An Improved View Synthesis of Light Field Images for Supporting 6 Degrees-of-Freedom

Sangwoon Kwak[1], Joungil Yun[1], Won-Sik Cheong[1], Jeongil Seo[1]

[1]Electronics and Telecommunications Research Institute, Korea, 34129
Keywords: View Synthesis, Light Field, Virtual Reality, 6DoF

ABSTRACT

In this paper, virtual view synthesis of sparse light field images is considered. We analyze the patch-wise 3D warping and blending methods of the conventional view synthesis, and propose an improved algorithm for supporting 6DoF. We suggest an enhancement for the super-pixel and additional blending weights, and present experimental results using multi-view contents of MPEG.

1 INTRODUCTION

The demand for immersive media contents that can provide high immersion feeling to users such as VR (Virtual Reality) is rapidly increasing, and studies on related core technologies is actively proceeding. In particular, based on LF (Light Field)[1], many researches make efforts to create 6 DoF (Degrees of Freedom) experience which includes translational motion in all directions, beyond the traditional 3 DoF which can only support rotational motion in a fixed position[2,3]. LF is a field of expressing intensities and directions of lights reflected from objects in a 3D space, and can be acquired by a rig of many cameras such as vertical/horizontal array structure or spherical omnidirectional structure. LF-based immersive media supporting 6 DoF can provide to users with high immersion and presence feeling, but it is difficult to directly acquire all viewport images in case of naturally captured images instead of computer graphics. Therefore, it is necessary to study the virtual view synthesis which can provide viewport images of arbitrary user's positions using only a limited number of acquired input LF images.

In this paper, we analyze the 3D warping and blending methods of virtual view synthesis. In order to improve the prior method and provide better 6 DoF experience, we propose the following improvements: adaptive super-pixel, weighted super-pixel and depth distribution-based blending weights. Finally, we present experimental results using test materials of MPEG(Moving Picture Experts Group)[4], and compare the quality of synthesized views to that of the reference view synthesizers of MPEG[5,6].

2 BACKGROUND

2.1 Virtual View Synthesis

Conventional virtual view synthesizing, generating an intermediate view of virtual position from acquired input views, can be operated by two common methods which are based on 3D warping and disparity. In the formal method, each pixel of input images in un-projected to 3D space and then projected to the virtual image coordinate using depth information and camera parameters, while pixels of input images are directly shifted according to their disparity information in the latter method. Although disparity-based method can synthesis virtual view rapidly and simply without camera parameters, but it cannot fully reflect the geometry information of the scene, so 3D warping-based methods[2,3,7] are mainly used for synthesizing views in arbitrary viewport positions.

The general process of the 3D warping-based view synthesis is shown in Fig. 1. Using the camera parameters, the input depth images are firstly warped to the virtual image plane in forward direction, and then median filtering is performed to simply fill a small crack-like hole that can occur when mapping pixels to the integer coordinate system. Since the values of depth pixels are relatively simple and have high continuity compared to that of the color pixels, it is possible to fill small cracks by simple filtering process. Using the warped depth maps partially refined by the filtering, the texture values are then obtained by backward warping to the input images. The backward warped texture images can be synthesized into one virtual view image in blending process by weighted average of the images, where the weights can include camera baselines or the depth value of each pixel. The post-processing is a process of improving the blended image, which mainly includes inpainting techniques for filling common hole by occlusion. Since the synthesis process described above is based on depth information, it is also called DIBR (Depth Image-Based Rendering).

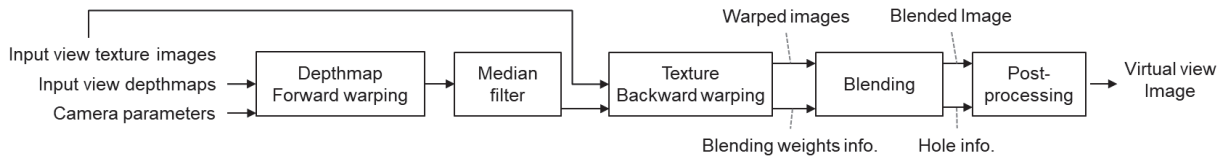


Fig. 1 Process of the virtual view synthesis using 3D warping

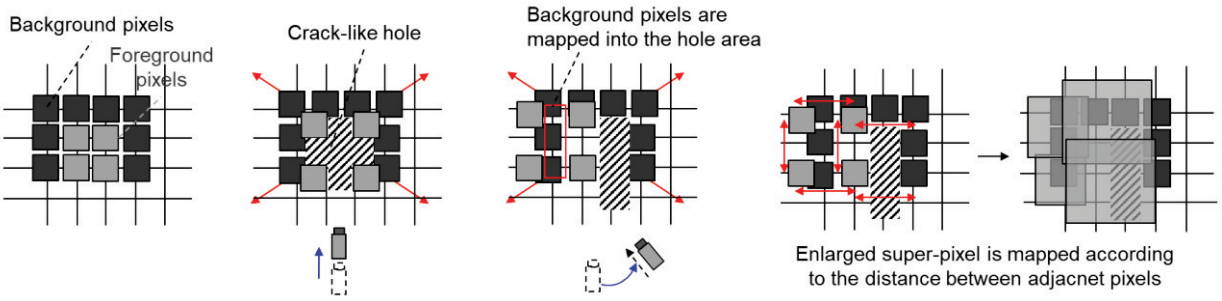


Fig. 2 The limitation of the pixel-wise warping and a schematic of the super-pixel method

2.2 Patch-wise Warping

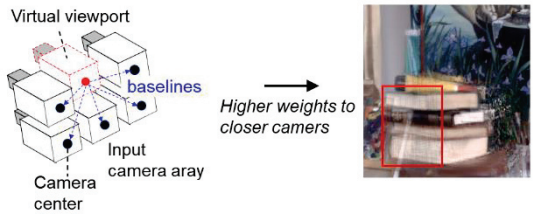
In the traditional stereoscopic, view synthesis algorithms are developed to generate only a simple intermediate view from left and right views. But recently, the demand for generating anonymous viewport images supporting 6 DoF from multi-view images such as light field is significantly increased, and the pixel-wise 3D warping addressed in 2.1 have limitations to sufficiently provide 6 DoF viewport to users. Fig.2 shows a problem that can occur in synthesizing 6 DoF viewport image which includes both rotational and translational motion. Assume that we have input depthmap as shown in Fig. 2 (a), where the grey pixels represent foreground and the black pixels represent background. When the virtual viewport moves to only forward direction as shown in Fig.2 (b), there occurs crack-like hole because the depth pixels of foreground become distant from each other and mapped into the virtual image plane which is the integer coordinate. These holes can be partially corrected by filtering or post-processing. However, in a case of 6 DoF motion, the background depth pixels can be mapped into the gap between foreground pixels generated by the spreading of them as shown in Fig.2 (c). Since this region is not treated as hole area, so it is left in the final synthesized view and causes unexpected quality reduction.

One of prior works to solve this problem is the super-pixel method[8]. It is a technique used in VSRS (View Synthesis Reference Software)[5] of MPEG, which maps enlarged pixels to the pixel position according to the interval of adjacent pixels in the warped depthmap. As shown in Fig. 2 (d), the pixels are expanded and mapped in a square shape having the length of each side as the maximum distance that the warped pixels are vertically and horizontally distant. However, since the conventional super-pixel method does not consider the depth values of the pixels and maps enlarged pixels only based on the distances in the virtual image, artifacts can occur when the depth values are largely different. If adjacent pixels at boundary region of an object are greatly distant from each other due to their depth difference, the corresponding region is not a gap should be filled with super-pixel but should be left in hole. However, as shown in Fig. 2 (d), excessively oversized super-pixel can be mapped into the region and generate artifacts. If the artifacts exist between the boundary region of warped images, blur may occur in the final blended image.

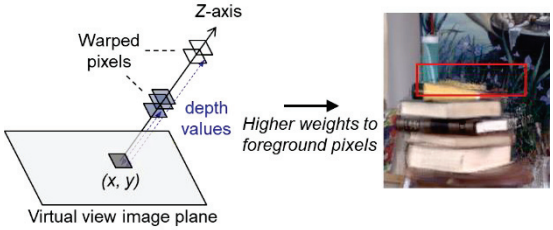
Another approach with similar purposes is triangle-based warping[2] which is used in RVS(Reference View Synthesizer)[6] of MPEG. This method composes triangles of three adjacent pixels in input images and fills out the gap between distant pixels in warped images by the triangles. It is useful to fill cracks smoothly using tri-linear interpolation of three adjacent pixels and reflects better the geometry information than super-pixel method. But it also has the similar problem, excessively stretching of triangles, when the depth difference is greatly large at boundary region.

2.3 Blending

For the blending method of warped images, weighted average is commonly used, and two weights considered in conventional blending are baseline between camera positions and depth values of warped pixels. In the formal, weights are inversely proportional to each baseline, so the pixels warped from the closer camera have higher weights. It can reduce artifacts due to the incorrectly warped objects from far distant cameras, but background pixels can be seen transparently as shown in Fig. 3 (a) if we do not consider the depth. In the depth value-based blending, higher weights are assigned to



(a) Blending weight based on baseline



(b) Blending weight based on depth values

Fig. 3 Conventional blending weights

the pixels of less depth values, so that foreground objects have the higher priority, but artifacts can occur due to incorrect warping. Of course, there can be a lot of another blending weights and they have this sort of pros and cons, so it can be considered that the weighted average of the blending weights to optimize the quality of blended image.

3 PROPOSED METHOD

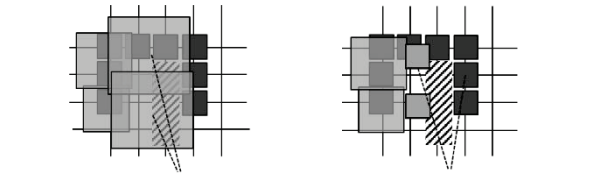
3.1 Adaptive Super-Pixel

In order to avoid the artifacts caused by the oversized super-pixel, we propose to skip mapping of super-pixel when the depth difference between neighboring pixels is larger than a certain threshold. As shown in Fig. 4 (a), we can remove excessively enlarged super-pixel by skipping when depth values differ greatly. Then, artifacts which were in warped image can be reduced and remained as hole area, while the desired effect of super-pixel inside the objects still maintains. Consequently, in blended image, blurring effect that cause quality reduction can be improved as seen in Fig. 4 (b).

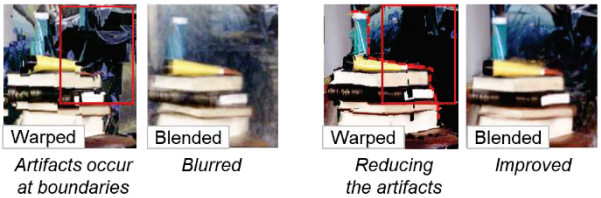
We also consider to apply a rectangle shaped super-pixel which is not fixed in square but varies by the distance between pixels. it allows super-pixel to have different shape depending on shape of objects, so that boundary noise can be suppressed.

3.2 Weighted Super-Pixel

For each pixel location in virtual image, the sizes of the super-pixels warped from input views can be different from each other. Although super-pixel is a useful complement for the problem of pixel-wise warping, but anyway it is approximately enlarged using information from adjacent pixels, so it is more inaccurate than a pixel which is directly warped from input view. In other words, for a specific pixel location in virtual image, if there are pixels that are directly warped from input view and approximately painted by super-pixel, it is more reasonable to trust the information of the former pixel. On this point, we propose to apply additional blending weight calculated by the size of pixels.



(a) Concepts (L: prior, R: proposed)



(b) Results (L: prior, R: proposed)

Fig. 4 Comparison of prior and proposed super-pixel

As shown in Fig.5 (a), for a pixel location (x, y), relatively larger pixels can be blended by lower weights than smaller pixels. Using this additional weight, it is possible to minimize the quality reduction of blended image by the approximately enlarged super-pixels.

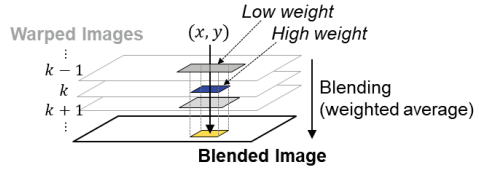
3.3 Blending Weight based on Depth Distribution

When several pixels which are warped to the same pixel position are gathered to similar depth values, it can be said that the pixels are more likely to be accurate than the other since they have mutual trust. On this point, we try to reflect the coincidence among depth values of warped pixels to the blending process, by applying depth distribution-based blending weights.

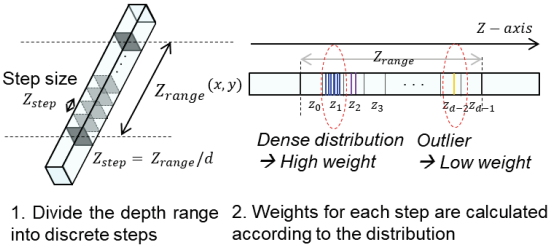
As illustrated in Fig. 5 (b), we first divide the whole depth range into a certain number of discrete steps, which are a sort of rooms to collect pixels having similar depth values. We then count how many warped pixels are in each step, and the weights can be obtained as follows:

$$w(x, y, z_i) = \frac{n(x, y, z_i)}{\sum_{i=0}^{d} n(x, y, z_i)},$$

where (x, y) is pixel position, z_i is the *i*-th depth step, and $n(x, y, z_i)$ is the number of pixels that are included in the step. In result, higher weights can be assigned to steps with larger number of pixels, which allows to reflect the distribution of depth in blended image, so that their information can be reinforced. On the contrary, allocating a lower weight to a step with small number of pixels can suppress or even erase outliers which are causes of quality reduction.

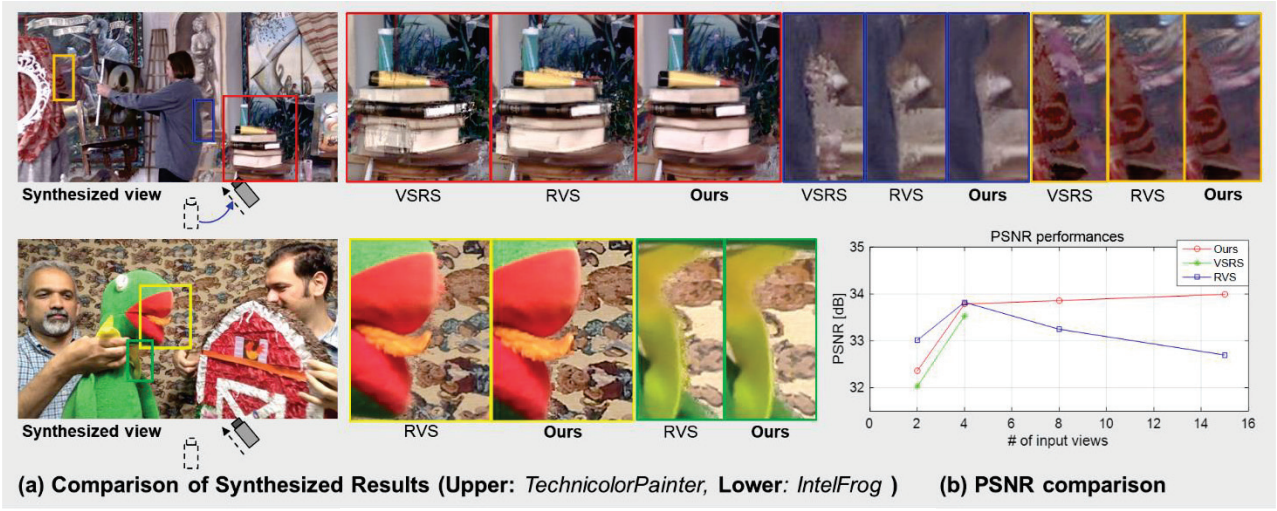


(a) Blending weight based on the size of super-pixel



(b) Blending weight based on depth distribution

Fig. 5 Proposed blending weights



(a) Comparison of Synthesized Results (Upper: *TechnicolorPainter,* **Lower**: *IntelFrog* **)** **(b) PSNR comparison**

Fig. 6 Experimental results

4 EXPERIMENTS AND DISCUSSION

In the experiments, two contents are used as input images which are test materials of MPEG[4]: *Technicolor Painter* and *IntelFrog*. *TechnicolorPainter* is a 4x4 arrayed sparse LF images with 7cm spacing between cameras, and *IntelFrog* is 13x1 multi-view images with 3.6cm spacing. Using the images, we compared the quality of virtual views with 6 DoF motion, synthesizing the views by our algorithm, VSRS[5] and RVS[6], respectively. Also, in order to compare objective performance, we synthesized a virtual view at the central input view's position and calculated PSNR (Peak Signal-to-Noise Ratio), changing the number of input images.

In Fig.6 (a), the comparisons of synthesized results are presented. Our method can improve the subjective quality of virtual views compared to the reference synthesizers, and especially, artifacts which were occurred near the boundaries of objects can be significantly reduced in the experiment of *TechnicolorPainter*. The PSNR performances are presented in Fig.6 (b), which are calculated at the view 5(central view) position, varying the number of input views: 2,4,8, and 15. Note that VSRS can only support 4 input views at most. In the figure, we can see that our method outperforms VSRS and has a crossing point with RVS. Since the triangle-based warping can smoothly fill out the gap between pixels by interpolation, so RVS performs better when using only two input views. However, when the number of input views increases, the quality is degraded rather than improved due to the blending with inaccurately warped pixels. Our algorithm solves this problem by depth distribution-based blending. It helps to reinforce the coincidence of depth values and suppress the outliers, so the result can be robust to incorrect warping. Therefore, the performance of our algorithm improved as the number of input views increases, and outperform RVS when the number of input views is larger than 4.

5 CONCLUSIONS

In this paper, we analyzed the virtual view synthesis of sparse LF images and proposed enhancements to improve the quality of synthesized view. By the experiments, we verified that the proposed algorithm can outperform the reference synthesizers of MPEG in terms of subjective and objective quality.

ACKNOWLEDGEMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. 2017-0-00072, Development of Audio/Video Coding and Light Field Media Fundamental Technologies for Ultra Realistic Tera-media)

REFERENCES

[1] G. Lee, E. Lee, W. Cheong, N. Hur, "Trend of Light-Field Image Acquisition and Representation Technology," Electronic and Telecommunications Trends, Vol. 31, June 2016.

[2] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, M. Domanski, "Multiview Synthesis – Improved view synthesis for virtual navigation," 2016 Picture Coding Symposium (PCS), Dec. 2016.

[3] S. Fachada, D. Bonatto, A. Schenkel, G. Lafruit, "Depth Image Based View Synthesis with Multiple Reference Views for Virtual Reality," 3DTV-CON 2018. June 2018.

[4] ISO/IEC JTC1/SC29/WG11, *Common Test Conditions for Immersive Video*, N18443, May. 2019.

[5] ISO/IEC JTC1/SC29/WG11, *Proposed View Synthesis Reference Software (pVSRS4.3)*, M44031, Oct. 2018.

[6] ISO/IEC JTC1/SC29/WG11, *Reference View Synthesizer (RVS) Manual*, W18068, Oct. 2018.

[7] D. Li, H. Hang, Y. Liu, "Virtual View Synthesis Using Backward Depth Warping Algorithm," 2013 Picture Coding Symposium (PCS), Dec. 2013.

[8] T. Tezuka, M. Tehrani, K. Suzuki, K. Takahashi, O. Fujii, "View synthesis using superpixel based inpainting capable of occlusion handling and hole filling," 2015 Picture Coding Symposium (PCS), June 2015.