Bayesian Early Mode Decision Technique for View Synthesis Prediction-enhanced Multiview Video Coding

Shadan Khattak, Thomas Maugey, Raouf Hamzaoui, Shakeel Ahmad, and Pascal Frossard

***Abstract*—View synthesis prediction (VSP) is a coding mode that predicts video blocks from synthesised frames. It is particularly useful in a multi-camera setup with large inter-camera distances. Adding a VSP-based SKIP mode to a standard Multiview Video Coding (MVC) framework improves the rate-distortion (RD) performance but increases the time complexity of the encoder. This paper proposes an early mode decision technique for VSP SKIP-enhanced MVC. Our method uses the correlation between the RD costs of the VSP SKIP mode in neighbouring views and Bayesian decision theory to reduce the number of candidate coding modes for a given macroblock. Simulation results showed that our technique can save up to 36.20% of the encoding time without any significant loss in RD performance.**

***Index Terms*—Multiview video coding, view synthesis prediction, fast mode decision.**

EDICS Category: IMD-CODE

I. Introduction

THE multiview plus depth representation is one of the most promising methods for providing multiview video services [1]. In a multiview plus depth representation, the information consists of multiple texture views together with their associated per-pixel depth maps. View synthesis uses the per-pixel depth maps and interpolation techniques to synthesize virtual views between camera views. Traditionally, it has been used to reduce the network and storage resource consumption of multiview video by providing N views (camera views plus synthesized views) at the decoder side while only K ($K < N$) camera views are captured, encoded, and transmitted. However, view synthesis can also improve the RD performance of MVC by providing new prediction modes for blocks to be encoded [2]. In particular, a VSP-based SKIP mode has been shown [3] to significantly improve the RD performance of MVC. Unlike the conventional SKIP mode, the VSP SKIP mode predicts a macroblock using a synthetic reference frame. However, the RD optimized framework of MVC already uses a computationally complex motion and disparity estimation process. Adding the VSP SKIP mode in this framework further increases the computational complexity of the encoding.

In this paper, we propose an early mode selection technique to reduce the time complexity of a VSP SKIP-enhanced MVC coder. We exploit the inter-view correlation between the RD costs of the VSP SKIP mode and use Bayesian decision theory to restrict the number of candidate coding modes that are tested during the encoding. In this way, motion and disparity estimation can be skipped for a large proportion of macroblocks. Our results show that the encoding time can be reduced by up to 36.20% compared to the latest version of the MVC Joint Multiview Video Coding (JMVC) reference software with integrated VSP SKIP mode, without any significant penalty on the RD performance.

No previous work has specifically addressed the problem of early mode selection for VSP SKIP-enhanced MVC. Most of the related work has focused on improving the quality of view synthesis [4], generating better depth maps [5], or pre- and post-processing of synthesized images for better prediction [6]. A number of fast algorithms [7], [8], [9], [10], [11], [12] have been proposed to reduce the time complexity of motion estimation, disparity estimation, reference frame selection, and mode decision processes in MVC. An early prediction technique for the Conventional SKIP mode in MVC was proposed in [13]. The method selects the Conventional SKIP mode for a macroblock if the corresponding macroblock identified by the Global Disparity Vector (GDV) [14] and its eight neighbouring macroblocks in a neighbouring view are encoded using this mode.

Bayesian decision theory was previously used for fast mode decision of H.264/AVC in [15] and Scalable Video Coding (SVC) in [16]. The techniques presented in these papers are not suitable for the proposed VSP-SKIP enhanced MVC coder as they do not exploit inter-view correlation. Moreover, our technique is unique in its use of the RD cost of a mode as the observed feature in the Bayesian decision rule.

The remainder of the paper is organized as follows. Section II presents the VSP SKIP-enhanced MVC coder considered in this paper and studies optimal coding modes for this coder. Section III proposes our Bayesian early mode decision technique. Section IV evaluates the performance of our method in terms of encoding time, bitrate, and peak signal-to-noise ratio (PSNR) and compares it to a baseline approach based on inter-view mode correlation. Conclusions are given in Section V.

II. Preliminaries

In this paper, we consider an extended MVC coder where a VSP SKIP mode is added to the existing eight Inter

S. Khattak, R. Hamzaoui, and S. Ahmad are with the Faculty of Technology, De Montfort University, Leicester, LE1 9BH, United Kingdom (e-mail: shadan.khattak@myemail.dmu.ac.uk, {rhamzaoui, sahmad}@dmu.ac.uk).

T. Maugey and P. Frossard are with the Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: {thomas.maugey, pascal.frossard}@epfl.ch).

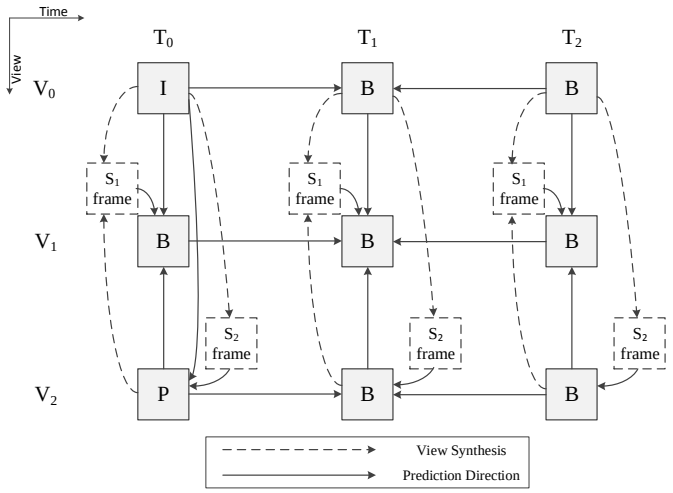

Fig. 1. Prediction structure using view synthesis. V_0, V_1, and V_2 represent camera views while S_1 and S_2 represent synthesized views. The dotted lines represent the reference view(s) for view synthesis and the solid lines refer to the prediction direction. The synthesized frames are used as reference frames for VSP prediction.

modes (Conventional SKIP, Inter16x16, Inter16x8, Inter8x16, Inter8x8, Inter8x4, Inter4x8, and Inter 4x4) and two Intra modes (Intra16 and Intra4) [10]. An example of the encoding structure with three camera views, V_0, V_1, and V_2, and two synthesized views S_1 and S_2 is illustrated in Fig. 1.

Conventional SKIP is a special mode in which the macroblock is reconstructed by motion-compensated prediction using a motion vector predicted from previously decoded motion vectors [17]. The VSP SKIP mode differs from the conventional SKIP mode in that the macroblock is reconstructed using the macroblock at the same position in a synthesized version of the current frame [2].

To synthesize the views required for VSP SKIP, we use the method proposed in [18], which is based on the image coordinate system, the camera coordinate system, and the world coordinate system. A pixel in the image coordinate system of the camera view is projected onto a pixel in the image coordinate system of the virtual view in two steps. First, using the intrinsic and extrinsic parameters of the reference camera and the depth information, the 3D point that corresponds to the pixel in the camera view is projected onto the world coordinate system. Then, from the world coordinate system it is projected onto the image coordinate system of the virtual view (using the camera parameters of the virtual view). When switching viewpoints, some background regions which are hidden behind foreground objects in the reference view, might appear in the virtual view and vice versa. This induces the hole problem. When a synthesized frame is created using only one reference frame, holes cannot be efficiently filled. This problem is solved by using two reference frames where the second reference frame is used to fill the holes.

Table I shows the proportion of modes selected for view V_1 in the settings of Fig. 1 and a VSP SKIP-enhanced JMVC 6.0 reference software [19]. We used two test sequences

TABLE I
PROPORTION (%) OF CODING MODES.

	Breakdancers sequence					
Mode\QP (Texture)	36	32	28	24	20	Avg.
VSP SKIP	40.36	34.50	28.82	24.69	20.16	29.70
Conventional SKIP	45.48	46.82	46.09	41.98	32.83	42.64
Inter16x16	11.60	14.37	17.73	20.62	23.64	17.59
Inter16x8	0.96	1.63	2.49	4.23	6.44	3.15
Inter8x16	0.97	1.46	2.75	4.33	6.19	3.14
Inter8x8	0.59	1.14	1.96	3.58	7.7	2.99
Intra modes	0.04	0.08	0.16	0.58	3.25	0.82
	Ballet sequence					
Mode\QP (Texture)	36	32	28	24	20	Avg.
VSP SKIP	29.91	26.54	23.46	20.76	18.28	23.79
Conventional SKIP	64.30	65.66	66.41	65.97	61.86	64.84
Inter16x16	4.80	6.16	7.61	9.38	13.68	8.33
Inter16x8	0.45	0.72	1.10	1.51	2.17	1.19
Inter8x16	0.36	0.60	0.91	1.48	2.25	1.12
Inter8x8	0.19	0.29	0.47	0.83	1.56	0.66
Intra modes	0.00	0.03	0.03	0.08	0.20	0.06

(Breakdancers and Ballet [20]) of resolution 1024x768 and five quantization parameter (QP) values (20, 24, 28, 32, 36). The number of frames in each sequence is 100, the Group of Pictures (GOP) size is 16, and the search range is $[\pm 16, \pm 16]$. In the table, Inter8x8 includes sub modes Inter8x4, Inter4x8, and Inter4x4. We observe that VSP SKIP and conventional SKIP are the dominant modes. The dominance is more prominent in the presence of large homogeneous regions (as in the Ballet sequence). We also observe that the other modes are rarely used.

III. Bayesian Early Mode Decision Technique

In the standard encoding scheme, the encoder considers all coding modes and selects one with minimum RD cost. The reference views (V_0 and V_2 in Fig. 1) are encoded first, followed by the bidirectionally predicted view (V_1 in Fig. 1). However, as observed in Section II, VSP SKIP or Conventional SKIP may be selected much more often than the other modes. In this section, we propose to exploit the correlation between RD costs across views and Bayesian decision theory to avoid testing unlikely coding modes during the encoding of V_1.

Let $m_1, m_2, \ldots, m_8$ denote VSP SKIP, Conventional SKIP, Inter16x16, Inter16x8, Inter8x16, Inter8x8, Intra16, and Intra4 modes, respectively, where, as before, Inter8x8 includes sub modes Inter8x4, Inter4x8, and Inter4x4. For a given macroblock in the predicted view (V_1), let $P(m_i|x)$ denote the a posteriori probability of selecting mode m_i given an observation x of the VSP SKIP RD cost of this macroblock in V_1. From Bayes theorem, we have $P(m_i|x) = \frac{P(m_i)p(x|m_i)}{p(x)}$ where $P(m_i)$ is the a priori probability of mode m_i, $p(x|m_i)$ is the conditional probability density function, and $p(x)$ is the mixture density function. Since $p(x) > 0$, Bayes decision rule implies that mode m_i should be selected if $P(m_i)p(x|m_i) > P(m_j)p(x|m_j)$. While the values of $P(m_i)$ and $p(x|m_i)$ are unknown in V_1, we can estimate them from their respective values in V_2. Fig. 2 shows that this approach is reasonable since the probability density functions $p(x|m_i)$ in V_1 and V_2 are similar. Here we used a lognormal distribution to model the probability density function of random variable x (Fig. 3).

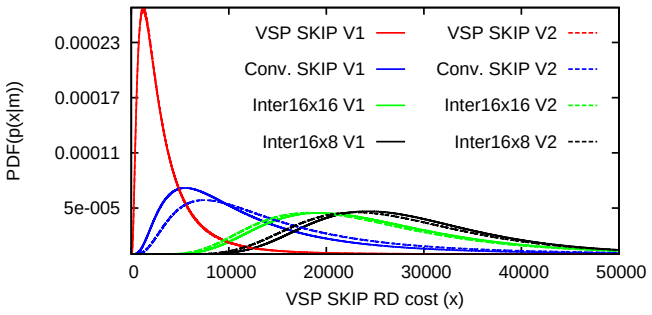

Fig. 2. Comparison of conditional probability density functions (PDFs) in V_1 and V_2 for the Breakdancers sequence (QP=36).

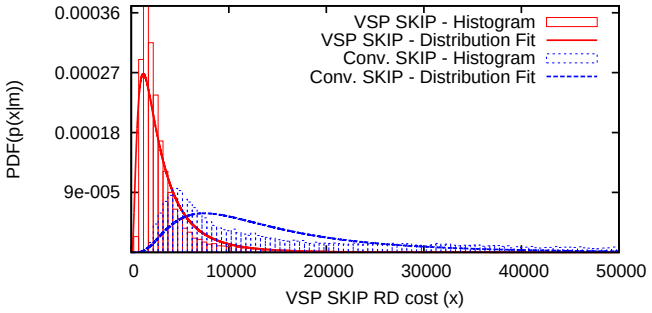

Fig. 3. Normalized histograms of VSP SKIP RD cost for different modes and lognormal distribution fit for the Breakdancers sequence (QP=36).

We considered several models and selected the lognormal one based on the Bayesian information criterion [22].

Bayes decision rule does not lead to perfect mode selection as its optimality holds in a probabilistic sense only. To account for the Bayes error and the fact that $P(m_i)$ and $p(x|m_i)$ are estimates from a different view, we introduce a tolerance threshold e and select not only the optimal mode m_i^* in Bayesian sense but also any other mode m_i such that

$$\frac{P(m_i^*)p(x|m_i^*) - P(m_i)p(x|m_i)}{\sum_{j=1}^{8} P(m_j)p(x|m_j)} \leq e \qquad (1)$$

Our algorithm is summarized in Algorithm 1.

IV. Experimental Results

We implemented our fast mode decision technique in the VSP SKIP-enhanced MVC coder described in Section II. As a baseline approach, we used a method that extends the inter-view correlation technique proposed in [13] by selecting a SKIP mode (Conventional SKIP or VSP SKIP) for a macroblock if the corresponding macroblock identified by GDV and its eight neighbouring macroblocks in V_2 are encoded using the same SKIP mode. The simulations were run on a machine with Intel Core i5 dual core 2.67 GHz CPU and 4 GB RAM.

Table II compares the performance of our technique and the baseline approach to that of the standard coder. The encoder settings were as for Table I. Parameters ΔPSNR, ΔR, and ΔT denote the increase in PSNR, increase in bitrate, and saving in encoding time, respectively, compared to the

Algorithm 1 Bayesian Early Mode Decision Technique

Pre-Processing:

1: Encode V_0.
2: Encode V_2.

Input: Threshold e.

Output: Modes to be checked for all macroblocks in V_1.

1: Calculate the a priori probabilities $P(m_i),\ i = 1, \ldots, 8$ in V_2.
2: Estimate the conditional probability density functions $p(x|m_i),\ i = 1, \ldots, 8$ in V_2.
3: **while** not all macroblocks in V_1 are encoded **do**
4: Determine the RD cost x of VSP SKIP for the current macroblock in V_1.
5: Determine $i^* = \arg\max_i P(m_i)p(x|m_i),\ i = 1, \ldots, 8$ from V_2.
6: **for** i = 1 to 8 **do**
7: **if** $\frac{P(m_i^*)p(x|m_i^*) - P(m_i)p(x|m_i)}{\sum_{j=1}^{8} P(m_j)p(x|m_j)} \leq e$ **then**
8: check mode i.
9: **end if**
10: **end for**
11: **end while**

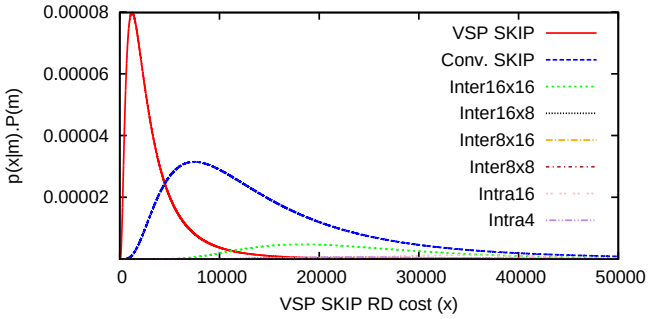

Fig. 4. Product of conditional PDFs and a priori probabilities of different modes for the Breakdancers sequence (QP=36).

standard coder. For our technique, all encoding steps are taken into consideration in the calculation of the encoding time. This includes the estimation of $P(m_i)$ and the fitting of the lognormal distribution model to the samples in V_2 (Steps 1 and 2 in the main part of Algorithm 1).

The table shows results that correspond to the Bayes decision rule ($e = 0$ in Algorithm 1) and to $e = 0.995$. For $e = 0$, our approach reduced the encoding time by 48.93% on average. However, the loss in rate-distortion performance was significant. For $e = 0.995$, the average saving in encoding time was smaller but still significant (29.91%) while the loss in rate-distortion performance was negligible. Further measurements showed that for this value of e, our algorithm found the optimal mode for 96.14% and 92.68% of the macroblocks for the Breakdancers and Ballet sequence, respectively.

Overall, the time saving for the Breakdancers sequence was slightly greater than for the Ballet sequence because the proportion of VSP SKIP coded macroblocks is higher in the Breakdancers sequence (Table I). The results also show that the reduction in encoding time, generally, increases as QP increases. This is because when QP increases, the quantization

TABLE II
COMPARISON OF THE BASELINE APPROACH AND THE PROPOSED METHOD WITH THE STANDARD CODER.

	Baseline			Proposed ($e = 0$)			Proposed ($e = 0.995$)		
QP	ΔPSNR	ΔR(%)	ΔT(%)	ΔPSNR	ΔR(%)	ΔT(%)	ΔPSNR	ΔR(%)	ΔT(%)
Breakdancers sequence									
20	-0.01	0.17	5.19	-0.77	12.10	44.81	-0.06	0.89	26.74
24	0.00	0.01	9.22	-0.63	14.22	46.89	-0.06	0.91	28.84
28	-0.01	0.04	9.57	-0.41	14.56	46.41	-0.07	0.29	29.64
32	-0.01	-0.17	13.39	-0.21	11.07	50.51	-0.07	0.14	34.01
36	-0.01	-0.03	14.91	-0.15	4.94	55.05	-0.05	0.06	36.20
Average	**-0.01**	**0.00**	**10.46**	**-0.43**	**11.38**	**48.73**	**-0.06**	**0.46**	**31.09**
Ballet sequence									
20	-0.02	0.02	9.85	-0.38	12.86	48.24	-0.04	0.43	22.32
24	-0.03	0.21	20.23	-0.32	13.70	49.65	-0.05	0.81	25.15
28	-0.07	1.05	27.79	-0.22	12.16	50.31	-0.07	0.77	27.27
32	-0.08	0.59	28.56	-0.15	9.37	50.44	-0.07	0.66	33.11
36	-0.07	0.43	27.60	-0.10	4.87	47.08	-0.03	0.26	35.82
Average	**-0.06**	**0.46**	**22.80**	**-0.23**	**10.59**	**49.14**	**-0.05**	**0.59**	**28.73**
Average	**-0.03**	**0.23**	**16.63**	**-0.33**	**10.98**	**48.93**	**-0.055**	**0.52**	**29.91**

becomes coarser, fewer details are preserved, and more large modes are selected. This benefits our method, which mostly predicts large modes such as VSP SKIP, Conventional SKIP, and Inter16x16 (Fig. 4).

The baseline approach reduced the encoding time by only 17.4% on average. Since the synthesized view S_2 is constructed using only one reference frame, its quality is lower than that of S_1, which is constructed from two reference frames (Fig. 1). Consequently, the contribution of VSP SKIP in V_2 is smaller than in V_1, and the VSP SKIP mode decisions in V_1 cannot be efficiently predicted using the VSP SKIP mode decisions in V_2.

We obtained similar results for two more test sequences (Poznan_Street and Poznan_Hall2 [21], both of resolution 1920x1088). For example, the average speedup for Poznan_Street and Poznan_Hall2 was 27.03% and 31.15%, respectively, with negligible loss in rate distortion performance.

V. CONCLUSION

We proposed an early mode decision technique for View Synthesis Prediction-enhanced Multiview Video Coding. Our method uses Bayesian decision theory to speed up the encoding by reducing the number of candidate coding modes. Our approach reduced the encoding time of the VSP SKIP-enhanced JMVC 6.0 by up to 36.20% while preserving the RD performance. We expect that more time savings can be achieved if we combine our method with techniques ([23], [24]) that can efficiently predict non-VSP coding modes. Our results also show that the methods based on inter-view mode correlation might not be suitable for predicting the VSP SKIP mode because of the difference in the quality of the synthesized reference frames in neighbouring views.

REFERENCES

[1] P. Aflaki, M. M. Hannuksela, D. Rusanovski, and M. Gabbouj, "Nonlinear depth map resampling for depth-enhanced 3-D video coding," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 87–90, Jan. 2013.

[2] C. Lee and Y. -S. Ho, "A framework of 3D video coding using view synthesis prediction," in *Proc. Picture Coding Symposium (PCS'12)*, Krakow, Poland, May 2012.

[3] S. Yea and A. Vetro, "View synthesis prediction for multiview video coding," *Signal Process.-Image Commun.*, vol. 24, no. 1–2, pp. 89–100, Jan. 2009.

[4] D. Tian, P. Lai, P. Lopez, and C. Gomila, "View-synthesis techniques for 3D Video," in *Proc. SPIE 7443*, 2009.

[5] Z. Ni, D. Tian, S. Bhagavathy, J. Llach, and B. Manjunath, "Improving the quality of depth image based rendering for 3D video systems," in *Proc. IEEE Int. Conf. Image Processing (ICIP'09)*, Cairo, Egypt, Nov. 2009.

[6] M. Solh and G. Alregib, "Hierarchical hole-filling for depth-based view synthesis in FTV and 3D Video," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 495–504, Jun. 2012.

[7] Z.-P. Deng, Y.-L. Chan, K.-B. Jia, C.-H. Fu, and W.-C. Siu, "Iterative search strategy with selective bi-directional prediction for low complexity multiview video coding," *J. Vis. Commun. Image Rep.*, vol. 23, no. 3, pp. 522–534, Feb. 2012.

[8] Y. Zhang, S. Kwong, G. Jiang, X. Wang, and M. Yu, "Statistical early termination model for fast mode decision and reference frame selection in multiview video coding," *IEEE Trans. Broad.*, vol. 58, no. 1, pp. 10–23, Mar. 2012.

[9] S. Khattak, R. Hamzaoui, S. Ahmad, and P. Frossard, "Fast encoding techniques for multiview video coding," *Signal Process.-Image Commun.*, vol. 28, no. 6, pp. 569–580, July 2013.

[10] H. Zeng, K.-K. Ma, and C. Cai, "Fast mode decision for multiview video coding using mode correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1659–1666, Nov. 2011.

[11] 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. Broad.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[12] 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 Technol.*, vol. 20, no. 6, pp. 925–930, June 2010.

[13] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "Early SKIP mode decision for MVC using inter-view correlation," *Signal Process.-Image Commun.*, vol. 25, no. 2, pp. 88–93, Feb. 2010.

[14] H.-S. Koo, Y.-J. Jeon, and B.-M Jeon, MVC Motion Skip Mode, ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-W081, Apr. 2007.

[15] I. E. G. Richardson, M. Bystrom, and Y. Zhao, "Fast H.264 skip mode selection using an estimation framework," in *Poc. Picture Coding Symposium (PCS'06)*, Beijing, China, Apr. 2006.

[16] C. H. Yeh, K. J. Fan. M. J. Chen, and G. L. Li, "Fast mode decision algorithm for scalable video coding using bayesian theorem detection and markov process,"*IEEE Trans. Circuits Syst. Video Technol.*,vol. 20, no. 4, Apr. 2010.

[17] A. Saha, K. Mallick, J. Mukherjee, and S. Sural, "SKIP prediction for fast rate distortion optimization in H.264," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, pp. 1153–1160, Aug. 2007.

[18] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *Proc. Picture Coding Symposium (PCS'06)*, Beijing, China, Apr. 2006.

[19] Y. Chen, P. Pandit, S. Yea, and C. S. Lim, Draft reference software for MVC (JMVC 6.0), London, U.K., Joint Video Team (JVT) Doc. JVT-AE207, Jul. 2009.

[20] L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, vol. 23, no. 3, Aug. 2004.

[21] M. Domanski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner "Poznan Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, Oct. 2009.

[22] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol., 6, no. 2, pp. 461–464, 1978.

[23] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Fast inter-mode decision in multi-view video plus depth coding," in *Proc. Picture Coding Symposium (PCS'12)*, Krakow, Poland, May 2012.

[24] L. Shen, Z. Liu, P. An, R. Ma, and Z. Zhang, "Low complexity mode decision for MVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 837–843, Jun. 2011.