ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

An Advanced Compression Algorithm Employing Expectation Vision Synthesis For FVV And 3D Videos

Praveen Kumar Lendale, Anurag University

Anil Kumar Gona Anurag University

Akku Madhusudhan Anurag University

Abstract: High standard virtual views should be integrated from the adjoining available views in order to produce greater realistic 3D experience for users with the help of free viewpoint video FVV and multiview video coding MVC. The present methods like view synthesis methods experience providing a lower quality as a result of holes produced by blockage and rounding the integer error by wrapping. They make use of temporal and spatial correlation in the depth maps in order to reduce holes in particular virtual view. Because of less similarity in the texture images and depth maps they experience the degradation of the quality in the surrounding and foreground regions. We use different models in Gaussian Mixture Modelling to distinct the foreground and the background pixels and to overwhelm the limitations of the existing techniques. Using the method of adaptive weighted average of intensities of pixels with the help of similar Expectation Maximization models the lost pixels after wrapping are retrieved. There is variation of the weights with respect to time for serving the changes caused by vital background and movement of the kinetic objects for view synthesis. On condition that presentation of intensities of pixels drops consequentially the proposed method introduces an adaptive approach for resetting Expectation Maximization modeling. Based on the executed outcomes the presented technique is proved to produce 5.39 to 6.59DB enhanced PSNR as compared to similar methods. To check the viability of presented vision synthesis method we make use of it as additional cited frame in moving estimate for MVC. The executed outcomes of the presented method prove to boost the PSNR by 3.14 to 5.12 DB with traditional 3 cited frames.

Index Terms—Vision synthesis, free viewpoint video, depth image based rendering, multiview video compression

I. INTRODUCTION:

FVV dragged a fair attention of users in past few years because it allows them to view a location from various angles [1]. Wide range of views with compact baselines is essential for enabling such luxury and it even expands storage data and transmission bandwidth. One of the practical methods for reducing storage and transmission bandwidth of the videos and relative depth maps is Depth image based rendering (DIBR) [1],[2]. Occlusions are the invisible sections of virtual view mostly caused due to obstacles mostly

front objects. Holes may also be created in the composite video due to the occlusions [3]–[6]. Wrapping of various views mostly lead to additional origin of errors by rounding of pixel location coordinates.

With the help of spatial and temporal correlation methods, the missing holes could be filled. In spatial domain, the spatial relationship of video is utilized for filling lost pixels. With the help of two adjoining cameras providing a wide angle of viewing, helps in reducing the holes [6]. The technique wraps

two adjoining views to get an outcome of single view and helps in reducing the holes. Because of bandwidth restriction limited views are transmitted. This results in losing the pixel information in rendered view [7]–[10]. Inpainting is one of the most popular methods for recovering the lost pixels by utilizing spatial connection in the absence of blur traces. It even degrades the quality of the view as it creates the difficulty in recognizing the background and foreground boundaries. This is the resultant of a lower spatial correlation between boundaries of foreground and background [2], [5].

The proposed method makes use of different Expectation Maximization to distinct the foreground pixels from the background. It even alters the intensities of the pixels accordingly. The errors occurred during the wrapping process are overcome using adaptive weighted average for generating the intensities of pixels. A reset mechanism is used to continue the applicability of the Expectation Maximization system. Vision synthesis methods are considered to be the most favorable tools for providing visions from multiview vision plus depth(MVD) and to carry the 3D video coding[1][18].

3D-HEVC gives a finest compression percentage for MVD details by utilizing the vision synthesis optimization (VSO) tool for coding [3]. The view provided is a result of encoding method for calculating RD performance. The method attains excessive compression capability, but it imposes a massive reckoning load to encoder. In an alternative method [22], depth and the vision synthesis distortion models were presented for reducing the evaluation complication in the absence of view rendering. The precision was declared to be less than supposed to be [22].the DBIR methods present an additional frame by utilizing disparity between the adjoining views. Because of higher resemblance of presented view synthesis to current view, the approach gives a

higher divination as compared to traditional three reference methods. For confirming the capability of the presented vision synthesis, a generated frame is used as supplemental reference frame in MVC for estimating the motion. The results of execution prove that the presented vision synthesis has the capacity to refine the PSNR as compared to the traditional methods. As the virtual view doesn't refer to any motion estimation, the time to compute four baseline frames is compared to the three baseline frames.

Two baseline frames are also used and are produced by the presented vision synthesis frame and the previous frame. The outcomes of the execution prove that the PSNR could be improved compared to three reference method.

The preparatory idea of vision synthesis technique proposed in [2]. The contemporary presentations in the paper are (i) adaptive weighting, (ii) adaptive reset policy for pixel modelling, (iii) a latest way for generating pixel intensity of the virtual view, (iv) vision synthesis using synthesized images, (v) initiating four reference and two reference methods in lieu of the standard three reference MVC.

The remaining paper gives section II that elaborates the presented vision synthesis proposal with adaptive weight hole stuffing method, section III focus on vision synthesis for MVC. Section IV and V describes experimental results and conclusion respectively.

II. PROPOSED VISION SYNTHESIS TECHNIQUE

In the recent techniques [13],[16] Expectation Maximization method was made to use for vision synthesis with the help of background frame. In the technique presented the models in Expectation Maximization help in separating the foreground and background pixels and the intensities of pixels are modified. it is made to happen with the help of intensities of similar model pixels. Apart from the background model it is done on models present in Expectation Maximization too. The pixels lost in background in the process of wrapping are retrieved by wrapped image and adaptive weighted average of intensities of pixels. The intrinsic attributes of Gaussian mathematical models turn to account for retrieving the occlusions. Expectation Maximization method is much beneficial in the framework of static backgrounds and in addressing the problems of pixel intensities during occlusions. It could even handle dynamic backgrounds with negligible changes to the approach. A reset mechanism is used to the presented method for handling more of dynamic scenarios in the situation where the relevancy is lost by current models.

The background is represented by pixel whose intensity is alike over a span of time as it is indicated to be a single model in the Expectation Maximization method. Pixels with various intensities at a position are elected with multiple Gaussian models.

The proposed method gives a superior pixel correlation, leading fines quality than the inpainting & other update methods. In particular positions the pixels are considered to be background pixel and at some positions they are considered to be foreground pixel [2]. Practically, the pixels after experiencing the foreground intensities even experiences the background intensities using the same, the pixels of foreground and background are found and the missing pixels are filled in the virtual view

In the method presented, with the help of correlated depth maps and various parameters of camera, the n^{th} image from a pair of adjoining views are encased to a virtual position and n^{th} image of the transitional view is generated

The wrapped images have holes present in them because of the occlusion. A blended image is obtained by blending a pair of wrapped images and count of holes is reduced. But all the missing pixel intensities cannot be recovered. In order to recover all the missing pixels, Expectation Maximization method is applied for modeling each and every pixel by the obtainable preceding frames of virtual view, indicate in Fig. 1.

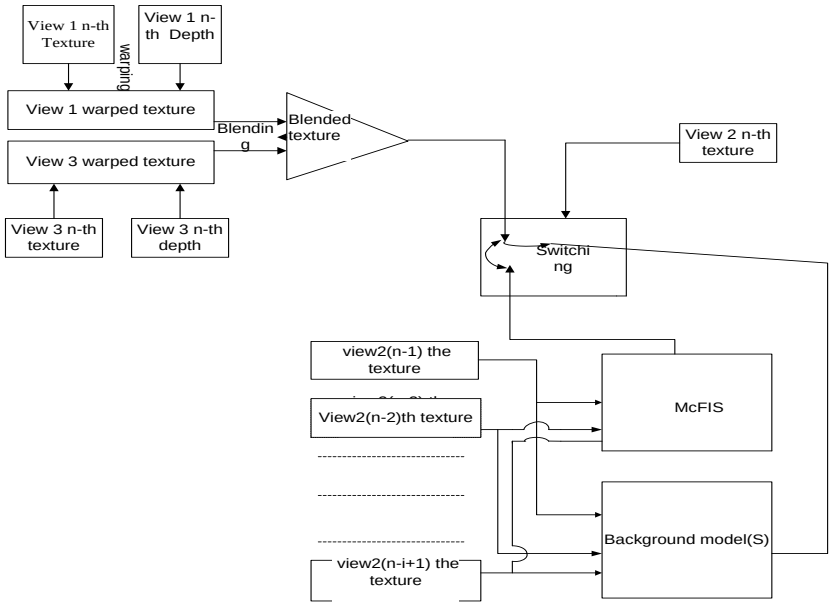

Fig. 1 Vision synthesis technique.

Our demonstration presumes to have 1-(n-1) th frames for Expectation Maximization, when the nth frames of a virtual view is generated where i= 2, 3, 4….N. Originally Expectation Maximization frames are used and later on synthesized frames are used.

Depending on the count of Expectation Maximization models pixels are arranged as fore ground and background pixels. Individually to retrieve back the lost pixels, most common frame in scene(MCFIS) IS GENERED FROM BACKGROUND MODEL[17].Using the adaptive intensities in between blended and learned background and foreground models ,the lost pixel intensities of background and foreground are filled up. The following sections explains the Expectation Maximization technique, interpolation of virtual view, hole filling method and picking the values of factor.

A. Interpolation of virtual view:

In the demonstration ot is assumed that a pair of texture images are transmitted by the sender The depth parameters are transformed from encrypted depth map Ω by [1][2][3]

$$z = \frac{z_{near} z_{far}}{\left(\frac{\Omega}{255}\right)(z_{near} - z_{far}) + z_{near}} \qquad (1)$$

Here, Z_{far} implies the farthest depth and Z_{near} implies the nearest depth.

B. The GMM technique*:*

The method is mostly applied in order to separate the foreground pixels from the background pixels. In this technique pixels are modeled individually with mixture of kth Gaussian distributions. The value set for k would be mostly 3. Assuming at time 't' the value of K-th Gaussian intensity would be $\eta_{k,t}$ mean would be equal to $\mu_{k,t}$, variance equals σ^2_{kk} and weight equals $w_{k,t}$. Therefore
$\sum_{k=1}^{k} w_{k,t}$ would be equal to 1

Once the initial parameters are set, the present pixels are matched with K-th Gaussian for the current viewing observation if $|X_t - \mu_{k,t}| \leq 2.5\sigma_{k,t}$ is satisfied averse to the existing model, X_t would be new intensity of pixel at time t.

The case where the model matches, Gaussian model would be upgrade to

$$\mu_{k,t} \leftarrow (1-\alpha)\mu_{k,t-1} + \alpha X_t; \qquad (2)$$

$$\sigma^2_{k,t} \leftarrow (1-\alpha)\sigma^2_{k,t-1} + \alpha(X_t - \mu_{k,t})^T(X_t - \mu_{k,t}) \qquad (3)$$

$$\omega_{k,t} \leftarrow (1-\alpha)\omega_{k,t-1} + \alpha, \qquad (4)$$

The weights of the remaining Gaussian models upgrade to

$$\omega_{k,t} \leftarrow (1-\alpha)\omega_{k,t-1} \qquad (5)$$

The weight values are normalized amid all the models such that $\sum_{k=1}^{k} w_{k,t}$=1

C. Hole filling:

A pixel is represented to be a static background pixel if it constantly experiences a single model even after a period of time in various frames. On the contrary, pixel if experiences multiple models, then it implies to be foreground pixel or the background pixel. Depending on the highest value of the weight a stable background is represented. Expectation Maximization has an intrinsic capability of capturing the intensities of background and foreground pixels. Hence the intensities of the lost pixels could be easily restored successfully by utilizing temporal correlation. The synthesized final images pixels intensities all considered to be the weighted average from blended one.

The particulars of interpolated recovering technique is explained in detail below

Case I

If pixel a single model for all the time for a particular assigned color, the previous value $(B^c_{k,t})$ is stored for ultimate image synthesis utilizing

$$\varphi^c_t = (\varepsilon_v - 0.5842)\Phi^c_t + (1.0 + 0.5842 - \varepsilon_v)B^c_{k,t}. \qquad (6)$$

Here ε_v implies weighing factor

Case II

Here pixel experiences multiple models for a particular assigned colour in a time period. It would either be considered as background or foreground pixel. Firstly, the holes are filled by applying inverse mapping, then the smallest variance between the intensities of pixel ϕ^c_t and recent values $B^c_{1,t} B^c_{2,t}$ and $B^c_{3,t}$ are found by

$$\triangle_1 = |\Phi^c_t - B^c_{1,t}|$$

$$\triangle_2 = |\Phi^c_t - B^c_{2,t}|$$

$$\triangle_3 = |\Phi^c_t - B^c_{3,t}|$$

$$\triangle = \min(\triangle_1, \triangle_2, \triangle_3) \qquad (7)$$

Background is represented by the pixels in the case $\Delta=\Delta^1$

D. Adaptive weighting factor

The quality of the virtual view depends on the values of factor ε_i,Hence it is important to discover the values of ε_i for various frames. The theory states that the condition where video has greater foreground areas in the company of rapid speed, the video must have greater count of pixels with several models in Expectation Maximization.A relation between weighting factor(ε_i) multiple models has been derived.

ε_i=f (A)

$$= \frac{A_2 + A_3}{A_1 + A_2 + A_3} \times 100$$

$$= A_{2.3} \qquad (8)$$

Here, A_1 A_2 and A_3 represent the count of pixels in model, a pair of models and three models respectively. $A_{2,3}$ represents the proportion of the count of the pixels possessing multiple models. It is stated that if there is increase in count of multiple models. The vision synthesis is also better.As the value of ε_i reaches 0.9 or above, reset is done.

The Fig. 2 indicate that, value of ε_i is near to 1 if the total of no of models reach 13% and above. When generating a virtual view it is compared with adapting weighing factor, the quality degradation isn't sacrificed when compared with maximum quality that could be achieved and the weights are set 0 to 1

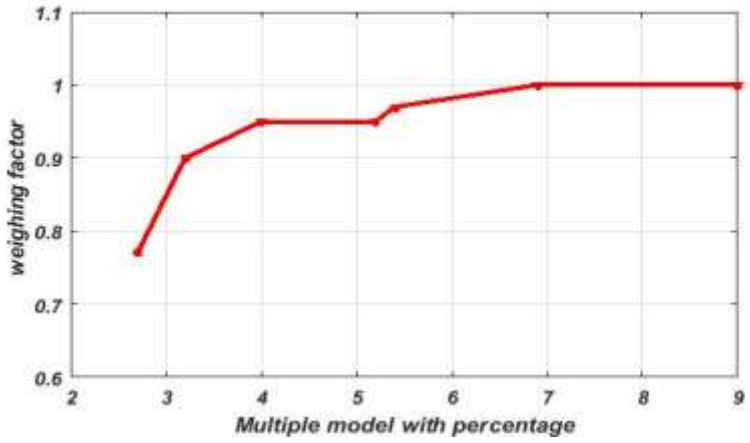

Fig. 2 Trend of weighing factor (ε_i)

III. VISION SYNTHESIS FOR MVC.

Adjoining views in a multiview video is represent using large number of cameras with slight variance in angles. Hence, disparities are present in the various views. For predicting the neighboring pixels from different samples of similar views, motion estimation method is applied. The above method is a complex and time consuming method ,Hence motion vector is an important feature of proposed research and makes the system less complex . The complexity could also be reduced by reducing the count of reference views. In Fig. 3 the beforehand encoded frames of adjoining views are used by 3 reference technique (1 and 2 reference frames). The preceding frames of present view are encoded individually [3] (reference frames). For finding the present block (X_c, Y_c) on adjoining reference views, disparity is used.

In lieu of the traditional approaches, the proposed view synthesis method is used for generalization of current frame and used as reference frame. Four references are used for picking blocks individually for encrypting the present frame of central view (Fig. 3). The current and the fourth reference frame results in better quality.

To prove the effective production of results two reference is also taken into consideration (Fig. 3) and two reference methods provides better tie for computation when compared with the three reference frame method.

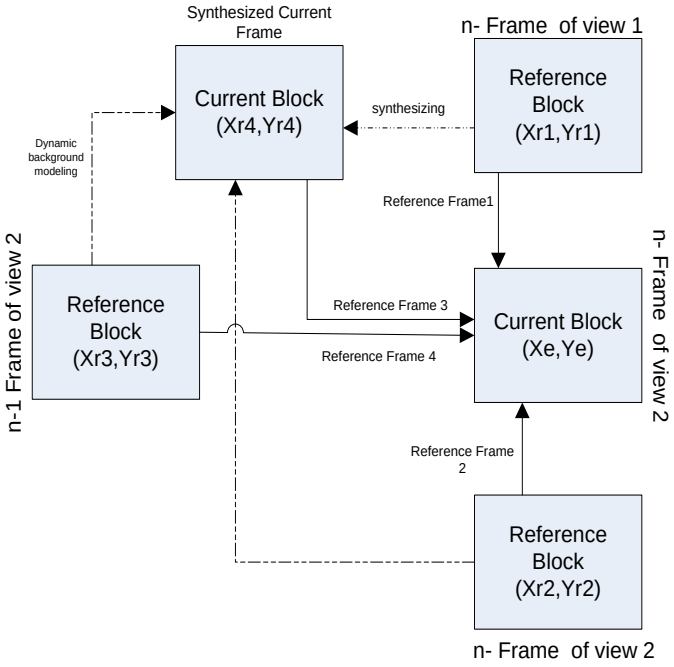

Fig. 3 MVC coding proposed method with the help of four reference such as (n-1)-th frame of view 2, n-th frame of view 1, n-th frame of view 3 and the virtual frame generated by the proposed method.

IV EXPERIMENTAL RESULTS

In our result the PSNR is used for measuring quality intensity variations of original and synthesized image pixels. Table 1 shows the input viewpoints (reference) virtual view points and the guideline of four sequences of video.

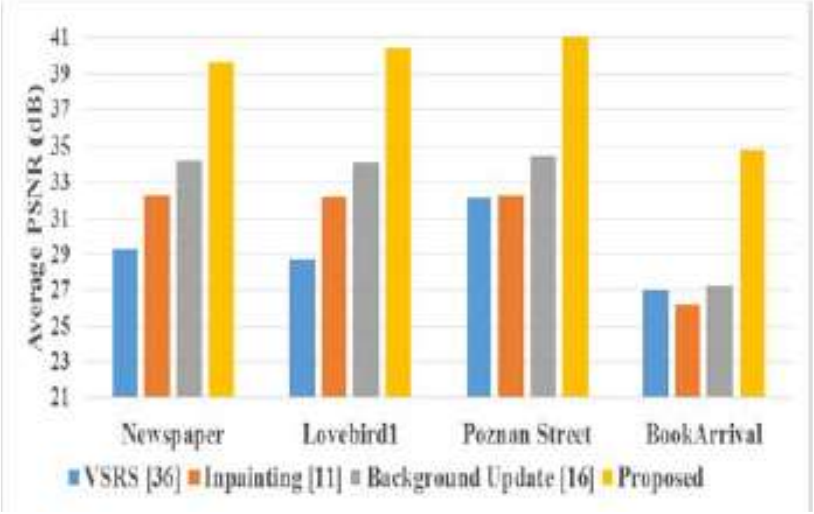

Fig. 4. Average PSNR comparison (100 frames).

TABLE I

TEST SEQUENCES, SYNTHESIZED VIEWPOINTS AND BASELINE

Sequences	Input Reference Viewpoint	Target Viewpoint	Baseline
Newspaper	6,2	*4*	*185.36*
Lovebird I	8,4	*6*	*148.13*
Poznan Street	*5,3*	*4*	*3.18*
Book Arrival	*10,6*	*8*	*2.32*

Improvement of VSRS and inpainting techniques are 9.72 and 8.25 dB respectively and are shown in Fig.4.

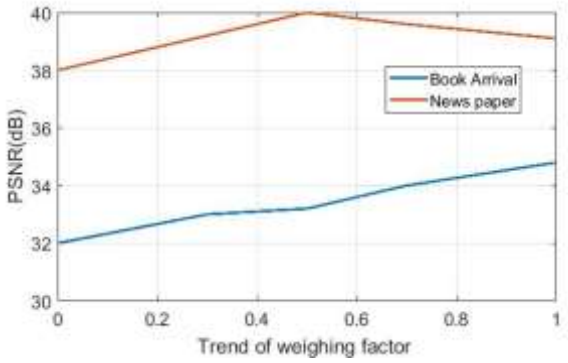

Fig. 5. PSNR Compare of the technique in Newspaper (NP) and Book Arrival (BA) video sequences.

Comparing the preliminary paper [2] to the proposed method, the results are better for our method and is shown in Fig. 5.

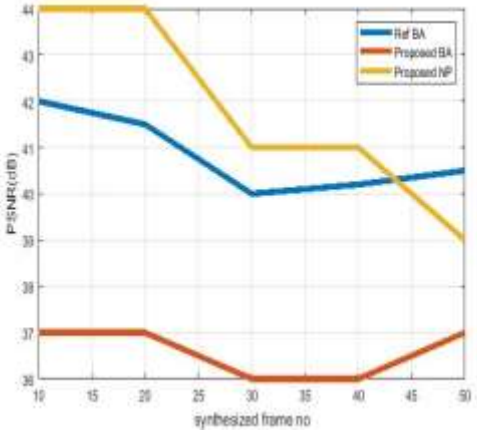

Fig 6. PSNR (dB) vs weighting factor (ε).

PSNR is analyzed for ε values for 0 to 1. It is observed

TABLEII

The Performance four And Two Reference

Sequences	Four Reference		TWO Reference	
	BD-PSNR (dB)	BD-BR(%)	BD-PSNR (dB)	BD-BR(%)
Newspaper	1.99	-32.01	1.01	-26.74
Lovebird 1	0.89	-22.45	0.82	-21.64
Poznan Street	0.68	-15.44	0.50	-12.56
Book Arrival	1.9	-39.05	1.82	-37.54
Balloons	0.95	-37.23	0.67	-34.76
Kendo	0.90	-22.87	0.73	-15.87
Poznan Hall2	0.90	-27.55	0.81	-23.10
Undo Dancer	1.52	-35.25	0.81	-23.12
Average	1.21	-28.98	0.89	-24.41

that every 30th frame of individual video sequence contains the synthesized final image

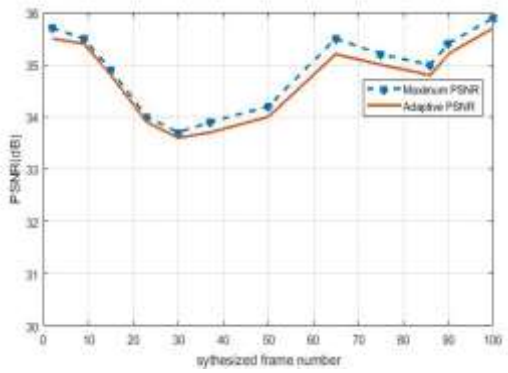

Fig 7 Maximum. PSNR (dB) vs Adaptive PSNR (ε).

The method improves 0.32 dB, 0.27 dB, 0.15 dB and 0.57dB. PSNR for book arrival, Poznan street, bird 1, newspaper respectively

Better subjective quality synthesized images are provided by the proposed method and is demonstrated in Fig. 8. subjective quality synthesized images are provided by the proposed method. The paper presented an advanced synthesis technique which efficiently uses temporal correlation in filling up the lost pixels the views are then interpolated using the adjoining images and respective depth maps. The interpolated image results in large number of hole. As spatial corresponding techniques suffer from low quality degradation because of low spatial correspondence of foreground pixels and background pixels, the presented method makes use of various models of Expectation Maximization. This helps in separating the foreground and background pixels perfectly. The lost pixels are retrieved from

3D-HEVC structure is used and HEVC is used for encoding. Table II gives the production of two reference and four reference methods.demonstrates the complexity and computational time for 3 reference and four reference MVC and synthesis virtual view.

V CONCLUSION

adaptive weighted average technique and wrapped images.The executed results show the technique proposed results in improvement of 9.72dB, 8.25dB and 6.15dB compared to other techniques. The effectiveness of the proposed system could be proved by usage of extra frame for MVC. Encoded frame's quality is improved by 0.73db as compared to traditional techniques. Inverse mapping technique [7] is compared to the proposed one and demonstrates that the proposed method results better PSNR when compared with traditional methods for 32×32 and 64×64

The method improves 0.32 dB, 0.27 dB, 0.15 dB and 0.57dB. PSNR for book arrival, Poznan street, bird 1, newspaper respectively

(a) *(b)* *(c)*

Fig8. (a) Original image, (b) inverse mapping techniques (c) proposed image

REFERENCES

[1] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[2] D. M. M. Rahaman and M. Paul, "Hole-filling for single-view plusdepth based rendering with temporal texture synthesis," in *Proc. IEEE Int. Workshop Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6, doi: 10.1109/ICMEW.2016.7574740.

[3] K. Müller *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 3366–3378, Sep. 2013.

[4] F. Zou, D. Tian, A. Vetro, H. Sun, O. C. Au, and S. Shimizu, "View synthesis prediction in the 3-D video coding extensions of AVC and HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1696–1708, Oct. 2014.

[5] G. Luo, Y. Zhu, Z. Li, and L. Zhang, "A hole filling approach based on background reconstruction for view synthesis in 3D video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1781–1789.

[6] D. M. M. Rahaman and M. Paul, "Free view-point video synthesis using Gaussian mixture modelling," in *Proc. IEEE Conf. Image Vis. Comput. New Zealand*, Nov. 2015, pp. 1–6.

[8] M. S. Farid, M. Lucenteforte, and M. Grangetto, "Depth image based rendering with inverse mapping," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process.*, Sep. 2013, pp. 135–140.

[9] C.-M. Cheng, S.-J. Lin, S.-H. Lai, and J.-C. Yang, "Improved novel view synthesis from depth image with large baseline," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[10] A. Oliveira, G. Fickel, M. Walter, and C. Jung, "Selective hole-filling for depth-image based rendering," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 1186–1190.

[11] D.-H. Li, H.-M. Hanh, and Y.-L. Liu, "Virtual view synthesis using backward depth warping algorithm," in *Proc. Picture Coding Symp.*, Dec. 2013, pp. 205–208.

[12] C. Yao, Y. Zhao, and H. Bai, "View synthesis based on background update with Gaussian mixture model," in *Proc. Pacific-Rim Conf. Multimedia*, 2012, pp. 651–660.

[13] Y. Gao, G. Cheung, T. Maugey, P. Frossard, and J. Liang, "Encoderdriven inpainting strategy in multiview video compression," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 134–149, Jan. 2016.

[14] C. Yao, Y. Zhao, J. Xiao, H. Bai, and C. Lin, "Depth map driven hole filling algorithm exploiting temporal correlation information," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 394–404, Jun. 2014.

[15] M. Paul, W. Lin, C.-T. Lau, and B. S. Lee, "A long-term reference frame for hierarchical B-picture-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1729–1742, Oct. 2014.

[16] C. Zhu and S. Li, "Depth image based view synthesis: New insights and perspectives on hole generation and filling," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 82–93, Mar. 2016.

[17] P. Pandit, A. Vetro, and Y. Chen, *Joint Multiview Video Model (JMVM) 7 Reference Software*, document N9579, MPEG of ISO/IEC JTC1/SC29/WG11, Antalya, Turkey, Jan. 2008.

[18] S. Ma, S. Wang, and W. Gao, "Low complexity adaptive view synthesis optimization in HEVC based 3D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 266–271, Jan. 2014.

[19] B. T. Oh and K.-J. Oh, "View synthesis distortion estimation for AVC- and HEVC-compatible 3-D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 1006–1015, Jun. 2014.

[20] A. I. Purica, E. G. Mora, B. Pesquet-Popescu, M. Cagnazzo, and B. Ionescu, "Multiview plus depth video coding with temporal prediction view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 2, pp. 360–374, Feb. 2016, doi: 10.1109/TCSVT.2015.2389511.

[21] P. Gao and W. Xiang, "Rate-distortion optimized mode switching for error-resilient multi-view video plus depth based 3-D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1797–1808, Nov. 2014.

[22] M. Haque, M. Murshed, and M. Paul, "Improved Gaussian mixtures for robust object detection by adaptive multi-background generation,"in *Proc. IEEE Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[23] M. Paul, "Efficient multi-view video coding using 3D motion estimation and virtual frame," *Neurocomputing*, vol. 175, pp. 544–554, Jan. 2016.

[24] *View Synthesis Reference Software 3.5*, document ISO/IEC JTC1/SC29/847 WG11 (MPEG), 2013.

[25] P. K. Podder, M. Paul1, and M. Murshed, "A novel motion classification based intermode selection strategy for HEVC performance improvement," *Neurocomputing*, vol. 173, pp. 1211–1220,Jan.2016