ICICS-PCM 2003
15-18 December 2003
Singapore

Stereoscopic Video Coding Based on Global Displacement Compensated Prediction

HuiZhu Jia[1,3] ; Wen Gao[1,2,3]; Yan Lu[2]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2] Department of Computer Science, Harbin Institute of Technology, Harbin, China
[3]Graduate School of Chinese Academy of Sciences, Beijing, China

Abstract

In this paper, we propose a novel algorithm for coding the stereoscopic video sequence based on the global displacement information between the left- and right-view images. MPEG-4 temporal scalable video coding scheme is employed as the baseline, in which the left- and right-view images are compressed as the base and enhancement layer, respectively. To further improve the coding efficiency, some intuitive properties of stereoscopic videos are utilized. One of the major properties lies in the global displacement between the left- and right-view images. In this paper, we propose coding the right-view image with global motion compensated prediction from the left-view image. A further main property of stereoscopic video arises from the fact that, objects simultaneously existing in the left- and right-view images usually result in the very close motion vectors. Based on the global motion information, the motion vectors of the right-view sequence can be easily predicted from that of the left-view sequence. Rate-distortion optimization is also used to select the coding mode. Experimental results show that, compared to the MPEG-4 temporal scalability algorithm for low bitrate stereoscopic video coding, the proposed algorithm can save the bitrate up to 20%.

1. Introduction

With the rapid development of multimedia and network, more and more images and videos are viewed and transmitted in digital format. 3D digital image/video is becoming popular because it can give the users the impression of 3D view of a visual scene. Stereoscopic video is one of the main categories of 3D video, which can provide more vivid and accurate information about the scene. Compared to the traditional 2D video, the stereoscopic video sequence normally requires the double bandwidth. Therefore, how to efficiently compress the stereoscopic video sequence has been involved as one of the major research fields recently.

In the past years, a number of stereoscopic image/video coding algorithms have been developed. Many of them investigated the efficient disparity estimation and compensated prediction schemes to improve the coding efficiency [1]-[3]. The additional redundancy associated with the similarity of the two images in a stereo pair was also exploited to further reduce the overall bitrate in some methods [4][5]. Recently, MPEG has started the work on 3D audio/video (3DAV) standardization, which consequently raises more attentions to the stereoscopic video applications.

To evaluate the potential stereoscopic coding methods, MPEG has defined the experimental conditions for explorations on 3DAV technologies, in which exploration experiments 3 (EE3) contributes to stereoscopic video coding [6]. So far, four promising methods based on MPEG-4 coding tools have been evaluated for stereoscopic video coding [7]. The first method (Coding_method_1) and the second method (Coding_method_2) exploit the MPEG-4 coding tool using multiple auxiliary component (MAC) for stereoscopic video coding. The difference is that the resident MPEG-4 coding tool using MAC is employed for the first method, whereas the extension of MPEG-4 coding tool using MAC with residual texture coding is used for the second method. The third method (Coding_method_3) proposes coding the left- and right-view video sequences independently only to investigate the coding gain of other methods. The fourth method (Coding_method_4) exploits the resident MPEG-4 coding tool using temporal scalability for stereoscopic video coding.

In [8], the performances of the four methods have been evaluated. The conclusion has been drawn that the fourth method using temporal scalable coding is the best in terms of both objective and subject visual quality. For the fourth method, the left- and right-view images are compressed as the base and enhancement layer, respectively. Left-view images are compressed using motion texture coding and right-view images are compressed using block-based motion/disparity compensated texture coding of MPEG-4. However, the performance of right-view image coding can be further improved by utilizing the global displacement information between the left- and right-view images.

In this paper, we propose an efficient stereo video coding algorithm due to its wide applications. Coding_method_4 of EE3, i.e. MPEG-4 coding tool using temporal scalability, is employed as the baseline. Currently, Coding_method_4 uses mature temporal scalability coding technique of MPEG-4 without considering the inherent similarity between left- and right-view images. In general, there is a little disparity between the left-view image and the corresponding right-view image, and the textures of both images are homogeneous. In order to further improve the coding efficiency, we propose to exploit the correlations in

a stereo pair with global displacement compensated prediction. Global motion compensation (GMC) technique is employed for this purpose. In addition, based on the global motion information, the motion vectors (MV) of the right-view sequence can be easily predicted from that of the left-view sequence. Therefore, a further coding mode with derived motion vectors is developed for coding the right-view video sequence.

The rest of this paper is organized as follows. Section 2 presents the global motion compensated predictive coding for the right-view video sequence in detail. Section 3 presents the motion vector prediction scheme based on the available global motion information. Section 4 presents the coding mode selection. Experimental results of the proposed algorithm and MPEG-4 temporal scalable coding algorithm are presented in Section 5, respectively. Finally, Section 6 concludes this paper.

2. GMC-Based Coding of Right-View Sequence

GMC-based video coding technology has been adopted in advanced simple profile of MPEG-4 standard [9], whereas it is neither used in B frame coding nor used in the enhancement layer coding. The main goal of GMC-based video coding is to encode the global motion in a frame with a small number of parameters. In this way, lots of bits for coding the motion vectors can be saved, and consequently the coding efficiency can be further improved.

Due to the inherent similarity of the stereo pair, the global displacement between the left- and right-view images is quite similar to the global motion between two consecutive images. Therefore, the GMC-based coding scheme can be directly used to exploit the correlations between the left- and right-view images. Traditional global motion model is competent for this purpose. However, by using a more sophisticated global motion model (e.g. affine or perspective motion model), the proposed algorithm can be easily extended for coding the multi-view video sequence.

2.1 GMC Framework

Figure 1 shows the proposed stereoscopic video coding structure by jointing GMC and temporal scalability techniques. As shown in Figure 1, the left-view sequence is encoded as the base layer, and the right-view sequence is encoded as enhancement layer. The 1-D illustration of the stereoscopic video is as follows: L0, R0, L1, R1, ... , Li, Ri, ...(Li denotes the ith left-view image, and Ri denotes the ith right-view image).

In the proposed system, A GOP contains the even number of images, in which only the first left-view image is coded as an Intra frame. The other left-view images are predictive encoded by referencing the coded left-view images. The first right-view image is predictive encoded by referencing the first left-view image. The other right-view images are predictive encoded by referencing either the corresponding left-view image or the previous right-view image.

In order to fully utilize the characteristic of stereoscopic video sequence, the new prediction mode, i.e. GMC-based prediction, is added to predictive encode the right-view image. As shown in Figure 1, GMC0, GMC1, etc, denote the position of the introduced GMC-based prediction mode. The global motion presented in a sequence is usually caused by the camera motion, which is equivalent to the very scheme of stereoscopic video generation. Therefore, the global motion information can properly represent the global displacement between a stereo pair.

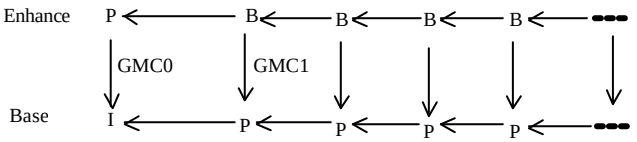

Figure 1: The proposed GMC coding scheme for right sequence coding

2.2 GMC Scheme Implementation

Since MPEG-4 has adopted GMC-based prediction for P frame coding in base layer, the baseline of this proposed GMC coding mode is developed from the MPEG-4 standard. Figure 2 shows the proposed coding structure. As shown in Figure 2, the left-view image is encoded with the resident coding structure in MPEG-4, whereas the right-view image is encoded with some extension tools. The global motion estimation (GME) and global motion compensation modules are incorporated into the framework of right-view image coding.

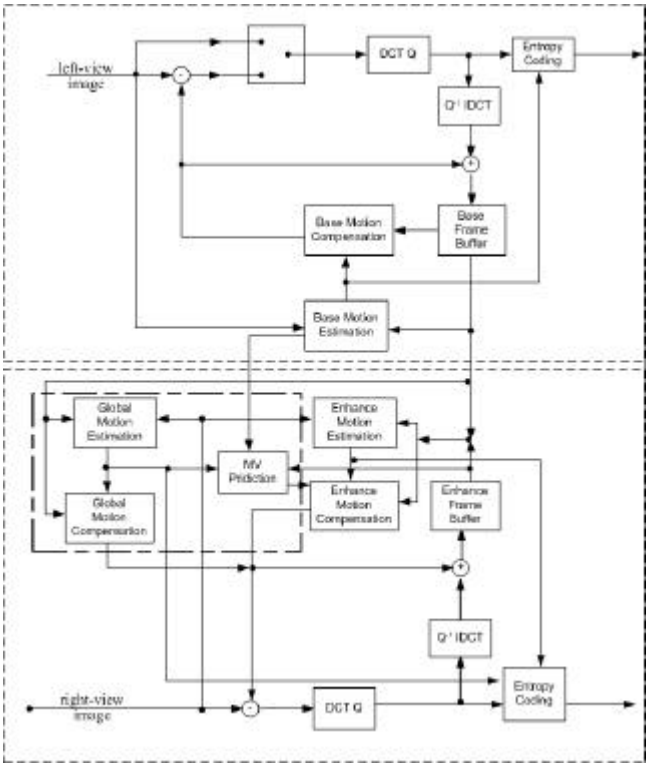

Figure 2: The proposed GMC coding scheme for right sequence coding based on MPEG-4

When the right-view image is input, GME between the right-view image to be encoded and the corresponding left-view image is first performed. Affine motion is employed as the global motion model in this paper. For each right-view image, only one set of motion parameters are encoded. The predicted GMC macroblock is obtained by warping the corresponding reconstructed left-view image in terms of the estimated global motion parameters. For each macroblock coding, the prediction can be due to either the local motion compensated prediction from the left-view and/or right-view images, or the global motion compensated prediction from the left-view image. The detailed coding mode selection scheme is described in the following subsection.

3. Motion Vector Prediction

Since global motion information has been utilized in the proposed scheme, a further improved method is to introduce a new coding mode based on the derived motion vectors from the available global motion information. The key problems of the MV prediction mode consist of how to predict motion vectors and how to select the least cost coding mode.

3.1 MV Prediction Framework

Figure 3 shows the proposed MV Prediction coding scheme for right-view sequence coding. GMC0 is the global motion information of the previous right-view image, which references its corresponding left-view image. GMC1 is the global motion information of the current right-view image, which references its corresponding left-view image. MV0 is the motion vector of the left-view image, which references its previous base layer image. MV1 is the texture motion information of the current enhancement layer image, which references its previous enhancement layer image and is computed by utilizing GMC0, GMC1, and MV0.

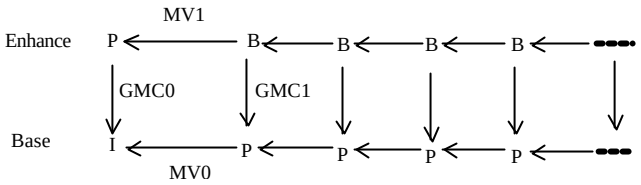

Figure 3: The proposed MV Prediction scheme for right sequence coding

Because there is a little disparity between the enhancement layer image and the corresponding base layer image included by GMC0 and GMC1, and MV0 includes the texture motion information between the two close base layer images, the prediction error between the two close enhancement layer images may be less by computing MV1 using GMC0, GMC1, and MV0 than direct motion prediction. Furthermore, the computed MV1 need not be coded. So it can improve the performance of the enhancement layer (right-view sequence) coding in a way. Its implement based on MPEG-4 temporal scalability is shown in Figure 2. However, one of the keys to realize this scheme is how to compute the MV1 by utilizing GMC0, GMC1, and MV0.

3.2 MV Computation

MV1 is computed by the method of finding special point position. In this paper, the special point is the center point of the current coding macroblock. The relation among GMC0, GMC1, MV0 and MV1 is shown in Figure 4, the top two grids denote the i-1th and ith frame of the enhancement layer, the bottom two grids denote the i-1th and ith frame of the base layer, respectively. Each little block in each grid denotes a macroblock. The gray block in the ith frame of the enhancement layer denotes the current coding macroblock, in which the little black square is the center point of the current coding macroblock.

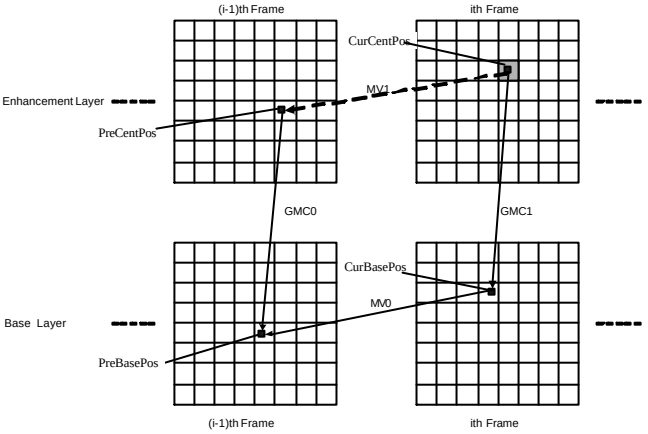

Figure 4: The relation among GMC0, GMC1, MV0 and MV1

The function of the arithmetic is computing MV1 by the GMC0, GMC1 and MV0. The arithmetic is as:

1) Compute the center point position of the current coding macroblock, CurCentPos.
2) Get the GMC1 vector of the current macroblock, and compute the corresponding position of CurCentPos in the corresponding base layer image (the ith frame of the base layer), CurBasePos, by it.
3) Get the MV0 vector of the macroblock in which CurBasePos is, and compute the corresponding position of CurBasePos in the i-1th frame of the base layer, PreBasePos, by it.
4) Get the GMC vector of the macroblock in which PreBasePos is, $GMC0^T$ vector, which is the GMC that the base layer image references the enhancement layer. Because the GMC0 vector is the GMC that the enhancement layer image references the base layer image, in order to compute PreCentPos by PreBasePos, the $GMC0^T$ vector of the macroblock responding to the PreBasePos position must be gotten. The $GMC0^T$ can be required by the GMC0, the arithmetic is as follows:

(1) The value of the $GMC0^T$ vector is the negative value of the GMC0 vector.
(2) Confirm the relation between $GMC0^T$ vector and macroblock in the i-1th frame of the base layer.

In this paper, we use the method of looping searching to get the position of the macroblock in the enhancement layer. Initially suppose the macroblock position of GMC0 vector in the enhancement layer is the same as the macroblock position of $GMC0^T$ vector in the base layer, then looping search the best macroblock position in the enhancement layer among the macroblocks around the supposing macroblock. So the $GMC0^T$ vector is required by this best macroblock position and the GMC0 vector.

5) Get the corresponding position of PreBasePos in the i-1th frame of the enhancement layer, PreCentPos, by the $GMC0^T$ vector.
6) Calculate MV1 vector of the current macroblock by PreCentPos and CurCentPos.

4. Coding Mode Selection

In this paper, two new prediction modes, i.e. GMC-based prediction and MV-based prediction, is added to predictive encode the right-view image. Thus, each macroblock of the right-view image can be predicted either from the corresponding reconstructed left-view image with global motion compensation or from the previously reconstructed right-view image with MV prediction or with local motion compensation (LMC) and/or the corresponding reconstructed left-view image with local motion compensation (LMC). The predictor leading to the lowest prediction error is selected for the current macroblock.

P-VOP prediction modes of enhancement layer include the existing LMC modes in MPEG-4 and GMC mode to be increased. The mode selection scheme between LMC mode and GMC mode is similar to the GMC/LMC decision scheme in MPEG-4 [9], as follows:

if(SADGMC-P<SAD16)
then GMC is selected
else LMC mode is selected

where SADGMC is defined as the sum of the absolute difference (SAD) in the luminance block when using GMC prediction, and SAD16 is defined as the sum of absolute difference for a 16x16 luminance block when the INTER mode is selected. P is an offset to favorite the choice of GMC mode because the motion vectors are not necessary to be encoded for GMC macroblock. P is defined as follows: If the previous criterion is INTER, P equals to NB*QP/64. If the previous criterion is INTER and the motion vector is (0,0), P equals to NB+2. NB indicates the number of pixels inside the macroblock.

B-VOP prediction modes of enhancement layer include the existing LMC modes predicted from the corresponding reconstructed left-view image or/and the previously reconstructed right-view image, GMC mode predicted from the corresponding reconstructed left-view image and MVP (Motion Vector Prediction) mode predicted from the previously reconstructed right-view image. In this paper, the mode decision for LMC modes for B-VOP prediction modes of enhancement layer uses the resident mode selection methods of MPEG-4, but the mode decision for B-VOP prediction modes of enhancement layer among LMC, GMC and MVP adopts the rate-distortion optimization (RDO) strategy. Perform mode selection by minimizing

$$MSE + \lambda Rate \quad (1)$$

where *MSE* is defined as the mean of the squared error between the original macroblock luminance signal and its reconstruction given as:

$$MSE = \left(\sum_{x=1,y=1}^{16,16} \left(s_Y[x, y] - c_Y[x, y, MODE \mid QP] \right)^2 \right) / 256 \quad (2)$$

and λ is the Lagrange multiplier for mode decision, in this paper, equals to $(0.85*2^{QP/3})^{1/2}$, *Rate* is the number of bits associated with choosing *MODE* and *QP*, including the bits for the macroblock header, the motion, and all DCT blocks. $c_Y[x, y, MODE \mid QP]$ and $s_Y[x, y]$ represent the reconstructed and original luminance values, *QP* is the macroblock quantization parameter.

5. Experimental Results

In order to evaluate the proposed stereoscopic video coding algorithm, some experiments have been performed. MPEG-4 temporal scalability algorithm is taken as a reference to investigate the coding gain. Two 3D stereoscopic test sequences, i.e. Soccer and Lineup, in YUV (4:2:0) format (720x480) are used in these experiments, which are provided by ETRI (Electronics and Telecommunications Research Institute). In both sequences, 10 seconds of video clips (600 frames) are processed.

The stereoscopic video sequences are coded using the proposed algorithm and MPEG-4 temporal scalability algorithm, respectively. The left- and right-view image is compressed by the same quantization parameter (QP), so that the coding quality of the left- and right-view image remains similar. Since the same group of QPs is used in both algorithms, the coding efficiency for the left-view is exactly identical. Therefore, we only compare the coding efficiency of both algorithms for the right-view.

Figure 5 shows average PSNR values of the reconstructed right-view image for Soccer sequence, which has high texture variation, partially fast motion, and camera zooming movement. Figure 6 shows average PSNR values of the reconstructed right-view image for Lineup sequence, which has lower texture variation, slow motion, and camera parallel movement. Bit-rate of x-axis is obtained by testing several arbitrary QP values between 1 and 30 for I, B, P types, respectively. Label 'MPEG-4' represents coding method using MPEG-4 temporal scalable video coding, and Label 'MVP&GMC' represents coding method using MV prediction and global displacement information, i.e. the proposed method.

In the comparison of MPEG-4 temporal scalable video coding method and the proposed algorithm, the proposed algorithm has higher PSNR values by about 0.4dB ~0.6dB at low bit-rates than MPEG-4 temporal scalability algorithm for Soccer sequence and about 1dB ~3dB at low bit-rates for Lineup sequence. The results show that, compared to the traditional MPEG-4 temporal scalability algorithm for low bit-rate stereoscopic video coding, the proposed algorithm can greatly improve the coding quality of right-view image. The better coding effect on Lineup than on Soccer is due to global motion prediction.

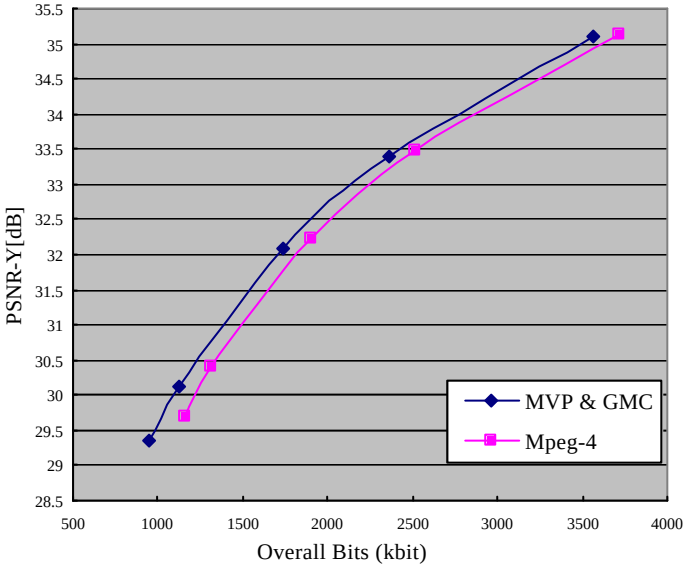

Figure 5: PSNR values of the reconstructed right-view image for Soccer

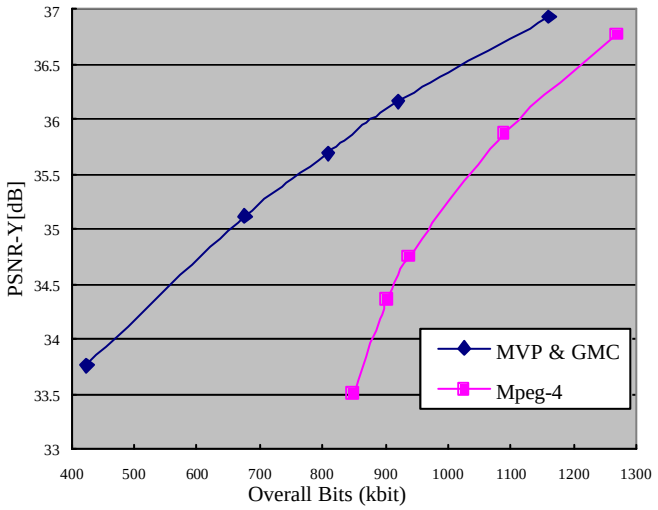

Figure 6: PSNR values of the reconstructed right-view image for Lineup

6. Conclusions

An efficient stereoscopic video coding scheme based on temporal scalability has been presented in this paper. Firstly, since the traditional temporal scalability technique does not consider the special relation of left- and right-view, the GMC prediction technique is proposed for the right-view sequence of the stereoscopic video coding to fully utilize the correlations between the right- and the left-view image. Secondly, since global motion information has been utilized in the proposed scheme, MV Prediction between the right-view images is also proposed to further improve the performance of the right-view sequence coding. The experimental results demonstrate that the proposed GMC and MV Prediction coding scheme can significantly improve the coding efficiency of the right-view sequence coding, namely, significantly reduce the bits for right-view sequence coding at the same quality.

7. Acknowledgements

The work in part has been supported by National Hi-Tech Research Program (863) of China (2002AA119010), National Fundamental Research and Development Program (973) of China (2001CCA03300).

References

[1] S. S. Intille and A. F. Bobick, "Disparity-space images and large occlusion stereo," M.I.T. Media Lab Perceptual Computing Group, Cambridge, MA, Tech.Tep.220, 1994

[2] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," IEEE Trans. Circuits Syst. Video Technol. vol.8, pp. 328-344, June 1998.

[3] L. Falkenhangen, R. Koch, A. Kopernik, and M. Strintzis, "Disparity estimation based on 3-D arbitrarily shaped regions," Digital Stereoscopic Imaging and Applications (DISTIMA), Tech. Rep. #R2045 /UH /DS /P /023 /b1 RACE Project R2045, 1994.

[4] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao, "Stereo without disparity gradient smoothing: A bayesian sensor fusion solution," in Proc. British Maching Vision conf, pp. 337-346, New York, 1992.

[5] N. Grammalidis, D. Beletsiotis and M. G. Strintzis, "Sprite generation and coding in multiview image sequence," IEEE Trans. Circuits Syst. Video Technol. vol.10, No. 2, March 2000.

[6] ISO/IEC WG11 MPEG Video & SNHC Group, "Description of exploration experiments in 3DAV," ISO/IEC JTC1/SC29/WG11 MPEG2002/N4929 Klagenfurt, July 2002

[7] ISO/IEC WG11 MPEG Video Group, "Description of exploration experiments in 3DAV," ISO/IEC JTC1/SC29/WG11 MPEG2002/N5169 Shanghai, China, October 2002.

[8] S. Cho, K. Yun, B. Bae, Y. Hahm, C. Ahn, Y. Kim, K. Sohn, Y. H. Kim, "Report for EE3 in MPEG 3DAV," ISO/IEC JTC1/SC29/WG11 M9186 Awaji, JP, December 2002

[9] ISO/IEC WG11 MPEG Video Group, "MPEG-4 video verification model version 16.0," ISO/IEC JTC1/SC29/WG11 MPEG00/N3312, Noordwijkerhout, March, 2000.