Fast Disparity and Motion Estimation Based on Correlations for Multiview Video Coding

Xiaoming Li, Debin Zhao, Siwei Ma, and Wen Gao

Abstract —*Recently, multiview video coding has attracted great attention from industries and research institutes. However, the heavy computational complexity limits its practical applications. In this paper, a fast disparity and motion estimation for multiview video coding is presented, based on the correlations between the neighboring cameras and between the motion and the disparity. In the proposed approach, first, a search region estimation is proposed to reduce the disparity estimation complexity according to that the camera set is usually fixed and therefore the disparity between the two neighboring views can be limited to an estimable range. Second, a motion vector derivation is given based on the geometric relationship between the motion and the disparity. In addition, an early termination scheme is provided to further reduce the number of reference frames. The experimental results show that roughly 50% time saving for disparity and motion computing can be reached when compared to the anchor in multiview video coding test model JSVM only with negligible coding efficiency loss*[1].

Index Terms —Multiview video coding, disparity estimation, fast motion estimation, geometric relationship.

I. INTRODUCTION

Multiview video is a group of video sequences captured by a set of cameras on different positions at the same time instance and from the same scene. It is much favored in the services such as free-viewpoint video (FVV), free-viewpoint television (FTV), and 3DTV [1], which expand the user's sensation far beyond what is offered by the traditional video. Multiview video coding (MVC) has attracted great attention from industries and research institutes. Recently, MPEG 3-D audio and video (3DAV) ad hoc group has been investigating the needs for standardization in 3DTV and FTV. The Joint Video Team (JVT) is also working on a MVC extension of H.264/AVC to meet the industry demand.

To improve the compression efficiency of multiview video coding, many algorithms for motion estimation have been proposed. A pixel-based hierarchical dynamic programming algorithm is used for disparity estimation [3]. A reference block-based depth estimation approaches for the view synthesis is presented in [4]. In [5], a temporal/inter-view prediction structure is proposed to reduce the view redundancy. HHI proposes a H.264/AVC based multiview video coding scheme which has achieved much better performance compared to the simulcast coding [6].

The basic coding scheme in [6], as shown in Fig. 1, uses the hierarchical B prediction structure for each view. The inter-view prediction is applied to every other view. All views can be classified into two categories: main view (such as S0, S2, S4, and S6), which needs motion-compensated prediction only and auxiliary view (such as S1, S3, S5, and S7), which can refer the main view as described in [7]. For the frames in the auxiliary view with inter-view prediction, the number of reference frames can be up to 4.

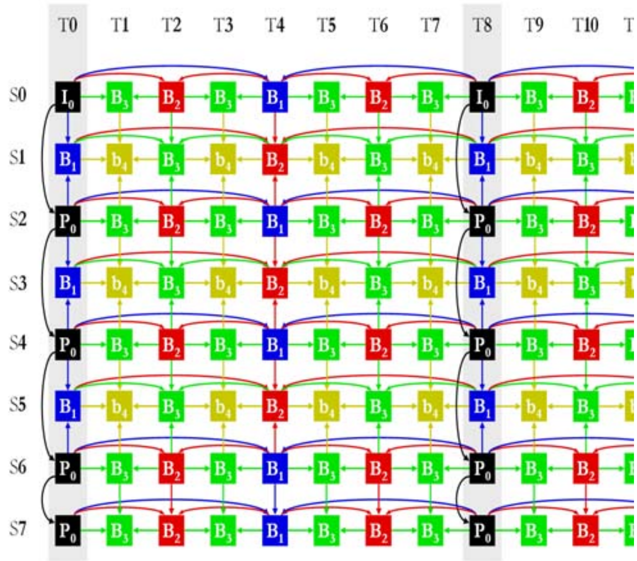

Fig. 1. HHI's multiview video coding scheme.

As we know that the computational complexity of motion estimation for traditional video coding is already very high, HHI's multiview video coding scheme further increases the computational complexity due to its disparity estimation. To reduce the computational complexity of disparity and motion estimation for multiview video coding, fast algorithms should be explored. In addition, reducing the number of reference frames can also speedup the prediction in MVC.

Usually, fast motion estimation algorithms, such as three-step search (N3SS) [8], four-step search (FSS) [9], diamond search (DS) [10], and hexagon-based search (HS) [11], aim at reducing the number of the searching points since the calculation of SAD

[1] This work was supported in part by National Science Foundation (60672088 and 60736043) and Major State Basic Research Development Program of China (973 Program, 2009CB320905).

X. Li, D. Zhao and W. Gao are with the Department of Computer Science and technology, Harbin Institute of Technology, Harbin 150001 China (e-mail: xmli@jdl.ac.cn; dbzhao@jdl.ac.cn; wgao@jdl.ac.cn).

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

for different points consumes the most time of the estimation, and these traditional approaches can also be used to reduce the computational complexity of disparity and motion estimation directly. However, these algorithms are investigated for the traditional video coding which only considers single view and do not explore the correlations between the neighboring views.

Recently, some fast motion estimation algorithms for MVC have been proposed. In [12], multiview camera geometry or the relationship between the disparity and motion vectors is utilized to estimate the reliability of the predicted vectors to reduce the search region. This reliability is decided by the difference between the motion vector predicted using median filtering, and the disparity vector obtained from the distance information between the cameras or the motion vector predicted by the joint disparity and motion estimation. When the reliability gets smaller, the search range will become larger. Considering that disparity estimation is different from motion estimation in MVC, an epipolar geometry-based fast disparity estimation algorithm is proposed in [13], which can reduce the computational complexity for inter-view prediction. It first transforms the commonly adopted median predicted search center to obtain its orthogonal projection epipolar search center on the corresponding epipolar line. Then, the disparity search is performed in an epipolar line-aligned search space. Furthermore, the set of camera positions can also be utilized to reduce the complexity of both disparity and motion estimation, but little work has been done to study this point.

A preliminary exploration of fast inter frame prediction for MVC has been done in our previous work [14], which is based on the relationship generated by the camera set. In [14], the computational complexity can be reduced greatly without noticeable loss of coding performance. However, the relationship between the motion and the disparity is not fully studied. Additionally, the theoretical analysis is not described.

The rest of the paper is organized as follows. Section II presents the fast disparity and motion estimation for MVC based on the correlations between the neighboring cameras and between the motion and the disparity. In the proposed approach, which assumes the aligned cameras are used, first, a search region estimation is proposed to reduce the disparity estimation complexity according to that the camera set is usually fixed and therefore the disparity between two neighboring views can be limited to an estimable range. Second, a motion vector derivation is given based on the geometric relationship between the motion and the disparity. In addition, an early termination scheme is provided to further reduce the number of reference frames. Section III provides the experimental results. The conclusions are given in the last section.

II. Fast Disparity and Motion Estimation for Multiview Video Coding

A. Search Region Estimation of Disparity for Multiview Video Coding

For multiview video captured by the aligned camera set in which the camera positions are fixed and parallel, as shown in Fig. 2, there exists strong relationship between the neighboring view videos. Therefore, the disparity between two frames in the neighboring views captured at the same time instance can be limited to an estimable region.

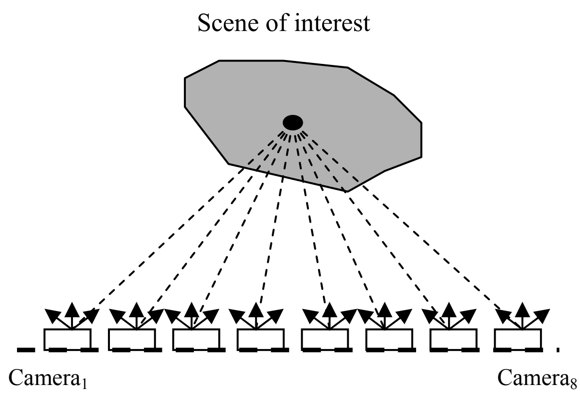

Fig. 2. Parallel multiview camera setup.

Considering one point *O* projected to the two frames in the neighboring views captured at the same time, as shown in Fig. 3, the positions of pixels *a* and *c* at the screen plane can be denoted as $|ab|$ and $|cd|$ respectively, where *c* and *a* are the projection points of *O* captured by the two neighboring cameras. c_1 and c_2 represent the positions of the two neighboring cameras. $|OH_1|$ is the distance between *O* and the screen plane, and $|OH_2|$ is the distance between *O* and the camera plane. Then the disparity value of *O* equals to $|cd|$-$|ab|$.

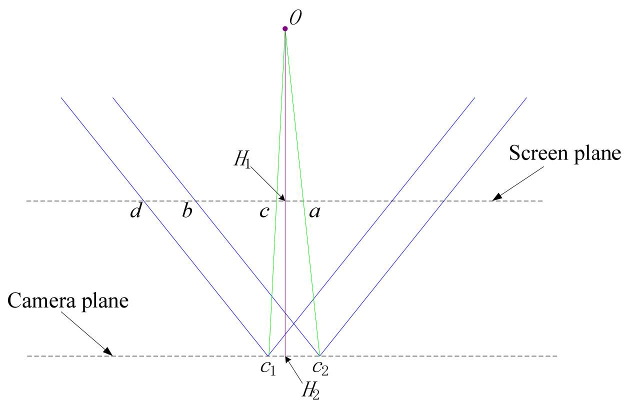

Fig. 3. The relationship between depth and disparity for parallel multiview video.

Since the two cameras in Fig. 3 are parallel, we can get that

$$|ac|/|c_1c_2| = |OH_1|/|OH_2| = (|OH_2| - |H_1H_2|)/|OH_2|, \tag{1}$$

and

$$|bd| = |c_1c_2|. \tag{2}$$

Then,

$$\begin{aligned} disparity(O) &= |cd| - |ab| = |bd| - |ac| \\ &= |bd| - |c_1c_2| \times (|OH_2| - |H_1H_2|) / |OH_2| \\ &= |c_1c_2| \times |H_1H_2| / |OH_2| \end{aligned} . \quad (3)$$

In (3), $|H_1H_2|$ is the intrinsic parameter of the camera, and $|c_1c_2|$ is determined by the positions of the two neighboring cameras. So the disparity of the object between two frames in the two neighboring views captured at the same time instance is inversely proportional to $|OH_2|$, or generally the object's depth. From (3), it can be seen that the disparity is near zero when $|OH_2|$, or the depth, is big enough. Considering the reference frame is two dimensions, the angle of the disparity vector can be derived by

$$\begin{aligned} Angle(\mathbf{disparity}(O)) &= \arctan(disparity_{Ver}(O) \\ &\qquad / disparity_{Hor}(O)) \\ &= \arctan((|c_1c_{2Ver}| \times |H_1H_2| / |OH_2|) \\ &\qquad / (|c_1c_{2Hor}| \times |H_1H_2| / |OH_2|)) \\ &= \arctan(|c_1c_{2Ver}| / |c_1c_{2Hor}|) \end{aligned} , \quad (4)$$

where $|c_1c_{2Hor}|$ and $|c_1c_{2Ver}|$ represent the horizontal and the vertical distance of the two neighboring cameras respectively. It can be seen that the angle of the disparity vector is only determined by the relative positions of the two neighboring cameras.

Based on the above analysis, it can be concluded that the disparity vector is concentrated in a region which is decided by the positions of the two neighboring cameras and the depth of the object. Fig. 4 shows the disparity vector distribution histogram for the sequence *Ballroom* (640x480). From Fig.4 we can see that nearly all of the vectors concentrate in one line. Since the relative positions of cameras are fixed, from (4) it can be concluded that the distributions of disparity vectors at different times are nearly the same.

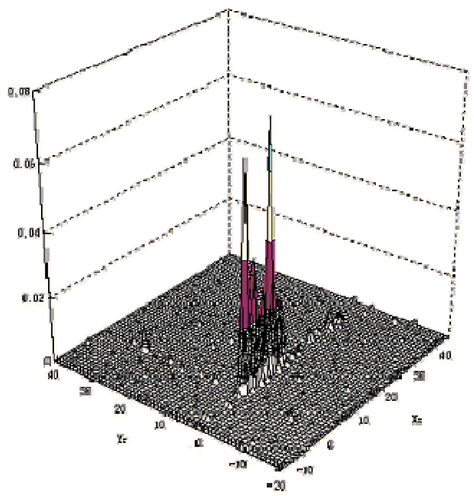

Fig. 4. Disparity vector distribution histogram for *Ballroom*.

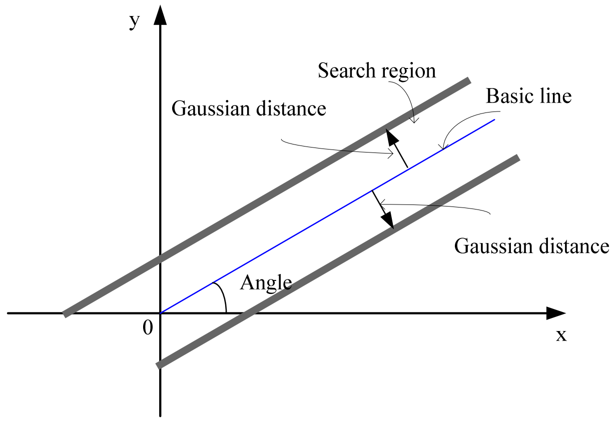

Fig. 5. The search region for the fast disparity estimation.

In the proposed method, the set of disparity vectors at times: T0 and T8 in Fig. 1 is utilized to generate the search region of disparity vectors at time Ti ($i \in \{1, 2, 3...7\}$). The proposed method can be described as follows. Firstly, an angle for disparity estimation is obtained based on the statistical information at times: T0 and T8. And then, a basis line is constructed by the angle and the disparity vector (0, 0). Finally, to increase the robustness of our approach [15], the search region for Ti ($i \in \{1, 2, 3...7\}$) is along the basis line within a certain Gaussian distance as shown in Fig. 5. That means the points whose distances to the basis line are smaller than the Gaussian distance compose of the search region, whereas the points whose distances are bigger than the Gaussian distance are not searched.

Fig. 6 gives the statistical result for the relationship between the Gaussian distance and the searching accuracy. The searching accuracy is defined as

$$Accuracy = (n / N) \times 100\% , \quad (5)$$

where n represents the number of 4x4 sub-blocks which have the same values of disparity vectors searched by the proposed approach and by the anchor in MVC test model JSVM, and N represents the total number of 4x4 sub-blocks for disparity prediction.

It can be seen from Fig.6 that with the Gaussian distance of 32 pixels, a high searching accuracy can be achieved (about 99.6%), so 32 pixels Gaussian distance is employed in our proposed method. To represent the searching complexity, the number of search points can be used. The search points are normalized considering the different complexity for different modes. For example, a search point for 16x8 mode is 1/2 point in statistics when assuming a search point for 16x16 mode is 1 point in statistics and so on. When the Gaussian distance equals to 32 pixels, the number of search points for the proposed approach is roughly 30% of that used in the anchor. So the complexity for the disparity estimation can be reduced greatly.

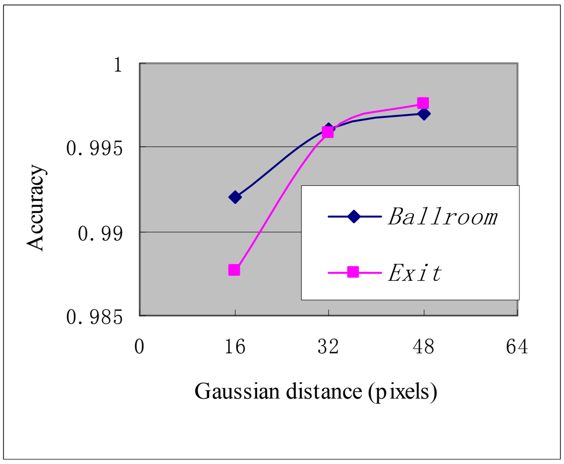

Fig. 6. Statistical result for the relationship between the Gaussian distance and the searching accuracy.

B. Motion Vector Derivation Based on the Geometric Relationship between the Motion and the Disparity

In this subsection, the motion vector is derived from the disparity and the corresponding motion vector in the reference view. The motion vector is usually determined by the displacement of an object at different times, whereas the disparity indicates the displacement of the object in two views at the same time. However, when the motion vector and the disparity correspond to the same content, all pixels with respect to this content can be found in both temporal and view neighboring frames which is described in [7]. Therefore, the motion vector can be derived. Fig. 7 provides the geometric relationship of the motion and the disparity in two views. In Fig. 7, f_t^v denotes the frame in view v at time t, where $v \in \{l, r\}$. There are four corresponding pixels with respect to the same object in two views at two times. Let P_t^v represent the corresponding pixel that belongs to f_t^v. $\overrightarrow{M}(P_t^v)$ and $\overrightarrow{D}(P_t^v)$ represent the motion vector and the disparity of pixel P_t^v respectively.

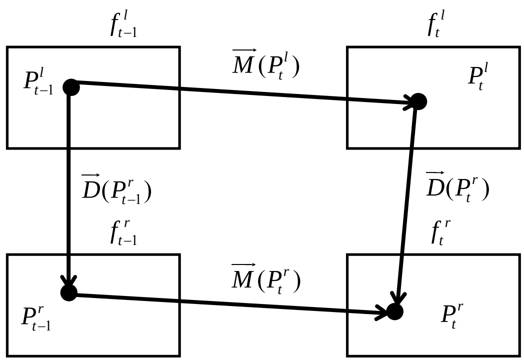

Fig. 7. Geometric relationship of the motion and the disparity in two views.

From Fig. 7, it can be seen that there exists a close relationship among $\overrightarrow{M}(P_t^r)$, $\overrightarrow{D}(P_{t-1}^r)$, $\overrightarrow{M}(P_t^l)$, and $\overrightarrow{D}(P_t^r)$, which is named as the stereo-motion consistency constraint or the loop constraint in [16]:

$$\overrightarrow{M}(P_t^r) + \overrightarrow{D}(P_{t-1}^r) - \overrightarrow{M}(P_t^l) - \overrightarrow{D}(P_t^r) \approx 0 \,. \tag{6}$$

From (6) we can conclude that $\overrightarrow{M}(P_t^r)$ can be derived if $\overrightarrow{D}(P_{t-1}^r)$, $\overrightarrow{M}(P_t^l)$ and $\overrightarrow{D}(P_t^r)$ are known. Moreover, $\overrightarrow{M}(P_t^l)$ is much correlated with $\overrightarrow{D}(P_t^r)$ as shown in Fig. 7. Based on this derivation, the fast motion estimation is described as follows. In (4), it is shown that the angle of disparity is determined by the positions of the two neighboring cameras, so it can be assumed that $\overrightarrow{D}(P_{t-1}^r)$ parallels to $\overrightarrow{D}(P_t^r)$. In the proposed approach, as shown in Fig. 8, $\overrightarrow{M}'(P_t^l)$, which is derived by $\overrightarrow{D}(P_t^r)$ and equals to $\overrightarrow{M}(P_t^l)$, is set as the predicted motion vector for $\overrightarrow{M}(P_t^r)$, and a basic line is obtained with the predicted motion vector and the angle of $\overrightarrow{D}(P_t^r)$. Considering the Gaussian noise, the search region for $\overrightarrow{M}(P_t^r)$ is along the basis line with a Gaussian distance of one pixel.

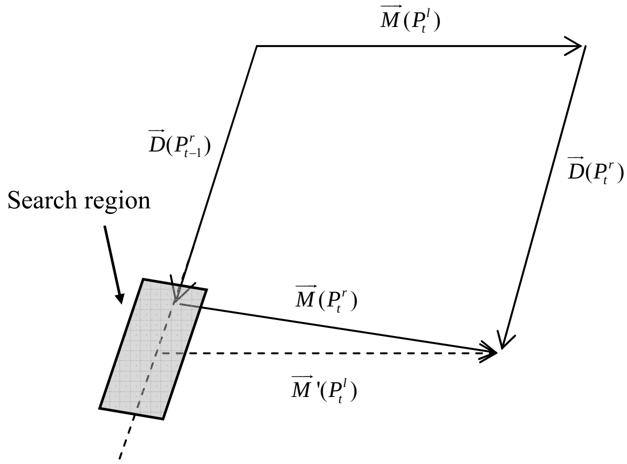

Fig. 8. Fast motion estimation based on the motion vector derivation.

TABLE I
STATISTICAL RESULT FOR FAST MOTION ESTIMATION BASE ON MOTION VECTOR DERIVATION

Sequence		*Ballroom* (640x480)	*Exit* (640x480)
Searching points ($x10^3$pixels/frame)	**anchor**	7718.76	4749.65
	proposed	388.27	389.28
Accuracy (sub-blocks/frame)		98.86%	99.01%

The statistical result is shown in Table I, where the definition of the searching complexity and the searching accuracy can be found in Subsection II-A. It can be seen from Table I that the accuracy can achieve more than 98% while roughly 8% search points are used.

C. Framework of the Proposed Fast Disparity and Motion Estimation

The proposed fast disparity and motion estimation contains three parts. The first part is the disparity estimation in the estimated search region. The second is the motion vector derivation based on the geometric relationship between the motion and the disparity. Additionally, in the third part an early termination scheme is provided to further reduce the number of reference frames. As the first two parts have been presented in Subsections II-A and II-B, in the following we mainly describe the third part.

It is well known that the motion vector comes from the motion of the object. Better performance can be achieved with intra-view prediction for homogeneous and stationary regions. However, for the regions with complex motions, intra-view prediction may not perform better than inter-view prediction. This is shown in Fig. 9. In Fig. 9, (a) and (c) are original frames. (b) and (d) show the blocks predicted with inter-view prediction in (a) and (c) respectively. From Fig. 9, it can be seen that most of the blocks belong to motion regions and that the distribution of those blocks has a spatial continuity.

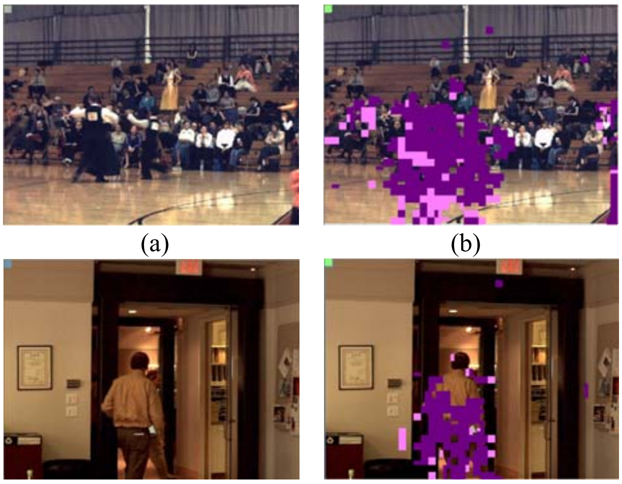

Fig. 9. The distribution of the blocks that refer frames in the neighboring view at the same time instance. (a) and (c) are original frames. (b) and (d) show the blocks predicted with inter-view prediction in (a) and (c) respectively.

Considering this continuity, in the proposed approach, an initial reference frame is selected by neighboring blocks. That is if most of the encoded neighboring blocks employ frame *F* as the reference frame, then *F* is as the initial reference frame of the current block. If the minimum cost in the searching region of the initial frame is good enough, other frames are not referred, and thus the number of reference frames can be reduced. To implement this early termination strategy, the minimum cost and the threshold for a block of size *mxn* should be defined. In H.264, the Lagrangian cost is defined as,

$$J(M,\mathbf{d}) = SAD_{B_{m\times n}}(\mathbf{d}) + \lambda(M)R(\mathbf{d}), \tag{7}$$

and

$$SAD_{B_{m\times n}}(\mathbf{d}) = \sum_{x=1,y=1}^{m,n} | f^c(x,y) - f^r(x+d_x, y+d_y)|, \tag{8}$$

where *mxn* is the size of the block which is decided by the block mode *M*, **d** represents the motion vector or the disparity, f^c and f^r are current and reference frames, respectively, λ is the lagrange multiplier, and $R(\mathbf{d})$ stands for the number of bits for coding. In the proposed approach, the minimum cost (*Cmin*) is defined as the minimum Lagrangian cost in a special searching region *Re* for the current block mode *M*,

$$Cmin = \min_{\mathbf{d}\in Re} J(M,\mathbf{d}) \cdot \tag{9}$$

In H.264, each mode such as 16x16, 16x8, 8x16, 8x8, and 8x4, 4x8, 4x4 should be implemented for prediction, so there is one best motion vector or disparity for each mode to a coded block, $\hat{\mathbf{d}}$, and the threshold in the proposed approach is set as,

$$Th_{(M,i)} = \begin{cases} 0 & i=0 \\ (\alpha \times \sum_{k=0}^{i-1} J_k(M,\hat{\mathbf{d}}))/i & \text{others} \end{cases}, \tag{10}$$

where *i* represents the macroblock number, a parameter alpha is multiplied as a tradeoff between the performance and the efficiency of the approach.

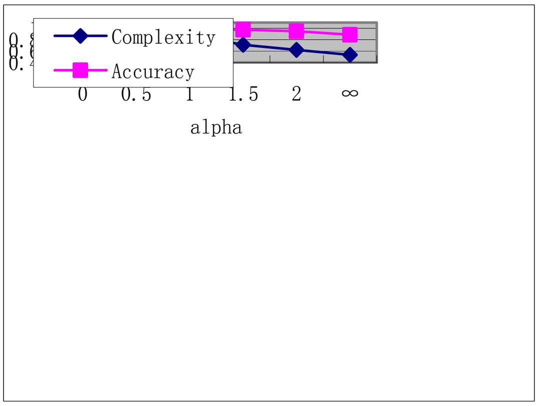

Fig. 10. Statistical result for the relationship between the parameter alpha and the performance of the proposed approach.

In (10), the threshold proposed is determined on the performance of coded blocks, and can be adjusted by the statistics. As Fig. 10. shows, as the parameter alpha becomes larger, both searching complexity and searching accuracy decline, where the definition of the searching complexity and the searching accuracy can be found in Subsection II-A, and 1.5 is selected as the value of this parameter to keep the accuracy while reducing the complexity greatly.

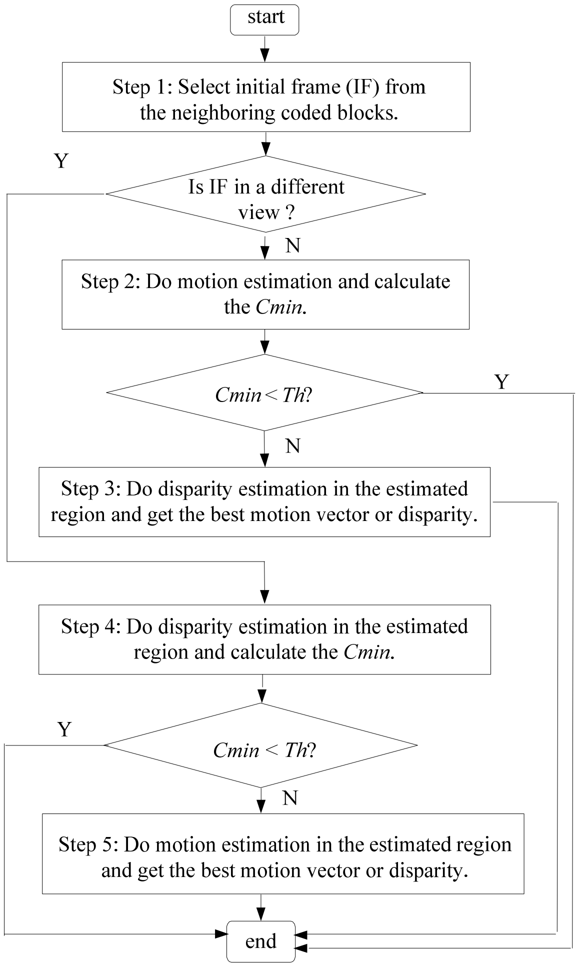

Fig. 11. Flowchart of the proposed fast disparity and motion estimation for MVC.

The framework of the proposed fast disparity and motion estimation is illustrated in Fig. 11. The initial reference frame is selected from the frames referred by neighboring blocks and the minimum cost is computed. If the initial reference frame is in the neighboring view, search the disparity with the fast disparity estimation, and the searching result for the disparity is stored in a buffer for fast motion estimation. If the minimum cost is higher than the threshold, fast motion estimation is implemented based on the derivation. However, if the initial reference frame is in the same view as the current frame, normal motion estimation for traditional video coding is employed, whereas the fast disparity estimation will be used if the minimum cost for motion estimation is worse than the threshold.

III. Experimental Results

The proposed algorithm has been implemented into MVC test model JSVM (Version 5_8) for verification. The proposed algorithm is compared with the anchor in JSVM, whose inter prediction is a trade-off between the complexity and the performance of the coding combining diamond search [10] and adaptive full search. The testing configuration is shown in Table II.

TABLE II
Configuration for the Experiments

Entropy Coding	CABAC
Transform	8x8 transform
FME	on
Hadamard Transform	on
Search Range	96
QP	27,29,30,32

In the experiments, three test sequences are used for performance comparison in terms of PSNR, bits difference and time saving. Among these three sequences, *Ballroom* (640x480) is for complex sequence and *Exit* (640x480) is for smooth sequence and *Race1* (640x480) is for sequence from moving set but fixed relative positions of cameras. The view S1 as shown in Fig.1, are used for the implementation of the approach, and the views S0 and S2 are as reference views. Corresponding to the three parts of the proposed approach, the experiments contain three parts and a complete simulation result. Table III-V give the comparisons between the anchor and the fast disparity estimation, between the anchor and the fast motion estimation, and between the anchor and the early termination scheme. The comparison between the anchor and the proposed fast disparity and motion estimation with early termination method is finally provided in Table VI. For the fast disparity estimation, it can be seen from Table III that the average time saving compared to the anchor for *Ballroom* , *Exit* and *Race1* is about 27.97% , 31.24% and 35.65% respectively. For the fast motion estimation, Table IV show that for *Ballroom* and *Race1*, 11.39% and 24.11% average time saving can be achieved whereas the average time saving for *Exit* is 6.82%, which is much smaller than the other two sequence. This is because that in the anchor the computational complexity of disparity prediction is greatly larger than that of motion prediction, especially for the smooth sequence. For the proposed fast disparity and motion estimation, we can get from Table VI that roughly 50% of the average time saving for three sequences can be reached compared to the anchor in MVC test model JSVM only with negligible coding efficiency loss.

TABLE III
Performance Comparison of Fast Disparity Estimation

QP	Performance	*Ballroom*	*Exit*	*Race1*
27	ΔPSNR (dB)	+0.00336	-0.00221	+0.00077
	ΔBits (%)	-0.14	-0.10	+0.10
	ΔTime (%)	-30.01	-31.38	-37.10
29	ΔPSNR (dB)	+0.00036	-0.00110	-0.01533
	ΔBits (%)	+0.05	-0.09	+0.58
	ΔTime (%)	-23.11	-31.38	-35.86
30	ΔPSNR (dB)	-0.00071	+0.00656	+0.01492
	ΔBits (%)	-0.54	0.00	+0.12
	ΔTime (%)	-29.52	-31.06	-35.79
32	ΔPSNR (dB)	+0.00006	-0.00121	+0.00743
	ΔBits (%)	-0.09	-0.26	-0.09
	ΔTime (%)	-29.26	-31.14	-33.85
Average	ΔPSNR (dB)	+0.00077	+0.00051	+0.001948
	ΔBits (%)	-0.18	-0.11	+0.18
	ΔTime (%)	-27.97	-31.24	-35.65

TABLE IV
Performance Comparison of Fast Motion Estimation

QP	Performance	*Ballroom*	*Exit*	*Race1*
27	ΔPSNR (dB)	-0.00444	-0.00221	-0.07041
	ΔBits (%)	+1.56	-0.10	-0.58
	ΔTime (%)	-10.56	-6.66	-25.61
29	ΔPSNR (dB)	-0.00689	-0.00793	-0.06018
	ΔBits (%)	+1.11	+1.63	+1.85
	ΔTime (%)	-13.53	-6.51	-24.52
30	ΔPSNR (dB)	-0.01350	-0.02140	-0.03653
	ΔBits (%)	-0.13	+1.84	+0.08
	ΔTime (%)	-10.65	-6.59	-23.35
32	ΔPSNR (dB)	-0.01390	-0.02383	-0.04727
	ΔBits (%)	+0.34	0.51	+1.51
	ΔTime (%)	-10.82	-7.5	-22.95
Average	ΔPSNR (dB)	-0.00968	-0.01384	-0.0536
	ΔBits (%)	+0.72	0.97	+0.72
	ΔTime (%)	-11.39	-6.82	-24.11

TABLE V
Performance Comparison of Early Termination

QP	Performance	*Ballroom*	*Exit*	*Race1*
27	ΔPSNR (dB)	+0.00164	-0.00415	-0.00966
	ΔBits (%)	-0.26	-0.50	+0.40
	ΔTime (%)	-31.09	-28.01	-27.71
29	ΔPSNR (dB)	+0.00029	-0.01592	-0.03598
	ΔBits (%)	-0.36	-1.18	+1.40
	ΔTime (%)	-31.34	-28.42	-27.48
30	ΔPSNR (dB)	-0.00131	+0.00222	-0.02336
	ΔBits (%)	-0.28	-0.35	-0.157
	ΔTime (%)	-31.27	-28.58	-26.87
32	ΔPSNR (dB)	-0.00675	-0.00270	-0.03864
	ΔBits (%)	+0.76	+0.09	+1.59
	ΔTime (%)	-31.87	-29.12	-27.95
Average	ΔPSNR (dB)	-0.00153	-0.00514	-0.02691
	ΔBits (%)	-0.04	-0.49	+0.81
	ΔTime (%)	-31.39	-23.53	-27.50

TABLE VI
Performance Comparison of Fast Disparity and Motion Estimation with Early Termination

QP	Performance	*Ballroom*	*Exit*	*Race1*
27	ΔPSNR (dB)	-0.01832	-0.00536	-0.02513
	ΔBits (%)	-0.47	-0.88	-1.22
	ΔTime (%)	-53.71	-48.85	-57.20
29	ΔPSNR (dB)	+0.01408	-0.00979	-0.02397
	ΔBits (%)	-0.26	-0.75	0.62
	ΔTime (%)	-53.98	-49.20	-54.69
30	ΔPSNR (dB)	-0.02804	-0.03176	0.00774
	ΔBits (%)	-0.84	-0.90	1.18
	ΔTime (%)	-53.47	-49.37	-51.87
32	ΔPSNR (dB)	-0.02742	-0.00288	-0.02884
	ΔBits (%)	-0.91	-0.90	+0.28
	ΔTime (%)	-53.93	-49.19	-53.43
Average	ΔPSNR (dB)	-0.01493	-0.01245	-0.01755
	ΔBits (%)	-0.62	-0.86	+0.22
	ΔTime (%)	-53.77	-49.15	-54.30

IV. CONCLUSIONS

This paper presents a fast disparity and motion estimation for MVC. Firstly, the disparity prediction complexity is reduced according to that the disparity between two neighboring views can be limited to an estimable range. Secondly, the motion vector is derived based on the relationship between the motion and the disparity. Additionally, an early termination scheme is provided to reduce the number of reference frames. The experimental results show that roughly 50% time saving for disparity and motion computing can be reached when compared to the anchor in MVC test model JSVM only with negligible coding efficiency loss.

REFERENCES

[1] "Report on 3DAV exploration," ISO/IEC JTC/SC29/WG11, N5878, July 2003.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13. , no. 7, pp. 1461-1473, Jul. 2003.

[3] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequence using joint 3-D motion/disparity compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7. , no. 2, pp. 312-327, Feb. 1997.

[4] I. Serdar, M. Emin, Y. Sehoon, and V. Anthony, "Depth estimation for view synthesis in multiview video coding", *Digital Object Identifier*, May, 2007.

[5] P. Merkle, A. Smolic, K. Muller, and T Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17. , no. 11, pp. 312-327, Nov. 2007.

[6] "Description of core experiments in MVC," ISO/IEC JTC1/SC29/WG11, MPEG2006/W7798, January 2006.

[7] X. Guo, Y. Lu., F. Wu, and W. Gao, "Inter-view direct mode for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp 1527-1532, Dec. 2006.

[8] R. Li, B Zeng, and M. L. Liou, "A new three-step search algorithm for block motin estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 4, pp. 438-443, Aug. 1994.

[9] L. M. Po and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 313-317, Jun. 1996.

[10] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation", *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 287-290, Feb. 2000.

[11] C. Zhu, X. Lin, and L. P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 349-355, May. 2002.

[12] Y. Kim, J. Kim, and K. Sohn, "Fast disparity and motion estimation for multi-view video coding," *IEEE Transaction on Consumer Electronics*, vol. 53, no. 2, pp. 712-719, May. 2007.

[13] J. Lu, H. Cai, J. Lou, and J. Li, "An epipolar geometry-based fast disparity estimation algorithm for multiview image and video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 6, pp. 737-750, Jun. 2007.

[14] X. Li, D. Zhao, X. Ji, Q. Wang, and W. Gao, "A fast inter frame prediction algorithm for multiview video coding," in *IEEE International Conference on Image Processing*, Sep. 2007.

[15] R. I. Hartley, "In the defense of eight-point algorithm," *IEEE Trans Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, Jun. 1997.

[16] W. Yang, K. Ngan, J. Lim, and K. Sohn, "Joint motion and disparity fields estimation for stereoscopic video sequence," *Signal Processing: Image Communication*, no. 3, Mar. 2005, pp. 265-276.

Xiaoming Li (S'07) received the B.S. degree in computer science from Harbin Institute of Technology, China, in 2005. He is currently working toward the Ph.D. degree at the Department of Computer Science and Technology, Harbin Institute of Technology. In 2006, he joined the Joint Research and Development Laboratory as a research assistant.

His research interests include digital image/video processing, multiview video coding and optical flow.

Debin Zhao received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, China, in 1985, 1988, and 1998, respectively. He was a research fellow in the Department of Computer Science, City University of Hong Kong, from 1989 to 1993. And now he is with the Institute of Computing Technology, Chinese Academy of Sciences.

His research interests include data compression, image processing, and human–machine interface. He has published 2 books and over 60 scientific papers.

Siwei Ma (S'03) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999. He is currently working toward the Ph.D. degree at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing.

His research interests include image and video coding, video streaming, and transmission.

Wen Gao (M'99) received the Ph.D. degree in computer science from Harbin Institute of Technology, China, in 1988 and the Ph.D. degree in electronics engineering, University of Tokyo, Japan, in 1991. He was a research fellow at the Institute of Medical Electronics Engineering, University of Tokyo, in 1992; a visiting scientist at the Robotics Institute, Carnegie Mellon University, in 1993; a visiting scientist at the MIT AI Lab, from May 1994 to December 1995. Now he is a professor of Peking University and the Institute of Computing Technology, Chinese Academy of Sciences and an adjunct professor in Hong Kong University of Science and Technology, China. He is the head of the Chinese National Delegation to MPEG working group (ISO/SC29/WG11) and chairs the Audio Video coding Standard (AVS) workgroup of China.

His research interests include pattern recognition, artificial intelligence, image understanding, data compression, hand gesture recognition, multimodal interface, and computer vision. He has published 4 books and over 400 scientific papers.