IEEE International Conference on Communications 2013: IEEE ICC'13 - Workshop on Immersive & Interactive Multimedia
Communications over the Future Internet

Low Complexity Disparity Estimation for Immersive 3D Video Transmission

Brian W. Micallef [1], Carl J. Debono [2], and Reuben A. Farrugia [3]
Department of Communications and Computer Engineering,
University of Malta, Msida, Malta
{[1] brian.micallef, [2] c.debono}@ieee.org, [3] reuben.farrugia@um.edu.mt

***Abstract*—Bandwidth-limited channels demand the transmission of the per-pixel depth maps with the texture data to provide immersive 3D video services that allow arbitrary 3D viewpoint reconstruction. This auxiliary depth data offers geometric information, which together with the multi-view and epipolar geometries, can be exploited during 3D video coding to calculate geometric positions for the search areas of disparity estimation. These positions represent a more accurate estimate match to compensate the current macro-block from than those provided by the median adopted by the H.264/MVC standard. The result is smaller search areas that reduce the encoder's computational requirement. In this work, we exploit this fact together with the largest depth variation within the macro-block to encode, to calculate and adaptively adjust these areas along the epipolar lines. The proposed solution achieves a speedup gain of up-to 32 times over the original disparity estimation, with negligible influence on the rate-distortion performance of 3D video coding. This highly reduces the computational cost of the H.264/MVC encoder and eases its need to be implemented on highly expensive systems that are otherwise necessary to meets the stringent latency requirement of broadcast transmissions. Moreover, it provides similar coding efficiencies required for such scenarios.**

***Index Terms*—3DV coding, fast and adaptive disparity estimation, geometric predictors, multi-view video coding.**

I. INTRODUCTION

THE latest developments in digital technology make 3D Videos (3DVs) easier to produce and display, increasing the need to create 3D TeleVision (3DTV) standards. This led to the Multi-view Video-plus-Depth (MVD) representation, standardized as a 3DV format [1], consisting of multiple texture videos coming from varies cameras that capture the same scene from different viewing angles, together with their geometrical per-pixel depth maps. Advanced 3D image reconstruction techniques, such as Depth Image Based Rendering [2], are then applied to reconstruct the scene from any arbitrary position between the fixed captured camera viewpoints. This feature provides viewpoint video information for both stereoscopic 3DTV [2]; to allow the depth perception of a scene, and for auto-stereoscopic 3DTV [3]; to allow smooth free-viewpoint navigation within a scene.

The large amount of data produced demands the use of efficient compression algorithms before transmission or storage. Currently, the standardized state-of-the-art multi-view video encoder available for commercial use is the H.264/AVC standard with the Multi-view Video Coding (MVC) extension [4]; as 3DV coding standardization is still under development [5]. It defines the extension of H.264/AVC to perform Disparity Estimation (DE) such that inter-view redundancy can be removed. This achieves higher coding efficiency compared to simulcast coding, at the expense of a substantial increase in the computational cost of the encoding system. The technique was originally developed to encode efficiently the texture Multi-View Videos (MVVs) [6], but eventually it was also adopted for depth map multi-view video coding [7], thus further increasing the computational demand of the 3DV coding system. This standard is already employed for 3D media delivery over Blu-ray^{TM} Disc technology and is expected to eventually be deployed over the Internet. Furthermore, immersive 3DVs (at least 3 Video plus 3 Depth – MVD3) will soon need to be broadcasted for home consumption over Digital Video Broadcasting (DVB) Terrestrial-T2 or Satellite-S2 networks. The MVC is currently the best encoding option for such broadcasting scenario since it offers a reduced bandwidth for 3DVs while still keeping its backwards compatibility with current 2D video receivers. Moreover, its bit-stream is compatible with many transmission channels, like all the second generation DVB standards [8]. Nevertheless, the high computational costs of H.264-MVC encoders require expensive systems with a huge computational power, to comply with the stringent low latency transmission requirements. Therefore, if the encoder's computations can be reduced with an insignificant loss in the compression efficiency, much less expensive systems would be required.

To reduce these computations, in our previous work, we proposed a solution that exploited the rich geometrical depth data provided through the MVD representation, together with the multi-view and the epipolar geometries, to accurately estimate an adequate fixed search area along the epipolar lines, for disparity estimation [9]. In this paper, we extend this idea and propose an adaptive search area along these geometric lines, which is proportional to the largest depth variation within the sub-Macro-block (MB) to encode. This is based on the fact that depth and disparity are directly proportional, thus, the higher the depth change, the higher the expected disparity and the larger the required DE's search area. This adaptive technique outperforms the speedup gain of our previous work [9], by about 1.5 times, where a fixed search area along the epipolar lines was considered. Moreover, combining these techniques we achieve a speedup gain of up-to about 32 times over the disparity estimation technique adopted by the standard H.264-MVC. These speedup gains were observed with a slight quality loss of about 0.04 dB from the former and 0.17 dB from the latter. Furthermore, these gains were achieved while encoding both the texture and the depth map multi-view videos, which is highly desirable for 3DV coding.

The rest of the paper is structured as follows; Section II describes the H.264-MVC extension to encode the texture and the depth map multi-view videos for 3DV coding. Section III discusses the exploitation of the available depth data, with the multi-view geometry, to improve the disparity estimation speed. It also presents the proposed technique that exploits the largest local depth variation with the current sub-MB, to obtain an adaptive search area. Section IV describes the simulation environment used, while Section V gives the results obtained. Finally, Section VI concludes the paper.

II. MULTI-VIEW VIDEO CODING

The H.264-MVC extension defines how the motion estimation technique within the H.264/AVC standard can be extended to additionally allow estimation of the current MB from the viewpoint reference frames, to exploit the inter-view redundancies. This is done by extending the buffers containing the reference frames within the CODEC, to include frames that are at the same time instance but representing different viewpoints, to obtain a hybrid motion/disparity estimation technique [6-7]. Since these frames can also contain high correlation with the current frame, as they represent the same objects but from distinct viewing angles, they provide more potential compensation replacements to estimate the current MB, thus improve the coding efficiency.

To find the best compensation replacement for the MB to encode, a MB can be divided into one of the seven combinations of smaller sub-MB parts [10] and the optimal compensation of each part must be identified through an exhaustive and a full search estimation process. This process estimates the block of pixels of the current sub-MB from every possible location within the search area and each reference frame, and a compromise between encoding the compensation position, the smallest distortion from the sub-MB to encode, and the optimal encoding mode is reached through minimizing the Lagrangian Rate-Distortion (R-D) cost function [6, 11]. Thus, to encode a MB, if efficient sub-MB estimates are identified, their translation vectors together with the blocks' difference; as residual data, are encoded. Otherwise, the MB is INTRA coded, both forming part of the video coded bit-stream. The higher the number of compensated MBs, the better exploited the redundancies within the videos from already encoded frames are, thus improving coding efficiency. These exhaustive searches make this process the most efficient but also the most computational intensive component of the video encoder [10]. Hence, extending it further to estimate the sub-MBs from viewpoint reference frames highly increases the probability of an efficient compensation replacement, at the expense of much more computational cost for the encoder. As a result, algorithms which lower this burden need to be identified for MVC [12].

Sub-optimal estimation techniques can be used to reduce these computations, such as the Diamond Search [13]. Furthermore, specific faster DE techniques, such as [14-16], have been proposed in the literature. Nevertheless, the 3DV representation contains a lot of useful geometrical depth information that is currently not being exploited. This data, together with the geometric properties of the multi-view camera system, can be used to geometrically calculate more accurate search areas, which can further improve the DE encoding speed.

III. ADAPTIVE DISPARITY ESTIMATION

The disparity estimation process exhaustively searches for the adequate block to compensate the current sub-MB from the viewpoint reference frames. These optimal replacements must contain small distortions from the MB to encode; to provide accurate compensations and reduce encoding of residual data. The smallest distortions are normally obtained from the corresponding multi-view areas, as they represent the same object in the viewpoint reference frames [17]. Hence, usually searching only around these similar areas can significantly reduce the exhaustive search and its computational time [9, 18-21]. These areas can be identified through the multi-view geometry [17] and the average depth value of the current sub-MB to estimate, with the depth values obtained from the depth map MVV. Since the majority of the depth pixel elements (pels) bias the average depth value, this can provide an accurate compensation area to estimate the sub-MB pels. Nevertheless, a significantly smaller search area of ±10 pels [9, 18] is still required to identify an optimal estimate for the whole block, as the average depth can only estimate a large part of its pels, but not necessarily the entire sub-MB. Thus, it only provides a good position where to search. The optimal R-D match still needs to be identified through DE and should lie around this region.

Furthermore, the DE will normally identify as optimal replacements a block along the epipolar lines [14], as these lines represent the relationship between the multi-view images [17]. Thus, our work in [9] proposed the use of the multi-view geometry to obtain the corresponding multi-view areas, but then search only along the epipolar lines, to reduce further the computational cost. Even so, a search area of ±10 pels, along these lines, was still maintained as the average depth value remained imprecise to obtain an accurate compensation match for the entire block, but the search area above and below these geometric lines was reduced to 3 pixel elements.

However, from the variation in the local depth map pixel elements, one can determine also the reliability of the average depth value to use for multi-view projection; to identify an appropriate area for the entire sub-MB compensation. This is because objects at different depth levels contain different disparity [17] and should be compensated from distinct locations from the viewpoint reference frame. Thus, the reliability of the average depth value can be determined by finding the largest depth change within the sub-MB to encode, and this can be used to estimate a proportional search area. This is based on the fact that depth and disparity are directly proportional, thus if a large depth difference results in the MB, there is more disparity between the compensation of the sub-MB pixels, and using the average depth value may be inaccurate to immediately identify an area to compensate the entire block. So the corresponding position can only be used to start the DE and the proposed fixed search area of ±10 pels is still required along the epipolar line [9], to find an optimal compensation block to efficiently encode the whole sub-MB. On the other hand, if the block contains small depth variation, the block is probably made of homogenous depth or representing the same object. Thus, there is smaller disparity between the optimal compensation of the sub-MB pixels. This means that the average depth value will be more reliable, and the search area can be reduced further. The average depth value and variance are calculated locally for every sub-MB and used to determine the relative multi-view position and search area, respectively. By using local values, the average depth value is more accurate to represent the equivalent sub-MB match, while the depth variation within the smaller sub-MBs reduces; allowing the use of small search areas that further reduce the disparity estimation computations.

Since the scene usually contains its objects and the background with homogeneous depths, a small search area together with the average depth value are often enough and used to immediately identify their sub-MB compensations. Then larger search areas are required and allowed only for the sub-MBs which contain higher depth variations, which usually occurs at the objects' boundaries. This is required since these sub-MBs can contain objects with different depths that can result in larger disparity to compensate the sub-MB parts. Thus, this method allows us to reduce the search area adaptively, depending on the variation within the sub-MB to encode, as shown in Figure 1, and consequently reduce the DE computations and duration.

A search area of ±3 pels around the epipolar line is needed due to its geometry inaccuracies [9, 24], and a minimum search area of ±3 pels along this search line is also required due to depth inaccuracies present in the original depth maps. Then the search area along the line can be varied from ±3 pels up-to ±10 pels depending on the largest depth change within the sub-MB to encode. If this search area exceeds the value of 10 pixel elements, then it is clipped, since this search area demonstrated to be large enough to use with geometric disparity estimation, to maintain coding efficiency [18-20]. Thus, it follows as:

$$\mathrm{search\_area} = \pm \min(3 + \alpha \cdot \Delta_{\mathrm{depth}},\ 10) \qquad (1)$$

where Δ_{depth} is the absolute difference between the statistical maximum and minimum depth value within the sub-MB. The parameter α needs to be obtained statistically for every sequence, as it is dependent on the statistical depth variations within its sub-MBs. To determine this parameter; the MBs in the frame are divided into 8×8 pels blocks, and their statistical maximum and minimum depth values are obtained and subtracted; to obtain their largest depth difference. Finally, the standard deviation of these depth differences is obtained, and α is calculated as:

$$\alpha = {}^{7}\!\big/_{\sigma_{\Delta_{\mathrm{depth}}}} \qquad (2).$$

This is done to allow only about 30% of the sub-MBs, those containing larger depth variation, to use the largest search area, while the other 70% use a smaller search area. This smaller search area is then adapted according to the depth variation within the sub-MB to encode and the spatial depth changes within the 3DV sequence itself. This is initially estimated from the first few depth frames and then updated for every frame, as the statistics are being collected while the MVD is being encoded.

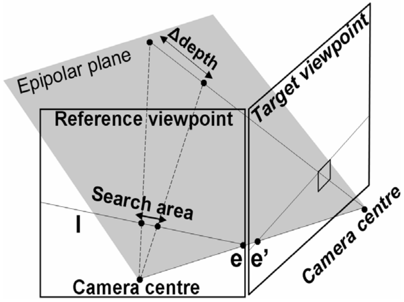

Fig. 1. The adaptive search area along the epipolar line *l*.

IV. SIMULATION OVERVIEW

The proposed adaptive DE was implemented within the Joint Multi-view Video Coding reference model (JMVC ver. 8.5 [22]); to evaluate its influence on its coding performance. The encoder was modified to obtain a geometrical DE with a search area of ±10, as in [18], and to perform this search along the epipolar lines, as in [9]. Then it was further updated to use the above algorithm to determine an adaptive search area along these geometric lines.

The *Breakdancers* [23], the *Ballet* [23] and the *Balloons* calibrated MVD sequences were used to assess the effectiveness of the proposed algorithm. The camera parameters supplied with the sequences were used as the same calibration accuracy required for the intermediate viewpoint rendering is enough for the required multi-view projection. However, if these parameters are not provided or the calibration accuracy is lost, these can be obtained through re-calibration [17, 24]. The first three viewpoints from both their texture and depth map MVVs were encoded. The 3DVs were encoded using only the disparity estimation extension for MVC, to determine the loss in its coding efficiency while using the proposed method. The disparity estimation was set up such that all frames in view 0 are Intra coded, all frames in view 2 are inter-view predicted from view 0, while all frames in view 1 are inter-view bi-predicted from views 0 and 2; for optimal inter-view coding efficiency [7, 12]. For the *Balloons* sequence, views 3-5 were used instead of views 0-2, respectively. Both the exhaustive search estimation and the Diamond Search (DS) were used to determine the optimal disparity vectors within the search area, while encoding both texture and depth map MVVs. The first twenty depth frames were used to determine initial α, which resulted in 0.88, 0.79, and 0.77, respectively, which accordingly relates to the 3D scenes as containing the lowest to the highest depth variations.

The simulations were carried out on a PC with an Intel® Core™ i7 @ 3.2GHz CPU, 6GB of RAM and running Microsoft Windows® 7 Ultimate x64. The 3D videos were encoded with the different MVC encoders, and their CPU usage duration was recorded. Afterwards, the overall speed-up gains of the proposed encoder were estimated. Finally, the original decoder was used to decode the bit-streams for objective evaluation.

V. SIMULATION RESULTS

Tables I and II give the Rate-Distortion performance results obtained using the adaptive disparity estimation, after encoding and decoding the first three viewpoints from the texture and the depth map MVVs, of the *Breakdancers* 3DV, respectively. These tables include the MVV distortion, as an average loss in the Luminance Peak-Signal-to-Noise-Ratio (PSNR), the percentage change in the total MVV bit-rate, and the speed-up gain in the MVC duration. These were calculated with respect to the original DE using the exhaustive search estimation, as this gives the optimal encoding efficiency but with the highest computational cost. The results include also the performances obtained when using the diamond search estimation. Fig. 2 and Fig. 3 illustrate the R-D performance curves obtained for texture and depth map multi-view video coding of the *Balloons* 3DV, respectively. Finally, Tables III-V provide the final summary of the average speed-up gains of the exhaustive search estimation (Speedup) and that of the Diamond Search (DS Speedup), together with the R-D performances in terms of Bjøntegaard Delta (BD) [25], for the *Breakdancers*, *Balloons* and *Ballet* 3DV coding.

TABLE I: R-D VALUES FOR TEXTURE MVC OF THE *BREAKDANCERS* 3DV.

QP	Original DE	Solution in [9]	Proposed DE	Original + DS	Solution in [9] + DS	Proposed DE + DS
28	39.63 dB	-0.012	-0.012	-0.009	-0.010	-0.011
	5203.45 kbps	+1.41%	+1.93%	-0.72%	+0.83%	+1.06%
	19.30 hrs	15.90×	22.60×	14.71×	38.31×	51.98×
32	38.34 dB	-0.061	-0.069	-0.028	-0.061	-0.048
	2840.59 kbps	+1.03%	+1.62%	-0.77%	+1.03%	+1.24%
	18.93 hrs	15.71×	23.52×	15.80×	37.23×	51.66×
36	36.86 dB	-0.105	-0.122	-0.048	-0.083	-0.087
	1804.25 kbps	+0.33%	+0.67%	-0.34%	-0.01%	+0.04%
	18.42 hrs	15.39×	23.14×	15.68×	39.89×	51.23×
40	35.10 dB	-0.108	-0.073	-0.054	-0.100	-0.108
	1239.93 kbps	-0.15%	+0.37%	-0.22%	-0.45%	-0.61%
	18.17 hrs	15.48×	22.57×	15.82×	39.71×	53.14×

TABLE II: R-D VALUES FOR DEPTH MAP MVC OF THE *BREAKDANCERS* 3DV.

QP	Original DE	Solution in [9]	Proposed DE	Original + DS	Solution in [9] + DS	Proposed DE + DS
36	41.13 dB	-0.040	-0.050	-0.026	-0.020	-0.020
	1088.93 kbps	+1.50%	+1.84%	-0.19%	-0.73%	+0.67%
	19.43 hrs	18.07×	25.36×	14.89×	39.06×	52.70×
39	39.22 dB	-0.157	-0.065	-0.022	-0.075	-0.086
	847.34 kbps	+1.13%	+1.48%	-0.45%	-1.06%	+1.02%
	21.18 hrs	18.54×	26.37×	15.37×	40.43×	55.63×
42	37.15 dB	-0.156	-0.127	-0.029	-0.080	-0.091
	677.84 kbps	+0.85%	+0.91%	-0.96%	-0.32%	+0.30%
	19.25 hrs	17.94×	25.29×	15.05×	40.24×	53.11×
45	35.25 dB	-0.136	-0.001	-0.004	-0.069	-0.068
	550.91 kbps	+0.86%	+1.35%	-1.11%	-0.91%	+0.38%
	19.11 hrs	17.84×	25.20×	15.28×	40.36×	54.34×

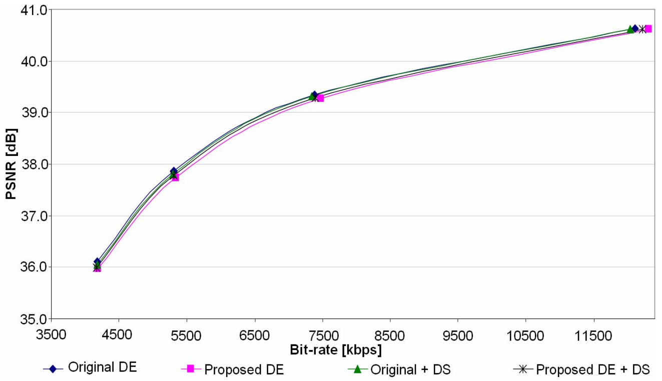

Fig. 2. R-D curves for texture MVC of the *Balloons* 3DV.

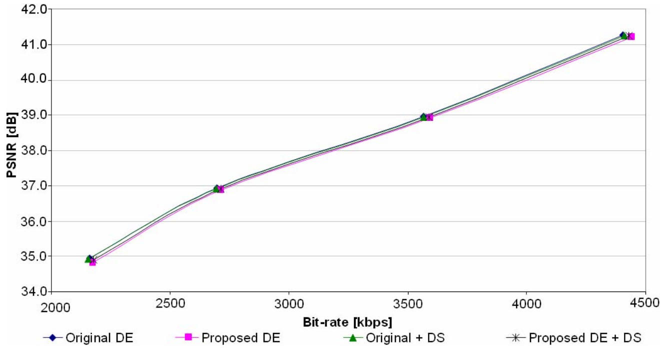

Fig. 3. R-D curves for depth map MVC of the *Balloons* 3DV.

TABLE III: SUMMARY OF ENCODING RESULTS FOR *BREAKDANCERS* 3DV CODING.

Video type	BD-PSNR	BD-bit-rate	Speedup	DS Speedup
Change with respect to the solution in [9]				
Texture	-0.006 dB	+0.27%	1.47×	1.34×
Depth	-0.023 dB	+0.26%	1.41×	1.35×
Change with respect to the original Search Estimations				
Texture	-0.077 dB	+2.48%	22.96×	3.52×
Depth	-0.168 dB	+2.10%	25.55×	3.65×

TABLE IV: SUMMARY OF ENCODING RESULTS FOR *BALLOONS* 3DV CODING.

Video type	BD-PSNR	BD-bit-rate	Speedup	DS Speedup
Change with respect to the solution in [9]				
Texture	-0.014 dB	+0.26%	1.50×	1.53×
Depth	-0.050 dB	+0.56%	1.46×	1.56×
Change with respect to the original Search Estimations				
Texture	-0.085 dB	+1.18%	33.23×	4.52×
Depth	-0.150 dB	+1.12%	35.19×	4.91×

TABLE V: SUMMARY OF ENCODING RESULTS FOR *BALLET* 3DV CODING.

Video type	BD-PSNR	BD-bit-rate	Speedup	DS Speedup
Change with respect to the solution in [9]				
Texture	-0.015 dB	+0.26%	1.67×	1.60×
Depth	-0.042 dB	+0.48%	1.56×	1.63×
Change with respect to the original Search Estimations				
Texture	-0.084 dB	+1.80%	35.60×	4.81×
Depth	-0.190 dB	+1.75%	37.40×	4.97×

These results show that using the proposed adaptive search area provides a further coding speedup gain of about 1.4 for the *Breakdancers* 3DV, 1.5 for the *Balloons* 3DV and 1.6 for the *Ballet* 3DV, over using a fixed reduced search area with the geometric techniques. Finally, these techniques together manage to provide an average speedup gain of 31.66 over the original exhaustive search estimation and of 4.4 over the original diamond search estimation. This causes a further average R-D degradation in terms of BD-PSNR, of about 0.012 dB for the texture and 0.038 dB for the depth map coding; from the solution provided in [9]. This translates to an overall BD-PSNR loss of 0.12 dB and 0.17 dB, respectively; compared to the performance of the original disparity estimations. Fixing the search area to the smallest value of ±3pels would give a higher and a more constant speed-up gain, but it would also add a further significant quality loss of about 3% BD-bit-rate over the presented results. Thus, using the adaptive search area helps us to obtain a good compromise between speed-up gain and coding efficiency. It is interesting to note that as the spatial depth variations within the video starts to decrease, as it can be seen when going from the *Breakdancers* 3DV; with more objects in the foreground, to finally the *Ballet* 3DV; with the least objects in the foreground, the speed up gain increases slightly while the coding efficiency loss remained almost the same. More objects in the foreground means that there are more regions with depth differences between the foreground and the background of the video, thus more depth variances that require larger search areas result. Large depth variation gives smaller α values that allow a corresponding adaptive and a linear gradual reduction in the search area between the large and the small disparity values, according to the video under consideration.

This fast disparity estimation technique can also be used with our fast geometric motion estimation techniques, proposed in [19-21], to finally obtain a low-computational multi-view video encoder, adequate for low-latency 3DV coding.

VI. CONCLUSION

In this paper, a method that exploits the largest local depth variation within the current block to encode; to identify a proportional adaptive search area for disparity estimation, while encoding 3D videos, was proposed and analyzed. This technique manages to improve the speedup gain by about 1.5 times over the previously proposed fixed search areas. This translates to a speedup gain of up-to 32 times over the original search estimations. These gains were obtained with a limited effect on the rate-distortion performance, in both texture and depth map multi-view video coding; thus it maintains the 3D video coding performance. These gains help to obtain the benefits of using the MVC techniques to encode the 3DVs for broadcast, while the 3DV coding computational requirements are reduced.

ACKNOWLEDGMENT

We would like to thank Microsoft Research (MSR) group and the Nagoya University for providing us with the calibrated Multi-view Video-plus-Depth test sequences. This research work was partially funded by the Strategic Educational Pathways Scholarship Scheme (STEPS-Malta) and by European Union - European Social Fund (ESF 1.25).

REFERENCES

[1] ISO/IEC MPEG, and ITU-T VCEG, "*Multi-view Video plus Depth (MVD) format for advanced 3D video systems*," Doc. JVT-W100, Apr. 2007.

[2] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing: Image Communication*, Feb. 2007.

[3] Y. Mori, N. Fukushima, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," in *Proc. of 3DTV Conf. 2008*, pp. 229-232, May 2008.

[4] ISO/IEC IS 14496-10, *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264, Version 13, Mar. 2010.

[5] ISO/IEC MPEG, and ITU-T VCEG, "*Call for Proposals on 3D Video Coding Technology*," Doc. N12036, Mar. 2011.

[6] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. on CSVT*, vol. 17, no. 11, pp. 1461-1473, Nov. 2007.

[7] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient compression of multi-view depth data based on MVC," in *Proc. of ICIP 2007*, pp. 201-204, Sept. 2007.

[8] C. Hellge, E. G. Torre, D. G.-Barquero, T. Schierl, and T. Wiegand, "HDTV and 3DVT services over DVB-T2 using multiple PLPs with SVC and MVC," in *Proc. of 61st Annual IEEE Broadcast Symposium*, Oct. 2011.

[9] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Fast disparity estimation for multi-view video plus depth coding," in *Proc. of VCIP 2011*, Nov. 2011.

[10] M. E. Al-Mualla, C. N. Canagrarajah, and D. R. Bull, *Video coding for mobile communications, efficiency, complexity, and resilience*, Elsevier Science, 2002, USA, pp. 93-200.

[11] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. CSVT*, vol. 13, pp. 688-703, Jul. 2003.

[12] ISO/IEC, "*Survey of algorithms used for multi-view video coding (MVC)*," Doc. N6909, Jan. 2005.

[13] S. Zhu, and K. –K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. on Image Process.*, vol. 9, no. 2, pp. 387-392, Feb. 2000.

[14] J. Lu, H. Cai, J. -G. Lou, and J. Li, "An epipolar geometry-based fast disparity estimation for multiview image and video coding," *IEEE Trans. on CSVT*, vol. 17, no. 6, pp. 735-750, June 2007.

[15] Y. T. Kim, J. Y. Kim, and K. H. Sohn, "Fast disparity and motion estimation for multiview video coding," *IEEE Trans. on Consumer Electronics*, vol. 53, no. 2, pp. 712-719, May 2007.

[16] X. Zhong, and Y. He, "Fast disparity estimation in MVC based on range prediction," in *Proc. of ICIP 2008*, Oct. 2008.

[17] R. Hartley, and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Press, 2003, UK, pp. 279-309.

[18] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Exploiting depth information for fast multi-view video coding," in *Proc. of PCS 2010*, Dec. 2010.

[19] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Exploiting depth information for fast motion and disparity estimation in multi-view video coding," in *Proc. of 3DTV Conf. 2011,* May 2011.

[20] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Exploiting depth information for efficient multi-view video coding," in *Proc. of ICME 2011*, Jul. 2011.

[21] B. W. Micallef, C. J. Debono, and R. A. Farrugia, "Fast multi-view video plus depth coding with hierarchical bi-prediction," in *Proc. of ISCCSP 2012*, May 2012.

[22] ISO/IEC MPEG, and ITU-T VCEG, "*Joint Multi-view Video Coding Model (JMVC 8.5)*," [Online]. Available: CVS: server:*garcon.ient.rwth-aachen.de*:/*cvs*/*jvt*, Jan. 2012.

[23] MSR 3DVs [Online]. Available: http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/.

[24] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A re-view," *Int. Jour. on Comp. Vision,* vol. 27, pp. 161-195, March 1998.

[25] G. Bjøntegaard, "*Calculation of average PSNR differences between RD-curves,*" Doc. VCEG-M33, Apr. 2001.