Object-centered View Synthesis using Learning-based Image Inpainting

Hong-Chang Shin, Gwangsoon Lee, Ho min Eum, and Jeong-il Seo

Electronics and Telecommunications Research Institute, Republic of Korea
Keywords: HMD, mobile, motion parallax, view synthesis, image inpainting.

ABSTRACT

This paper presents an object-centered view synthesis technique using multilayer concept. we divide the image into multiple layers based on depth information and then provide different motion parallaxes for each layer depending on the depth. When the disocclusion region appears due to motion parallax, the uncovered region is filled by using a learning-based image inpainting.

1 INTRODUCTION

Mobile devices have developed so rapidly in the past decade. But there are still many practical problems to solve. The main problem is that mobile devices have fewer resources than desktop devices so that multiview content creation and conversion of live-action video in a mobile environment still remain a challenge. It is hard to deal with multiple high-resolution images for practical applications such as a real-time multiview rendering system, mobile-based virtual reality system and so on.

Recently, the development of VR (Virtual Reality) technology has been accelerated by the spread of HMD [1]. VR devises releases to data have been studied in various aspects such as user's movement recognition, display resolution, and user interface. One of the problems to be solved is that it should be able to support high degrees of freedom. Since the user can see the omnidirectional view image through the HMD, it basically supports 3 degrees of freedom (3DoF) corresponding to three directions of up, down, left, and right rotations around the user. It can support up to 6 degrees of freedom by supporting additional 3 degrees of freedom (3DoF), which are the up and down, the left and right, and the forward and backward movements.

In this paper, we present an object-centered view synthesis technique that can increase the degree of freedom by supporting limited translational motion by providing motion parallax to scenes in the existing 3DoF devices.

2 ALGORITHM

The existing 3DoF only supports rotational movement in three directions about the user. To support translation, it must be able to provide motion parallax. When the user moves left and right while looking at an object, the movement of the object is changed due to the distance from the user to the object. This is called motion parallax.

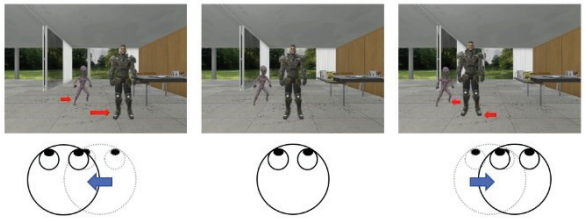

Fig. 1 Motion parallax due to user's movement

Fig. 1 shows motion parallax when the user's head moves from side to side. When the user moves, the uncovered region behind the object is revealed. The revealed region has to be filled with the pixel information on other viewpoint images during view blending phase. After that, there are still remained holes. The holes are not visible at any viewpoint images. The holes are filled with pixels generated by image inpainting technique. These view synthesis techniques are very time-consuming and require many data, multiple views, and depth information.

This paper presents a view synthesis technique using a multilayer concept. To provide motion parallax in the scene, the object of interest has to be divided. To do this, we divide the image into multiple layers based on depth information and then provide different motion parallax for each layer depending on the depth.

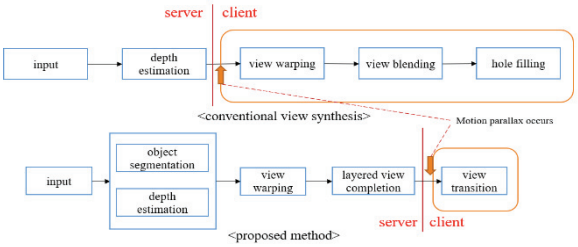

Fig. 2 Comparison with conventional view synthesis process

Fig. 2 shows the flowchart of the method we propose. The proposed method is different from conventional view synthesis technique. Considering server-client structure. the server transmits the layered images and the depth information generated to the client. The layered images can be generated through the layered view generation process such as scene layering, depth estimation, and image inpainting. At the client, it could be shown that

motion parallax can be provided through the proposed view synthesis technique having a low complexity.

2.1 Multi-layered view generation

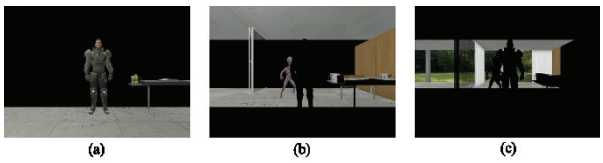

Fig. 3 Multilayering the scene

Fig. 3 shows the result of multilayering the scene using depth. The number of layers is 3. In Fig. 3, (b) and (c) include the region occluded by (a). The occluded region can be filled with the pixels that are visible at other viewpoint images by view warping.

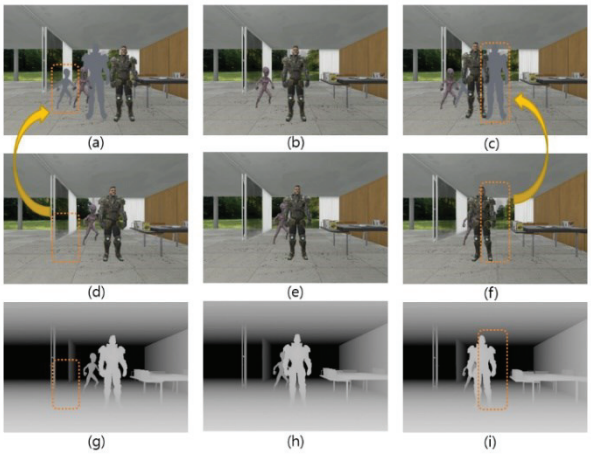

Fig. 4 Filing occluded region through view warping

In Fig. 4, (d) to (f) are images captured by multi-view cameras. In (a) and (c), each grey-colored region is an uncovered region when the object moves sideways. The uncovered region is visible from (d), which is one of the reference viewpoint camera around the center camera (b) so that it is possible to fill the region. Likewise, the grey-colored region of (c) can be filled using (f).

2.2 Image inpainting

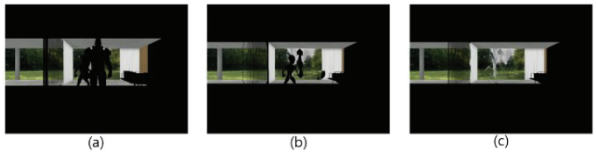

Fig. 5 Filing remained hole through learning-based image inpainting

After the previous process, there are still remained a region. Fig. 5(b) shows the result of the hole filling through view warping. Fig. 5(b) still has holes. The remained uncovered region is filled by image inpainting phase. Fig. 5(c) shows the result of image inpainting technique[3].

Image inpainting techniques have been studied to reconstruct missing parts of images and videos. Recently, Generative Adversarial Networks(GAN)-based image inpainting technique have shown good result. In this paper, we used generative image inpainting [3]. Basically, this technique has an encoder-decoder structure. First, it generates an interpolated region and then refine it by passing them through the encoder-decoder network. In the refinement module, it is possible to generate accurate and clean images through the image generation by referring to the adjacent patches surrounding the missing region.

3 EXPERIMANTAL RESULTS

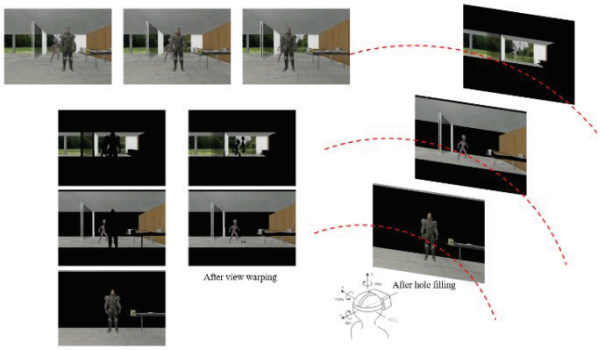

Fig. 6 Providing motion parallax through object-centered view synthesis

Fig. 6 shows the concept of providing motion parallax through object-centered view synthesis. Scene images acquired through Multiview camera is divided into 3 layers by using depth information. In each layer, the occluded region covered by the foreground objects is firstly filled through view warping with reference view image and depth information. After that, the hole region is still remained. Since the region is an invisible area in the given reference camera, it is filled using image inpainting. When all layered images are fully generated, motion parallax due to user's movement can be provided by moving the layered images.

In this paper, the proposed method is able to provide motion parallax with only multi-layered images. The scene having continuous depth is divided into several discrete depth layers. For this reason, depth discontinuity phenomenon occurred in a region where having continuous depth such as the floor. In addition, when the object having almost no continuous depth value and the floor having continuous depth were divided into different layered images, the unrealistic floating-object effect was found. However, this problem can be fixed by view warping based on the depth corresponding to the background.

The reason why we propose this method is to provide

motion parallax technique that has low complexity in display devices such as HMD. This method considers a server/client system that divides the scene into several layers, fills the occluded region corresponding to the background, and then transmits them with depth information. In a client, only the view warping is performed based on the depth information by considering the user's movement to provide motion parallax.

One of the techniques that greatly affect the quality of the results of the proposed technique is the image inpainting technique that fills the hole region.

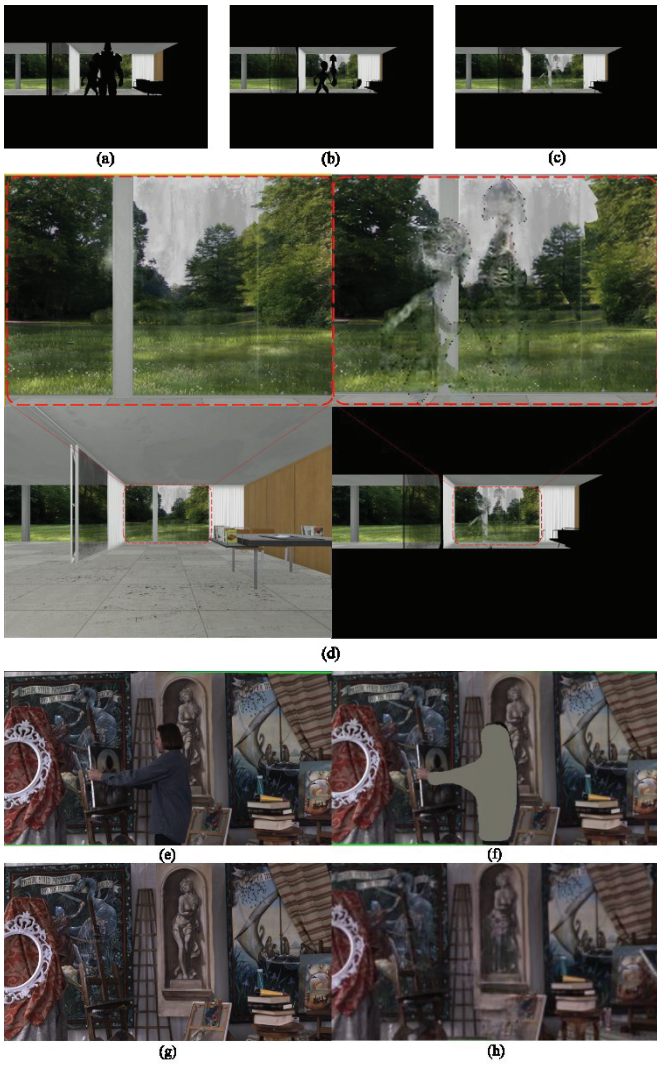

Fig. 7 results of hole filling

Fig. 7 shows the result of the hole filling using generative image inpainting [3]. Fig. 7(d) is an enlarged image of (c). It also shows the comparison result with the ground truth view. In Fig. 7, (e)-(h) also shows the result of hole filling: (e) is the source image with an object. (f) is the image that the object is removed. (h) is the result of image inpainting of (f). It can be seen that a reasonable result is obtained comparing with the ground truth (g).

4 CONCLUSIONS

In this paper, we experimented that we can increase the omnidirectional freedom by providing motion parallax generated by the user's movement with object-centered view synthesis in limited degrees of freedom that support only 3 degrees of freedom in the existing VR environment. As a result, a phenomenon in which depth is discretized by dividing a region having a continuous depth into multiple layers has been found, but this can be solved through depth-based image warping.

Besides, in consideration of the server-client structure, the server transmits each layer image and depth information generated through the layered view generation process such as scene layering, depth estimation and image inpainting to the client. It was shown that it is possible to provide motion parallax through the object-centered view synthesis technique which has low implementation complexity.

As a future work, we will find a feasible way to segment the objects of interest from the scene and also continue research about a learning-based image inpainting technique which considers spatiotemporal consistency between multiple views.

ACKNOWLEDGEMENTS

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory) and 'The Cross-Ministry Giga KOREA Project' of The Ministry of Science, ICT and Future Planning, Korea (GK19C0200, Development of full-3D mobile display terminal and its contents).

REFERENCES

[1] Virtual Reality – Ecosystem & Standards Workshop, "3GPP activities around VR," Dec. 2017.

[2] H.-C. Shin, G.-S. Lee, and Namho Hur, "View interpolation using a simple block matching and guided image filtering," IEEE 3DTV-Conference (3DTV-CON), pp.1-4, (2014).

[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," IEEE Computer Vision and Pattern Recognition(CVPR), (2018).