Low-Complexity Multiview Video Coding

Shadan Khattak
and Raouf Hamzaoui
and Shakeel Ahmad
Faculty of Technology
De Montfort University
Leicester, UK
shadan.khattak@myemail.dmu.ac.uk, {rhamzaoui, sahmad}@dmu.ac.uk

Pascal Frossard
Signal Processing Laboratory (LTS4)
Ecole Polytechnique Federale de Lausanne (EPFL)
Lausanne 1015 - Switzerland
pascal.frossard@epfl.ch

***Abstract*— We consider the problem of complexity reduction in Multiview Video Coding (MVC). We provide a unique comprehensive study that integrates and compares the different low complexity encoding techniques that have been proposed at different levels of the MVC system. In addition, we propose a novel complexity reduction method that takes advantage of the relationship between disparity vectors along time. The relationship is exploited with respect to the motion activity in the frame, as well as with the position of the frame in the Group of Pictures. We integrate this technique into our unique comprehensive framework and evaluate the performance of the resulting system in different setups. We show that the effective combination of complexity reduction techniques results in saving up to 93% in encoding time at the cost of only 0.08 dB in peak signal-to-noise ratio (PSNR) and 1.64% increase in bitrate compared to the standard MVC implementation (JMVM 6.0).**

I. Introduction

Multiview video (MVV) is a technology that uses multiple cameras to simultaneously capture a scene from different view points. MVV is used in applications such as 3D Television and Free View-point Television (FTV) [1]. While MVV gives a richer viewing experience than conventional video, it requires more storage space and more bandwidth. One straightforward solution to this problem is to encode each view with the H.264/AVC codec [2]. However, this approach does not exploit the similarity of content between neighboring views. This led to the development of MVC [3], the multiview extension of H.264/AVC. In MVC, reference frames for block matching are taken from neighboring views (Disparity Estimation) as well as across the temporal axis (Motion Estimation). A typical prediction structure of MVC is presented in Fig. 1. Here S_0, S_1, and S_2 represent three views while $t_0, t_1, \ldots, t_8$ represent nine successive frames. In each view, the first frame of the GOP is said to be at Temporal Level 0 (TL0). All the frames that use frames at TL0 as references belong to Temporal Level 1 (TL1). Similarly, the frames that use frames at TL1 as references belong to Temporal Level 2 (TL2), etc. A GOP of length l has $\log_2(l) + 1$ TLs.

Unfortunately, the encoding complexity of MVC is very high, making it unsuitable for applications such as live 3D TV or immersive teleconferencing. Therefore, reducing the time complexity of the MVC encoder is very important. While several techniques have been proposed for fast MVC, the potential benefit of combining them has not been studied. Moreover, none of the existing techniques exploits the relationship between disparity vectors across the temporal axis.

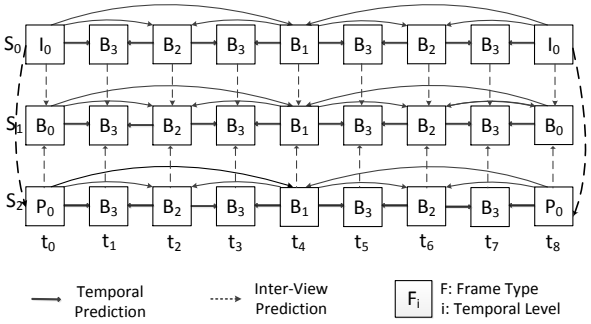

Fig. 1. Typical MVC prediction structure.

This paper addresses these two issues. We studied existing fast MVC techniques and identified the combination that results in largest complexity reduction. Our methodology was based on selecting methods for which gains add up. Moreover, we observed that the Disparity Vectors (DVs) of macroblocks located at the same position in temporally successive frames do not change significantly in regions with low or no motion activity. We also observed that as the temporal distance between a macroblock and its reference block decreases, the distance between DVs of corresponding macroblocks decreases as well. Using these observations, we devised an algorithm that chooses the Previous Disparity Vector (PDV) as the search center for disparity estimation (DE) instead of the Median Prediction Vector (MPV) as in the standard approach. For a given macroblock, PDV is the disparity vector of the corresponding macroblock in the temporally preceding frame. The new search center makes the search process start closer to the best match, allowing the search range to be reduced according to the motion activity of the macroblock. We integrated our DE algorithm in our effective combination of state of the art techniques, achieving an average encoding time saving over

JMVM 6.0 of about 93% at the cost of only 0.08 dB loss in PSNR and 1.64% increase in bitrate.

II. Multi-level complexity reduction

Previous complexity reduction techniques for fast MVC exploit redundant computations in mode size decision, prediction direction, and reference frame selection.

Shen et al. [4] found a high correlation between modes of neighboring views and devised a mode decision strategy for motion estimation (ME) and DE in which only mode size 16x16 is used if (i) the corresponding macroblock in the lower neighboring view and its eight spatial neighbors are encoded with SKIP or INTER 16x16 modes. In addition, mode sizes 16x8 and 8x16 are used if (ii) the corresponding neighboring macroblock or one of its eight spatial neighbors is encoded using INTER 16x8 or INTER 8x16 modes. Finally, all mode sizes are used if (iii) the corresponding neighboring macroblock or one of its eight spatial neighbors is encoded using INTER 8x8 or a smaller mode. They called the regions associated with macroblocks of type (i) Simple, type (ii) Medium, and type (iii) Complex. Simple, Medium, and Complex types correspond to regions with increasing motion activity since mode sizes reflect motion. In [5], they observed that the best match for a macroblock was found through DE in regions with high motion activity. So they devised a method which first identifies the motion activity of a region by considering the motion vectors of 4x4 sub-blocks of the corresponding macroblock in the neighboring view and its eight spatial neighbors. They identified three regions in a frame: homogeneous motion regions, medium homogeneous motion regions, and complex motion regions. DE was disabled in homogeneous motion regions. They combined this approach with the approach in [4] to increase the overall speed-up. The results were further improved in [6] by reducing the size of the search range for both ME and DE in regions with homogeneous motion.

Inter-view SKIP mode correlation was exploited by Shen et al. [7]. Zhang et al. [8] noted that because of the high spatial correlation inherent in a macroblock, there is a high probability that the smaller macroblock partition sizes eventually select the same reference frame and prediction direction as the ones selected by the largest macroblock partition size. So they proposed an algorithm for B-pictures in which the smaller mode sizes follow the decisions of the higher mode sizes for selecting the best prediction direction and reference frame.

Our first contribution was to combine the five methods proposed in [4], [5], [6], [7], and [8]. We selected these methods because they gave the highest speed-up for their respective target areas in the MVC encoding process. We found that by combining these methods, the speed-up gains add up without rate-distortion performance penalty.

III. Low-complexity disparity estimation

While DE consumes as much time as ME, the probability that it is used for prediction is low. For example, in regions with low motion activity, DE was used for prediction only about 5% of the time, while this percentage increased to 35% in regions with fast motion [6]. In this section, we propose a low-complexity DE algorithm. The main idea is to adjust the search range for DE according to the temporal level of the frame and the type of the macroblock.

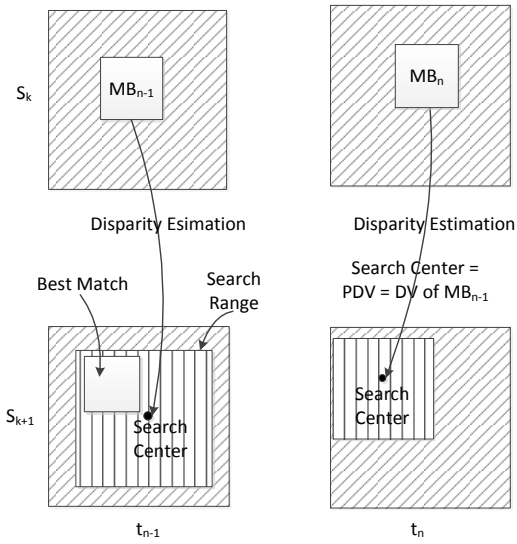

Fig. 2. Search center in PDV-DE.

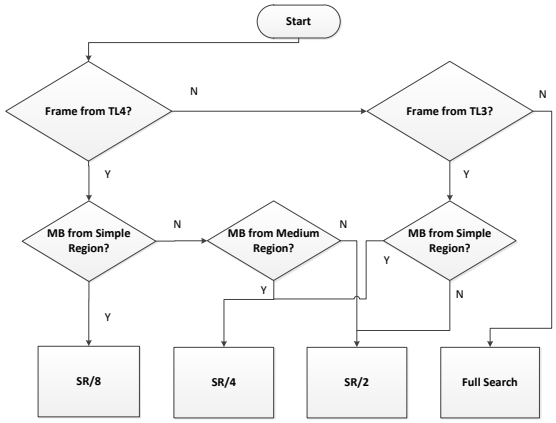

Fig. 3. Disparity vector search in PDV-DE.

A. Proportion of regions at different temporal levels

The search for the best macroblock starts in the search center. If the search centre is close to the best match, one can reduce the size of the search range and still find the best match. Because one does not know a priori where the best match for the current macroblock will be found, JMVM uses Median Prediction (MP). In median prediction, the median of the vectors of the top, top-right, and left macroblocks is used as the search center. The same procedure is used for both ME

TABLE I

PERCENTAGE OF MACROBLOCKS FOR WHICH THE BEST DISPARITY VECTOR WAS FOUND WHEN PREVIOUS DISPARITY VECTOR IS USED AS SEARCH CENTER. RESULTS ARE SHOWN FOR TL4 AT VARIOUS QP VALUES AND SEARCH RANGES (SR).

	Simple			Medium			Complex		
	SR/8	SR/4	SR/2	SR/8	SR/4	SR/2	SR/8	SR/4	SR/2
Ballroom 20	88.76	91.61	95.27	80.75	86.14	91.43	75.62	81.79	89.23
24	88.60	91.83	95.46	75.51	82.03	89.09	76.40	82.23	89.19
28	86.85	90.21	94.87	76.03	82.88	89.61	74.94	80.31	87.89
32	86.60	90.76	95.23	77.20	83.76	89.99	73.46	80.08	87.88
36	87.43	91.60	95.82	70.47	78.85	86.81	68.40	75.84	85.59
Exit 20	85.15	91.30	96.10	79.49	85.26	91.65	81.77	88.36	94.69
24	88.79	93.33	97.21	84.84	89.95	93.02	82.09	87.58	92.68
28	91.15	84.27	97.46	81.67	86.76	91.85	80.52	86.62	92.34
32	92.51	95.36	98.06	82.69	88.38	92.48	75.83	84.09	91.12
36	92.92	95.41	98.06	72.31	78.08	88.85	74.53	81.13	89.15
Average	88.87	91.57	96.35	78.09	84.21	90.48	76.36	82.80	89.98

TABLE II

PERCENTAGE OF MACROBLOCKS FOR WHICH THE BEST DISPARITY VECTOR WAS FOUND WHEN PREVIOUS DISPARITY VECTOR IS USED AS SEARCH CENTER. RESULTS ARE SHOWN FOR TL3 AT VARIOUS QP VALUES AND SEARCH RANGES (SR).

	Simple			Medium			Complex		
	SR/8	SR/4	SR/2	SR/8	SR/4	SR/2	SR/8	SR/4	SR/2
Ballroom 20	74.07	87.83	93.81	67.70	82.94	90.72	57.21	73.69	87.05
24	74.04	87.65	93.60	64.31	83.04	92.05	54.27	70.32	85.92
28	73.48	87.01	93.72	52.77	74.47	90.00	54.57	68.59	84.82
32	73.69	87.74	94.34	52.39	73.04	88.26	54.13	68.45	85.75
36	75.54	88.92	95.09	51.39	70.24	85.91	56.23	70.63	85.50
Exit 20	66.41	74.44	91.25	65.61	70.95	90.23	61.25	68.35	85.87
24	70.23	75.86	92.12	67.72	73.02	83.86	62.98	67.40	83.24
28	73.66	78.11	93.15	70.00	74.55	85.15	61.70	67.35	81.88
32	75.68	79.72	93.79	68.75	73.05	85.94	63.00	70.70	84.25
36	77.41	81.21	94.51	62.92	67.42	82.02	56.12	63.78	82.15
Average	73.42	82.85	93.54	62.36	74.27	87.42	58.15	68.93	84.64

and DE. However, the nature of disparity is different from that of motion. Indeed, even in the presence of motion, the source of disparity, the camera arrangement, usually is fixed. Thus, disparity is not as difficult to predict as motion.

To check this assumption, we used PDV instead of MP as the search center and determined the proportion of macroblocks that found their best match at the search centre, in 1/8th of the search range, 1/4th of the search range, and half of the search range. We found that the way the best match is spread across the search range depended on the temporal level of the frame as well as on the region type. At TL4, the best match is found earlier than at TL3 (Table I and II). Also in Simple regions, the best match is found earlier than in Medium and Complex regions.

B. Previous disparity vector based disparity estimation

Based on our findings, we propose a new search strategy for DE called Previous Disparity Vector-based Disparity Estimation (PDV-DE). During DE, the search center is set to PDV and two conditions are checked: (i) Does the frame belong to TL3 or TL4? (ii) Does the macroblock belong to a Simple, Medium or Complex region? According to the answers, different search ranges are selected (see Fig. 2 and Fig. 3).

IV. RESULTS

We compared the MVC methods in [4], [5], [6], [7], [8] and combined them with and without our low-complexity disparity estimation technique (PDV-DE). Note that the method in [6] is a combination of the techniques in [4], [5], and another technique (see Section II). Fig. 4 and 5 show the percentage encoding time saving with respect to JMVM 6.0 [9]. In the figures, label [6]+ [7] + [8] denotes the combination of the five methods without PDV-DE, while PDV-DE MVC is the system obtained by combining the five methods with PDV-DE.

Three views of the standard test sequences Ballroom and Exit were used. These sequences are recommended by the Joint Video Team (JVT) for testing new algorithms [10]. The GOP size was 16, and the maximum search range was 64. For each sequence, results were obtained for Quantization Parameter (QP) values 20, 24, 28, 32, and 36. The simulations were run on a machine with Intel Core i5 dual core 2.67 GHz CPU and 4 GB RAM.

Combining the methods in [4], [5], [6], [7], and [8] increased the encoding time saving over the method in [6], the currently best published result, by more than 10%. Adding PDV-DE further increased the time saving by 1.5%. This cor-

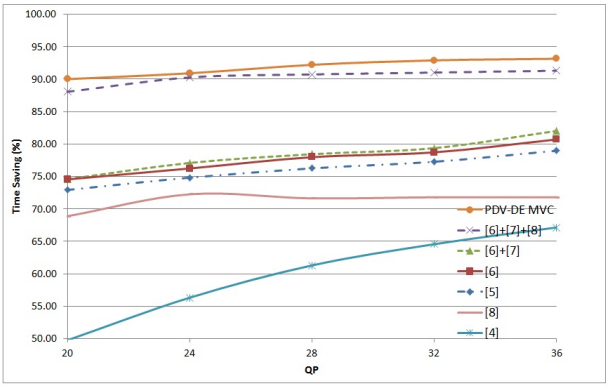

Fig. 4. Encoding time saving compared to JMVM 6.0. for Ballroom.

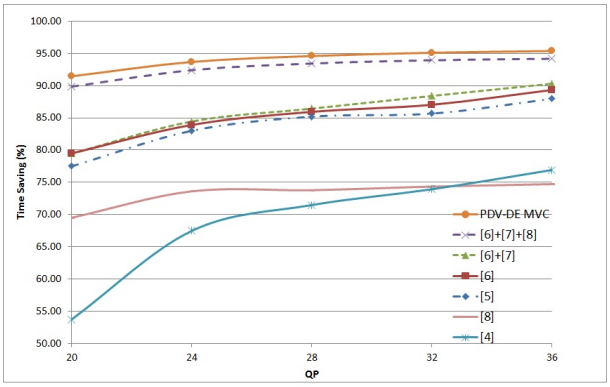

Fig. 5. Encoding time saving compared to JMVM 6.0. for Exit.

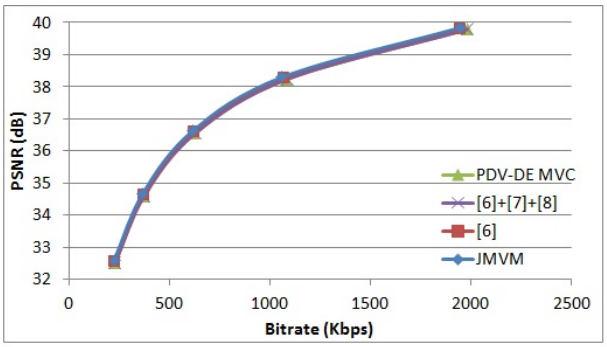

Fig. 6. Rate-distortion performance for Ballroom.

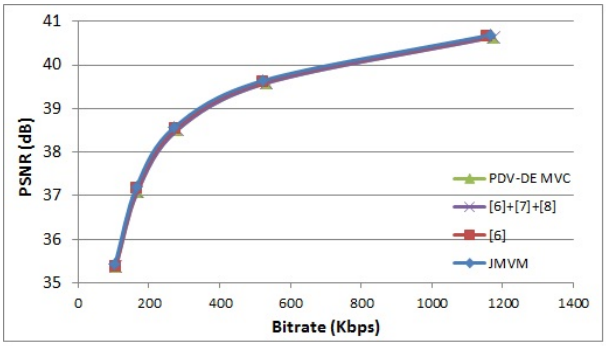

Fig. 7. Rate-distortion performance for Exit.

responds to an encoding time reduction of 92.94% compared to the standard implementation of JMVM 6.0. It is a significant gain, achieved at the cost of only 0.06 dB in PSNR and 1.72% increase in bitrate. Compared to JMVM 6.0, PDV-DE saved on average 35.28% of the encoding time at the cost of only 0.01 dB in PSNR and 0.07% increase in bitrate (the PDV-DE curve was not included in Fig. 4 and 5 to better show the differences between the other curves). Fig. 6 and 7 show the rate-distortion performance of the methods.

Note that our system was particularly useful in scenarios involving large disparities. For example, the time saving for Exit was over 2% greater than that for Ballroom, which has less disparity.

We obtained similar results for test sequences Vassar and Race1 [10]. Compared to JMVM 6.0, PDV-DE MVC saved on average 93.22% of the encoding time with only 0.09 dB reduction in PSNR and 1.56% increase in bitrate.

V. Conclusion

We identified five low-complexity MVC techniques that target different areas of speed-up and combined them in a unique framework with a disparity estimation technique that exploits the correlation between disparity vectors across the temporal axis. The resulting MVC scheme reduced the complexity of JMVM 6.0 by about 93% with negligible PSNR loss and bitrate increase. In this work, we searched for the motion and disparity vectors separately. We plan to improve the results by modeling this relationship.

References

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video - technologies, applications and MPEG standards," in Proc. ICME 2006, Toronto, Ontario, July 2006.

[2] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audio Visual services," 2005.

[3] M. Flierl and B. Girod, "Multi-view video compression - exploiting inter-image similarities," IEEE Signal Proc. Mag. vol. 24, pp. 66–76, Nov. 2007.

[4] L. Shen, T. Yan, Z. Liu, Z. Zhang, P. An, and L. Yang, "Fast mode decision for multiview video coding," in Proc. IEEE ICIP, Cairo, Egypt, pp. 2593–2596, Nov. 2009.

[5] L. Shen, Z. Liu, S. Liu, Z. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," IEEE Trans. Broadcasting, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[6] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "View-adaptive motion estimation and disparity estimation for low-complexity multiview video coding," IEEE Trans. Circuits Syst. Video Tech. vol. 20, pp. 925–930, 2010.

[7] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "Early SKIP mode decision for MVC using inter-view correlation," Signal Processing: Image Communication, vol. 25, pp. 88–93, Feb. 2010.

[8] Y. Zhang, S. Kwong, G. Jiang, and H. Wang, "Efficient multi-reference frame selection algorithm for hierarchical B pictures in multiview video coding," IEEE Trans. Broadcasting, vol. 57, no. 1, pp. 15–23, 2011.

[9] ISO/IEC JTC/1 SC29/WG11 and ITU-T SG16 Q.6, "JMVM 6.0 software," JTV-Y208, Oct. 2007.

[10] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, "Common test conditions for multiview video coding," JVT-T207, Klagenfurt, Austria, July 2006.