Inter-View Direct Mode for Multiview Video Coding

Xun Guo, Yan Lu, Feng Wu, *Senior Member, IEEE*, and Wen Gao

***Abstract*—Global disparity between views is usually caused by the displacement between cameras, which can be accurately represented by a global geometric transformation. In this paper, we first propose an inter-view motion model in terms of the global geometric transformation to represent the motion correlation between two adjacent views. Specifically, the motion vector of a pixel in one view may be directly derived from that in another view according to the inter-view motion model. Further, we propose an inter-view direct mode to signal the decoder that the motion of a macroblock (MB) can be achieved from the coded view without any coding bits. The proposed inter-view direct mode is further incorporated in the existing multiview video coding (MVC) schemes (i.e., AVC-based MVC and 4-D wavelet-based MVC), working together with the other classical coding modes. The mode selection at each MB is accomplished with the rate-distortion optimization technique. The proposed inter-view direct mode can significantly reduce bits to code motion vectors especially at low bit rates, thus improving the coding efficiency.**

***Index Terms*—Direct mode, inter-view correlation, motion vector prediction, multiview video coding (MVC).**

I. INTRODUCTION

IN MODERN video coding schemes, motion prediction is becoming more and more accurate for better coding efficiency. For example in MPEG-4 AVC/H.264 standard [1], a macroblock (MB) in P-picture may have up to 16 different motion vectors, while an MB in B-picture may even have up to 32 different motion vectors in the case of bidirectional prediction. However, it also implied that many bits have to be assigned to code these motion vectors. Although motion vectors are relatively efficiently coded considering that they are differentially coded compared to a motion vector predictor, taken as the median value of the motion vectors of the spatially adjacent blocks, MPEG-4 AVC/H.264 standard still introduces two direct modes: SKIP and DIRECT within P- and B-pictures respectively, for a further reduction to the amount of bits for coding motion vectors [2].

SKIP mode assumes that the current MB has motion vectors equal to the median predictor. Furthermore, no residual information is transmitted for a SKIP MB. DIRECT mode can infer motion in either spatial or temporal prediction. Spatial prediction for DIRECT MBs is relatively similar to that of SKIP MBs although in this case up to two different motion vectors are needed for bidirectional prediction. In the temporal prediction, motion is inferred through the consideration of motion data in previously decoded pictures and is based on the assumption that an object is moving with relatively constant speed. In this case motion vectors are derived by temporally interpolating the motion vectors of a co-located MB in the forward reference.

SKIP and DIRECT modes are also applicable in multiview video coding (MVC). Some MVC schemes based on H.264 have been proposed in [3]–[5], by virtue that the existing multiple reference structure in the standard can be borrowed to alternatively utilize motion-compensated prediction (MCP) and disparity-compensated prediction (DCP). Recently, MPEG has started the MVC standardization process. Accordingly, a number of MVC techniques have been proposed to MPEG [6]–[8]. In particular, an AVC-based MVC scheme with hierarchical B-prediction structure and inter-view texture prediction is developed in [6]. It has been selected as the reference solution in the forthcoming core experiments in MPEG 3-DAV group. Besides the investigation on AVC-based MVC schemes, several efforts have been invested in developing high dimensional wavelet coding for multiview video [9], where temporal and view filters replace MCP and DCP, respectively. These filters also need precise motion vectors between frames and between views so that the wavelet decomposition along motion trajectories can result in high energy-compacted sub-bands.

Then, a question arises here. Besides the traditional SKIP and DIRCT modes, do we have more efficient way to compress motion vectors in the MVC? The common idea in most of the investigations on MVC is how to efficiently exploit the correlations between adjacent views for pixel prediction [10], [11]. It enlightens us whether the view correlation can be also used to code the motion vectors. As we know, the disparity between views is caused by the locations and angles of two cameras. It can be accurately represented by a global geometric model. When the motion vectors in one view have been estimated and coded, the motion vectors in another view should be able to be derived from that in the coded view through a certain model supposing only simple local motion exists in the video.

Upon the above idea, in this paper we propose an inter-view motion model to describe the motion correlations between video sequences from different views, based on which the motion vector of a pixel in a view is directly calculated from that of a coded view. Since pixels are usually processed at MB unit in video coding schemes, we further propose an inter-view direct mode to signal the decoder that motion vectors of a MB can be generated from the coded view without any coding bits. The proposed mode is also applicable in MVC schemes containing the classical coding modes. Furthermore, we have incorporated the view direct mode into the AVC-based MVC scheme in [6] and the 4-D wavelet-based MVC scheme in [9], respectively. Both schemes demonstrate the advantages of the proposed technology.

Manuscript received December 05, 2005; revised March 31, 2006. The work was performed at Microsoft Research Asia, Beijing, China, and was supported in part by the National Science Foundation of China under Grant 60333020. This paper was recommended by Associate Editor O. C. Au.

X. Guo is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: xguo@vilab.hit.edu.cn).

Y. Lu and F. Wu are with Microsoft Research Asia, Beijing 100080, China (e-mail: yanlu@microsoft.com; fengwu@microsoft.com).

W. Gao is with the School of Electronic Engineering and Computer Science, Peking University, Beijing 100080, China (e-mail: wgao@jdl.ac.cn).

Digital Object Identifier 10.1109/TCSVT.2006.885724

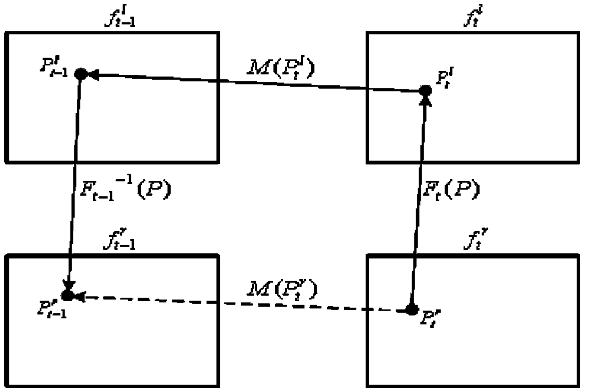


Fig. 1. Geometric relationship of pixels and motion vectors in two views.

The rest of this paper is organized as follows. Section II presents the inter-view motion model to derive the motion of a view from that of a coded view. Section III discusses the proposed inter-view direct mode and its applications in the MVC schemes. In Section IV, experimental results are given. Finally, Section V concludes this paper.

II. INTER-VIEW MOTION MODEL

In this section, we will model the relationship between the temporal motions at different views based on a geometric model. The motion of an object usually indicates the displacement of the object in two successive frames, which can be represented by such motion models as the block-based translational model with pixel matching between the two frames. Moreover, we know that a pair of frames at the same time instant but from different views also corresponds to the same content. Similar to the motion model, the disparity model is used to indicate the displacement of an object in two view frames. Therefore, it is possible to find all corresponding pixels with respect to the same content in both temporally and view neighboring frames. Intuitively, we can deduce the inter-view motion correlation based on the pixel mapping along both temporal and view directions.

We take the case of two-view video sequences (i.e., the stereoscopic video) as an example to describe the proposed inter-view motion model. Above all, we define some notations as used in Fig. 1. Let f_t^v, $v \in \{l, r\}$, represent the frame at time t in view v, and hence f_t^l and f_t^r indicate a stereo pair at time t. We assume that the temporal motion of every pixel in the view l is already known. Now the problem is how to derive the temporal motion of every pixel in the view r. In this paper, we still employ the translational model to represent the temporal motion because it has been extensively used in existing hybrid video coding schemes. Let $P_t^r = (x, y)^T$ represent the location of one pixel in f_t^r, and let P_t^l, P_{t-1}^r and P_{t-1}^l represent the locations of its corresponding pixels in f_t^l, f_{t-1}^r and f_{t-1}^l, respectively. As shown in Fig. 1, $M\left(P_t^l\right)$ denotes the motion vector of P_t^l in f_t^l pointing to P_{t-1}^l in f_{t-1}^l, i.e., $M\left(P_t^l\right) = P_{t-1}^l - P_t^l$. Then, the problem becomes how to derive $M\left(P_t^r\right)$ from $M\left(P_t^l\right)$, i.e., the motion vector of P_t^r in f_t^r.

Towards the above goal, we also need get the disparity-based pixel mapping in the stereo pair. Let $F_t(P)$ indicate the mapping of pixel P from f_t^r to f_t^l. It is well known that the disparity is usually very compact and can be represented by several global disparity models [12], [13]. In this paper, we employ the six-parameter affine transformation to represent the global disparity. Therefore, we define $F_t(P)$ as

$$F_t(P) = A_t P + C_t \tag{1}$$

where

$$A_t = \begin{pmatrix} a_{1,t} & a_{2,t} \\ b_{1,t} & b_{2,t} \end{pmatrix} \text{ and } C_t = \begin{pmatrix} c_{1,t} \\ c_{2,t} \end{pmatrix}. \tag{2}$$

Here A_t is a 2×2 transform matrix and C_t is an offset vector. Given the pixel P_t^r in f_t^r, we can get its corresponding pixel P_t^l in f_t^l through the above affine transformation, i.e., $P_t^l = F_t\left(P_t^r\right)$. Since the motion vector of P_t^l in f_t^l is already known, we can also easily find its corresponding pixel P_{t-1}^l in f_{t-1}^l by

$$P_{t-1}^l = M\left(P_t^l\right) + P_t^l. \tag{3}$$

In addition, we assume that the global disparity in the stereo pair at time $t-1$ is represented by the affine transformation $F_{t-1}(P)$ as well, which indicates the mapping of pixel P from f_{t-1}^r to f_{t-1}^l. $F_{t-1}(P)$ may have transform matrix and offset vector different to $F_t(P)$. Therefore, we define it as

$$F_{t-1}(P) = A_{t-1} P + C_{t-1}. \tag{4}$$

As for the pixel mapping from f_{t-1}^l to f_{t-1}^r, we need the inverse of $F_{t-1}(P)$, i.e.,

$$F_{t-1}^{-1}(P) = A_{t-1}^{-1} P - A_{t-1}^{-1} C_{t-1}. \tag{5}$$

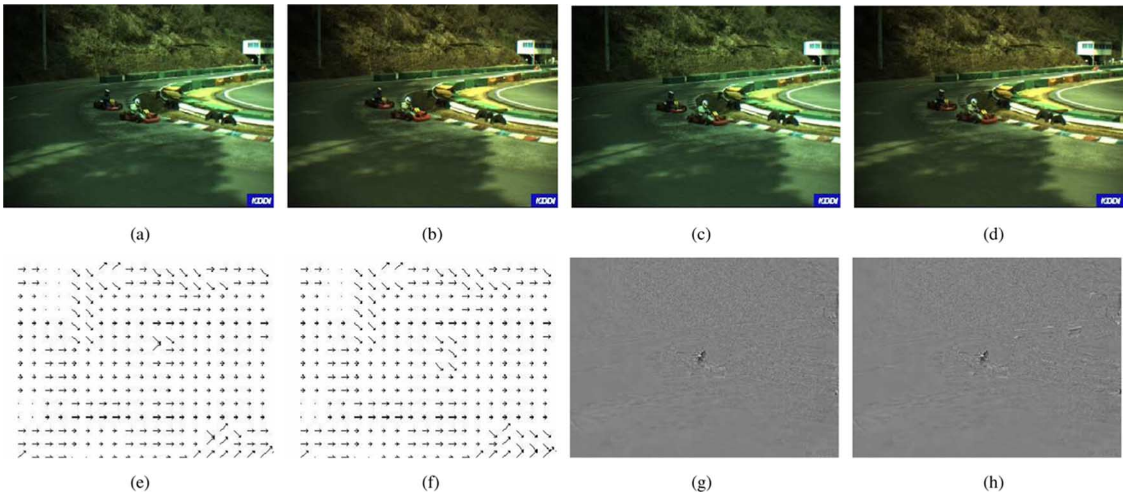
Then, given pixel P_{t-1}^l in f_{t-1}^l, we can get its corresponding pixel P_{t-1}^r in f_{t-1}^r through the above affine transformation model, i.e., $P_{t-1}^r = F_{t-1}^{-1}\left(P_{t-1}^l\right)$. Thus, we can derive the motion vector of P_t^r in f_t^r from $M\left(P_t^r\right) = P_{t-1}^r - P_t^r$. Combining (1), (4), and (5), we have

$$M(P_t^r) = A_{t-1}^{-1} M\left(P_t^l\right) + A_{t-1}^{-1} C_t - A_{t-1}^{-1} C_{t-1} + A_{t-1}^{-1} A_t P_t^r - P_t^r. \tag{6}$$

Consider that the relative locations and angles of the cameras are usually fixed. In this case, the global disparity models of the two successive stereo pairs are almost identical. Then, $F_t(P)$ and $F_{t-1}(P)$ have the same parameters, i.e., $A_{t-1} = A_t$ and $C_{t-1} = C_t$. Thus, we can further simplify (6) as

$$M\left(P_t^r\right) = A_t^{-1} M\left(P_t^l\right). \tag{7}$$

Both (6) and (7) clearly indicate that we can deduce the temporal motion of one view from another if the disparity between views



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2. Illustration of inter-view motion model. (a) First frame of the left view. (b) First frame of the right view. (c) Second frame of the left view. (d) Second frame of the right view. (e) Motion field from the block-based motion estimation. (f) Motion field from the inter-view motion model. (g) Prediction errors in terms of motion field in (e). (h) Prediction errors in terms of the motion field in (f). (Color version available online at: http://ieeexplore.ieee.org.)

and the temporal motion of the latter view are available. In particular in the case of (7), we can deduce the motion vector of every pixel in the view to be coded by simply scaling the motion vector of its corresponding pixel in the coded view.

Here we give an example to illustrate the accuracy of the derived motion vectors based on the proposed inter-view motion model. Above all, we explain some practical issues in the test of the inter-view motion model. We employ the block-based full-search motion estimation scheme to achieve the motion vector of every pixel in the left view. Secondly, we employ the global motion estimation algorithm proposed in [14] to calculate the affine transformation parameters between every stereo pair. Thirdly, since the location calculated in (1) may correspond to a fractional position or a position outside the image, we force the motion vector at this location to be the one at the closest integer pixel. Finally, we round the derived motion vector to quarter pixel precision due to the consideration of the tradeoff between prediction accuracy and complexity of interpolation.

Fig. 2 shows the original images, motion fields and prediction errors in terms of two successive stereo pairs in a multiview video. Note that the distributions of the prediction errors from the inter-view motion model and from the direct motion estimation are very similar. The corresponding PSNR is also calculated in terms of the motion-compensated prediction errors. The PSNR from the inter-view motion model is 32.16 dB, and the PSNR from the direct motion estimation is 32.34. In other words, the objective evaluation in terms of PSNR also shows the accuracy of the inter-view motion model. Note further, that the relatively larger prediction errors from the inter-view motion model are usually clustered in some local regions where complicated motion exists.

III. INTER-VIEW DIRECT MODE

As discussed above, the motion of a view can be derived from another view through (6) or (7), depending on whether the parameters of the global geometric model in two successive frame pairs are equivalent or not. However, in the practical MVC schemes, obviously not all pixels can derive their motion vectors with the inter-view motion model. It is no doubt that the proposed method should be jointly utilized with some existing methods. Accordingly, we propose a new direct mode namely inter-view direct mode for MVC. The proposed coding mode is integrated into the AVC-based MVC scheme and the 4-D wavelet-based MVC scheme, respectively.

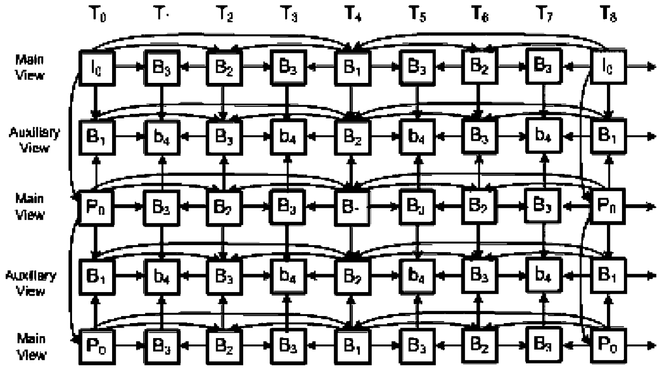


Fig. 3. Employed MVC structure based on [6].

A. AVC-Based MVC

In recent years, a number of MVC algorithms based on the AVC standard have been proposed [4]–[8]. They have the common feature of adaptively using MCP and DCP for predictive coding. In particular, the scheme in [6] has shown significant coding gains over the simulcast anchors, in which a prediction structure with hierarchical B-pictures is employed. Since the inter-view motion correlation has not been utilized in this scheme, we further extend it by incorporating the proposed inter-view direct mode. Above all, we give a brief overview of the employed AVC-based MVC scheme. As shown in Fig. 3, the basic coding scheme uses the hierarchical B-prediction structure for each view. The inter-view texture prediction is also applied to every second view. In summary, this MVC scheme mainly uses the existing AVC techniques to exploit temporal and inter-view dependencies.

TABLE I
USAGE PERCENTAGES OF MCP AND DCP MODES

	with DCP modes			w/o DCP modes	
	inter-view direct	other MCP	DCP	inter-view direct	other MCP
Race1	34.0%	50.3%	15.7%	35.7%	64.3%
Flamenco2	13.3%	78%	8.7%	14%	86%

In terms of the above MVC scheme, all MB coding modes can be classified into two categories: MCP and DCP. In fact, the MCP modes are often selected, especially when the luminance and/or chrominance variance across views exists and the multiview video is not rectified. In the original scheme in [6], only the temporal texture correlation is utilized when an MCP mode is selected. As discussed previously, even in the case of temporal MCP, the corresponding motion vector can possibly be derived from the other views due to the existence of inter-view motion correlation. Therefore, the proposed inter-view direct mode can be taken as a new MCP mode and be incorporated into the above MVC scheme. The rate-distortion optimization technique in [15] can be employed for coding mode selection. In addition, the proposed new coding mode used in an auxiliary view references another coded view (i.e., the main view) for motion vector derivation. Accordingly, we classify all views into two categories, e.g., taking every second view as an auxiliary view as shown in Fig. 3.

Some practical issues about how to derive the motion vectors in terms of the inter-view motion model have been addressed in Section II. Here we only address some practical issues about how to derive the prediction directions when the inter-view direct mode is used. For an MB coded with inter-view direct mode in a P-picture, every pixel in the MB only corresponds to the forward prediction. For an MB coded with inter-view direct mode in a B-picture, every pixel corresponds to a pixel (after rounding) in the main view. Thus, the pixel in the current MB has the same prediction directions as the corresponding pixel in the main view, including forward, backward, and bidirectional predictions. Note further, that it is unnecessary to code the prediction directions when it is coded with inter-view direct mode, because they are also derived from the main view.

B. 4-D Wavelet-Based MVC

In our previous work, a 4-D wavelet-based MVC algorithm has been proposed [9]. The basic idea is to decompose the multiview video using the lifting-based high dimensional wavelet transform. An example decomposition structure is shown in Fig. 4. Since the temporal correlation is usually stronger than the inter-view one, several levels of motion-compensated temporal filtering (MCTF) are performed first, followed by a few levels of disparity-compensated view filtering (DCVF). The coefficient frames in every temporal-view subband are further decomposed using the 2D spatial digital wavelet transform (DWT). As shown in Fig. 4, the 2-level MCTF, 1-level DCVF, and 2-level 2D spatial DWT are subsequently performed. The decomposed temporal-view-spatial subbands are coded with the entropy coding algorithm developed from 3-D-ESCOT in [16].

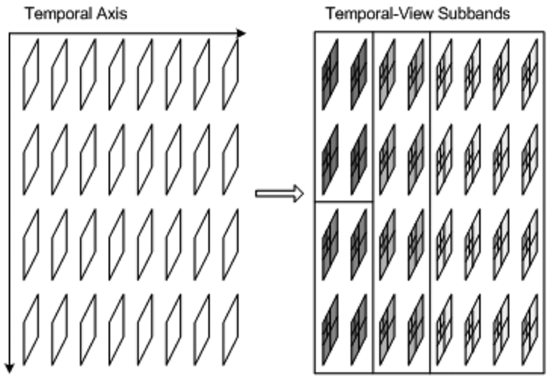


Fig. 4. Illustration of the 4-D wavelet decomposition.

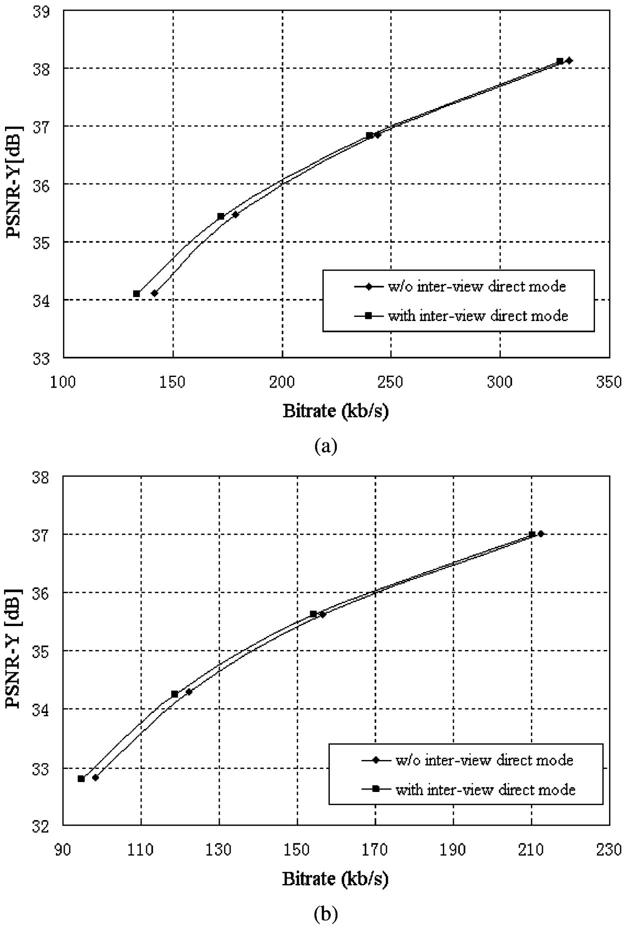
In general, the MCTF dominates the overall coding performance, even though the DCVF can further improve the coding efficiency. The original MCTF only exploits the temporal texture correlation. To jointly utilize the inter-view motion correlation and the temporal texture correlation, we propose to incorporate the inter-view direct mode in MCTF in MVC. The motion vector derivation process for a MB with inter-view direct mode is the same as before. After the motion vectors are derived, the prediction and updating steps in the lifting-based MCTF are performed on the view video along motion trajectories in the temporal direction. Thanks to the energy-distributed updating (EDU) technique [17] adopted in the MCTF, the derived motion vector need not be inversed in the updating step. For more details about the other parts of the 4-D wavelet-based MVC scheme, we refer to [9].

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed inter-view direct mode in terms of the AVC-based MVC scheme and the 4-D wavelet-based MVC scheme, respectively. The results on *Race1, Flamenco2,* and *Flamenco1* sequences provided by KDDI Lab [18] are shown in this paper. All multiview videos have the frame size of 320×240 and frame rate of 30 fps, representing the simple and complex local disparities, respectively. The testing results are presented as follows.

A. AVC-Based MVC

The performance of the original AVC-based MVC scheme has been reported in [6]. In this section, we only present the improvement of the proposed coding technique. For each multiview video, all views are coded with the hierarchical B structure as shown in Fig. 3, and the overall performance is given in terms of the rate-distortion curves. Above all, we get some statistics on the usage percentages of the MCP and DCP modes. Since the proposed inter-view direct mode is only used in auxiliary



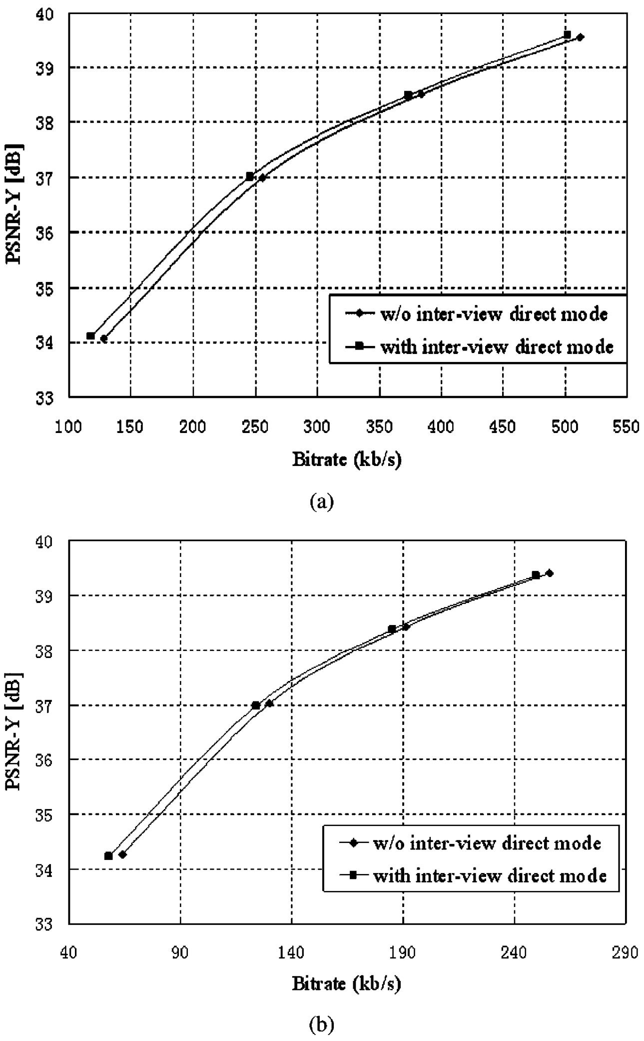
(a)

(b)

Fig. 5. Results of the AVC-based MVC. (a) *Race1*. (b) *Flamenco2*.

views, the statistics is done on all auxiliary views. The statistics in Table I shows that the MCP modes are selected with a large ratio. For example, for *Race1*, 84.3% MBs are coded with MCP modes, including 34.0% MBs coded with inter-view direct mode. The multiview video has not been rectified, which may more or less decrease the number of DCP MBs. However, the major reason is that the inter-view texture correlation is usually not as strong as the temporal one. In addition, the percentages of MBs coded with inter-view direct mode change little between the MVC with and without DCP modes, which indicates that the inter-view direct mode has most probably replaced some traditional MCP modes rather than the DCP modes.

Fig. 5 shows the overall coding results. We can observe that the proposed inter-view direct mode can always improve the coding efficiency. The bit rate saving is up to 6% for *Race1* and 4% for *Flamenco2*. Note that the coding gain depends on the percentage of the bits for coding motion vectors and the prediction structures. As shown in Fig. 5, the coding gain at the high bit rate is not as obvious as that at the low bit rate, because the amount of bits for coding motion vectors increases slower than that for coding prediction errors. Furthermore, as shown in Fig. 5, the inter-view direct mode works better for *Race1* rather than for *Flamenco2*. The reason lies in that the global disparity is less effective for the regions with local disparity. In this case, the derived motion vectors in terms of the inter-view direct mode may not be accurate enough. The complex motion field may also influence the accuracy of global disparity estimation.



(a)

(b)

Fig. 6. Results of the 4-D wavelet-based MVC. (a) *Race1*. (b) *Flamenco1*.

B. 4-D Wavelet-Based MVC

Although the 4-D wavelet-based MVC framework supports the content-adaptive wavelet decomposition structure, we employ the fixed decomposition structure in our test, i.e., 4-level MCTP followed by 2-level DCVF, as used in [9]. Considering the efficiency of the DCVF decomposition, all eight views are coded for each multiview video. The performance of the original 4-D wavelet-based MVC scheme compared with the H.264-based simulcast coding has been reported in [9]. In this paper, we only present the evaluation of the proposed inter-view direct mode in the 4-D wavelet-based MVC scheme. In particular, the coding results of the 4-D wavelet-based MVC with and without inter-view direct mode are reported, respectively.

Since the 4-D wavelet-based MVC scheme supports the SNR scalability, we create only a single bitstream for each test and get the coding results by truncating the bitstream at the targeted bit rates. To evaluate the saving of bit rate, the selection of the targeted bit rates aims at achieving the similar PSNR in terms of the two coding schemes. Fig. 6 shows the rate-distortion (R-D) curves. We can observe that the proposed inter-view direct mode can achieve up to up to 8% bit rate saving for *Race1* and 9% bit rate saving for *Flamenco1*. Note further, that the same set of motion bits has been used in decoding the bitstream truncated at any targeted bit rate. In other words, the portion of motion bits also decreases at high bit rates, which leads to the coding gain not as obvious as that at the low bit rates. Nevertheless, the coding results have shown that the proposed inter-view direct mode can achieve coding gains in all cases.

C. Discussion of Complexity

The complexity issues in terms of the proposed inter-view direct mode in both the AVC-extension and 4-D wavelet-based MVC schemes are discussed as follows. At the encoder, the proposed algorithm does increase the complexity to some extent due to the following two aspects. First, the global disparity estimation between two adjacent views is performed in addition to the traditional block-based disparity estimation, which increases the complexity in disparity and motion search. As we know, the global disparity information is also helpful for enhancing inter-view texture prediction accuracy. Second, the proposed algorithm adds a new candidate coding mode (i.e., the proposed one), which increases the complexity in coding mode selection. However, the complexity of a typical AVC encoder is inherently high due to the adoption of some efficient coding techniques. Therefore, the increased encoding complexity due to the proposed coding mode is acceptable. At the decoder, the complexity introduced by the proposed coding mode is comparable to the traditional temporal direct mode in H.264. The only difference lies in the methods of deriving the motion vectors. With the proposed coding mode, the block-level motion vector is derived after being scaled twice. Actually, the temporal direct mode in H.264 also involves the scaling operations. Therefore, the decoding complexities with and without the proposed mode almost remain the same.

V. CONCLUSIONS

In this paper, the temporal motion correlation between views in the multiview video has been investigated, based on which an inter-view motion model is proposed. In particular, the affine transformation model is employed to represent the global disparity between any two views. Thus, the temporal motion of one view can be derived from that in another view supposing the affine transformation parameters are known. Due to the consideration of the bit saving of motion information in coding, we further propose an inter-view direct mode based on the inter-view motion model. The proposed new coding mode has been incorporated in the existing MVC schemes, including the AVC-based MVC and the 4-D wavelet-based MVC. Experimental results have shown that the proposed inter-view direct mode can achieve coding gains in all cases in both MVC schemes

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] A. M. Tourapis, F. Wu, and S. Li, “Direct mode coding for bipredictive slices in the H.264 standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 119–126, Jan. 2005.

[3] A. Vetro, W. Matusik, H. Pfister, and J. Xin, “Coding approaches for end-to-end 3-D TV systems,” presented at the Picture Coding Symp., San Francisco, CA, Dec. 2004.

[4] X. Guo and Q. Huang, “Multiview video coding based on global motion model,” *Lecture Notes Comput. Sci.*, vol. 3333, pp. 665–672, Dec. 2004.

[5] G. Li and Y. He, “A novel multiview video coding scheme based on H.264,” in *Proc. ICICS-PCM 2003*, Singapore, Dec. 2003, pp. 493–497.

[6] K. Mueller, P. Merkle, A. Smolic, and T. Wiegand, “Multiview coding using AVC,” presented at the MPEG2006/M12945, 75th MPEG Meeting, Bangkok, Thailand, Jan. 2006, Doc. m12945.

[7] Y.-L. Lee, J.-H. Hur, D.-Y. Kim, and Y.-K. Lee, “H.264/MPEG-4 AVC-based multiview video coding,” presented at the MPEG2006/M12871, 75th MPEG Meeting, Bangkok, Thailand, Jan. 2006, Doc. m12871.

[8] J.-R. Ohm, “Submissions received in CfP on multiview video coding,” presented at the MPEG2006/M12969, 75th MPEG Meeting, Bangkok, Thailand, Jan. 2006, Doc. m12969.

[9] W. Yang, Y. Lu, F. Wu, J. Cai, K. Ngan, and S. Li, “4-D wavelet-based multiview video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, 11, Nov. 2006.

[10] M. Siegel, S. Sethuraman, J. McVeigh, and A. Jordan, “Compression and interpolation of 3-D stereoscopic and multiview video,” in *Proc. SPIE Stereoscop. Displ. Virtual Reality Syst.*, 1997, vol. 3012, pp. 227–238.

[11] B. L. Tseng and D. Anastassiou, “Multiviewpoint video coding with MPEG-2 compatibility,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 4, pp. 414–419, Jul. 1996.

[12] D. V. Papadimitrion and T. J. Dennis, “Three dimensional parameter estimation from stereo image sequences from model-based image coding,” *Signal Process.: Image Commun.*, vol. 7, pp. 471–487, 1995.

[13] R. S. Wang and Y. Wang, “Multiview video sequence analysis, compression, and virtual viewpoint synthesis,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 397–410, Apr. 2000.

[14] F. Dufaux and J. Konrad, “Efficient, robust, and fast global motion estimation for video coding,” *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 497–501, Mar. 2000.

[15] T. Wiegand and B. Girod, “Lagrange multiplier selection in hybrid video coder control,” in *Proc. Int. Conf. Image Process.*, Greece, 2001, vol. 3, pp. 542–545.

[16] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, “Three-dimensional embedded subband coding with optimized truncation (3-D Escot),” *J. Appl. Comput. Harmonic Anal.*, vol. 10, pp. 290–315, May 2001.

[17] R. Xiong, F. Wu, J. Xu, S. Li, and Y.-Q. Zhang, “Barbell lifting wavelet transform for highly scalable video coding,” presented at the Picture Coding Symp. 2004, San Francisco, CA, Dec. 2004.

[18] R. Kawada, “KDDI multiview video sequences for MPEG 3-DAV use,” presented at the MPEG2004/M10533, 68th MPEG Meeting, Munich, Germany, Mar. 2004, Doc. m10533.