MRF-based Depth Map Interpolation using Color Segmentation

Jae-Il Jung and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Korea
E-mail: {jijung, hoyo}@gist.ac.kr Tel: +82-62-715-2258

Abstract— **A depth map is an essential data for high-quality 3D video services, but resolutions of the depth maps captured by commercially available depth cameras are smaller than those of the corresponding color images due to technical limitations. In this paper, we propose a new interpolation method for the depth map, considering discontinuity information of the corresponding color image. We define a Markov random field model for a segmented color image and a low-resolution depth map. We fill holes in the depth map, and match boundaries between the color image and the depth map. Experimental results show our method enhances the resolution of the depth map significantly.**

I. INTRODUCTION

The 3D movie 'Avatar' is the biggest box office hit in history, and other 3D movies such as 'Alice in wonder world' and 'Cloudy with a chance of meatballs' are following the flow of the box office hit. As 3D movies become a long-term box office success, interest on 3D video services rapidly increases.

The current 3D videos provide 3D feeling based on binocular depth cues, thus stereoscopic images captured with two cameras are widely used for 3D video generation. Recently, Moving Picture Experts Group (MPEG) has initiated a work aimed specifically toward free viewpoint system for viewing natural video, allowing the user to interactively control the viewpoint and to generate new views of various positions from dynamic scenes [1].

Fig. 1. Hybrid camera systems consisting of color and depth cameras

In order to acquire the views to allow high quality rendering of the scenes from any angle, the geometric information of the scene, called the depth map, is essential. Although it is commonly estimated by the stereo matching algorithm, its accuracy is not guaranteed for various scenes. To get accurate depth maps, the hybrid camera system has been proposed [2]. Figure 1 shows the hybrid camera systems of Realistic Broadcasting Research Center (RBRC), and they

consist of color cameras and depth cameras which simultaneously capture the color image and the corresponding depth map.

The hybrid camera system has the advantage, but it also has the problem that the resolution of depth maps captured by depth cameras is smaller than that of the corresponding color images due to technical limitations of the depth cameras. Figure 2 shows the resolution difference between the color image and the depth map of the hybrid camera system.
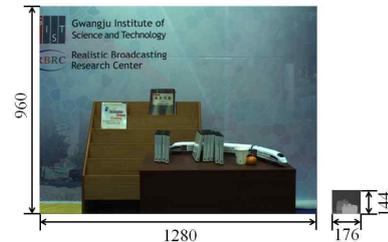
Fig. 2. Resolution difference between the color image and the depth map

Since the quality of depth maps is very important for image-based rendering, low-resolution depth maps should be enhanced. To solve this problem, various approaches have been proposed. In the beginning of the research, very simple approaches were exploited such as bilinear, nearest-neighbor, and bicubic interpolation methods [3]. 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 the adaptive weighting function according to the color image gradient [4]. They proposed the depth smoothness prior using the weighting factor reflecting color differences. Yang *et al.* presented the new post-processing step using the bilateral filter [5]. 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 they also do not consider the severe boundary mismatch problem between color images and depth maps.

In this paper, to enhance low-resolution depth maps with considering the boundary mismatch problem, we employ the color image segmentation and design a new posterior model

19

based on the MRF model.

II. MRF-BASED DEPTH INTERPOLATION USING COLOR SEGMENTATION

We propose a new interpolation method for the low-resolution depth map to effectively enhance the resolution. At first, the low-resolution depth is warped onto the position of the color image, and the color image is divided into small segments. Based on these two inputs, we design a new MRF model with considering color image discontinuities and the boundary mismatch problem. The solution is globally optimized with the graph cut algorithm.

A. 3D Warping

In the hybrid camera system, color images and depth maps are captured at the same time but at different positions. To get the depth map on the same position of the corresponding color image, we warped the depth map by the 3D warping technique [6].

At first, we carry out the camera calibration to obtain camera parameters for both the color and depth cameras. The camera parameter describes the relationship between the camera and world coordinates. It consists of an intrinsic parameter A and two extrinsic parameters: rotation matrix R and translation vector t. With the parameters, we can project the depth map onto the any position. We put the depth values of the depth map into the world coordinates using (1).

$$X_r = R_r^{-1} \cdot A_r^{-1} \cdot x_r \cdot d_r(x_r) - R_r^{-1} \cdot t_r \qquad (1)$$

where X_r represents the position in the real world coordinates for a pixel x_r in the depth map, and $d_r(x_r)$ has the return value of the corresponding depth value of x_r. After this backward projection, we re-project it onto the position of the color image using (2).

$$x_t = P_t X_r \qquad (2)$$

where x_t is the corresponding position of x_r in the depth map and, it is projected by using the projective matrix $P_t = A[R|t]$. Repeating the backward and forward projection, all values of the depth map are mapped onto the color image position. However the 3D warping technique cannot guarantee that the position of the depth map is perfectly matched with the position of the color image.

Figure 3 shows the structure after the 3D warping process. The gray nodes A and black nodes C mean warped depth values and color values, respectively. The white nodes B are the unknown depth values to be estimated. As can be seen in the figure, the warped depth values are very sparse and located at irregular positions. In addition, the depth values of node A are not accurate due to the boundary mismatch problem induced by the 3D warping process. Therefore, it is not straightforward to estimate proper depth values for B nodes.

B. Color Image Segmentation

Our approach co-aligns boundaries by considering the discontinuities of the color image. Although there is the previous approach that exploits gradient values of color image, it is not enough, especially for warped depth images. Thus we divide the color image into segments so that each segment has the similar color distribution [7].

The segments act like a criterion when we distinguish the valid depth value from its whole neighbors, and our MRF model is also based on the segments.
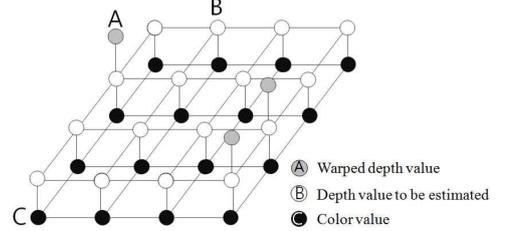
Fig. 3. Structure after the 3D warping

C. Markov Random Field Modeling

The warped depth map generated in the previous section contains many holes and mismatched boundaries. To fill the holes and to match the boundaries, we propose a MRF model considering correlation between color images and depth maps. The MRF theory provides a convenient and consistent way of modeling context-dependent entities such as image pixels and correlated features [8].

In the beginning, we compute the posterior distribution from the prior and the likelihood. According to the Bayes rule, the posterior probability can be computed by (3).

$$P(I_d \mid I_c, D_w) = \frac{P(I_c, D_w \mid I_d) P(I_d)}{P(I_c, D_w)} \qquad (3)$$

where $P(I_d)$ is the prior probability of the depth map I_d, $P(I_c, D_w \mid I_d)$ is the conditional probability density function of the color image I_c and the warped depth map D_w, called the likelihood function of I_d for fixed I_c and D_w. $P(I_c, D_w)$ is the density of I_c and D_w, which is a constant when I_c and D_w are given. $P(I_d|I_c,D_w)$ is proportional to the joint distribution, and the Maximum a posterior estimate (MAP) is equivalently found by

$$I_d^* = \arg\max_{I \in \hat{I}} \{ P(I_c, D_w \mid I_d) P(I_d) \} \qquad (4)$$

The posterior term is proportional to the exponential function of the energy function U as (5).

$$P(I_c, D_w \mid I_d) P(I_d) \propto e^{-U(I_d|I_c,D_w)}$$
$$U(I_d \mid I_c, D_w) = U(I_c, D_w \mid I_d) + U(I_d) \qquad (5)$$

Therefore, the MAP estimate is equivalently found by minimizing the energy function U as (7).

$$I_d^* = \arg\min_{I \in \hat{I}} U(I_d \mid I_c, D_w) \qquad (7)$$

With this property, we design the posterior energy function by considering the properties of the segmented color images and warped depth values. The posterior energy function consists of two terms for the likelihood and prior probabilities

$$U(I_d \mid I_c, D_w) = \sum_{i \in L} f_1(d_i, d_{wi}) + \sum_i \sum_{j \in N(i)} f_2(d_i, d_j) \qquad (8)$$

where i is the current position, and $N(i)$ means the neighbors of i. d and d_w are current and warped depth values, respectively. L stands for the position having a warped depth value.

If there is no mismatched depth value, the depth interpolation can be simply achieved. In this case, the accurate depth values can be calculated by referring to only neighbors in the same segment. In practice, several mismatched depth values caused by the 3D warping process exist, as shown in Fig. 4.
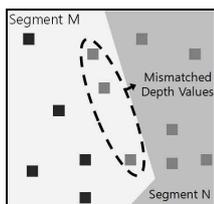
Fig. 4. Mismatched depth values near the boundary

In order to reduce the effects of these values, we model the likelihood energy function f_1 as (9). It penalizes the difference between warped depth values and current depth values, only when the position i has a warped depth node.

$$f_1(d_i - d_{wi}) = \begin{cases} w \mid d_{wi} - d_i \mid & if \mid d_i - d_{wi} \mid < T \\ \mid d_{wi} - d_i \mid & otherwise \end{cases} \qquad (9)$$

where T is the threshold value for depth difference, and w is the weighting factor for the small difference region. This value is set with the greater value than one, and it disturbs the mismatched depth values change slightly to minimize the prior energy. Figure 5 is the shape of the proposed likelihood energy function. The shape means the function gives less penalty to large difference values which are commonly induced by the boundary mismatch problem.
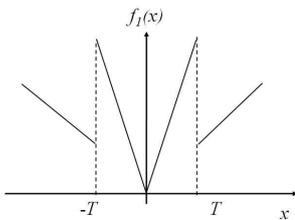
Fig.5. Shape of the proposed likelihood energy function

The prior energy function f_2 penalizes the violation of smoothness caused by the difference between a current depth value and neighbors. We use pair-site clique potentials for depth maps to be estimated. It only penalizes when depth values, d_i and d_j, belong to the same segment as (10).

$$f_2(d_i, d_j) = \begin{cases} \mid d_i - d_j \mid & if\ i, j \in a\ segment \\ 0 & otherwise \end{cases} \qquad (10)$$

D. Optimization

In general, minimizing the posterior energy function given in (8) is very difficult. Recently, several new algorithms for global optimization have been developed to efficiently solve the energy minimization problems. We exploit the graph cut algorithm [9] to approximate the global minimum of the proposed MRF model.

Figure 6 shows the energy fluctuation according to the number of iteration. Before the optimization, the prior energy is very small, while the likelihood energy has the greater value. During the optimization, the likelihood energy becomes smaller while the prior energy becomes greater. It means that mismatched depth values near boundaries are corrected. It increases the likelihood energy but decreases the prior energy slightly.
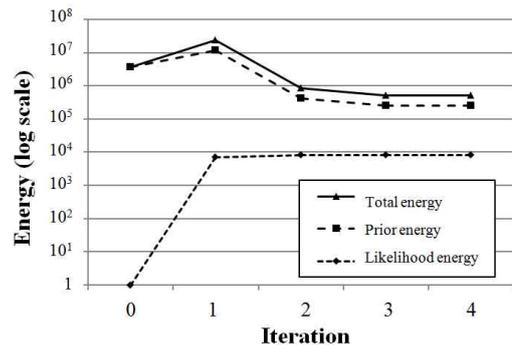
Fig. 6. Energy fluctuation during optimization

III. EXPERIMENTAL RESULTS

(a) (b) (c)
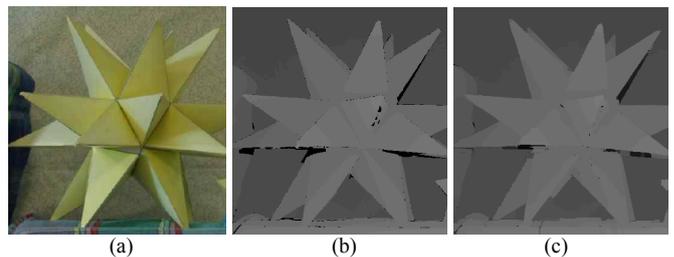
Fig. 7. Experimental result: (a) input color image, (b) ground truth depth map, and (c) our result

Figure 7 demonstrates the experimental result on a part of "Moebius" image provided by Middlebury. Fig. 7(a) is the original image, and Fig. 7(b) is the ground truth depth map. After down-sampling the ground truth depth map by the factor of 4, we applied our algorithm. As you can see Fig. 7(c),

the edge components of the depth map are successively preserved, and the small holes in the ground truth depth map are also filled with proper depth values.

We also took the test with images captured by the hybrid camera system. Figure 8(a) shows the warped depth values on the color image. There exist mismatched depth values near boundaries. Figure 8(b) is the simple result that only fills the holes with neighboring depth values. As you can see, the edge components are blurred, and the boundaries are not matched. On the other hand, our result in Fig. 8(c) shows the clear and well-matched boundaries.
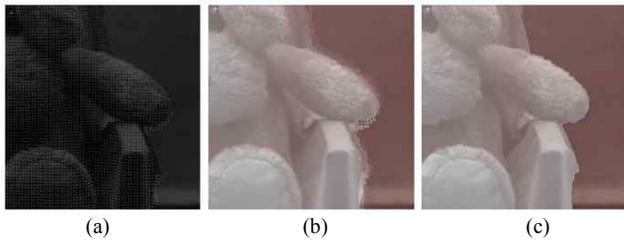
Fig. 8. (a) Warped depth values on the input color image, (b) interpolated depth map using neighboring values, and (c) our result

Figure 9 and Fig. 10 show rendering results. Fig. 9(a) and Fig. 10(a) are input color images, and Fig. 9(b) and Fig. 10(b) are their corresponding depth maps. Fig. 9(c) and Fig. 10(c) demonstrate interpolated depth maps. Fig. 9(d) and Fig. 10(d) show 3D mesh models. Our algorithm successively enhances the depth map from low- to high-resolution, and matches the boundaries between the color image and depth map.
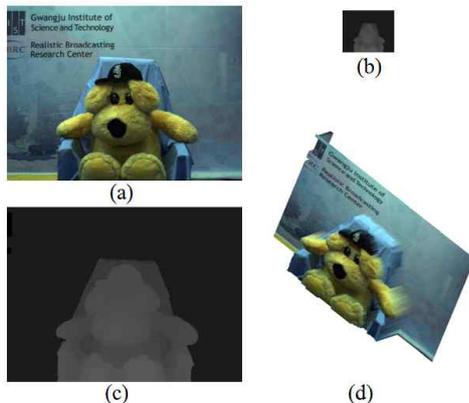
Fig. 9. Rendering result of "dog" image: (a) color image, (b) original depth map, (c) enlarged depth map, and (d) 3D mesh model

IV. CONCLUSIONS

The 3D video service is attracting much attention, and the geometric information of scenes is very important to render 3D scenes. Although accurate depth maps can be captured with the hybrid camera system, the depth map resolution is not sufficient due to technical limitations. In this paper, we have proposed the depth map interpolation method using color segmentation and the MRF model. To consider discontinuities of color images and the boundary mismatch

problem, we have designed the prior and likelihood energy function. From the experimental results, we can confirm that our proposed algorithm enhances the resolution of depth maps and fills the hole caused by up-sampling with reliable depth values. It also solves the boundary mismatch problem induced by the inaccurate 3D warping process. With the interpolated depth map, we can successively generate 3D mesh model.
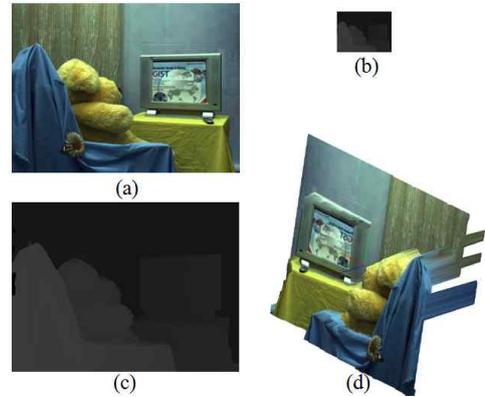
Fig. 10. Rendering result of "dog & TV" image: (a) color image, (b) original depth map, (c) enlarged depth map, and (d) 3D mesh model

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 N9784, "Introduction to 3D Video," May 2008.
[2] Y. Kang, E. Lee, and Y. Ho, "Multi-Depth Camera System for 3D Video Generation," Proceedings of International Workshop on Advanced Image Technology, pp. 44(1-6), Jan. 2010.
[3] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The art of Scientific Computing, 2nd ed.*, Cambridge university press, 1992.
[4] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," Advances in Neural Information Processing Systems, vol. 18, pp. 291-298, Dec. 2005.
[5] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," International Conference on Computer Vision and Pattern Recognition, June 2007.
[6] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," SPIE Stereoscopic Displays and Virtual Reality System XI, vol. 5291, pp. 93-104, Jan. 2004.
[7] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, May 2002.
[8] S. Li, *Markov Random Field Modeling in Image Analysis, 3rd Ed.,* Springer, 2009.
[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 11, Nov. 2001.