Multi-view Image Generation Using Single-view Color Image and Low-resolution Depth Map

Jae-Il Jung and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712 Korea
Email: {jijung, hoyo}@gist.ac.kr

***Abstract*— Among various 3D video formats supporting multi-view navigation, single viewpoint video-plus-depth format consisting of a high-resolution color image and low-resolution depth map receives a lot of attention. This format provides high compression efficiency, but it needs additional tasks: depth map enhancement and virtual view synthesis. In this paper, we propose a multi-view generating system from the single-view format. The proposed system refines an initial depth map with consideration of color and cumulative gradient values of a color image. After enhancing the initial depth map, the proposed system automatically generates multi-view images by using 3D warping, color mapping, and hole filling. The experimental results show that the proposed system can generate high quality multi-view images although input depth maps have low-resolution and contain errors around boundaries.**

Keywords—multi-view image; depth map; free viewpoint TV; view synthesis; 3D TV

I. INTRODUCTION

The 3D video service is attracting much attention due to its various applications including movie, broadcasting, and game. The current service is based on stereoscopic images captured by two cameras at different positions. Such images are able to show an immersive scene, but they only can provide a fixed viewpoint. As an alternative, a multi-view image becomes popular providing freedom of viewpoint selection [1]. Moving Picture Experts Group (MPEG) has initiated a work aimed specifically toward a free viewpoint system based on multi-view images [2].

Although the multi-view provides both 3D perception and free view navigation, it still has some limitations to be directly used for 3D video and free viewpoint television (FTV) systems. The performance of multi-view video systems significantly depends on the number of original views. That is, the system must capture a large number of views so as to render a realistic 3D scene with multiple viewpoints. However, it is difficult to acquire so many views in practical settings; cameras are still too bulky and too expensive.

Recently algorithms for synthesizing virtual images at desired viewpoints have been developed. Among them, a depth image-based rendering (DIBR) is widely used in multi-view images [3]. DIBR is associated per-pixel depth information. This approach consists of the following two steps: At first, original image points are projected into the 3D world by utilizing the respective depth data. Thereafter, these 3D space points are re-projected into a desired viewpoint. Since the performance of DIBR depends on the quality of depth maps, high quality depth maps are essential to acquire the views to allow high quality rendering of the scenes from any angle.

In general, depth estimation methods can be classified into two categories: passive depth sensing and active depth sensing. The former calculates depth information indirectly from 2D images captured by two or more cameras [4]. The later usually employs physical sensors, such as laser, infrared ray (IR), or light pattern, to directly obtain depth information [5]. To combine the advantages of both approaches, a hybrid camera system consisting of multiple video cameras and one or more time-of-flight (TOF) camera have been introduced [6].

However, the quality of depth maps acquired by TOF cameras is not satisfactory due to a physical limit of depth sensor. The resolution of such depth maps is much smaller than that of the corresponding color images. For instance, SR4000 developed by Mesa Imaging and KINECT developed by Microsoft provide depth maps only up to 176×144 and 640×480, respectively.

The low-resolution depth maps significantly degrade the quality of multi-view images synthesized by DIBR. Therefore we propose a multi-view generating system including a depth enhancement process in this paper. The proposed system refines input depth maps with consideration of their corresponding color images, and then automatically generates virtual views at desired viewpoints.

II. MULTI-VIEW IMAGES AND DEPTH ENHANCEMENT

In this section, we briefly introduce approaches for acquiring multi-view images and for enhancing the resolution of depth maps.

The easiest way to acquire multi-view images is to directly capture images by multiple cameras at different positions. Figure 1 shows the several multi-view cameras with different arrays. According to applications, we select a proper camera array. Such a method guarantees the high quality of each image,

but the amount of data to be stored or transmitted is extremely large.

Figure 1. Multi-view cameras: Parallel, Convergence, and 2D arrays

Recent approaches send alternative image formats to synthesize virtual images at the decoder side with the lower amount of data: stereoscopic image, single-view color image and depth map, and multi-view single image and depth map. When generating a multi-view image from a stereoscopic image, users should estimate depth information for their selves. It is a difficult and time-consuming process for the users. The other approaches sending depth maps with color images can lift the burden at the decoder side. Especially, we can additionally reduce the amount of data by sending a color image and low-resolution depth map.

As we mentioned in the previous section, the quality of synthesized images depends on the quality of depth maps, since the depth values are used for calculating pixel positions in the synthesized image. Therefore the low-resolution depth map should be enhanced before we synthesize virtual images.

Various approaches have been proposed to effectively enhance the depth resolution. In the beginning of the research, simple approaches were exploited such as bilinear, nearest neighbor, and bicubic interpolation methods. Although these algorithms provide reliable results, their results include lots of errors around boundaries. It is because they interpolate depth values without the consideration of color discontinuities.

On the other hand, Diebel *et al.* proposed an interpolation method using the Markov random field (MRF) and designed an adaptive weighting function according to a color image gradient [7]. They suggested a depth smoothness prior using the weighting factor reflecting color differences. Yang *et al.* presented a new post-processing step using the bilateral filter [8]. This method iteratively refines the input low-resolution depth map, in terms of both its spatial resolution and depth precision. Both algorithms show the better results than the results of the previous simple algorithms, but their performance decreases when the objects having similar colors are contiguous.

III. High-Resolution Color Image and Low Resolution Disparity Image

We propose a technique for generating a multi-view image. The most important parts are how to enhance an initial low-resolution depth map and how to synthesize virtual images. Figure 2 shows the overall procedure of the proposed system.

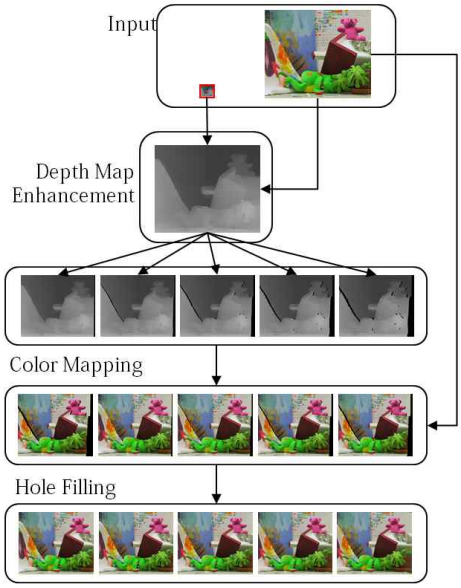


Fig. 2. The overall procedure of the proposed multi-view generating system

After enhancing the initial depth map, we synthesize virtual depth maps at desired viewpoints and map color values to them. Since the synthesized images have holes caused by viewpoint shifting, we fill the holes by the inpainting technique.

A. Depth Map Enhancement

In order to use the DIBR technique, the resolution of a depth map has to be equal to that of a color image. Therefore it is essential to effectively interpolate the initial depth map. At first, we simply interpolate the initial depth map by using the nearest neighbor algorithm and refine it. The enhanced depth value d_e at a position (x, y) can be obtained by

$$d_e(x,y) = \frac{\sum_{\tilde{x}=x-b}^{x+b}\sum_{\tilde{y}=y-b}^{y+b} w(x,y,\tilde{x},\tilde{y})d_i(\tilde{x},\tilde{y})}{\sum_{\tilde{x}=x-b}^{x+b}\sum_{\tilde{y}=y-b}^{y+b} w(x,y,\tilde{x},\tilde{y})} \quad (1)$$

where b is a support size and d_i means a depth value of the initial depth map. A weight w considers two properties; The pixels in the same object have similar colors and depth values in a local region, and the objects are distinct from each other with noticeable edges. These two weights can be express as

$$w(x,y,\tilde{x},\tilde{y}) = w_c(x,y,\tilde{x},\tilde{y})w_g(x,y,\tilde{x},\tilde{y}) \quad (2)$$

where w_c is for the first property and can be expressed as

$$w_c(x,y,\tilde{x},\tilde{y}) = \exp\left(-\frac{\{I(x,y)-I(\tilde{x},\tilde{y})\}^2}{\gamma_c}\right) \quad (3)$$

where I represents the intensity value and γ_c is a constant parameter for color sensitivity. The second weight w_g can be expressed as

$$w_g(x,y,\tilde{x},\tilde{y}) = \exp\left(-\frac{\sum_{(\tilde{x},\tilde{y})\in path(x,y)-(x',y')} gradient(\tilde{x},\tilde{y})}{\gamma_g}\right) \quad (4)$$

where $gradient(\tilde{x},\tilde{y})$ is the gradient value of the color image at position $(\tilde{x},\tilde{y})$. We add all gradient values on the shortest path from (x, y) to $(\tilde{x},\tilde{y})$. γ_g is a constant parameter for gradient sensitivity.

By using (2), we can adaptively allocate proper weights to reference pixels in the local supports. Figure 3 shows the textures and allocated weights. Even though each texture has different shape and color, the allocated weights reflect them well. Especially, although there are many objects with similar colors in Part B, the proposed method allocates higher weights to the same object that contains the center position. We can enhance the initial depth map by applying (1) to all pixels.

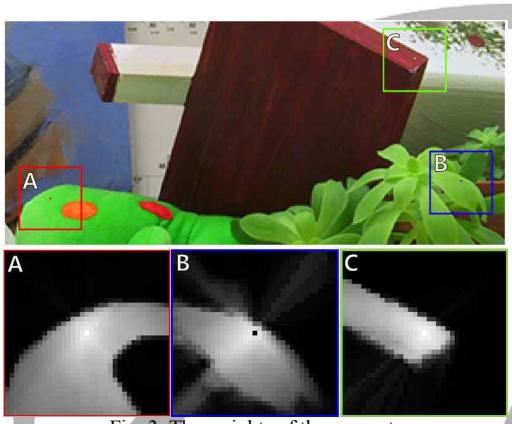


Fig. 3. The weights of the supports

B. View Synthesis

After enhancing the depth map, we generate multi-view images using DIBR. If all camera parameters of an input image are given, we calculate the corresponding pixel position between input and virtual images [9]. When a point **M** in world coordinate is projected to a camera, a pixel **m** in the image can be found by (5). The representations of a single point $\mathbf{M} = [X\ Y\ Z\ 1]^T$ and a projected point $\mathbf{m} = [x\ y\ 1]^T$ are the homogeneous form. The relation between **m** and **M** can be expressed as

$$\mathbf{m} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]\mathbf{M} \quad (5)$$

where **A** is the intrinsic camera parameter, and **R** and **t** are the extrinsic camera parameters. With (5), we project a pixel $\mathbf{m_d}$ in the enhanced depth image to the world coordinate using (6).

$$\mathbf{M_d} = \mathbf{R_d}^{-1}\cdot\mathbf{A_d}^{-1}\cdot\mathbf{m_d}\cdot d(\mathbf{m_d}) - \mathbf{R_d}^{-1}\cdot\mathbf{t_d} \quad (6)$$

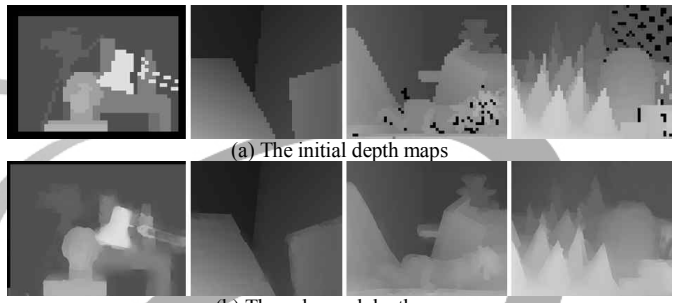
where the representations of $\mathbf{A_d}$, $\mathbf{R_d}$, and $\mathbf{t_d}$ stand for camera parameters of the input view. $d(\mathbf{m_d})$ is a depth value of the pixel of $\mathbf{m_d}$. After projection of $\mathbf{m_d}$, we re-project $\mathbf{M_d}$ onto the desired view position using (7).

$$\mathbf{m_t} = \mathbf{A_t}[\mathbf{R_t}|\mathbf{t_t}]\mathbf{M_d} \quad (7)$$

where the representations of $\mathbf{A_t}$, $\mathbf{R_t}$, and $\mathbf{t_t}$ represent camera parameters of the desired view. After that, we map color information to the synthesized depth image and fill the holes caused by viewpoint shifting with the inpainting algorithm.

IV. Experimental Results

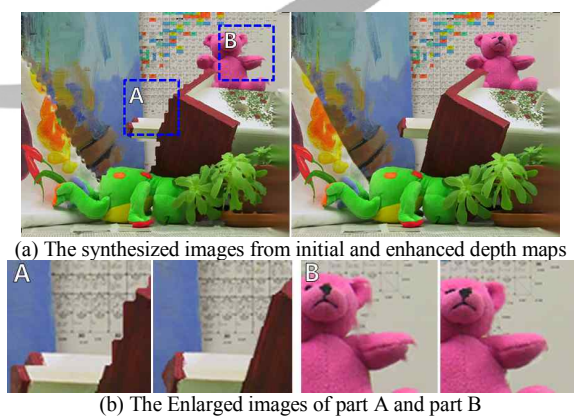
In order to evaluate the performance of the proposed system, we experimented on three single-view images, *tsukuba*, *cone*, and *teddy*. We downloaded these test images and their corresponding depth maps from the website of Middleburry. The original depth maps were down-sampled by a factor of 8 for each axis and up-sampled again with the nearest neighbor algorithm.


(a) The initial depth maps

(b) The enhanced depth maps

Fig. 4. Comparison of the depth maps: *tsukuba*, *venus*, *teddy*, and *cone*

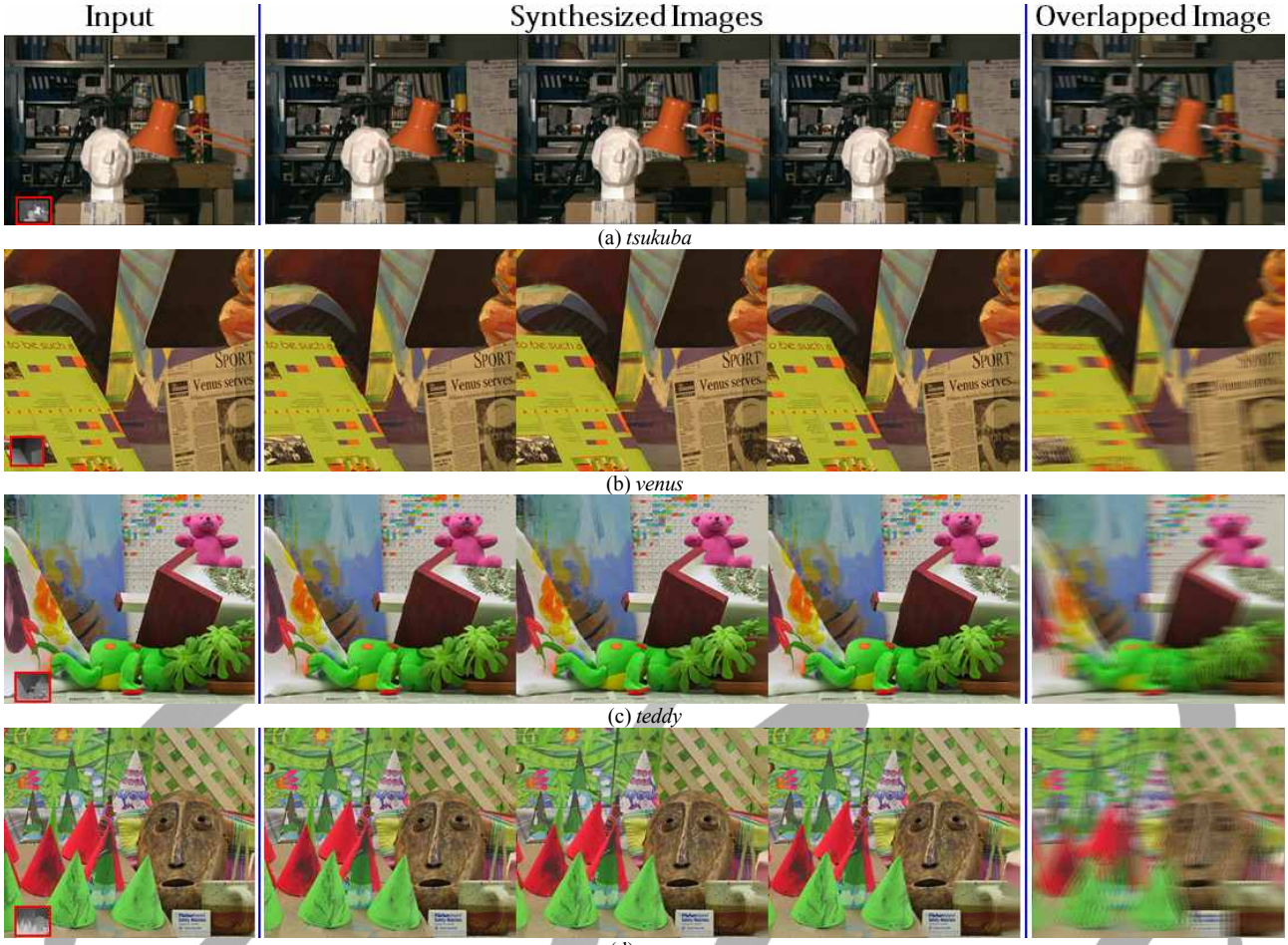
Figure 4(a) and Fig. 4(b) show the initial and enhanced depth maps, respectively. The initial depth maps have very rough boundaries due to their low resolution, but the enhanced depth maps have well matched edges with color images. Figure 5 demonstrates the comparison of synthesized images from the initial and enhanced depth maps. While the initial depth map causes serious distortions around boundaries, the proposed method improves the viewing quality of the synthesized images. Figure 6 shows the input color images, depth maps (red-boxes), synthesized images, and overlapped images. Even though the input depth maps are much smaller than the color images, the proposed algorithm can generate the natural multi-view images with proper view intervals.



(a) The synthesized images from initial and enhanced depth maps

(b) The Enlarged images of part A and part B

Fig. 5. Comparison of the quality of the synthesized images



(a) *tsukuba*

(b) *venus*

(c) *teddy*

(d) *cone*

Fig. 6. The results of multi-view generation: input(color image and depth map), synthesized images, and overlapped images.

V. CONCLUSION

In this paper, we have presented a new approach to generate a multi-view image from a single-view color image and low-resolution depth map. We have enhanced the resolution of the depth map with consideration of colors and cumulative gradient values. Thereafter, we have generated the multi-view image by depth map warping, color mapping, and hole filling. Experimental results have shown that our scheme produced more reliable high-resolution depth maps and multi-view images compared with the previous methods. The proposed multi-view generation system could provide high quality multi-view images with small amounts of data. Therefore, our proposed system could be useful for various 3D multimedia applications and displays.

ACKNOWLEDGMENT

This research is supported by MCST and KOCCA in the CT Research & Development Program 2012.

REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multi-View Imaging and 3DTV (Special Issue Overview and Introduction)," *IEEE Signal Processing Magazine,* vol. 24, no. 6, pp. 10-21, Nov. 2007.

[2] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video - technologies, applications and MPEG standards," *IEEE International Conference on Multimedia and Expo* Toronto, Canada, pp. 2161-2164, July 2006.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, May 2004.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision,* vol. 47, no. 1, pp. 7-42, April 2002.

[5] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 195-202, June 2003.

[6] E. Lee and Y. Ho, "Generation of multi-view video using a fusion camera system for 3D displays," *IEEE Transactions on Consumer Electronics,* vol. 56, no. 4, pp. 2797-2805, Nov. 2010.

[7] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," *Advances in neural information processing systems,* vol. 18, no. pp. 291-298, 2006.

[8] Y. Qingxiong, Y. Ruigang, J. Davis, and D. Nister, "Spatial-Depth Super Resolution for Range Images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 2007.

[9] R. Hartley, A. Zisserman, and I. ebrary, Multiple view geometry in computer vision vol. 2: Cambridge Univ Press, 2003.