Multiview Synthesis – improved view synthesis for virtual navigation

Adrian Dziembowski, Adam Grzelka, Dawid Mieloch, Olgierd Stankiewicz, Krzysztof Wegner, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poland
{adziembowski, agrzelka, dmieloch, ostank, kwegner}@multimedia.edu.pl, marek.domanski@put.poznan.pl

Abstract — **In the paper, we propose a new method for the virtual view synthesis called as Multiview Synthesis. In our approach, virtual views are synthesized using two neighboring real views, but the disoccluded areas are not inpainted, but filled by the information from the further real views – additional views, traditionally not included in the view synthesis. The whole synthesis is performed with triangles rather than with individual pixels. In the proposed Multiview Synthesis, additional steps of adaptive color correction, blurred edge removal and spatial edge blurring are also included. The experimental results show that both the objective and subjective quality of the synthesized views is significantly higher than the quality of views obtained using the state-of-the-art MPEG reference view synthesis software.**

I. Introduction

Virtual navigation in 3D scenes is essential for free viewpoint television [1] and augmented reality. In order to display the views from virtual viewpoints, virtual view synthesis has to be performed [2]. The virtual view synthesis is also exploited in order to increase compression efficiency for multiview video [3-5]. The abovementioned applications are the main focus of this work.

In practical multiview systems, the number of used cameras should be limited [6], therefore distances between cameras are larger than in most experimental systems [1]. Consequently, this leads to large areas which are occluded in neighboring cameras and, as a result, the process of virtual view synthesis becomes difficult. Moreover, if cameras of the system are sparsely located around a scene, the colors of pixels in objects strongly depend on view direction, because of different illumination.

The popular method of virtual view synthesis implemented as MPEG reference software VSRS [7] is using only the information from the two nearest neighboring cameras. Such approach results in virtual navigation with notable changes of virtual view's color characteristics and lack of spatial consistency of occluded areas in different viewpoints.

Other known methods are based on extraction of background of a scene from multiple views [8] or with the use of inter-frame information in time domain [9]. Knowledge about the background of the scene allows proper synthesis of disoccluded areas. Nevertheless, some areas of a virtual view can be still not synthesized, so one of the hole filling methods has to be used [10-12,22,23]. In the aforementioned methods of virtual view synthesis different color characteristics of the acquired views are not taken into account. The known methods of multiview color correction are based on either depth information [13] or on feature extractors [14,15]. Unfortunately, the state-of-the-art methods are developed and tested almost exclusively for real cameras located on a line [e.g. 21,22,24]. Here, we propose a new method that we call Multiview Synthesis. It is capable of providing high quality free navigation from the multiview video acquired by cameras sparsely located around a scene in arbitrary positions.

II. Overview of the proposal

A. Novelty

Novelty of this proposal consists in the design of the whole processing path that results in improved quality of the final virtual views. In particular, the path includes improved backward synthesis with triangles rather than for individual pixels. Moreover, there are steps of multiview disocclusion filling, adaptive color correction blurred edges removal and spatial edge blurring (performed respectively before and after the projection of points from the real views to the virtual one).

B. Brief description

The main part of the virtual view is synthesized based on information from the nearest views to the left and to the right. Remaining, non-synthesized areas are projected from further views. Therefore, in some scenes, information from all of the real views might be used. In the mentioned step, in order to avoid discontinuities in the virtual view, the whole synthesis process is conducted with triangles instead of single pixels. Also, in order to provide better subjective quality, the pixels form the close neighborhood of spatial edges in the scene are not used for the synthesis. After merging pixels projected from different views a novel color correction is applied to eliminate color artifacts in the synthesized virtual view.

For the sake of high naturalness of the obtained virtual views, the spatial edges in the virtual view are blurred. At the end, remaining holes in the virtual view are filled up using the depth-based inpainting.

III. Proposed view synthesis

A. Triangle-based projection

In the first step of the virtual view synthesis the information from only two neighboring real views is used. This step is 'bi-directional' because of choosing the nearest left and the nearest right real views for the reference for synthesis. Such an approach result in limited (comparing to single-view based view synthesis) amount of occlusions in the synthesized view.

In the proposed method, the backward-type synthesis [16] was used. At first, the depth map of the virtual view is created – all the pixels from the real views are projected onto the virtual one, but only the depth value of each pixel in virtual view is calculated. In this operation, the color information is omitted. In the second stage of the backward synthesis, the depth map of the virtual view is used to project the color from the real views. That approach allows to remove small holes or depth artifacts in the depth maps of the virtual view because of the possibility of filtering the depth before color projection.

Typically [7,11,16], in the view synthesis all the pixels are projected from the real views to the virtual view independently. Because of finite resolution of the images, this approach causes discontinuities and thus holes (Fig. 2a) in the synthesized view (Figs. 1a and 1b). Three pixels: A, B and C, neighboring in the real view (Fig. 1a), after projection to the virtual view are dispersed (Fig. 1b).

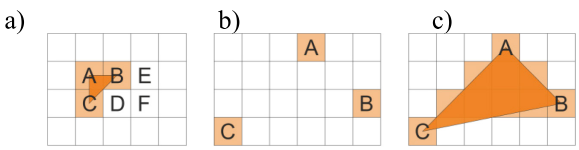

Figure 1. Projection of pixels A, B, C from real view (a) to virtual view using single pixels (b) and using triangles (c)

In the proposed method, groups of 3 pixels within overlapping triangles are projected instead of single pixels. In this approach, consecutively triangles ABC, BCD, BED, EDF, etc. are projected. At first, three vertices of each triangle are synthesized from the real views. Then, all the pixels inside projected triangle in the virtual view are estimated (Fig. 1c). Color of the pixels inside the triangle being synthesized are linearly interpolated basing on the colors and positions of the respective vertices.

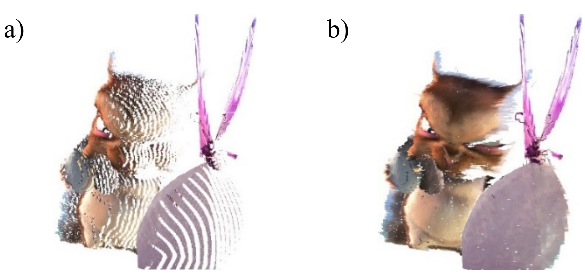

Figure 2. Projection of single pixels (left) vs. triangles (right)

In Fig. 2 shown are the result of triangle-based synthesis compared to projection of single pixels. Presented example is an extreme one because the angle between the optical axes of virtual camera and the real camera used to synthesis was almost 80° – view 6 was projected to the view 84. However, even in such a difficult case, most of the holes (areas non-synthesized in the virtual view) were filled using projection of triangles.

B. Multiview disocclusion filling

In the result of the first step of view synthesis, there are some holes, where none pixel was projected, neither from the left nor from the right real view. These non-synthesized regions represent disoccluded areas. Typically [7,10-12], these regions would be inpainted. In the proposed, Multiview Synthesis approach, non-synthesized areas are filled using information from the remaining views.

The proposed approach provides high quality increase, especially for the scenes with high dynamic range of depth, characterized by many occlusions, e.g. in Big Buck Bunny Flowers sequence [17]. The advantage of multiview disocclusion filling is presented in Fig. 3. Virtual view V is synthesized in the position between real views 0 and 1. If the information of view 2 would be omitted, the green box occluded by orange and blue spheres in the neighboring real views would not be synthesized and the area between two spheres in the view V would be inpainted using color of the background. By using the information from all of the real views, all the objects in view V will be synthesized. In the case of complicated scene (many objects, occlusions and wide range of the depth – Big Buck Bunny Flowers Arc test sequence), proposed solution provides halved percentage of non-synthesized regions in synthesized view, decreasing this area from 10% to less than 5%.

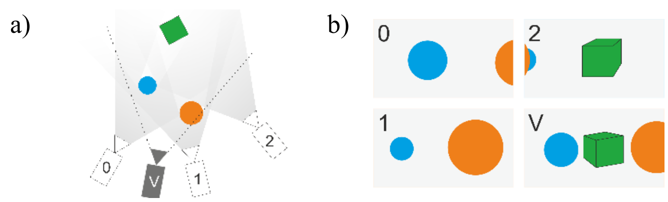

Figure 3. Motivation of multiview disocclusion filling: a) scene arrangement, b) views captured by cameras 0, 1, 2 and V

The main disadvantage of the proposed multiview disocclusion filling is the necessity of having good quality depth maps, because depth artifacts cause projecting of improper pixels from the real views to the virtual view. To allow disocclusion filling operate properly for poor quality depth maps, the confidence condition is verified. If the pixel in the virtual view was captured by less than two cameras, it is rejected as uncertain.

C. Color correction

A synthesized virtual view consists of pixels projected from different real views, which may have inconsistent colors. The first reason of that is that the sensors (Fig. 4a) of the real cameras (even the same model from the same manufacturer) are slightly different because of imperfect production. Secondly, the illumination of the same object is different in different views. The differences are more significant for the sparse distribution of the cameras around a scene, like in our scenario. Also, the light reflections are different in each view (Fig. 4b). Another, minor, but also noticeable problem is the glow effect (either natural or created in CGI postprocessing) in the real views. It increases the naturalness of real images, but creates non-existent edges in the virtual view (Fig. 4c).

The color artifacts, like the abovementioned ones, significantly reduce subjective quality of synthesized views, especially in uniform areas (smooth background, big objects, etc.). Therefore, they should be eliminated.

Various color correction techniques may be used to remove color artifacts, e.g. [13-15]. Here we propose a novel adaptive color correction method which uses so called color correction maps. Such map is used to perform color correction of pixels projected from given reference view in order to adjust them to the colors in the final virtual view. For each pixel, such map define scaling factors for given color component (RGB).

For the sake of brevity, we present color correction algorithm for a simplified case of two reference views only and single color component only. Obviously, pixels synthesized from more views in RGB color space can be corrected in a similar way, but using more color correction maps. The algorithm can be divided into 5 following steps:

1. Creation of color correction maps, separately for the left and the right real view – for each pixel projected from both the left and the right view the ratio of pixel's color from each real view to its blended color is calculated.
2. For each pixel synthesized from only one neighboring real view, its pixel neighborhood is being considered. In that neighborhood the mean value from corresponding color correction map is estimated (but only for pixels with similar color and depth to the pixel being considered).
3. The color correction map for the given pixel is updated by the calculated mean ratio.
4. Color of the given pixel is multiplied by the updated value from color correction map.
5. Steps 2 – 5 are repeated until no pixel value is changed in the iteration over the whole image.

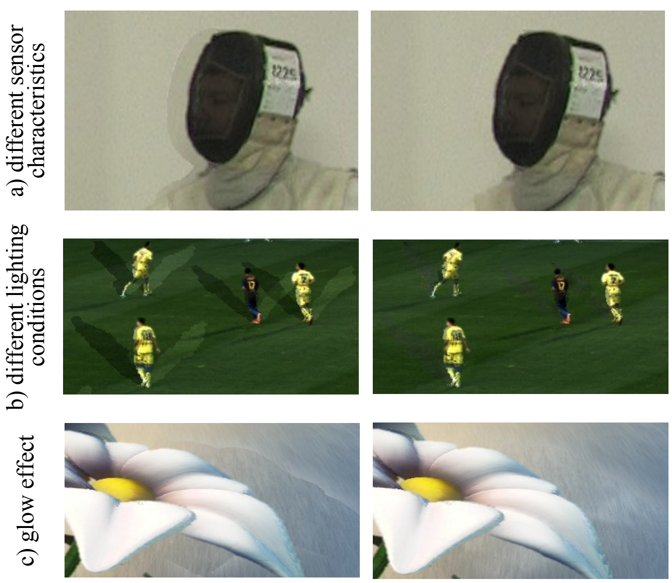

Figure 4. Different reasons of color artifacts: fragments of Poznań Fencing2, Soccer Arc and BBB Flowers before (left) and after correction (right)

D. *Blurred edges removal*

According to the principles of the optics, the edges of some objects be captured by a camera are blurred. On the other hand, the depth of these objects has to be sharp in every one pixel – it has to be determined as a pixel representing the background or the foreground object.

Both mentioned constraints imply that close to the spatial edges in the scene pixels have specified depth and blended color. Thus, in view synthesis operation, wrong (blended) colors are projected to the virtual view. It causes synthesis of non-existing artificial edges (Fig. 5a). Similar effect can be noticeable in the virtual view in the situation, when the edge of the same object in the depth map and the texture is shifted.

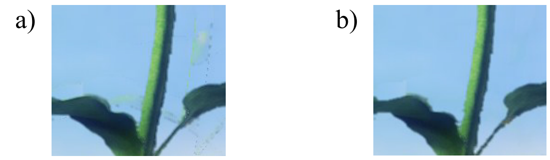

Figure 5. Fragment of Big Buck Bunny Flowers before (a) and after (b) blurred edges removal

In order to decrease these artifacts, the proposed algorithm omits all the pixels placed in the immediate neighborhood of spatial edges in the scene (areas with significant depth difference). In the virtual view, these areas are simply interpolated. Result of enabling this technique is presented in Fig. 5b.

E. *Spatial edge blurring*

Because of projecting pixels to the virtual view using sharp depth maps, the objects in synthesized view have unnaturally sharp edges (Fig. 6a). It was proposed to add slight blur to the spatial edges in the scene. Therefore, in the regions with significant depth difference between neighboring pixels, the virtual view is filtered using simple low-pass filter. Remaining areas of the view are retained unchanged. Fig. 6b contains the fragment of synthesized view after the operation of spatial edge blurring. The edge of the flower was blurred, but all the details of the flowers remained unchanged.

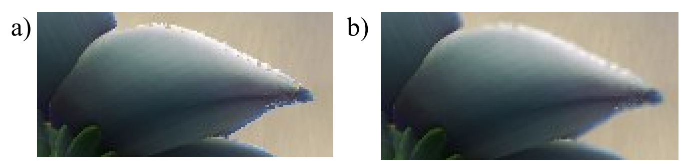

Figure 6. Fragment of Big Buck Bunny Flowers before (a) and after (b) spatial edge blurring

IV. EXPERIMENTS

A. *Experiment setup*

In the experiment, the quality of proposed algorithm of multiview synthesis (hereafter named MVS), compared to MPEG's reference software (VSRS), was evaluated. During the tests, both the subjective and objective quality were measured.

To estimate objective quality PSNR measurement was used: luminance PSNR value was calculated for pairs of real (reference) view and the virtual view synthesized at the position of the reference. Subjective quality measurement was performed on the group of 10 experts. Each viewer rated the quality of synthesized views on the scale of 0 to 10, where 10 is the perfect quality – indistinguishable from the reference.

In order to perform the experiments 6 multiview test sequences were chosen: 2 synthetic ones (Big Buck Bunny Butterfly and Flowers [17]) and 4 natural ones (Poznań Blocks2 [18], Poznań Fencing2 [18], Soccer Arc [19] and Soccer Linear [20]), acquired by real multicamera systems. Most of the sequences (5 of 6) are characterized by an arc camera arrangement. Only in Soccer Linear sequence cameras are arranged linearly.

The test sequences from Poznań contain 10 real views. 9 of them (0 – 3, 5 – 9) were used to synthesize virtual view at the position of real view 4. For that view, the objective and subjective quality of virtual view synthesis were measured.

In remaining test sequences only 6 real views were used for synthesis process. In Soccer sequences, virtual view at the position of view 4 was synthesized using views 1 – 3 and 5 – 7, in both BBB sequences views 6, 19, 32, 58, 71 and 84 were synthesized to view 45.

B. Results

Objective and subjective quality of virtual views synthesized using proposed method and reference software were estimated. The results of performed experiments are presented in Table I.

TABLE I. OBJECTIVE AND SUBJECTIVE QUALITY OF VIRTUAL VIEW SYNTHESIS USING VSRS AND MVS (PROPOSED)

Sequence name	Virtual view quality			
	Objective [dB]		Subjective (scale: 0 – 10)	
	VSRS	MVS (proposed)	VSRS	MVS (proposed)
BBB* Butterfly	32.04	33.56	5.48	7.48
BBB* Flowers	22.68	25.81	4.23	7.30
Poznań Blocks2	29.25	29.61	7.00	7.50
Poznań Fencing2	28.33	28.82	6.40	7.75
Soccer Arc	22.79	23.48	3.15	4.33
Soccer Linear	35.00	34.98	8.50	8.38
Average	**28.35**	**29.38**	**5.79**	**7.12**

* – Big Buck Bunny

At first, let us consider the objective quality of synthesis, presented also in Fig. 7. The average quality gain of using the proposed MVS over VSRS algorithm is slightly greater than 1 dB. The quality of views synthesized using MVS is higher for 5 of 6 test sequences. Only for Soccer Linear there is an unnoticeable quality decrease. It is the consequence of the simplicity of Soccer Linear sequence: linear camera arrangement, few occlusions, similar color characteristics of each view, small objects (players) on smooth area. Other sequences were more difficult to synthesize, what caused quality increase while using proposed method.

It should be noted that the greatest quality increase was achieved for synthetic test sequences – the ones with perfect (synthetic) depth maps. For BBB Butterfly and Flowers the average gain of using the proposed MVS is 2.3 dB, for 3 natural arc sequences (Soccer Linear is omitted) it is only 0.5 dB. It shows that proposed techniques function the better, the higher is the quality of the depth maps.

Results of subjective quality measurement from Table I (including confidence intervals) are presented in Fig. 8. On average, proposed method is more than 12% better than the

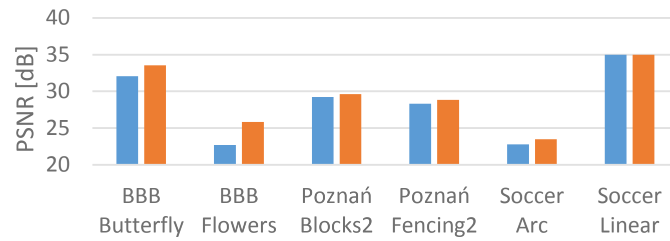

Figure 7. Comparison of objective quality of VSRS (blue) and the proposed MVS (orange) measured as PSNR in dB

reference software (gain of 1.33 in an 11-step scale). Just like for objective quality, MVS is a better choice for 5 of 6 test sequences – the only exception is, again, Soccer Linear.

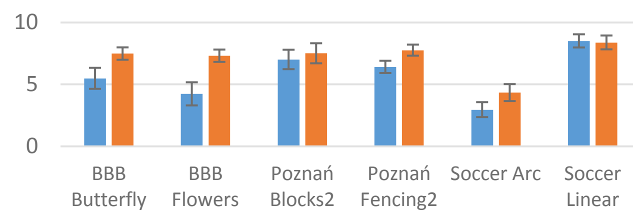

Figure 8. Comparison of subjective quality of VSRS (blue) and the proposed MVS (orange) in 11-step scale

For 4 of 6 test sequences the gain of using the proposed MVS was statistically significant (significance level: 5%). All of these sequences are not easy to synthesize, even while using perfect depth maps. BBB Flowers and Butterfly (especially the first one) are characterized by many occlusions, what allows the technique of multiview disocclusion filling operate in relatively large areas. The grass in Soccer Arc, mowed in different directions, has anisotropic characteristics, what causes the necessity of color correction. The same operation allows to increase the quality for Poznań Fencing2 – sequence with uniform white background, where even slight difference in characteristics of camera sensors is noticeable for a viewer.

Quality difference between view synthesized using MVS and VSRS for Poznań Blocks2 was also noticeable (0.5 on average), but not significant statistically.

Fig. 9 contains fragments of all used test sequences. In the bottom row results obtained using proposed method are shown. Above, the same fragments synthesized using reference method – VSRS. To allow the quality assessment, at the top row the fragments from reference views are presented.

V. CONCLUSIONS

In the paper a novel method of virtual view synthesis was presented. Performed experiments show that both the objective and subjective quality of views synthesized using proposed method are better than using MPEG's reference software (VSRS) [7]. The attained objective PSNR quality is 1 dB higher. The attained subjective gain is almost 1.5 in 11-step scale, compared to using VSRS.

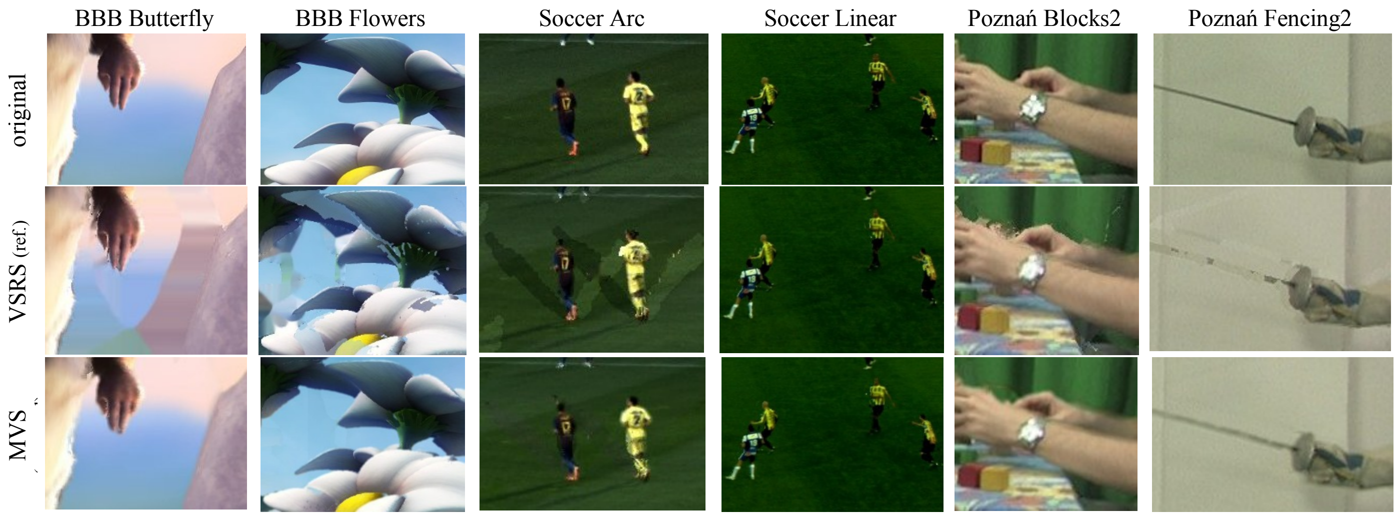

Figure 9. Fragments of original sequences (top row) compared to fragments synthesized using VSRS (middle) and MVS (bottom).

Described steps of multiview disocclusion filling, color correction, blurred edges removal and spatial edge blurring allow to achieve acceptable subjective quality for free navigation purposes (above 7 in 11-step scale). Therefore, presented method may be applied in FTV systems [1].

The second application of multiview synthesis is video compression. Better synthesis algorithm allows to predict interview information with more accuracy, what increases the compression efficiency. To reduce multiview bitstreams, the synthesis algorithm used in 3D-encoders [3-5] can be replaced by the proposed one, what will be a subject of our future works.

ACKNOWLEDGEMENT

Research project was supported by The National Centre for Research and Development, Poland. Project no. TANGO1/266710/NCBR/2015.

REFERENCES

[1] M. Tanimoto, M. P. Tehrani, T. Fujii, T. Yendo "FTV for 3-D spatial communication", Proc. IEEE, vol. 100, no. 4, pp. 905-917, 2012.

[2] M. Domański, M. Gotfryd, K. Wegner, "View synthesis for multiview video transmission", The Int. Conf. Image Proc., Computer Vision, and Pattern Recognition, 2009.

[3] M. Domański, O. Stankiewicz; K. Wegner; M. Kurc; J. Konieczny; J. Siast; J. Stankowski; R.t Ratajczak; T. Grajek., "High Efficiency 3D Video Coding Using New Tools Based on View Synthesis", IEEE Trans. Image Proc., vol. 22, 2013, pp. 3517-3527.

[4] K. Müller et al., "3D High-Efficiency Video Coding for multi-view video and depth data", IEEE Trans. Image Proc., vol. 22, 2013, pp. 3366-3378.

[5] J. Stankowski et al., "3D-HEVC extension for circular camera arrangements", The 3DTV Conference, 2015.

[6] M. Domański et al., "A Practical Approach to Acquisition and Processing of Free Viewpoint Video", Picture Coding Symposium, 2015.

[7] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domański, "Enhanced view synthesis reference software (VSRS) for Free-viewpoint Television", ISO/IEC JTC 1/SC 29/WG 11, MPEG M31520, 2013.

[8] G. Luo, Y. Zhu, "Foreground Removal Approach for Hole Filling in 3D Video and FVV Synthesis", IEEE Trans. Circuits Syst. Video Techn., 2016, IEEE Early Access Articles.

[9] H. Liu, X. Li, K. Chen, "Global-background based view synthesis approach for multi-view video", The 3DTV Conference, 2012.

[10] K. J. Oh, S. Yea, Y. S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-D video", Picture Coding Symposium, 2009.

[11] I. Ahn, C. Kim, "A Novel Depth-Based Virtual View Synthesis Method for Free Viewpoint Video", IEEE Trans. Broadcasting, Vol. 59, pp. 614-626, 2013.

[12] C. Zhu, S. Li, "Depth Image Based View Synthesis: New Insights and Perspectives on Hole Generation and Filling", IEEE Trans. Broadcasting, Vol. 62, pp. 82-93, 2015.

[13] F. Shao, G. Jiang, M. Yu, "New Color Correction Method of Multi-view Images for View Rendering in Free-viewpoint Television", WSEAS Trans. Computers, Vol. 7, pp. 569-578, 2008.

[14] S. A. Fezza, M. C. Larabi, K. M. Faraoun, "Feature-Based Color Correction of Multiview Video for Coding and Rendering Enhancement", IEEE Trans. Circuits Syst. Video Techn., Vol. 24, pp. 1486-1498, 2014.

[15] K. Yamamoto, M. Kitahira, H. Kimata, T. Yendo, "Multiview Video Coding Using View Interpolation and Color Correction", IEEE Trans. Circuits Syst. Video Techn., vol. 17, 2007, pp. 1436-1449.

[16] L. Du-Hsiu, H. Hsueh-Ming, L. Yu-Lun, "Virtual view synthesis using backward depth warping algorithm", Picture Coding Symposium, 2013.

[17] P. Kovacs, "[FTV AHG] Big Buck Bunny light-field test sequences". ISO/IEC JTC1/SC29/WG11, Doc. MPEG M35721, Feb. 2015.

[18] M. Domański, et al., "Multiview test video sequences for free navigation exploration obtained using pairs of cameras", ISO/IEC JTC1/SC29/WG11, Doc. MPEG M38247, May 2016.

[19] P. Goorts, S. Maesen, M. Dumont, S. Rogmans, P. Bekaert, "Free viewpoint video for soccer using histogram-based validity maps in plane sweeping", Int. Conf. Computer Vision Theory Appl., 2014.

[20] P. Goorts, M. Dumont, S. Rogmans, P. Bekaert, "An end-to-end system for free viewpoint video for smooth camera transitions", International Conf. 3D Imaging, 2012.

[21] W. Sun, O. Au, L. Xu, Y. Li, W. Hu, "Seamless view synthesis through texture optimization", IEEE Trans. Image Proc., vol. 23, pp. 342-355, 2014.

[22] T. Tezuka, M. Tehrani, K. Tkahashi, T. Fuji, "View synthesis using superpixel based inpainting capable of occlusion handling aand hole filling", Picture Coding Symposium, pp. 124-128, 2015.

[23] M. Köppel, K. Müller, T. Wiegand, "Filling disocclusions in extrapolated virtual views using hybrid texture synthesis", IEEE Trans. Broadcasting, vol. 62, 2016, pp. 457-469.

[24] L. Toni, G. Cheung, P. Frossard, "In-network view synthesis for interactive multiview video systems", IEEE Trans. Multimedia, vol. 18, 2016, pp. 852-864.