transactions on Broadcasting*, vol. 49, no. 1, pp. 14-21, March 2003.

[5] A.Puri, R.V.Kollarits, and B.B.Haskell, "Basics of Stereoscopic Video, New Compression Results with MPEG-2 and a Proposal for MPEG-4", *Signal Processing: Image Communication,* vol. 10, pp. 201-234, 1997.

[6] Sriram Sethuraman, M.W.Siegel, and Angel G.Jordan, "A Multiresolution Framework for Stereoscopic Image Sequence Compression", *International Conference on Image Processing,* IEEE proceedings of, pp. 361-365, 1994.

[7] Anthony Mancini, "Disparity Estimation and Intermediate View Reconstruction for Novel Applications in Stereoscopic Video", Master Thesis, McGill University, Feb.1998.

[8] W.Yang, K.N.Ngan, and K.H.Sohn, "Edge-Preserving Regularization of Displacement Fields and Joint Optimization of Disparity and Motion", *International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 629-634, Dec.2003.