

Performance Improvement of Segmentation-Based Depth Representation in 3D Imagery by Region Merging

Clifford De Raffaele¹, Kenneth P. Camilleri¹

¹Department of Systems and Control Engineering
University of Malta
Malta

cderraffaele@ieee.org, kenneth.camilleri@um.edu.mt

Reuben A. Farrugia² and Carl J. Debono²

²Department of Communications and Computer Engineering
University of Malta
Malta

reuben.farrugia@um.edu.mt, c.debono@ieee.org

Abstract—The feasible implementation of immersive 3D video systems entails the need for a substantial reduction in the amount of image information necessary for representation. Multiview image rendering algorithms based on depth data have radically reduced the number of images required to reconstruct a 3D scene. Nonetheless, the compression of depth maps still poses several challenges due to the particular nature and characteristics of the data. To this end, this paper outlines a depth representation technique, developed in our earlier work, that exploits the correlation intrinsically present between color intensity and depth images capturing a natural scene. In this technique, a segmentation-based algorithm that is backwards compatible with conventional video coding systems is implemented. The effectiveness of our previous technique is enhanced in this contribution by a region merging process on the segmented regions, which results in a decrease in the amount of information necessary for transmission or storage of multiview image data by a factor of 20.5 with respect to the reference H.264/AVC coding methodology. This is furthermore achieved whilst maintaining a 3D image reconstruction and viewing quality which is quasi identical to the referenced approach.

Keywords—3D scene processing; depth-map representation; multiview images; segmentation-based video coding.

I. INTRODUCTION

Multimedia representation has experienced relentless progression in the past decade whilst striving to enrich quality of experience [1]. Simultaneously, the augmentation in customer expectations has been mainly directed towards more immersive multimedia experiences with the introduction of three-dimensional (3D) media gaining prominence [2] as technological advancements in consumer electronics and the cost reduction of capturing hardware equipment have been registered [3]. This proved evermore influential since the realization of such a multimedia system involves the simultaneous capturing of a unique scene from multiple viewpoints via distributed cameras [4].

The implementation of this technology however necessitates more advancements than ones in acquisition hardware and auto stereoscopic or holographic rendering displays. The sheer volume of data obtained from the capturing cameras still presents a noteworthy challenge in the image processing and transmission domains [5]. Apart from the

physical restrictions that limit the spatial proximity of cameras [6], feasibility constraints also imply that the amount of scene sampling hardware used in 3D systems is truncated, since this would present a linear reduction in the amount of raw video data that would require processing [7]. Conversely however, the availability of multiple scene viewpoints in a 3D system is a paramount user requirement and the continuation of 3D views is a factor directly responsible for the vividness of the user's experience [8].

These contradicting and critical constraints for multiview systems have been effectively addressed by the utilization of Depth Image Based Rendering (DIBR) techniques. These techniques make use of a limited number of strategically placed cameras around the site and their respective information on scene depth to generate a continuous 3D scene. The depth information can be obtained either by using specialized hardware that samples the Time-of-Flight measurements in a scene [9] or by employing software correspondence algorithms that derive the disparity and consequently depth information from multiple, spatially distinct, image frames [10]. Subsequently, virtual viewpoints in the scene are rendered using Intermediate View Reconstruction (IVR) techniques [11] that perform a 3D warp on the color intensity image based on the respective depth values of the pixels [12], as illustrated in Figure 1. These methodologies are thus able to provide the user with the ability to view the scene from arbitrary perspectives with a specific range limit hence generating the fundamental requirement of multiview images [13].

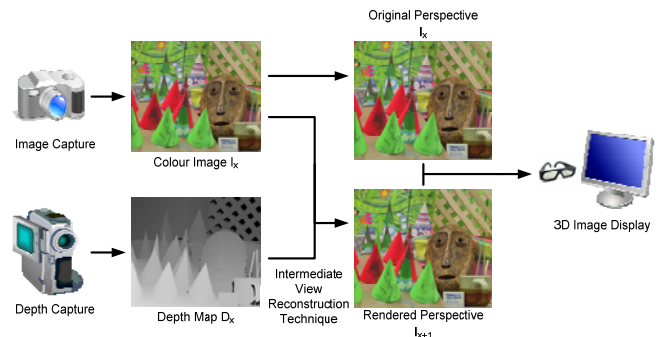


Figure 1. Block diagram overview of the 3D scene reconstruction technique employing a depth image based rendering approach

Foundational work to represent the particular format of data needed by this multimedia technique was devised by the European ATTEST project [14], which combined color images and depth data together into a multiview plus depth (MVD) representation format [15]. This approach also indicated that depth information can withstand additional compression due to the peculiar distribution of individual pixel values constituting a depth map. Consequently, it was proposed that MVD data is processed by the standard H.264/AVC [16] codec, with depth data undergoing further compression such that depth information constitutes a ratio of 10% to 20% when compared to color intensity images [17].

Making use of conventional image compression methods however is not optimal for depth information. This leads to several artifacts being present on reconstructed depth maps since these methodologies are particularly tuned to exploit and remove psychovisual redundancies in color and texture during compression whilst maintaining the perceived visual fidelity to the user [18]. In contrast, depth map information is not directly perceived by the viewer, but employed within IVR algorithms to generate new image perspectives. Thus, in order to yield high quality rendering, this type of depth information is especially sensitive to the preservation of accurate data at depth discontinuities and boundaries, whilst on the other hand being able to rationally tolerate inexact values in uniform depth regions with minimal influence in reconstruction quality [19].

These distinctive issues, together with the unique properties inherent to depth maps have invoked substantial interest within the academic and industrial community to successfully exploit the available redundancies, whilst maintaining accurate depth representation [20]. Hence, numerous algorithms have been proposed within the image and video processing domain to code depth information. These range from shape adaptive discrete wavelet transform (SA-DWT) [21] to exploiting the correspondence between depth and motion vectors [22] by implementing temporal MPEG standards [23]. The scenario of multiview has also been tackled from the geometric modeling perspective, whereby several attempts have been made to represent depth maps by means of mesh-based coding algorithms [24] or as voxel-based octrees [19]. Predominant techniques within the image processing discipline such as JPEG-2000 and region-of-interest (ROI) coding algorithms have been successfully able to take advantage of the spatial similarities present in depth maps [25]. This has led to accurate results being obtained by employing region-based techniques [26], however, these algorithms were not feasible for implementation due to the extensive amount of data required to represent the region contours by chain-coding techniques [27] and thus lose practical attractiveness when compared to the widely adopted standardized H.264/AVC codec [14]-[16].

Following the aforementioned limitations, this paper presents a novel coding methodology for depth information which makes use of the ingrained redundancies present between color and depth images of a natural scene. The spatial consistency of such data is taken advantage of by means of a segmentation-based approach which is able to greatly reduce the amount of data necessary to represent depth maps whilst

still being able to preserve the critical accuracy required to reconstruct high quality 3D images.

The remainder of this paper is organized to provide a description of the technique for representing depth information together with the enhancements proposed in this work in Section II. The algorithm implementation and simulations performed on real-world multiview images is described in Section III. Finally, Section IV outlines a brief conclusion of the presented work.

II. DEPTH MAP REPRESENTATION TECHNIQUE

Our algorithm makes use of the widely-agreed fact that depth maps can be described as being composed of smoothly varying regions enclosed within sharp contours arising from object boundaries in a scene [28 - 30]. These unique properties enable the characteristics of depth maps to be effectively represented by means of a segmentation process that partitions the image into several arbitrary shaped disjointed regions of pixels that contain quasi-homogeneous values. This attribute implies that a singular 8-bit value representing the median depth of pixels inside a spatial segment can be used to encode the representation of depth pixel values for each unique segment. Whilst reconstructing the depth information however, the decoder lacks the necessary information to generate the outline profile of each individual region. Providing the exact contours for boundary reconstruction of each segment is on the other hand prohibitively expensive due to the sheer amount of data required, which would in turn reduce the viability of employing the segmentation algorithm [27]. Our earlier work [38], overcomes such a problem by exploiting the degree of correspondence present between the depth image and the color intensity image which are both characterizing a unique scene from an identical viewpoint and timestamp. This technique considers the color intensity image to be provided to the decoder by means of traditional coding algorithms. Subsequently, the color image, being present at the encoder and decoder of the system, can be utilized at both ends to extract the contours of each region by means of an identical segmentation procedure. As a result, the only additional coding data necessary to be provided to the decoder, in order to reconstruct the depth map, is a short sequence of depth values, each corresponding to a specific segmented region as obtained at the encoder, conveyed as side information.

The technique postulates that since any object captured in an image would contribute to a depth boundary, the corresponding region can be rendered by at least one segment on the color intensity image. This postulate is generally well satisfied in practice since the contours in a depth map represent the occluding edge between a foreground object and a distinct background article. If both objects have dissimilar appearance characteristics such as color, texture or illumination values, these measures will be exploited by the segmentation process and resulting boundary would be detected through the image [31]. It may be noted however, that in the incidence of two distinct and adjacent objects in a scene having similar appearance characteristics which, the foreground and background regions in the image may be merged into a single under-segmented region across the depth boundary. Nevertheless, this issue can be mitigated by the fact that if the

two objects have distinct depth values and thus constitute a depth boundary, they are physically located at different locations in the scene, and thus are typically subject to dissimilar illumination from both natural and artificial light sources. This discrepancy would thus be evident in image components which are illumination dependent as can be witnessed in Figure 2.

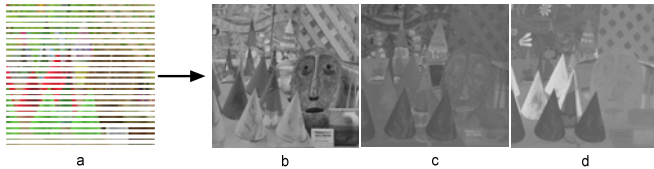


Figure 2. RGB colour image (a) is decomposed into an illumination dependent colour space before segmentation on the (b) luma component, (c) blue-difference chroma component (d) red-difference chroma component

The depth representation technique entails the need for objects in a scene to be segmented independently in a number of unique regions from their adjacent objects. Hence, this requires a segmentation algorithm that is able to generate contiguous regions within the image that correspond to a unique and distinct label. Consequently, algorithms such as boundary-based segmentation [32] and K-means clustering [33] techniques cannot be suitably applied. The required properties for segmentation can be fulfilled by employing a Normalised Graph-Cut approach [34], [35]. Further detail and elucidation on the manner in which the segmentation algorithm was employed is documented in [38].

Although the reconstructed depth map accuracy, and subsequently, the quality of the rendered image depend on the performance of the segmentation algorithm, this content-dependent uncertainty is mitigated by executing the aforementioned segmentation algorithm on each of the illumination color-space components. This technique, apart from ensuring that segment boundaries are identified by the algorithm in at least one image component, aids the framework in achieving a segmentation quality which is not content-dependent.

The utilization of segmentation procedures on individual components of the image potentially yields to segmentation effects that result in very small isolated regions as evidenced in the black-colored regions from the segmented component of Figure 3b. These instances, which are generated due to noise artifacts present in the image components [36], are made more frequent following the combination of the individual segmentation results for each color component as described further down and as illustrated in Figure 3d. Moreover, since these partitions are only a few pixels wide in nature they provide negligible information to the representation technique. Nevertheless, the additional distinct regions generated by these artifacts require the proposed framework to necessitate more data, since each small partition will be allocated a distinct depth value. This in turn reduces the effectiveness of the same methodology in curtailing the amount of information required for depth map representation.

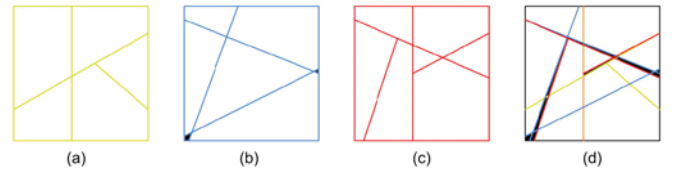


Figure 3. (a-c) Separate segmentation results on the individual image components are then combined to derive the (d) final segmentation map. Black partitions represent regions with insignificant information that will be merged in the proposed framework

To this end, this paper further improves the results attained by the segmentation-based approach defined in our previous work [37] by filtering out these segments and merging the latter to the neighboring regions that retain the closest colorimetric properties. The threshold for minimum amount of pixels required in a region prior to integration was heuristically determined based on the statistical distribution of the segments in the images [38], and an iterative algorithm was developed to ensure that pixels were properly connected prior to being assigned to a neighboring large region. Consideration was done at each epoch on the amount of neighboring regions a candidate pixel possessed, and neighborhood evaluation was only performed once more than two-thirds of the adjacent pixels have been already allocated to pertinent regions. This approach ensured the connectedness of individual segments whilst providing more data to help ascertain the correct selection of a neighboring region.

III. TECHNIQUE IMPLEMENTATION AND RESULTS

The proposed 3D data representation technique was implemented in a manner to ensure backward compatibility with traditional 2D image frameworks. As a result, the system considers that the color intensity image is compressed at the encoder and subsequently decoded back following a transmission or storage procedure using standard multimedia compression codecs. To validate the proposed algorithm effectiveness and quality, the standard baseline images of *cones*, *sawtooth* and *teddy* were employed. These were captured at a resolution of 450×375 pixels and needed 57.2kB, 49.9kB and 44.3kB of data respectively to be represented with the state of the art H.264/AVC encoding scheme [39]. The proposed technique hence presumes that the decoder has reconstructed the color intensity image at the receiver and does not hold any information relevant to the depth map prior to segmentation as delineated in Section II.

To further supplement implementation feasibility whilst availing of the illumination dependent nature in which the data is formatted during the decoding process, the proposed technique employs the segmentation algorithm directly on the $YCbCr$ color model information which describes the luminance and chroma components of the image respectively as illustrated in Figure 2. The normalized graph cut methodology is hence executed simultaneously at the encoder and decoder entities with parallel implementation performed on each channel component of the image individually. To reduce the computational complexity and abridge the processing time necessary, the segmentation methodology is executed to perform only 50 maximum energy cuts on each image component. The results of these individual segmentations are

successively amalgamated in a conjunctive manner such that overlapping regions are fragmented into individual segments that have common support as shown in Figure 4.

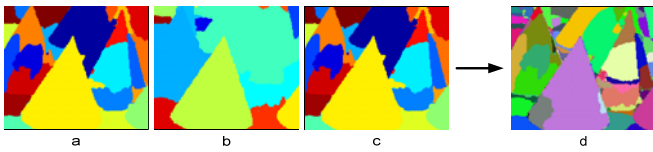


Figure 4. (a-c) Segmentations performed separately on all 3 components of the image colour-space are then utilized to derive the (d) final segmentation region map by combining the diverse partitions.

It is instantly apparent in Figure 4d however, that the over-segmentation resulting from the region combination of each image component is not optimal and a number of distinct redundant partitions containing a very limited number of pixels are developed. As described in Section II these partitions are merged to their neighboring segment if they contain less than ten pixels. To ascertain connectedness, pixel assignment evaluation is performed on the nine adjacent pixels of each candidate, and color based merging is completed only once a pixel has more than two-thirds neighborhood connectedness. The resulting segments after deduction of the insignificant regions is performed are illustrated in Figure 5.

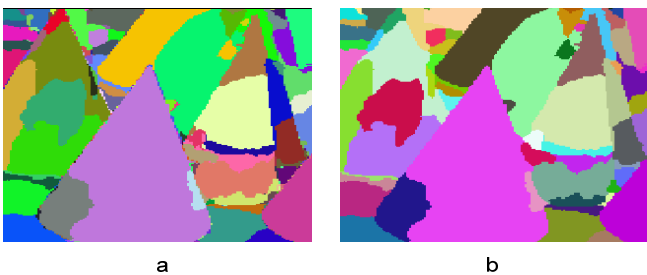


Figure 5. (a) Combined segmentations performed on individual illumination components, (b) Resulting segments following proposed merging algorithm.

The proposed merging procedure augmented significantly the effectiveness of the depth representation technique since only significant regions were considered for depth allocation. Henceforth, the depth representation is achieved by allocating each segment a respective depth value based on the median depth of region. Thus, the encoder is able to utilize the aforementioned list and transmit this data to the decoding entity as side information. Furthermore, owing to the quantized nature of depth values, standard Huffman coding algorithms are used to reduce the final size of the side information. The proposed depth representation technique thus stipulates that the decoder, subsequent to executing an identical segmentation procedure, assigns to each derived region the respective depth value from the ordered side information. The resulting reconstructed depth map, shown in Figure 6c is successively utilized in 3D enabled displays to generate a spatially shifted virtual viewpoint, which together with the original color image yield a stereo image pair as depicted in the procedure of Figure 1.

Since decoded depth maps are not directly observed by viewers during 3D vision, comparison metrics performed on depth maps fail to provide an effective and reliable measurement of the quality of depth images. Hence, as concurred in literature [19], [23], a legitimate comparison of depth maps can be obtained when considering objective quality metrics, such as peak signal-to-noise ratio (PSNR), on the virtual reconstructed viewpoints. These images were generated using a standard IVR algorithm [40], which employed the decoded color image together with a respective depth map for reconstruction.

The baseline images utilized were captured from an N -camera multi-view system, where a number of images I_k , where $k=1,2,\dots,N$, are available for each given scene such that each image I_k , represents the view from the k^{th} camera respectively. As a baseline reference for comparison on our system, the decoded view of image I_k is rendered for the $k+1^{\text{th}}$ camera to obtain a reconstructed image I_{k+1}' using an H.264/AVC decoded depth map with a quantization parameter to limit the data rate at 15% of the color image. This process is repeated using the proposed compressed depth map to obtain I_{k+1}'' . Both I_{k+1}' and I_{k+1}'' are then compared to an I_{k+1} image which was directly captured by a camera within the multi-view system, to estimate the rendered view quality employing the proposed depth map representation technique with respect to benchmark H.264/AVC depth image as shown in Figure 6.

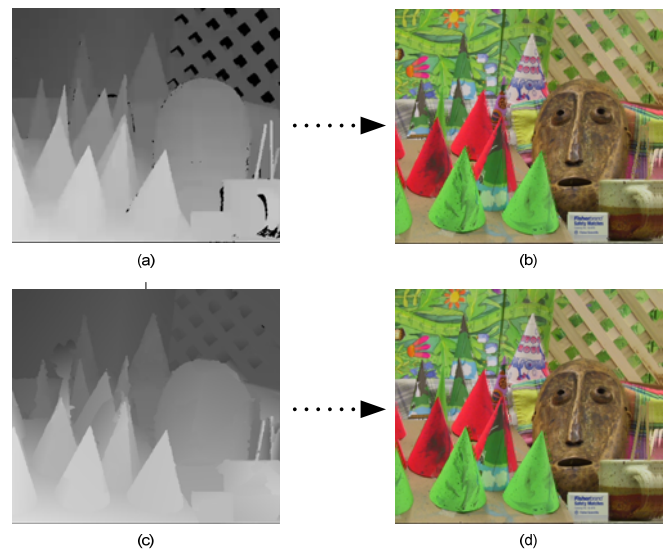


Figure 6. (a) reference depth map compressed via H.264/AVC, (b) reference virtual-view rendering, (c) proposed depth map representation technique, (d) virtual-view rendered utilizing the proposed depth image.

Performing a subjective evaluation between the resulting images in Figures 6(b) and 6(d) immediately exposes the fact that the proposed depth representation technique is able to provide a rendered view quality which is comparable to that generated from depth map encoded using the reference H.264/AVC coding methodology. The insensitivity of the proposed methodology to the segmentation quality has been demonstrated by the consistent quality results that were also derived when utilizing the described representation technique

for depth images of other well-known baseline image sets as shown in Table I. Furthermore, to additionally expose the data rate reduction resulting from representing the depth image via the novel technique, Table I also delineates a comparison between the H.264/AVC compression of depth images using a quantization parameter of 34 such that the data rate for depth information is limited to 15% of that required for the color image.

TABLE I. DATA RATE AND PSNR COMPARISON OF DEPTH MAP REPRESENTATION TECHNIQUES ON RECONSTRUCTED IMAGES

Image	Depth Representation			
	No. of Regions	H.264/AVC depth map at 15% color data rate (kB)	Proposed Method (kB)	Coding Gain (%)
Cones	498	8.44	0.39	2164
Sawtooth	345	4.11	0.25	1644
Teddy	399	7.24	0.32	2263
Average	414	6.59	0.32	2059
Image	Rendering Quality			
	H.264/AVC I_{k+1} ' PSNR (dB)	Proposed Method I_{k+1} ' PSNR (dB)	PSNR Difference (dB)	
Cones	29.92	30.04	+0.12	
Sawtooth	33.24	33.46	+0.22	
Teddy	31.88	32.27	+0.39	
Average	31.68	31.92	+0.24	

The values expressed in Table I indicate that the IVR rendered images reconstructed from depth maps represented by the proposed technique are objectively comparable to those generated via the H.264/AVC compressed depth images. This is established by the measured PSNR values which differ only by 0.2 dB. Moreover, this quality consistency was obtained under the exceptional interpolation performed by the IVR algorithm in mapping the virtual views onto a neighboring camera. This abnormally distant interpolation from the original capturing camera location would potentially further amplify discrepancies between depth images. As outlined in Section II, this depth representation methodology requires that the multi-component segmentation algorithm is executed also at the decoder, thus increasing the computational complexity at this end of the system. This burden is however well alleviated by the compression gains achieved by the proposed technique. In fact, as shown in Table I, the depth map compression results demonstrate that the novel depth representation technique results in coding gain factor of 20.5 compared to standard H.264/AVC, requiring on average only 414 bytes to represent a dense depth map with resolution of 450×375 . Moreover, when evaluated against our previous contribution [38], the enhancements proposed in this paper yielded a data rate decrease of over 27% whilst preserving the difference in objective quality across the different images to 99.7%.

IV. CONCLUSION

This paper has described a segmentation-based depth representation technique which exploits the inherent correspondence present between color intensity and depth map images captured from a natural scene by means of a multiview system. The effects of over segmentation artifacts were analyzed, and an algorithm which merges small regions based on their colorimetric features was proposed. This enhancement is integrated in the presented depth representation technique and implementation results show that a data rate reduction by a factor of 20.5 is achieved with respect to the standardized H.264/AVC compression algorithm. This is done whilst maintaining a quasi-identical quality in the reconstructed 3D virtual views, as testified by the consistent PSNR values when compared to the reference H.264/AVC depth map compression scenario.

REFERENCES

- [1] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W.I. Jsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek and I. Sexton, "An evolutionary and optimised approach on 3D-TV", in Proc. of Int. Broadcast Conf., pp. 357–365, Amsterdam, The Netherlands, 2002.
- [2] A. Kubotka, A. Smolic, M. Magnor, M. Tanimoto, T. Chen and C. Zhang, "Multiview imaging and 3DTV", in IEEE Signal Process Magazine, vol. 24, no. 6, pp. 10–21, Nov. 2007.
- [3] S.U. Kum and K. Mayer-Patel, "Intra-stream encoding for multiple depth streams", in Proc. of the 16th Int. Workshop on Network and Operating Systems Support for Digital Audio and Video, pp. 62–67, Newport, USA, May 2006.
- [4] A. Smolic and P. Kauff, "Interactive 3D video representation and coding technologies", in IEEE Special Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, Jan. 2005.
- [5] A. Vetro, S. Yea, M. Zwicker, W. Matusik and H. Pfister, "Overview of Multiview Video Coding and Anti-Aliasing for 3D Displays", IEEE Int. Conf. on Image Processing, vol. 1, pp. 1-17 - I-20, Sep. 2007.
- [6] J. Carranza, C.M. Theobalt, M. Magnor, and H.P. Seidel, "Free-viewpoint video of human actors," in ACM Trans. on Graphics, vol. 22, no. 3, pp. 569–577, Jul. 2003.
- [7] S.U. Yoon, E.K. Lee, S.Y. Kim and Y.S. Ho, "A framework for multi-view video coding using layered depth images", in Proc. of the Pacific Rim Conf. on Multimedia, pp. 431–442, Jeju Island, Korea, 2005.
- [8] J. Starck, J. Kilner and A. Hilton, "Objective quality assessment in free-viewpoint video production," in Proc. 3DTV Conf.: The true vision – capture, transmission and display of 3D video, Istanbul, 2008, pp. 225–228.
- [9] H.T. Nguyen and M.N. Do, "Image-based rendering with depth information using the propagation algorithm", in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, Mar 2005.
- [10] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", in Int. Journal of Computer Vision, vol. 47, no. 1, pp. 7–42, Jun. 2002.
- [11] S. Seitz, and C. Dyer, "View Morphing," in Proc. of ACM SIGGRAPH, pp. 21–30, Aug. 1996.
- [12] C. Zhang, L. Huo, C. Xia, W. Zeng and W. Gao, "A virtual view generation method for free-viewpoint video system," in Proc. of the Int. Symp. on Intelligent Signal Processing and Communication Systems, Xiamen, pp. 361–364, 2007.
- [13] S.C. Chan, H.Y. Shum and K.T. Ng, "Image-based rendering and synthesis", in IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 22–33, 2007.
- [14] A. Redert, M. Op de Beeck, C. Fehn, W. Ijsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton and P. Surman, "ATTEST – Advanced Three-Dimensional Television Systems Technologies", in Proc. of 3DPVT, p.p. 313-319, Italy, Jun 2002.

- [15] P. Kauff, A. Smolic, P. Eisert, C. Fehn, K. Müller and R. Schäfer, "Data Format and Coding for Free Viewpoint Video", in Proc. of Int. Broadcast Conf., Amsterdam, The Netherlands, Sep. 2005.
- [16] T. Wiegand, G.J. Sullivan, G.Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard" in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 560-576, Jul. 2003.
- [17] C. Fehn, K. Hopf and Q. Quante, "Key technologies for an advanced 3D-TV system", in Proc. SPIE 3-DTV Video Display, pp. 66-80, 2004.
- [18] R. Krishnamurthy, B.B. Chai, H. Tao and S. Sethuraman, "Compression and transmission of depth maps for image-based rendering", in IEEE Int. Conf. on Image Processing, Oct. 2001.
- [19] B.B. Chai, S. Sethuraman and H.S. Sawhney, "A depth map representation for real-time transmission and view-based rendering of a dynamic 3D scene", in Proc. of 1st Int. Symposium on 3D Data Proc. Visualization and Transmission, pp. 107-114, Jun 2002.
- [20] M. Flierl, A. Mavlankar and B. Girod, "Motion and disparity compensated coding for multi-view video", in IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, pp. 1474 - 1484, Nov 2007.
- [21] M. Maitre and M.N. Do, "Depth and depthcolour coding using shape-adaptive wavelets", in Journal of Visual Communication and Image Representation, vol. 21, no. 5-6, 2010.
- [22] X. Cao, Y. Liu and Q. Dai, "A flexible client-driven 3DTV system for real-time acquisition, transmission, and display of dynamic scenes", in EURASIP Journal on Advances in Signal Processing, vol. 2009.
- [23] D. Tzovaras, N. Grammalidis and M.G. Strintzis, "Disparity field and depth map coding for multiview image sequence compression", in Proc. of the Int. Conf. on Image Processing, pp. 887-890, Lausanne, Sep 1996.
- [24] S.Y. Kim and Y.S. Ho, "Mesh-based depth coding for 3D video using hierarchical decomposition of depth maps", in IEEE Int. Conf. on Image Processing, pp. 117-120, Oct. 2007.
- [25] M.J. Gormish, D.Lee, and M.W. Marcellin, "JPEG-2000: overview, architecture and applications, proceedings" in Proc of Int. Conf. on Image Processing, Sep 2000.
- [26] L.S. Karisson and M. Sjöström, "Region-of-interest 3D video coding based on depth images", in Proc. of 3DTV Conf: The true vision-capture, transmission and display of 3D video, pp.141-144, May 2008.
- [27] D.V.S.X. De Silva, W.A.C. Fernando and S.L.P. Yasakethu, "Object based coding of the depth maps for 3D video coding", in IEEE Transactions on Consumer Electronics, vol. 55, no. 3, pp. 1699-1706, Aug 2009.
- [28] C. Cigla, X. Zabulis and A.A. Alatan, "Region-based dense depth extraction from multi-view video", in Proc of Int. Conf. on Image Processing, pp. 213-216, Texas, Sept 2007.
- [29] W. Yang and K.N. Ngan, "Unsupervised Multiple Object Segmentation of Multiview Images", in Proc. of Advanced Concepts for Intelligent Vision Systems, pp. 178-189, 2007.
- [30] B. Goldlucke and M.A. Magnor, "Joint 3D reconstruction and background separation in multiple views using graph cuts", in Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 683-688, 2003.
- [31] B. Goldlucke and M.A. Magnor, "Joint 3D-reconstruction and background separation in multiple views using graph cuts", in IEEE Computer Vision and Pattern Recognition, pp. 683-688, 2003.
- [32] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, "Contour detection and hierarchical image segmentation", in Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, pp. 898-916, May 2011.
- [33] A.Z. Chitade and S.K. Katiyar, "Colour based image segmentation using k-means clustering", in Proc. of Int. Journal of Engineering, Science and Technology, vol 2, no. 10, pp. 5319-5325, 2010.
- [34] J. Shi and J. Malik, "Normalised cuts and image segmentation", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. 2000.
- [35] T. Cour, F. Benezit and J. Shi, "Spectral segmentation with multiscale graph decomposition", in IEEE Computer Vision and Pattern Recognition, pp. 1124-1132, California, 2005.
- [36] H. Talbi, M. Batouche and A. Draa, "A quantum-inspired evolutionary algorithm for multibjective image segmentation", in Proc. of the 31st World Academy of Science, Engineering and Technology, pp. 205-210, 2007.
- [37] D.N. Chun and H.S. Yang, "Robust image segmentation using genetic algorithm with a fuzzy measure", in Trans. on Pattern Recognition, vol. 29, no. 7, pp.1195-1211, 1996.
- [38] C. De Raffaele, K.P. Camilleri, C.J. Debono and R.A. Farrugia, "Efficient multiview depth representation based on image segmentation", in Proc. of IEEE Picture Coding Symposium (PCS 2012), pp. 65-68, Krakow, Poland, May 2012.
- [39] "Advanced video coding for generic audiovisual services," Int. Telecommun. Union-Telecommun. (ITU-T) and Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC) JTC 1, Recommendation, H.264 and ISO/IEC 14 496-10 (MPEG-4) AVC, 2003
- [40] D. De Silva, W. Fernando, and H. Kodikaraarachchi, "A new mode selection technique for coding depth maps of 3D video", in IEEE Int. Conf. on Acoustics, Speech and Signal Proc., pp. 686-689, Mar. 2010.