Efficient Up-sampling Method of Low-resolution Depth Map Captured by Time-of-Flight Depth Sensor

Yun-Suk Kang and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
Gwangju, 500-712, Republic of Korea
E-mail: {yunsuk, hoyo}@gist.ac.kr

Abstract— **In this paper, we propose an efficient up-sampling method of the low-resolution depth map that is obtained by a time-of-flight (TOF) depth sensor. After we capture the color images and TOF depth maps simultaneously using a fusion camera system, each pixel of the low-resolution depth map is relocated to the corresponding color image positions by 3D warping, and the warping error is eliminated in each color segment. Then, we employ a Markov random field (MRF) model to produce a high-resolution disparity map using the warped values and color segment information. Experimental results show that the proposed method efficiently up-sample the low-resolution depth maps compared to the other up-sampling approaches.**

I. INTRODUCTION

Recently, three-dimensional television (3DTV) and 3D video contents have become attractive media since they provides immersive and realistic images to viewers [1]. The basic principal of 3DTV and 3D video contents is to be composed of multiple viewpoints for a scene unlike the conventional two-dimensional (2D) videos. Since two or more viewpoints include the parallax among the adjacent views, viewers recognize the sense of distance or 3D feeling by watching the 3D video using the stereoscopic or auto-stereoscopic displays. In general, the more number of viewpoints gives more natural and immersive scene within the technical limitation of the displays.

However, it is difficult to capture a number of viewpoint images due to following problems. The amount of high-resolution video data from multiple viewpoints is too huge to efficiently encode, transmit, and decode. The camera system to capture the multi-view video is also hard to construct. Moreover, the volume of each camera makes difficult to arrange the cameras densely, which is related on the quality of 3D video.

Depth information of the scene can be the solution of the above problems. In general, a depth map is 8-bit gray level image that represents the range to each object of the scene from the camera. Therefore, if we have the color image and the corresponding depth map of the scene, we can reconstruct images at novel viewpoints using the depth image-based rendering (DIBR) technique [2]. Finally, we can obtain multiple viewpoint images without the less number of cameras. Also, it makes simple to encode, transmit, and decode processes.

In order to calculate the depth map is very important since the quality of the reconstructed novel view image is highly related on the depth quality. The representative method to calculate the depth map is stereo matching [3]. Stereo matching calculates the disparity values of each pixel by searching the best matched pixel in the reference view. However, stereo matching not only often fails to calculate the proper disparity values for the color-mismatched, non-textured and occluded regions, but also requires long processing time.

On one hand, there are depth sensors based on the time-of-flight (TOF) technique, which measure the range from the sensor to objects in the scene. The measured range information is then quantized and represented in the image in real-time. However, since the output depth map is noisy and the resolution is very small, it is required to enhance the depth map to use for 3D applications. Moreover, the depth sensors operate based on the light signal, there is lots of limitations for capturing.

In this paper, we explain an efficient up-sampling for a low-resolution depth map from a TOF depth sensor. After capturing the color image and depth map, we calibrate each camera. Then, the TOF depth values are 3D warped to the color image position. The color image is segmented, and the warping noise is eliminated in each segment. Finally we up-sample and obtain the high-resolution depth map using a Markov random field (MRF) model based on the warped depth value as a measurement and the color segment information.

II. TOF DEPTH SENSOR CHARACTERISTICS

Figure 1(a) shows the TOF depth sensor Swiss Ranger SR4000 that is used in our experiment [4]. It is composed of the illumination cover and the optical filter. The illumination cover emits the light signal to the scene, and the optical filter, that receives the arriving emitted light signal. The minimum and maximum capturing ranges are about 0.3m and 5.0m, respectively.

Figure 1(b) and Fig. 1(c) show the output images from the TOF depth sensor. They are the depth and intensity images, respectively. Each pixel of the depth map represents the depth index having values from 0 to 255 which are quantized values

of the depth of the scene. The intensity image is similar to the gray image of the scene.

(a) SR4000 (b) Depth map (c) Intensity map

Figure 1. TOF depth camera and captured images

However, as shown in the captured images, there are noise and distortion in the TOF depth map. There are a median-value noise at the boundary regions and a large amount of the lens radial distortion. In the proposed method, we correct this lens distortion and reduce noise for the low-resolution depth map before all the processes.

III. PROPOSED METHOD

In this section, we explain our proposed method to up-sample the low-resolution depth map. After capturing, we correct the lens distortion of the low-resolution depth map. Then, camera calibration is performed to estimate camera parameters. Using the camera parameters, the TOF depth map is warped to the color image position and the noise is eliminated in each segment. They are used as a measurement for MRF model. Finally, we up-sample the TOF depth map to the high-resolution using the MRF model considering the TOF depth values and color segment. Figure 2 shows the whole process of the proposed method.

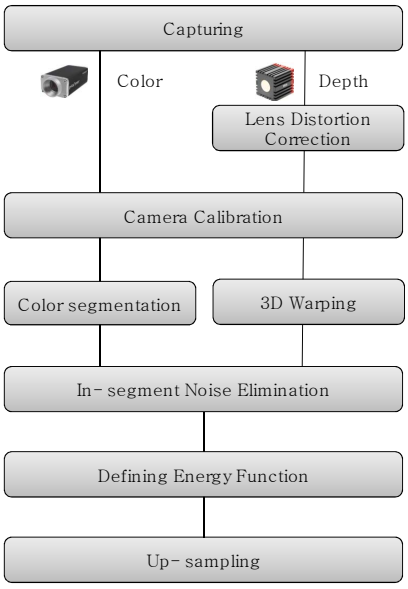
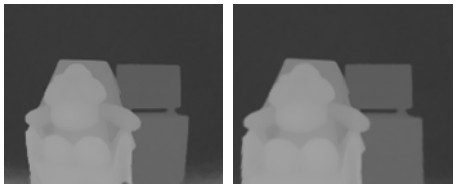


Figure 2. Procedure of the proposed method

A. Capturing, Correction, and Calibration

We captured the test images; high-resolution color image and low-resolution TOF depth map. The TOF depth sensor is located below the color camera. As shown in Fig. 1(b), there is lens radial distortion in the TOF depth map. Therefore, the distortion is corrected as Fig. 3 [5], and then we estimate the camera parameters.



(a) Before (b) Corrected

Figure 3. Lens distortion correction

The camera parameters are the principal information of the camera for 3D image processing and application. They are composed of the intrinsic and extrinsic parameters. The intrinsic parameters represent the camera's internal features such as the focal length, and they are represented as an intrinsic matrix **A**. The camera's orientation and location information are appeared in the extrinsic parameters. The extrinsic parameters are the rotation matrix **R** and translation vector **t**. The camera projection matrix **P** is then composed with these parameters, as Eq. (1). These parameters are essential for 3D warping step.

$$\mathbf{P} = \mathbf{A}[\mathbf{R}|\mathbf{t}] \tag{1}$$

Camera calibration is the process to compute the camera parameters using the captured grid-pattern images [6]. In the proposed method, we perform camera calibration of the depth cameras after correcting the lens distortion of the depth maps. Camera calibration toolbox for MATLAB [7] is used for camera calibration process.

B. Color Segmentation

Color segmentation is the process that divides the image into regions that have similar characteristics. In general, the region with the similar color characteristics by color segmentation has not a radical depth change or depth discontinuity but smooth depth change or the same depth. Also, the boundary between two adjacent segments has a possibility to be the depth discontinuity. In the proposed method, we segment the color image using mean-shift color segmentation algorithm [8].

C. 3D Warping

The TOF depth map is warped to each color image positions by 3D warping technique. For 3D warping, the camera parameters, and real depth values of the scene are required. Eq. (2) describes the 3D warping of the TOF depth pixel to the color image position.

$$\widetilde{\boldsymbol{m}_C} = \mathbf{A}_C\mathbf{R}_C\mathbf{A}_T^{-1}\mathbf{C}_T^{-1}\widetilde{\boldsymbol{m}_T} - \mathbf{A}_C\mathbf{R}_C\mathbf{A}_T^{-1}\mathbf{t}_T + \mathbf{A}_C\mathbf{t}_C \tag{2}$$

The subscripts *C* and *T* mean the color and TOF depth, respectively. $\boldsymbol{m}$ is the pixel position at (u, v) scaled by the real

depth Z, as shown in Eq. (3). The pixel positions are represented by the homogeneous representation.

$$\tilde{\boldsymbol{m}} = Z[u \quad v \quad 1]^{\mathrm{T}} \tag{3}$$

Figure 4 explains this process, and also shows the generated holes during the warping process due to the resolution difference between the color and TOF depth map.

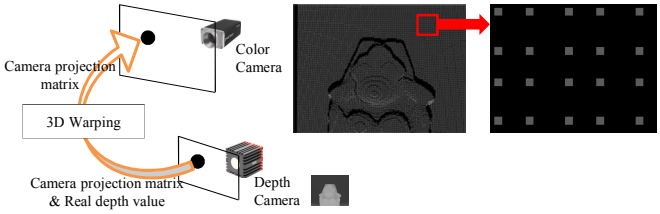


Figure 4. 3D warping of the TOF depth map

D. *In-segment Noise Elimination*

As mentioned in Section II, there are some median depth values at the depth discontinuity regions of the TOF depth map. We define these values as noises that are belong to neither foreground nor background. Figure 5 shows these noises, and the warped noises could be located at wrong positions. These values can decrease the quality of up-sampled depth maps since the MRF-based up-sampling uses the warped depth values as the measurement for data term.

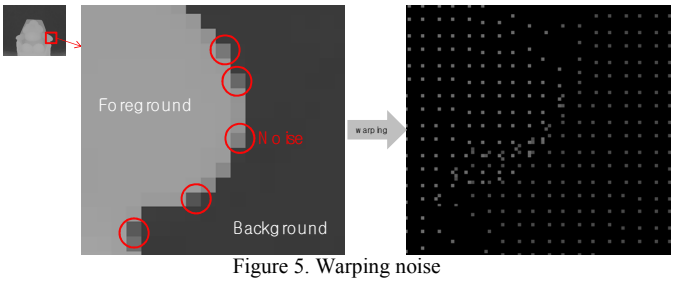


Figure 5. Warping noise

In order to eliminate these noises, we define minimum and maximum bounds for each segment to eliminate the noises. Figure 6 shows the color segment and warped depth values in the segment. For n-th segment, we calculate the average of the warped depth value in this segment, $d_{avg,n}$. We also calculate the standard deviation σ_n for n-th segment. Then, the minimum and maximum bounds for n-th segment as Eq. (4) and Eq. (5), respectively.

$$B_{min,n} = \left\lfloor d_{avg,n} - \frac{{\sigma_{d,n}}^2}{2} \right\rfloor \tag{4}$$

$$B_{max,n} = \left\lfloor d_{avg,n} + \frac{{\sigma_{d,n}}^2}{2} \right\rfloor \tag{5}$$

Using Eq. (4) and Eq. (5), we eliminate the noises in each segment. The warped depth values which are in between the minimum and maximum bounds are survived, and the other values are eliminated. These bounds permit the smooth change on the basis of $d_{avg,n}$. Therefore, the outlier values are filtered out in each segment.

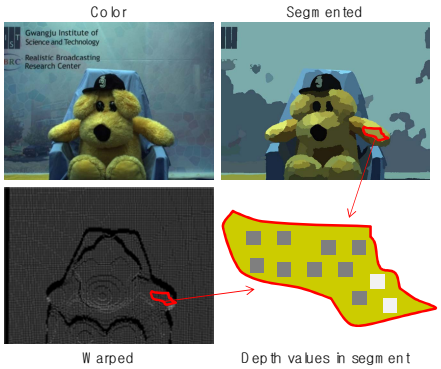


Figure 6. Segment and noises

E. *Up-sampling*

We then design the energy function E using the data term D and smoothness term V as Eq. (6), where y_i and z_i are label and measurement for i-th pixel, respectively. The measurement is the warped TOF depth values after in-segment noise elimination. $N(i)$ means the neighborhood of i-th pixel.

$$E = \sum_{i \in L} D(y_i, z_i) + \sum_{i} \sum_{j \in N_V(i)} V(y_i, y_j) \tag{6}$$

The data term is calculated as Eq. (7), where L means the pixel position with the measurement.

$$D(y_i, z_i) = (y_i - z_i)^2 \tag{7}$$

The smoothness term is calculated as Eq. (8), where $w_{c,ij}$ and $w_{s,ij}$ are weights considering color and segment, respectively. The color weight shown in Eq. (9) compares the color values I of the adjacent pixels i and j. The constant c controls the influence. Eq. (10) indicates the segment weight that maintains the smoothness cost when two adjacent pixels are in the same segment, and vice versa. S means the segment index for a pixel, and C_{seg} is a small coefficient.

$$V(y_i, y_j) = w_{c,ij} w_{s,ij} (y_i - y_j)^2 \tag{8}$$

$$w_{c,ij} = e^{-c(I(i)-I(j))^2} \tag{9}$$

$$w_{s,ij} = \begin{cases} 1 & if\ S(i) = S(j) \\ C_{seg} & otherwise \end{cases} \tag{10}$$

Then the energy function is optimized using the belief propagation method [9], and then calculates the proper depth

value for each pixel in the high-resolution. The number of iteration for belief propagation is adjusted according to the resolution difference.

IV. Experimental Results

We captured three sets of color image and TOF depth map as shown in Fig. 7. The resolutions of the color and TOF depth map are 1024x768 and 176x144, respectively. The vertical distance between the color camera and TOF depth sensor is about 7.5cm.

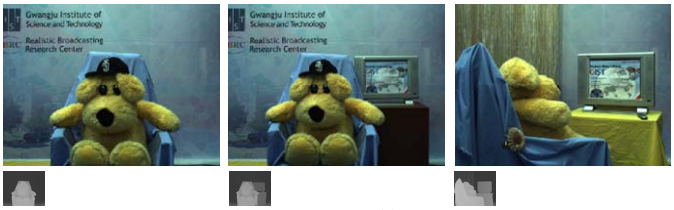


Figure 7. Captured images

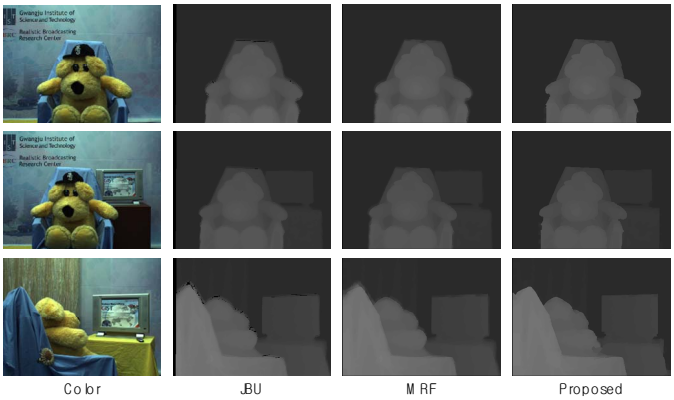


Figure 8. Up-sampled depth maps

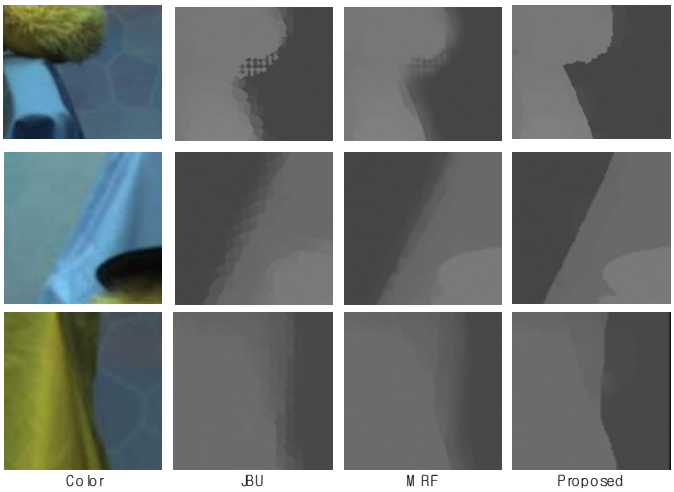


Figure 9. Comparison in detail

Figure 8 shows the up-sampled depth maps by using the proposed method, joint bilateral filter [10], and MRF model [11]. The up-sampled results by the proposed method have clean boundary regions with less noise compared to the other methods. Figure 9 shows the detail of the up-sampled depth maps. The proposed method efficiently separated the foreground and background without noises, and the object boundary regions of the up-sampled depth map matched to the color images.

V. Conclusion

In this paper, we presented an efficient up-sampling method to generate the high-resolution depth map using color and TOF depth cameras. The enhanced TOF depth map is warped to the color image position and used as a measurement for MRF-based up-sampling. To generate clean and accurate boundary regions, we perform in-segment noise elimination to the warped depth values. Using these warped depth and color segment information, we up-sample the correct depth maps from the low-resolution depth maps. The results by the proposed method outperforms in terms of the noise and object boundary regions.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009228).

References

[1] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies," Proceedings of the IEEE, Spatial Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, pp. 99-110, 2005.

[2] C. Fehn, "Depth-image-based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV," Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 5921, pp. 93-104, 2004.

[3] J. Sun, N.N. Zheng, and H.Y. Shum, "Stereo Matching Using Belief Propagation," IEEE Transactions of Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 787-800, 2003.

[4] SR4000 User Manual, Mesa Imaging AG.

[5] A. Wang, T. Qiu, and L. Shao, "A Simple Method of Radial Distortion Correction with Centre of Distortion Estimation," Journal of Mathematical Imaging and Vision, vol. 35, no. 3, pp. 165-172, 2009.

[6] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.

[7] http://www.vision.caltech.edu/bouguetj, Camera Calibration Toolbox for MATLAB.

[8] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 603-619, 2002.

[9] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Belief Propagation for Early Vision," International Journal of Computer Vision, vol. 70, no. 1, pp. 41-54, 2006.

[10] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint Bilateral Upsampling," ACM Transactions on Graphics, vol. 26, no. 3, pp. 1-5, 2007.

[11] J. Diebel and S. Thrun, "An Application of Markov Random Fields to Range Sensing," Proc. of Advances in Neural Information Processing Systems, vol. 18, pp. 291-298, 2006.