An Enhanced Multi-view Video Compression Using the Constrained Inter-view Prediction

Sunghwan Chun
Dept. of computer Science & Engineering
Pusan National University
Busan, Korea
+82-51-510-2877
rainingman23@gmail.com

Seoyoung Lee
Dept. of computer Science & Engineering
Pusan National University
Busan, Korea
+82-51-510-2877
memoiry@gmail.com

Kwangmu Shin
Dept. of computer Science & Engineering
Pusan National University
Busan, Korea
+82-51-510-2877
sin@pusan.ac.kr

Kidong Chung
Dept. of computer Science & Engineering
Pusan National University
Busan, Korea
+82-51-510-2877
kdchung@pusan.ac.kr

ABSTRACT

We propose a method that uses the restricted inter-view prediction for multi-view video coding. In the multi-view video, there exists occluded area because of the locations and angle of cameras. This increases the computational complexity, as it still uses both reference pictures for predicting the area which is not shown in the current frame. In this paper, we proposed a method that does not use the interview prediction in cases when macroblocks are occluded. Experimental results show that benefits can be obtained compared with the conventional approaches.

Categories and Subject Descriptors

I.4.2 [Computing Methodologies]: Image Processing and Computer Vision – Compression (Coding)

General Terms

Algorithms, Experimentation

Keywords

Multi-view Video Coding, Global Disparity Vector, Restricted Inter-view Prediction

1. INTRODUCTION

Recently, several types of technologies have been researched on the multimedia fields. Above all, Free-Viewpoint Video (FVV) [1] , [2], Free-Viewpoint Television (FTV) [3] and 3DTV [4], [5] (display technology that enables depth perception for the viewer) are typical applications. FVV is expected to be the next generation visual application. It enables users to watch a static or a dynamic scene from different viewing angles. These applications can be addressed by the key technology named multi-view video coding (MVC).

A multi-view video is a group of videos captured by multiple cameras of the same scene. MVC is used to compress the video signals captured by two or more cameras.

Generally, most of the video compression standards, including ITU-T H.263 [6], MPEG-4 [7] and the JVT codec [8] which is a joint standard of ITU-T Recommendation H.264 and ISO/IEC MPEG-4 Part 10, use a block-based motion estimation/compensation. In recent years, MPEG has investigated the needs for standardization in the area of 3D and free viewpoint video in a group called 3DAV. The results of in-depth investigating of MVC approaches were promising and MPEG decided to issue a "Call for Proposals" for MVC technologies along with related requirements. As a consequence, a new standard [9] for multi-view video coding is developed by the Joint Video Team (JVT) of VCEG and MPEG.

MVC, which has been developed to improve coding efficiency of multi-view video sequences, uses inter-view correlation between cameras. In MVC, the key point is how to use temporal, spatial and inter-view correlations. For these features, a well known approach is to apply block-based adaptive selection of intra coding, motion/disparity compensation-based coding [10], [11], [12]. In a MVC, motion estimation removes the temporal redundancy while disparity estimation removes the inter-view redundancy. Above all, disparity compensated prediction is a well known technique for exploiting the redundancy between different views. This prediction mode provides gains when the temporal correlation is lower than the spatial correlation because of the fast motion, occlusions and so forth. Therefore, MVC system can bring higher coding efficiency than the conventional coding standards such as H.264/AVC by adaptively selecting the prediction mode. However, MVC has a problem of having higher complexity compared to the conventional scheme due to additional prediction mode**,** i.e., interview prediction mode occupies computational complexity considerably.

In this paper, we propose a multi-view video coding method, which reduces the computation complexity by constraining the unpredictable region which is occurred by the location and angle between cameras. In section 2, we review the prediction structure on multi-view video coding and related work. Our proposed method is described in section 3 and the experimental results are presented in section 4. Finally, Section 4 concludes this paper.

SAC'09, March 8-12, 2009, Honolulu, Hawaii, U.S.A.

2. MULTI-VIEW VIDEO CODING

2.1 Multi-view Video Coding Structure

By now, several MVC schemes have been proposed. They are adaptively using motion estimation and disparity estimation for predictive coding. Among them, the basic coding scheme of HHI [13], [14] as shown in Fig.1, uses the hierarchical B prediction structure for each view. Additionally, the inter-view prediction is applied to every 2nd view, i.e. S1, S3 and S5 in Fig. 1. All views can be classified into two categories: main view (such as S0, S2, S4, S6) which needs motion compensated prediction only and auxiliary view (such as S1, S3, S5, S7) which can reference the main view. The frames in the main views are predicted by motion compensated prediction and those in the auxiliary views are predicted adaptively by motion estimation or disparity estimation.

Consequently, This HHI coding structure's performance is better than simulcast coding structure that doesn't use the inter-view prediction. However, HHI coding scheme has many computation complexity because of the more additional prediction. Thus, a trade-off relationship exists between performance and computation complexity.

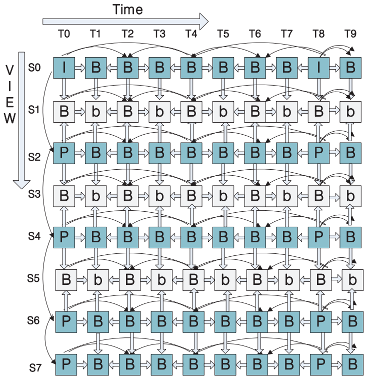

Figure 1. Multi-view Video Coding Scheme

2.2 Global Disparity

Multi-view video consists of several videos captured by multiple cameras which are aligned in a parallel or a convergent array. There exists a global disparity between adjacent views because of the camera's location and angle. Figure 2 shows the global disparity between "ballroom" view 0 and view 1. Although we captured the same scene, we can see that view1 is the result of shifted right of view 0 because of the location and angle between cameras.

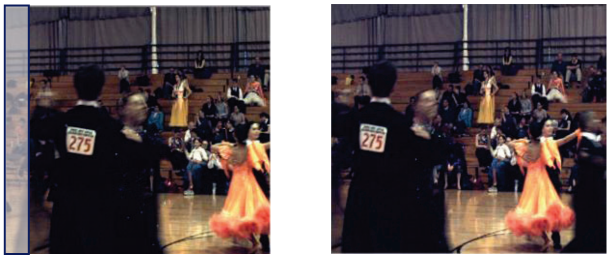

Figure 2. Global Disparity for View 0 and View 1 in the Ballroom sequence

There are several methods for calculating global disparity. One of them is a motion skip mode which employs the Mean Absolute Difference (MAD) per pixels obtaining the global disparity vector (GDV) of the anchor picture [15]. After that, the GDV is applied to the non-anchor picture.

2.2.1 Global Disparity Calculation (Anchor Picture)

The motion skip mode [16] is motivated by the idea that there is a similarity in motion information between the corresponding macroblocks in the neighboring two views.

In this mode, motion information of the current macroblock is inferred from the corresponding macroblock in the picture with the same temporal index of the adjacent view. The motion skip mode uses the MAD to calculate the GDV described in Fig. 3.

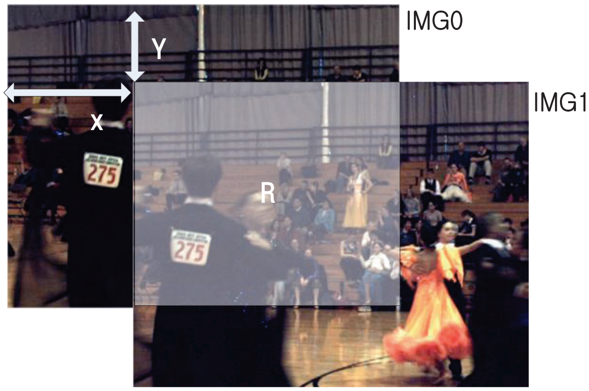

Figure 3. Global Disparity Calculation (Anchor Picture)

GDV is expressed in the following equation [15].

$$(g_x, g_y)_{MAD} = \min \left[\frac{1}{R} \sum_{i,j \in R} | img0(i,j) - img1(i-x, j-y) | \right] \quad \mathbf{(1)}$$

According to Fig. 3, img0 and img1 are two images for calculating GDV. R represents number of pixels which is overlapped area. The GDV (g_x,g_y) in the frame is a GDV using MAD. These GDVs are inserted into the slice header in the anchor picture.

2.2.2 Global Disparity Calculation (Non-Anchor Picture)

In a non-anchor picture, global disparity vector is derived from anchor picture's global disparity vector on in Fig.4. GDV is calculated from anchor pictures located at the both sides of GOP is used as a global disparity vector for non-anchor pictures.

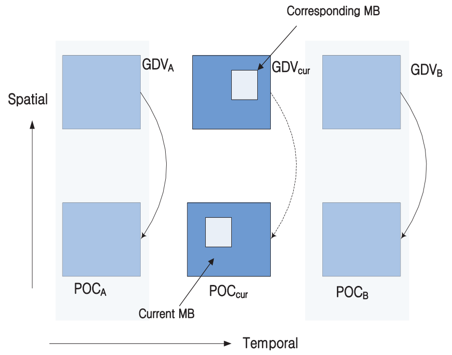

Figure 4. Global Disparity Calculation (Non-Anchor Picture)

$$GDVcur = GDV_A + \left[\frac{POCcur - POC_A}{POC_B - POC_A} \times (GDV_B - GDV_A) \right] \qquad \textbf{(2)}$$

In equ. 2 [16], GDV of the current frame uses a picture order count and a GDV which are obtained from anchor picture.

3. THE RESTRICTED INTER-VIEW PREDICTION STRUCTURE

3.1 Global Disparity Vector

As we know, Motion Skip mode's scheme can get the GDV using the two adjacent views. Also, GDV is very accurate because of calculating per pixels in the two frames. But our proposed method gets the GDV using calculating per Blocks in the two frames, not per pixels. The reason is that both calculation methods have similar RD performance and our method improves encoding time compared to the motion skip's method**.** We briefly explain our scheme below. First, we have to get the GDVs in the anchor pictures. In this paper, we get the displacement which is macro block 16 X 16 sizes that selected inter-view mode. And this DV is divided into a number which denotes the number of MBs selected as16 X 16 inter-view modes.

$$(g_x, g_y)_L = \frac{1}{C}\sum (dx, dy)_L, \quad (dx, dy)_L \in 16 \times 16\_DV_L \qquad \textbf{(3)}$$

$$(g_x, g_y)_R = \frac{1}{C}\sum (dx, dy)_R, \quad (dx, dy)_R \in 16 \times 16\_DV_R \qquad \textbf{(4)}$$

In Equ. 3, 4, (g_x, g_y) represents GDV. L, R is left reference frame and right reference frame respectively. In other words, in the current view, we derive GDVs from the left reference frame and right reference frame. These GDVs are used for current frame. C represents a number of MBs that are selected 16 X 16 inter-view modes. 16 X 16_DV$_L$ and 16 X 16_DV$_R$ represent disparity vectors which are selected 16 X 16 interview modes. In the non-anchor frame, uses the GDV which derives anchor frame like motion skip mode's non-anchor frame.

3.2 Restricted inter-view scheme

The restricted inter-view prediction scheme proposed in this paper is shown in Fig. 5.

Fig. 5. (a) represents the frame which is encoded in B-frame and (b) and (c) represents the frames which are referenced by the frame in Fig. 5. The square region which has a white vertical line on the right side of the frame at (a) does not have corresponding region in the left frame at (b). Also, the square region which has a vertical line on the left side of the frame at (a) does not have corresponding region in the right frame at (c). Therefore, macroblocks which of view 0 and view 2 belong to such regions at Fig.5 (a) can be predicted by referencing to the corresponding MBs of frames of only one view instead of referencing two frames.

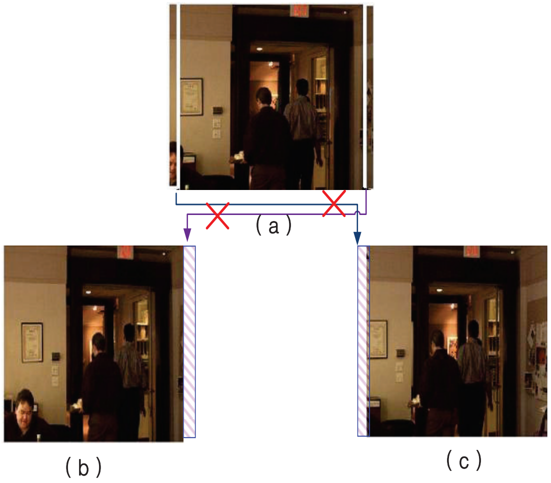

Figure 5. (a) is a view 1 which uses on inter-view prediction, (b) and (c) are reference frames of view 0 and view 2

3.3 The process of constrained inter-view prediction

We briefly describe our scheme in Fig. 6. First, in Step 0, GDVs in the anchor frame are derived. Second, in step 1, restricted macroblocks are set. In Step 2, restricted interview prediction is performed. This process of restricted inter-view prediction is recursive per GOP which have anchor pictures and non anchor pictures.

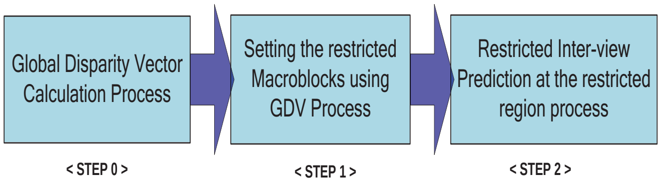
Global Disparity Vector Calculation Process
Setting the restricted Macroblocks using GDV Process
Restricted Inter-view Prediction at the restricted region process
< STEP 0 >
< STEP 1 >
< STEP 2 >

Figure 6. Proposed Algorithm Process

Step 0: We get the sum of disparity vectors which are selected as 16 X 16 inter-view modes in the anchor frame. After this, the sum is divided by a number which denotes a number of blocks selected as 16 X 16 interview mode. This process is applied for both reference frames to get two GDVs which are list 0 GDV and list 1 GDV.

Step 1: After Step, we can set restricted macroblocks in the non-anchor frame. Accurate GDV is a key point for a good performance.

Step 2: We already set the restricted macroblocks in the non-anchor frame. Restricted macroblocks can be predicted from macroblocks of a frame which has a corresponding region. The result reduces a computation complexity at the prediction process

4. EXPERIMENTAL RESULTS

To evaluate our proposed algorithm, **experiments** have been done considering both complexity and the quality of video. We carried out our experiment on two video sequences named "Ballroom" and "Exit", which were provided by Mitsubishi Electric Research Laboratories. The "Ballroom" sequence has a fast motion while the "Exit" sequence shows less motion relative to the previous sequence. These data were originally captured by 8 cameras. Both videos are rectified and are VGA (640 X 480) each consisting of 250 frames shown in Table 1.

Table 1. Experimental Sequence Features

Sequence	Resolution	Camera Arrangement	Frame Number
Ballroom	640 x 480 25fps	8 cameras, 1-D parallel	250
Exit	640 x 480 25fps	8 cameras, 1-D parallel	250

We carried out our experimentation based on JMVM 7.0, which is tested on Intel Pentium-IV based computer with 4GB RAM and Windows XP Professional operating system. Here are the experiment parameter settings: (FramesToBeEncoded = 121; SearchRange = 32, 64, 96; QP value: 34; GOP size: 12). Also, in the GOP structure, hierarchical- B is used in the temporal direction and only Intra mode is applied in the view direction.

Table 2. Encoding Time Comparison

Search-Range Sequence		32	64	96	Avg.Time Saving
Ballroom	JMVM(sec)	8725	14729	23688	3.5%
	Proposed(sec)	8512	14207	22787	
Exit	JMVM(sec)	8140	13885	20988	3.7%
	Proposed(sec)	8035	13354	20036	

Table 2 shows the encoding time of our proposed method relative to JMVM in the "Ballroom" and the "Exit" sequences. Through the various search ranges, we can see that our proposed method has reduced computation time.

Fig. 7 shows the encoding time of frames of the JMVM and our proposed method in the Ballroom sequence. Our scheme reduces encoding time about 5 ~ 8 second per frame related to JMVM.

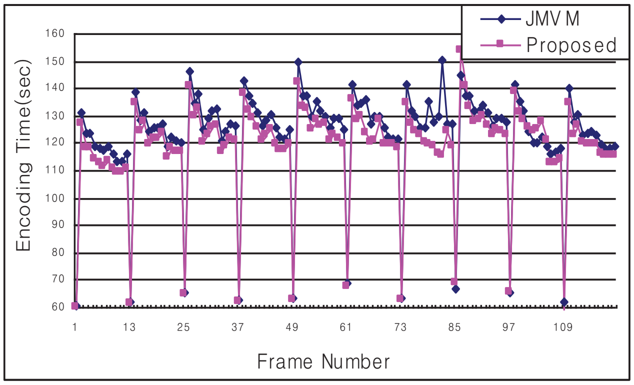

Figure 7. Encoding Time Comparison per frame in the Ballroom sequence

Table 3, 4 shows PSNR and Bitrates according to the various search ranges. The results are nearly equal when comparing the proposed algorithm with JMVM. The reason is that the prediction process is not carried out for the region which is not needed to be predicted.

Table 3. PSNR and Bitrates Comparison in the Ballroom sequence

Search Range	PSNR(dB)		Bitrates(bits/sec)	
	JMVM	Proposed	JMVM	Proposed
32	31.8482	31.8452	190.1868	191.6843
64	31.8642	31.8602	183.9471	185.9355
96	31.8650	31.8590	183.8542	185.3636

Table 4. PSNR and Bitrates Comparison in the Exit sequence

Search Range	PSNR(dB)		Bitrates(bits/sec)	
	JMVM	Proposed	JMVM	Proposed
32	34.7129	34.7101	77.2926	77.5041
64	34.7263	34.7190	75.8793	76.5851
96	34.7289	34.7196	76.0612	76.4579

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a restricted inter-view prediction method for multi-view video coding. Conventional JMVM predict the reference frame using full search block matching without considering displacement of the camera's distance. But our proposed scheme does not employ a inter-view prediction for the region which is not in the reference frame considering displacement of camera's distance. From the experimental results we can see that proposed method reduces a computation time with only negligible loss of performance. Consequently, our proposed scheme is effective for reducing computation complexity using restricted inter-view prediction without degraded quality. Although limitation of our proposed method is only possible to predict different views using full search mode, our scheme have a further improvement in time without quality degradation.

In the future work, we'll need to carry out experiments which get the more accurate GDV for a better performance. And we'll conduct to verify other sequences as well as two sequences which were conducted.

6. REFERENCES

[1] M.Tanimoto.: FTV (Free Viewpoint Television) creating ray-based image engineering. Proc. IEEE Int'l Conf. Image Proc., vol. 2, pp.25-28, Genoa, Italy, Sept. (2005)

[2] A. Smolic, K. Muller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and Thomas Wiegand.: 3D Video and Free Viewpoint Video –Technologies. Applications and MPEG Standards, ICME 2006, IEEE International Conference on Multimedia and Expo, Toronto, Ontario, Canada, July (2006)

[3] ISO/IEC JTC1/SC29/WG11 M11259.: FTV (Free Viewpoint Television): achievements and Challenge. October (2004)

[4] ISO/IEC JTC/SC29/WG11, N5878.: Report on 3DAV Exploration. July 2003

[5] A. Smolic, and P. Kauff.: Interactive 3D Video Representation and Coding Technologies. Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, Jan (2005)

[6] ITU Telecom. Standardization Sector.: Video Codec Test Model Near-Term, Version 10 Draft 1. H.263 Ad Hoc Group, April (1998)

[7] ISO/IEC JTC1/SC29/WG11 N3056.: Information Technology- Coding of Audio-Visual Objects Part2: Visual Amendment 1: Visual Extensions, Dec. (1999)

[8] T.Wiegand.: Final draft international standard for joint video specification H.264. In JVT of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, Mar (2003)

[9] ISO/IEC JTC1/SC29/WG11 W8019.: Description of Core Experiments in MVC. April (2006)

[10] H.Kimata, M.Kitahara, K.Kamikura, and Y. Yashima.: Multi-view video coding using reference picture selection for freeviewpoint video communication. PCS2004, (2004)

[11] S. C. Chan, K. T. Ng, Z. F. Gan, K. L. Chan, and H. Y. Shum.: The plenoptic video. IEEE Trans., Circuits Syst., Video Technol., vol. 15, no. 12, pp. 1650-1659, (2005)

[12] Survey of algorithms used for multi-view video coding (mvc). Document N6909 MPEG Hong Kong Meeting, (2005)

[13] ISO/IEC JTC1/SC29/WG11.: Description of Core Experiments in MVC. MPEG2006/W7798, January (2006)

[14] Kwangmu Shin, Seoyoung Lee, Sungmin Kim, Kidong Chung.: An Improved GoGop Structure for Multi-view Video Coding in H.264/AVC. Korean Institute of Information Scientists and Engineers, Proc. of The 34st KIISE Fall Conference, vol. 34, no. 2, pp. 383-387, (2007)

[15] Kwan-Jung Oh, Yo-Sung Ho.: Global Disparity Compensation for Multi-view Video Coding. Journal of The Korean Society of Broadcast Engineers, vol.12, no.6, (2007)

[16] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 JVT-W081.: MVC Motion Skip Mode. April, (2007)