Depth Map Refinement Using Superpixel Label Information

Su-Min Hong and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea
E-mail: {sumin, hoyo}@gist.ac.kr

***Abstract*— In this paper, we propose a depth map refinement method to reduce the mismatch depth values along object boundaries. We design a filter which has segmentation, distance and color similarity weighting function. Our segmentation weighting term is defined based on SLIC superpixels because it delivers sufficient results at a very high speed. We give a penalty factor for the neighborhood pixels that are not within the same superpixel. Experimental results show the proposed method solve the problem that depth value propagation from one region to another region. So, the proposed method efficiently enhance the depth map quality. Experimental results shows the proposed method outperforms the conventional algorithms in terms of the RMSE for all the tested images. Therefore, they are expected to be used for various applications of 3D video processing**

I. INTRODUCTION

Accurate and high quality depth information is an issue of the importance in a number of applications, including FTV, image based rendering, 3DTV, view synthesis, among many others [1]. Various methods for acquisition of depth information have been researched and can be classified into two types: a passive sensor based method and an active sensor-based method. Passive depth sensing uses multi-view images to calculate depth information. The active depth sensing obtain the depth information using various types of sensors, such as, laser, infrared ray (IR), or light patterns. One of the most popular depth camera Kinect measure the range from the camera to an object in the captured scene using time-of-flight (ToF) technology. Although ToF depth camera make the depth map in real-time, there still have some problems such as boundary noise, distortion, and low resolution of the depth map.

Former image filter, such as bilateral filter (BF), can be directly used own image for depth map refinement. But filter based methods has problem like over smooth at the depth discontinuity regions. A joint bilateral upsampler (JBU) removes the over smooth at the depth discontinuity region by adding additional information [2].

In this paper, we propose a novel enhancement algorithm that takes associated color information into account to enhance the quality of edges in a depth map. We use the edge information found in the corresponding color image via a superpixels segmentation and compute a new refine depth map D_p which stores robust edge information corresponding to the color image.

II. DEPTH MAP REFINEMENT

A. Superpixels Based Image Segmentation

Superpixels provide an effective initial information from which to use local image features. They find redundancy in the image and greatly reduce the complexity of subsequent image processing tasks. They have proved increasingly useful for applications such as object localization, skeletonization, body model estimation, image segmentation, and depth estimation. To use the superpixels efficiently, they must be fast, easy to use, and produce high quality segmentations. We used a SLIC superpixels (simple linear iterative clustering) algorithm because it makes sufficient results for our purpose in a low computational time [3]. SLIC superpixels generates labels by clustering pixels based on their color similarity and closeness in the image plane. This is formulated in the five-dimensional [*labxy*] vector space, where [*lab*] is the pixel color intensity in CIELAB color space, which is widely considered as perceptually uniform for small color distances, and *xy* is the pixel coordinate. While the maximum possible distance between two image points in the CIELAB space (assuming sRGB input images) is limited, the spatial distance in the *xy* image plane depends on the image size. It is impossible to simply use the Euclidean distance in this 5D space without normalizing the spatial distances. To cluster pixels in this 5D space, they introduce a new distance measure that considers superpixel size. Using it, SLIC apply color similarity as well as pixel proximity in this 5D space such that the expected cluster sizes and their spatial extent are approximately equal.

First, SLIC method takes as input a desired number of approximately equally-sized superpixels *K*. For an image with *N* pixels, the approximate size of each superpixel is therefore *N/K* pixels.

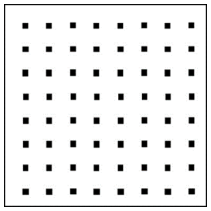


Fig. 1 Superpixel cluster centers

For roughly equally sized superpixels $S = \sqrt{N/K}$ represents a superpixel center at every grid interval. First step of their algorithm is choose the K superpixel cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ with $k=[1, K]$ at regular grid intervals S. Fig. 19 shows an example of the superpixel cluster centers. Since the spatial size of any superpixel is roughly S^2 (the approximate area of a superpixel), we can assume that pixels that are associated with this cluster center lie within a $2S \times 2S$ area around the superpixel center on the xy plane. This becomes the search area for the pixels nearest to each cluster center.

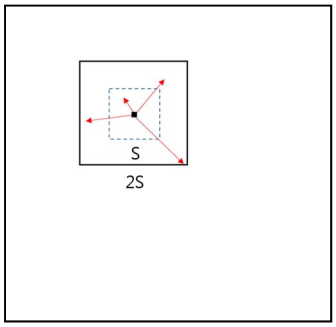


Fig. 2 Approximate area of the superpixel

Euclidean distances in CIELAB color space are perceptually useful for small distances. If spatial pixel distances exceed this perceptual color distance limit, then they begin to outweigh pixel color similarities. Therefore, they use a distance measure D calculated as follows:

$$\begin{aligned} \boldsymbol{d_C} &= \sqrt{(\boldsymbol{l_j} - \boldsymbol{l_i})^2 + (\boldsymbol{a_j} - \boldsymbol{a_i})^2 + (\boldsymbol{b_j} - \boldsymbol{b_i})^2} \\ \boldsymbol{d_s} &= \sqrt{(\boldsymbol{x_j} - \boldsymbol{x_i})^2 + (\boldsymbol{y_j} - \boldsymbol{y_i})^2} \\ \boldsymbol{D} &= \sqrt{\boldsymbol{d_c}^2 + \left(\frac{\boldsymbol{d_s}}{\boldsymbol{S}}\right)^2 \boldsymbol{m^2}} \end{aligned} \tag{1}$$

In (1), D means the sum of the lab distance and the xy plane distance normalized by the grid interval S. And m introduced in D is the control factor which allowing us to control the compactness of a superpixel. The greater the value of m, the more spatial proximity is emphasized and the more compact the cluster. This value can be in the range [1, 20]. They choose m = 10 for all the experiments. This roughly matches the empirical maximum perceptually meaningful CIELAB distance and offers a good balance between color similarity and spatial proximity. After that, they begin by sampling K regularly spaced cluster centers and moving them to seed locations corresponding to the lowest gradient position in a 3 ×3 neighborhood. This is done to avoid placing them at an edge and to reduce the chances of choosing an error pixels. Image gradients are calculated as:

$$\begin{aligned} \boldsymbol{G(x,y)} = &\|\boldsymbol{I(x+1,y)} - \boldsymbol{I(x-1,y)}\|^2 \\ &+ \|\boldsymbol{I(x,y+1)} - \boldsymbol{I(x,y-1)}\|^2 \end{aligned} \tag{2}$$

where I(x, y) is the lab vector corresponding to the pixel at position (x, y), and $\|\cdot\|$ is the L2 norm. This takes into account both color and intensity information. Each pixel in the image is associated with the closest cluster center whose search area overlaps this pixel. After all the pixels are grouped with the closest cluster center, a new center position is calculated as the average *labxy* vector of all the pixels belonging to the cluster. They then repeat the process of associating pixels iteratively with the nearest cluster center and recalculating the cluster center until convergence.

At the end of the SLIC, a few stray labels may remain, that is, a few pixels in the vicinity of a larger segment having the same label but not connected to it. While it is unique, this may arise despite the spatial proximity measure since their clustering does not perfectly enforce connectivity. However, SLIC make connectivity in the last step of their algorithm by relabeling disjoint segments with the labels of the largest neighboring cluster. Last step is *O(N)* complex and takes less than 10% of the total time required for segmenting an image.

Fig. 3 Result of the SLIC superpixels

B. Depth Refinement filter

In this chapter, we describe our depth refinement filter and our confidence weighting scheme for defining the weights W_p. Our refining filter modifies the conventional joint bilateral filter (JBF). Refined depth value D_p is computed by:

$$D_p = \frac{1}{W_p} \sum_{q \in \Omega} W_p D_q \tag{3}$$

The value of W_p defines the spatial coherence of neighborhood pixels. Our confidence weighting is decomposed into three terms based on segmentation (w_s), distance (w_d), and color similarity (w_c).

$$W_p = w_s w_d w_c \tag{4}$$

Our first term is defined based on color segmentation using the SLIC superpixels to segment an image into super pixels as shown in Fig 18. We tried different superpixel-segmentation methods including Mean-Shift but in the end because it delivers sufficient results for our purpose at a very high speed. For the neighborhood pixels that are not within the same super pixel, we give a penalty term defined as:

$$w_s = \begin{cases} 1 & if \quad S_{lab}(p) = S_{lab}(q) \\ p & otherwise \end{cases} \tag{5}$$

where S_lab(·) is the segmentation label, p is the penalty factor with its value between 0 and 1. In our implementation, we empirically set it equal to 0.3. The obtained segmentation is then projected into the depth stream. In an ideal depth map, the edges of the segmentation would coincide with the edges in the depth map. Unfortunately, the depth edges and color edges are not correspond, so this problem should be solved. Fig. 4 shows an example of the segmentation weighting term in our proposed method.

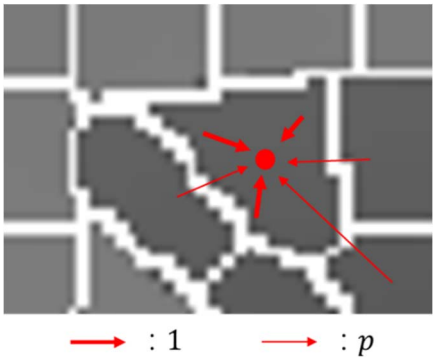
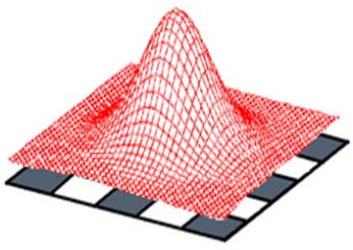


Fig. 4 Segmentation weighting

The distance weighting term and color similarity term are similar to conventional JBF. Two weighting functions w_d and w_c assign higher values to points closer to the central point p and to points with color intensity similar to $\tilde{I}_p$. Typically, the weights w_d and w_c are assigned according to Gaussian functions, respectively, with variance σ_d and σ_c. The distance between the coordinate points and between triplets in the color space are often computed according to the L2 norm.

$$\begin{aligned} w_d &= G_{\sigma_d}(\| p - q \|) \\ w_c &= G_{\sigma_c}(|\tilde{I}_p - \tilde{I}_q|) \end{aligned} \tag{6}$$

In other words, the distance weighting term gives higher weights to spatially close pixels and color weighting term gives higher weights to pixels with similar intensities. Fig. 5 shows an example of distance and color weighting terms.



120	120	50	50	90
120	120	50	50	50
120	120	120	120	120
30	30	120	120	120
30	30	30	30	30

Fig. 5 Distance and color similarity

III. Experiment Results

In order to evaluate the proposed depth upsampling algorithm, we have conducted experiments on 10 test image sets provided by Middlebury Stereo [4]. These are composed of color images and ground-truth depth maps. For objective evaluation, we performed experiment to fill out black areas of the error depth map provided by Middlebury Stereo. Note that the black areas are caused by inherent infrared sensor problem. We added some holes in the test image sets, which were similar with real captured depth maps.

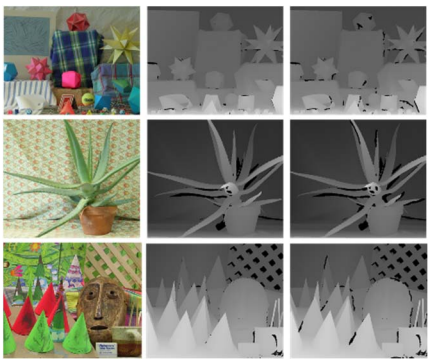


Fig. 6 Noisy Middlebuty dataset: color image, ground truth, noisy depth map

In order to evaluate the proposed preprocessing algorithm, we have conducted experiments on 3 test image sets provided by Lai et al: “Kitchen,” “Meeting 1,” and ”Meeting 3”. These are composed of color images and depth maps as shown in Fig. 7 [5]. This dataset was captured using a Kinect style 3D camera that records synchronized and aligned 640x480 RGB and depth images at 30 Hz.

Fig. 7 Image sets for experiments: “Kitchen” and “Meeting 1”

For the refining of the depth maps, the number of superpixels were set to 2000. Also, the length of one side in the local window for the filtering, variance of the distance weighting term σ_d and variance of the color weighting term σ_c were set to 11, 2.5 and 30. These fixed parameters were empirically defined. The refined depth maps by the proposed preprocessing filter using these parameters are shown in Fig. 8.

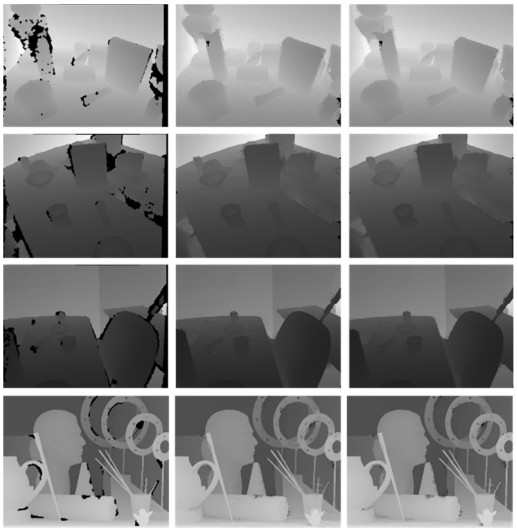


Fig. 8 Refined depth maps: JBF and proposed method

Fig. 9 shows detail of the refined depth maps of "Meeting 1," "Kitchen," and "Art". The expanded depth map shows proposed method can solve the depth propagation problem of the conventional refining method JBF. The refined results by the proposed algorithm have clean boundary regions with less noise compared to the JBF. The proposed method efficiently separated each objects without noises, and the object boundary regions of the refined depth map well matched compared with the JBF.

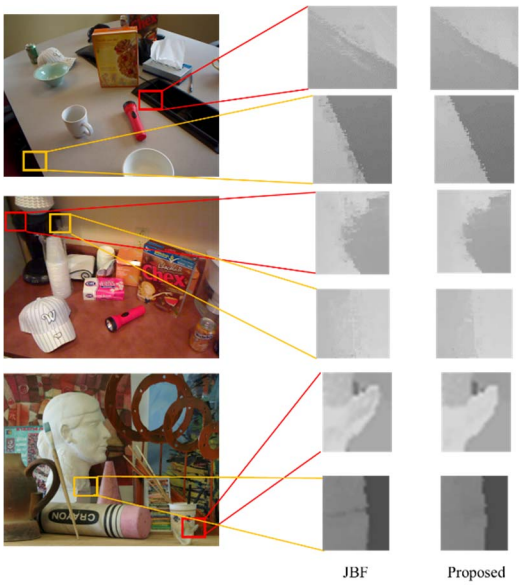


Fig. 9 Comparison of refined depth maps in detail

For objective evaluation, we compared the proposed algorithm with two other depth refining approaches: bilateral filter (BF) and joint bilateral filter (JBF). The comparison results using root-mean-square error (RMSE) with respect to the ground-truth depth maps. Table 1 shows quantitative comparison on the Middlebury datasets.

TABLE I
COMPARISON OF ROOT-MEAN-SQUARE ERROR

Dataset	BF	JBF	Proposed
cones	5.18	4.19	4.01
tsukuba	5.54	5.23	4.98
venus	1.32	0.98	0.86
books	3.46	1.99	1.76
bowling	4.57	3.86	3.44
baby	1.95	1.41	1.26
art	3.05	2.42	2.01
moebius	2.54	2.08	1.84
monopoly	2.84	1.91	1.77
aloe	4.84	4.18	3.59

IV. CONCLUSIONS

In this paper, we propose a novel depth map refining algorithm that takes associated color information into account to enhance the quality of edges in the depth map. We use the edge information found in the corresponding color image via a superpixels segmentation and compute a new representative depth map which stores robust edge information corresponding to the color. In the proposed method, confidence weighting is decomposed into three terms based on segmentation, distance and color similarity. With the proposed scheme, we could improve the depth accuracy along the object boundary. In comparison with conventional refining filter JBF, our preprocessing filter decreased the RMSE 0.27% in average error rate, and for the five most reduced error rates we achieved 0.38%.

ACKNOWLEDGMENT

This work was supported by the National Research Foundat ion of Korea(NRF) Grant funded by the Korean Government(MSIP)(No. 2011-0030079)

REFERENCES

[1] 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," in Proc. of IEEE International Conference on Multimedia and Expo, pp. 2161-2164, July 2006.

[2] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," ACM Transactions on Graphics. vol 26, no. 3, pp. 1-6, July 2007.

[3] R. Achanta, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels," EPFL, Lausanne, Switzerland, Tech. Rep. 149300, 2010.

[4] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 519-528, June 2006.

[5] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multiview RGB-D object dataset," in Proc. ICRA, pp. 1817–1824, May 2011.