HIGH QUALITY NOVEL VIEW SYNTHESIS BASED ON LOW RESOLUTION DEPTH IMAGE AND HIGH RESOLUTION COLOR IMAGE

Jui-Chiu Chiang, Zheng-Feng Liu, and Wen-Nung Lie

Department of Electrical Engineering, National Chung Cheng University
168, University Road, Ming-Hsiung, Chia-Yi, 621, Taiwan, ROC.

ABSTRACT

In this paper, a new technique to generate high resolution depth image is proposed. First, a low resolution depth map is obtained by the time-of-flight depth camera. Then a high resolution depth map for a given view is generated by depth warping followed by depth value refinement taking into account the color information at the given view. The edge in the final depth map is then processed by bilateral filtering for edge preserving. With the color image and the corresponding depth image, novel view synthesis can be carried out by depth image based rendering (DIBR). Experimental results show that the depth map generated by the proposed technique is able to ensure novel view images with high quality.

Index Terms— novel view synthesis, DIBR, depth camera

1. INTRODUCTION

With the advance in acquisition and display technologies, entertainment ensuring higher perceptual realism is desired and becomes feasible recently. FTV (Free view-point television) [1] is recognized as the next generation of TV. It offers human not only the stereoscopic perception, but also various interactions and possibilities. To provide the audience with arbitrary view on demand, novel views have to be rendered in the display. Novel view synthesis techniques have been developed for many years, where image based rendering (IBR) [2-3], depth image based rendering (DIBR) [4-5] and ray-space method [6-7] are the main technologies to be adopted. In IBR, no depth image is required and the challenge of IBR is to find out the correspondence points in given images pair. Then novel view images can be synthesized by proper interpolation. However, the image quality of the novel view could be limited due to insufficient geometric information. Ray-space is the representation of light rays in three-dimensional space and virtual view rendering for arbitrary positions is performed without any geometry information. However, it usually needs to capture the scene with dense cameras. DIBR attracts lots of attention recently. DIBR needs color images with corresponding depth information to render the novel view-point image. The image quality of novel view based on DIBR could be much improved, compared to IBR methods, under the assumption that the depth image is reliable. Although DIBR has the potential to offer novel views with high quality, several problems during DIBR

realization have to be managed, such as hole, disocclusion and occlusion. To cope with these problems, many research works are presented [8-10].

The main functionality supported by FTV is the free navigation by seamless novel view generation, while stereoscopic perception is the main target of 3DTV. 3DTV offers human the stereoscopic experience and became available in the market since 2010. There are several 3D representation formats, including stereo image pair, frame compatible stereo image format and video plus depth. There are two main advantages provided by video plus depth format. First, the depth image is represented as gray level and usually the required bitrate is much less than color video due to the homogeneous property. Second, a stereo image pairs with flexible disparity range can be generated in the receiver side by DIBR synthesis technique.

Both 3DTV and FTV could be realized by DIBR and it is imperative to acquire accurate depth maps for rendering stereo image pair and novel view-point image with satisfactory perception. Generally, the depth image can be captured directly or estimated by stereo matching algorithm [11]. Usually, the estimated depth maps suffer from inaccurate results in the occlusion regions, as well as textureless regions. Although the captured depth image ensures higher accuracy compared to estimated depth image, the resolution is usually too small to be coordinated with the color image directly for DIBR synthesis. There are several works concentrating on high resolution depth image generation under the scenario that a hybrid camera system (including both color camera and depth camera) is available [12-15]. In [12], an initial depth map for the left image is computed by a stereo matching algorithm. Then the ROI (region of interest) in the depth map is refined by warping the depth values obtained from the depth camera to the left image.

In this paper, we propose a new technique to generate high resolution depth map from a low resolution depth image and a high resolution color image. Different from the work in [12] where the depth map is first estimated by stereo matching and the depth information provided by depth camera is used to refine the ROI, the stereo matching is not performed in the proposed scheme and one color image is required in the proposed method. Instead, the color image will be used to refine the depth information obtained by the time-of-flight depth camera.

The rest of this paper is organized as follow. Section 2 introduces the technique to generate the preliminary high resolution depth map using the low-resolution depth map. Section 3 details the method to refine the depth map considering the color information. The experimental results are presented in section 4 and section 5 draws the conclusion.

2. GENERAITON OF A PRELIMINARY HIGH-RESOLUTION DEPTH MAPS

In our image capture system, there are one color camera and one time-of-flight depth camera in parallel setting. The captured depth video and color video can be seen as the video plus depth pair for 3DTV application. Note that, the resolution of the depth image is smaller than the color image, and the challenge is to make the resolution of the depth image the same as the color image for rendering purpose. And the same time, how to ensure accurate depth map is also important in providing high-quality novel view image. To facilitate the following procedures, camera calibration has to be employed first for the two cameras to obtain the extrinsic parameters for each camera.

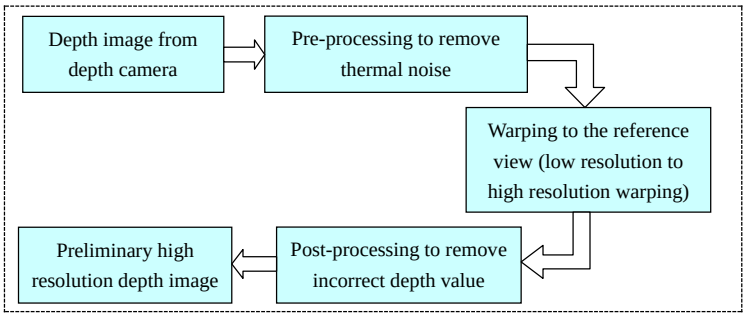

Figure 1. Block diagram of preliminary high-resolution depth map generation.

2.1 Pre-processing of the depth map

To realize the novel view synthesis, a color image and the associated depth map are needed. In our camera setting, the view capturing color image has no captured depth image. Thus, the depth image captured in the center view has to be warped into the reference view (i.e., left or right view). Figure 1 presents the procedures to generate high resolution depth map for the view with color image.

Usually, the sensing results of the time-of-flight depth camera are not always reliable. Optical noise, as well as the material of objects in the scene may cause incorrect depth values. To resolve this problem, the captured depth map is pre-processed with a 3×3 median filter. Furthermore, the lens distortion in the depth map has to be removed before warping to the reference view.

2.2 Low resolution to high resolution warping

After removing the noise in the depth map, 3D warping is performed to get the depth map of the reference view. However, the resolutions for the depth camera and color camera are not the same, and the depth map in the reference view will be incomplete when the low resolution depth map is warped directly. It means many pixels in the reference map have no assigned depth value. To overcome this problem, each pixel in the source view will be warped into a block with size 5×5 in the reference view. It means that in the reference view, in addition to the pixel location determined by the warping procedure, the surrounding 24 pixels will have the same depth value. Besides, if a pixel in the reference view has more than one warped depth value, the maximum value is used.

2.3 Post-processing by morphological operation

After warping, the obtained high resolution depth image may have incorrect values, especially around the edges. To resolve this problem, morphological opening is performed to get the preliminary high resolution depth image.

3. GENERATION OF HIGH QUALITY HIGH-RESOLUTION DEPTH MAP

The preliminary depth map may have some remained errors after morphology operation. To enhance the quality of the depth image, the color image at the reference view

will be used. Figure 2 illustrates the flowchart of color-image guided depth map refinement.

First, a binary representation of the preliminary depth map $D(x,y)$ is obtained, represented as $D_b(x,y)$ after comparing to a given threshold as following,

$$D_b(x,y) = \begin{cases} 1, & if \quad D(x,y) \geq TH_d \\ 0, & if \quad D(x,y) < TH_d \\ -1, & if \quad D(x,y) = -1 \ (\text{"}hole\text{"}) \end{cases} \tag{1}$$

where $D(x,y) = -1$ denotes the situation that the location (x,y) has no depth value after warping and morphological opening (i.e., it can be seems as "hole"). Note that, the foreground object usually have higher value in depth map, and correspond to $D_b=1$.

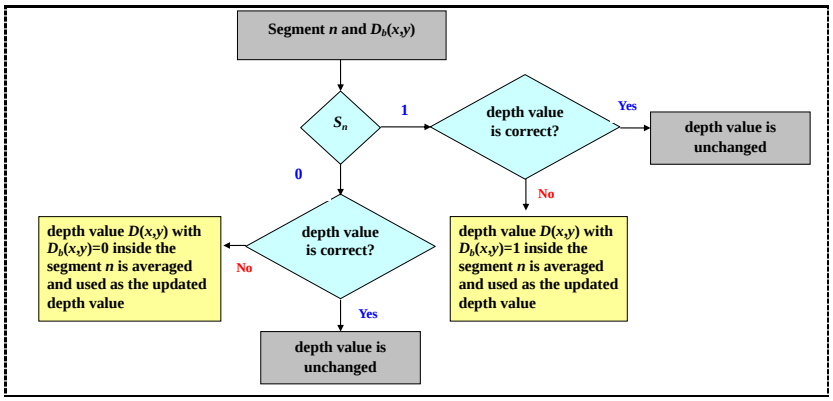

Figure 2. Flowchart of color-image guided high-resolution depth map refinement.

The segmentation results of the color image in the reference view are used to refine the depth map. Mean-shift algorithm [16] is adopted to segment the color image into several groups. The application scenario in this paper is a simple scene with two depth layers; one for the foreground and the other background. Thus, each segment n will be labeled as "foreground" or "background" according to the rule as below,

$$S_n = \begin{cases} 1, & \dfrac{D_n}{C_n} \geq TH \quad (\text{Foreground}) \\ 0, & \dfrac{D_n}{C_n} < TH \quad (\text{Background}) \end{cases} \tag{2}$$

where D_n and C_n denote, respectively, the number of pixel with value $D_b=1$ inside the segment n and the total pixel number inside the segment n.

After labeling each segment as foreground or background, the correctness of each depth value inside each segment will be verified. If $S_n=1$, the segment n is recognized as a foreground region and the depth value $D_b(x,y)$ for the pixel inside this segment should be equal to 1. If the depth value $D_b(x,y)$ is not equal to 1, it is treated as an incorrect depth value, and the depth value $D(x,y)$ with $D_b(x,y)=1$ inside the segment n is averaged and used as the updated depth value for this pixel. The verification of depth values correctness in the segment with $S_n=0$ is similar. There is no more hole in the depth map after this process.

To preserve the object edges and enhance the quality of depth image, a bilateral filter combining domain and range filtering is performed on the depth map. The two filter coefficients are determined by the relative distance Δd and the intensity difference ΔI in the color image, as below:

$$D_{bf}(x,y) = \frac{\sum_{s=-m/2}^{m/2}\sum_{t=-n/2}^{n/2}\left(\omega_{bf} \times D(x+s, y+t)\right)}{\sum_{s=-m/2}^{m/2}\sum_{t=-n/2}^{n/2}\omega_{bf}} \qquad \omega_{bf} = 2^{-(\Delta d + \Delta I)}, \tag{3}$$

where *m* and *n* denote the window size of filtering. Note that, to avoid the interference between different color segments, not all the image, but the edge and surrounding pixels are processed by the bilateral filter.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm, we have carried out the algorithm on the real captured images. The resolution of the captured depth image and color image are 176×144 and 640×480, respectively. Several scenes are captured and experimental results of one scene with different time slot images are presented here due to limited space..

4.1 Depth Map Generation

Figures 3 illustrates the results of preliminary depth map generation, where Figure 3(a) is the color image in the reference view. It shows that many pixels in Figure 3(b) are represented as green due to no assignment during warping process. Figure 3(c) presents a more complete depth map by the proposed warping technique where one pixel in the source view will be mapping to a block in the target view. However, the depth values around the edge seem incorrect and can be resolved by morphological opening operation, as shown in Figure 3(d) & (f).

The refined depth maps after considering the color image information in the same view is illustrated in Figure 4. The segmentation results on the color image are shown in Figure 4(a). Then each segment will be classified as foreground or background and used to judge the correctness of the depth value inside each segment. The pixel represented as red in Figure 4(b) is the incorrect pixel and will be updated, as shown in Figure 4(c). The depth image after bilateral filtering is shown in Figure 4(d), where the edge is more sharped. As highlighted in Figure 4(f), the profile of the edge is clearer than that in Figure 4(e). Figure 5 shows a series of depth map generated by the proposed scheme.

4.2 Novel View Synthesis

To evaluate the correctness of the depth image, a novel view synthesis be performed, which can be carried out by one color image and the associated depth map. Typically, the synthesis quality will be better for the scenario with multiple color images and depth maps. Note that, the proposed algorithm can be easily applied to the scenario with two or more color cameras. Then each color image will have a corresponding high resolution depth map for rendering purpose.

If the novel view synthesis relies on only one color image and the corresponding depth map, the difficulty to fill the disocclusion region is higher than the scenario with multiple color images and depth maps. To tackle this problem, the depth map generated by the proposed scheme can be post-processed by Gaussian filtering. In this way, the transition between foreground and background is smoother, and satisfactory image quality is revealed on the synthesized image after simple hole filling. Figure 6 demonstrates the synthesized results using various depth maps.

4.3 Execution time analysis

Here, the execution time for each step for generating the final depth map is analyzed, as shown in Table 1. Intel Core 2 Duo Q6600 2.33 GHz, and DDR2 800 4GB are used for the simulation. Table 1 indicates that the color segmentation is the most time-consuming part in the proposed scheme.

Table 1. Execution time for depth map generation

	Procedure	time(sec)
Pre-processing	Depth Warping	0.117
	Color segmentation	0.742
Post-processing	Bilateral filtering on edges	0.352
	Gaussian filtering	0.104
Novel view synthesis by DIBR (including hole filling)		0.144

5. CONCLUSION

A hybrid camera system consisting of high-resolution color camera and low-resolution depth camera is considered in this paper. The goal is to render a high quality novel view image and an algorithm for accurate depth map generation is proposed. A preliminary depth map is built using the depth image captured by the depth camera and refined after taking the color information into consideration. The color segmentation results will guide whether the depth value is correct and appropriate update is realized accordingly. Experimental results show that a high quality novel view mage can be rendered using the depth map generated by the proposed scheme.

6. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 M8595, "FTV-Free Viewpoint Television," 2002.

[2] S. M. Seiz and C. R. Dyer, "View Morphing," *Proc. of SIGGRAPH*, pp. 21-30, 1996.

[3] P. E. Debevec, Y. Yu and G. Borshukov, "Efficient View-dependent Image-based Rendering with Projective Texture Mapping," *Proc. of Eurographics Rendering Workshop*, pp. 105–116, 1998.

[4] C. Fehn, "A 3D-TV Approach Using Depth-Image-Based-Rendering (DIBR)," *Proc. of Visualization, Imaging, and Image Processing,* pp. 482-487, 2003.

[5] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," in *SPIE-IS&T*, Vol. 5291, pp. 93-104, Jan. 2004.

[6] T.Fujii, T.Kimoto, M.Tanimoto, "Ray Space Coding for 3D Visual Communication", *Proc. of Picture Coding Symposium*, pp. 447–451, Mar. 1996.
[7] T. Naemura, M.Kaneko, and H. Harashima,"3-D Visual Data Compression Basedon Ray-Space Projection", *Proc. of SPIE VCIP*, pp. 413–424, 1997.
[8] I. Daribo, C. Tillier, B. Pesquet-Popescu, "Distance Dependent Depth Filtering in 3D Warping for 3DTV," *Proc. of IEEE Int'l Conf. on Multimedia Signal Processing* (MMSP), pp.312-315, Chania, Crete, Greece, Oct. 2007.
[9] C.-M. Cheng, S.-J. Lin, S.-H. Lai and J.-C Yang, "Improved Novel View Synthesis from Depth Image with Large Baseline," *Proc. of IEEE Int'l Conf. on Pattern Recognition*, 2008.
[10] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View Generation with 3D Warping using Depth Information for FTV," *Signal Processing Image Communication*, vol. 24, no. 1, pp. 265–272, 2009.
[11] J. Sun, N.N. Zheng, and H.Y.Shum, "Stereo Matching using Belief Propagation," *IEEE Trans. of Pattern analysis and Machine Intelligence*, vol. 25, no. 5, pp. 787-800, 2003.
[12] Sung-Yeol Kim, Eun-Kyung Lee, Yo-Sung Ho, "Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera," *IEEE Trans. on Broadcasting*, Vol. 54, No.4, pp.732-740, Dec. 2008.
[13] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of Time-of-flight Depth and Stereo for High Accuracy Depth Maps," *Proc. of IEEE Int'l Conf. Computer Vision Pattern Recognition*, 2008.
[14] Bogumil Bartczak and Reinhard Koch, "Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views," *Proc. of International Symposium on Visual Computing* (ISVC), Las Vegas, Nevada, Dev. 2009.
[15] Qingxiong Yang, Ruigang Yang, James Davis, David Nistér, Spatial-Depth Super Resolution for Range Images, CVPR 2007.
[16] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. of Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790-799, Aug, 1995.

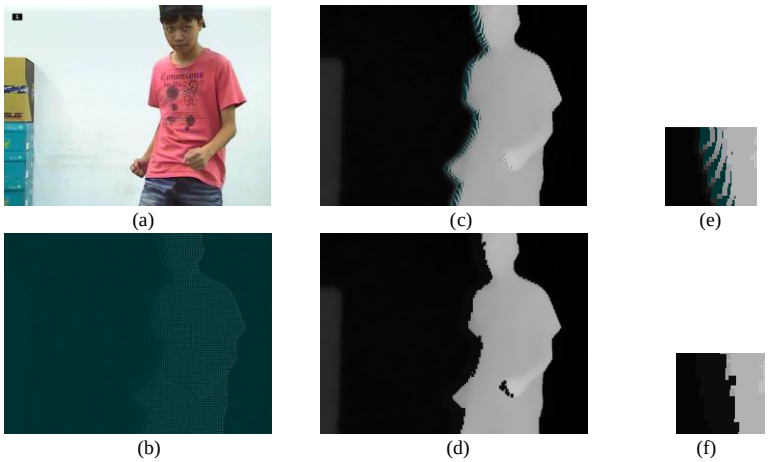

Figure 3. Illustration of preliminary depth map generation. (a) color image in the left view, (b) the warped depth image in the left view (pixel without intensity assignment is marked as green) (c) the depth map after applying one pixel to 5×5 block warping, (d) the depth image after morphological opening, (e) region before morphological opening, (f) region after morphological opening.

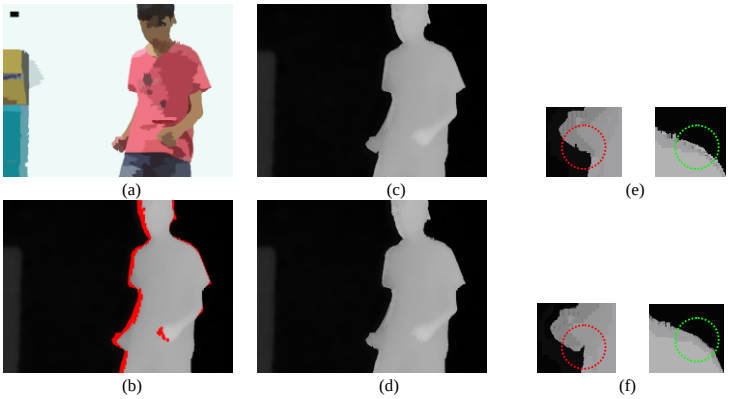

Figure 4. Illustration of depth map refinement (a) color segmentation results,(b) the high resolution depth image after taking the color information into consideration (the pixel marked in red is labeled as incorrect pixel), (c) the depth image after correction, (d) the depth image after bilateral filtering on the edges, (e) region before bilateral filtering, (f) region after bilateral filtering.

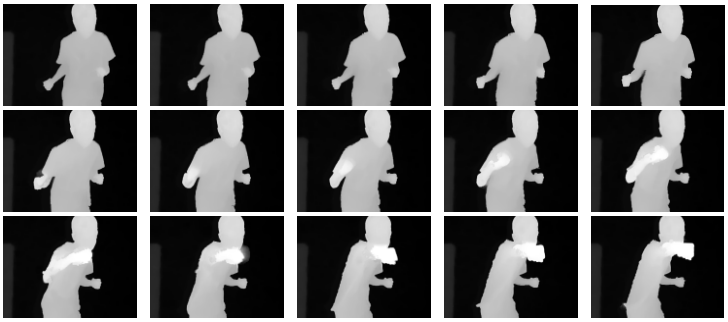

Figure 5. A series of the generated depth maps.

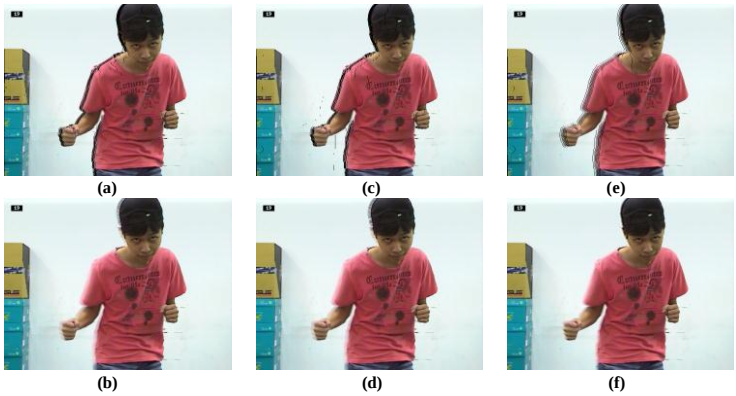

Figure 6. Virtual view images using (a) depth image with bilateral filtering on whole frame (without hole filling),(b) depth image with bilateral filtering (after hole filling), (c) depth image with bilateral filtering on edges (without hole filling), (d) depth image with bilateral filtering on edges (with hole filling), (e) depth image with bilateral filtering on edge, followed by Gaussian filtering (without hole filling), (f) depth image with bilateral filtering on edge, followed by Gaussian filtering (after hole filling).