# Efficient Depth Map Upsampling Method Using Standard Deviation

Su-Min Hong(✉) and Yo-Sung Ho

School of Information and Communications,
Gwangju Institute of Science and Technology (GIST), 123 Cheomdan-Gwagiro,
Buk-Gu, Gwangju 500-712, Republic of Korea
{sumin,hoyo}@gist.ac.kr

**Abstract.** In this paper, we present an adaptive multi-lateral filtering method to increase depth map resolution. Joint bilateral upsampling (JBU) increases the resolution of a depth image considering the photometric property of corresponding high-resolution color image. The JBU uses both a spatial weighting function and a color weighting function evaluated on the data values. However, JBU causes a texture copying problem. Standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data in image. Therefore, it includes an edge information in the each kernel. In the proposed method, we decrease the texture copying problem of the upsampled depth map by using adaptive weighting functions are chosen by the edge information. Experimental results show that the proposed method outperformed compared to the other depth upsampling approaches in terms of bad pixel rate.

**Keywords:** Depth map upsampling · Joint bilateral upsampling

## 1 Introduction

Accurate and high resolution depth sensing is an issue of the importance in a number of applications, including 3DTV, image based rendering, view synthesis, among many others. 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 [1]. In the active sensor-based methods, depth information is obtained from the object directly using physical sensors, such as infrared ray (IR) sensors [2]. Recently, KINECT depth camera and Time-of-Flight (ToF) range camera became a popular alternative for dense depth sensing. Although ToF cameras can capture depth information for object in real time, but are noisy and subject to low resolutions. For example, the current ToF camera 'Mesa Imaging SR4000' provides a 176 × 144 depth map up to 30 frames per second. For the actual utilization, we need the same resolution of the depth image and color image [3]. Therefore, an efficient depth map upsampling method is necessary.

Several techniques have been developed to solve problems of the depth map captured by ToF cameras. Gaussian filtering and other smoothing filtering have bad


Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 33–40, 2015.
DOI: 10.1007/978-3-319-24078-7_4

results in depth discontinuous region. Therefore, many depth map upsampling methods based on edge preserving.

A joint bilateral upsampling (JBU) removes the over smooth at the depth discontinuity region by adding additional information [4]. The information is an original color image used in depth estimation. JBU uses the two weighting functions with respect to photometric similarity of the neighbor pixels. However, there are two problems on JBU: using the color image as a guide cause edge blurring some regions and texture copy from the color image to the depth map. To solve these problems, noise-aware filter for depth upsampling (NAFDU) is adaptively apply the color weighting function based on edge information [5]. Also, Markov random field (MRF) based method is proposed in [6]. MRF depth map upsampling defined the probability model by MRF using color and depth information, and the model is optimized by conjugate gradient to provide upsampled depth map.

In this paper, we propose a depth map upsampling method that uses the edge information to decide the weighting functions. To get the edge information, we calculate the standard deviation value in each kernel. Using this value, we apply the blending function for depth map upsampling. Blending function defines the how much contributes to the each weighting functions. By applying blending function to the depth map upsampling, we obtain the high resolution depth map without texture copying of color guide image. In addition, the proposed upsampling method makes high-resolution depth map while protecting edge region.

## 2 Related Work

Low resolution of depth map provided by ToF cameras cause the problems in some 3D applications making. Using the color image information for the upsampling is a valid approach to enhance and improve the ToF data. Kopf et al. presented joint bilateral upsampling (JBU), a modified version of the joint bilateral filter for the depth map upsampling considers guide color image information [4]. In the JBU, the upsampled depth map is made via the Gaussian weighted sum of neighbors within a filter kernel. Assume that there are input low resolution depth map $I$, output high resolution depth map $\tilde{S}$ and high resolution color image $\tilde{I}$. Formally, the depth value $\tilde{S}_p$ at $p$ in an upsampled depth map $\tilde{S}$ is computed by JBU as Eq. (1).

$$\tilde{S}_p = \frac{1}{k_p} \sum\nolimits_{q_\downarrow \in \Omega} I_{q\downarrow} f\left(\left\|p_\downarrow - q_\downarrow\right\|\right) g\left(\left\|\tilde{I}_p - \tilde{I}_q\right\|\right) \tag{1}$$

where $p_\downarrow$ and $q_\downarrow$ denote the pixel coordinate in a low resolution depth map and $p$ and $q$ are denote the corresponding coordinates in the high resolution color image. $\Omega$ represent the spatial neighborhood around $p$, and $k_p$ is a normalizing factor. In this equation $f$ and $g$ respectively represent the spatial filter and range filter that have Gaussian distribution, and $\|\cdot\|$ is an Euclidean distance operator. The final goal of JBU is upsample of low resolution depth map while protecting the edge information.

Although the JBU can upsampled the depth map with protect the edge region, it has texture copy from the color image to the upsampling result. Texture copying problem is caused by the different discontinuity information between the color and depth data. Also, it is found in if depth sensor contains a lot of random noise. To solve this problem, the Noise-Aware Filter for Depth Upsampling, shortly NAFDU, was proposed [5]. NAFDU is a kind of adaptive multilateral filter that blends two range filters of color and depth information. Upsampled depth value $\tilde{S}_p$ is computed by NAFDU as Eq. (2).

$$\tilde{S}_p = \frac{1}{k_p}\sum\nolimits_{q_\downarrow \in \Omega} I_{q_\downarrow} f\left(\|p_\downarrow - q_\downarrow\|\right)\left[\alpha(\Delta\Omega) g\left(\|\tilde{I}_p - \tilde{I}_q\|\right) + (1-\alpha(\Delta\Omega)) h\left(\|I_{p_\downarrow} - I_{q_\downarrow}\|\right)\right] \tag{2}$$

where $\alpha$ denotes the blending function and $\Delta\Omega$ represent the depth difference between the minimum and maximum depth values in the each filter kernel. Blending function of NAFDU is defined by Eq. (3).

$$\alpha(\Delta\Omega) = \frac{1}{1 + e^{-\varepsilon(\Delta\Omega - \tau)}} \tag{3}$$

The blending function is dependent to the depth value difference ($\Delta\Omega$). So, if the $\Delta\Omega$ is increased, $\alpha$ is close to 1 and NAFDU works like JBU. Otherwise, NAFDU use the depth information in the range filter, so it works like bilateral upsampling.

## 3 Efficient Depth Map Upsampling Using Edge Information

The proposed method modifies the conventional NAFDU. First of all, we compute the standard deviation in filter kernel. Based on standard deviation value, we estimate the edge region in the image. According to the edge information, we apply the blending function for adaptive weighting. For the edge region, our proposed method works like JBU. Otherwise, proposed method applies the bilateral upsampling using the median filtered depth map. Figure 1 shows the overall structure of our proposed method.

### 3.1 Edge Region

The goal of depth map upsampling is to estimate a high quality and high resolution depth map. It is difficult to expect a good result of depth map upsampling without considering the edge region. In our algorithm, we use the standard deviation to get the edge information. In statistics, the standard deviation ($\sigma$) is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A standard deviation close to 0 indicates that the data points tend to be very close to the mean of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values [7]. Standard deviation in the filter kernel is defined by Eq. (4).

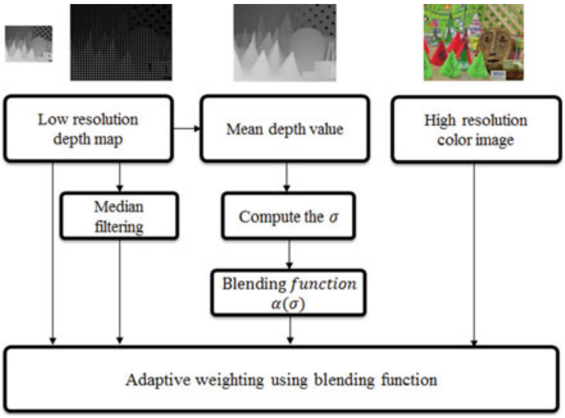



**Fig. 1.** Overall structure of the proposed method

$$\sigma_p = \sqrt{\frac{\sum_{i=1}^{n}\left(q_i - M_p\right)^2}{n}} \tag{4}$$

where $p$ denote the center pixel in the filter kernel and $q$ represent the neighbor pixels in the filter kernel. $n$ is the number of pixels in the filter kernel. Also, we can get the mean value in the filter kernel using Eq. (5).

$$M_p = \frac{1}{n}\sum_{i=1}^{n} q_i \tag{5}$$

The standard deviation is a measure of how spreads out pixel values are. In addition, it can represent the dispersion of a set of pixel values from the filter kernel. So, we can decide which region is the edge by using standard deviation.

### 3.2 Blending Function

Our proposed method use the adaptive weighting functions based on edge information. Blending function is given by Eq. (6).

$$\alpha(\mathrm{x}) = \frac{1}{1 + e^{-\varepsilon(x-\tau)}}, \mathrm{x} = \frac{\sigma}{\sigma_{max}} \tag{6}$$

blending function is the distribution of the 0 to 1 [5]. Here x is the ratio between the result of Eq. (4). and the maximum standard deviation in each kernel. The largest standard deviation possible will be found where the largest possible gaps are. So, we can calculate the largest standard variation very easy. The maximum standard deviation is calculated as in Eq. (7) [8].

$$\sigma_{max} = \sqrt{(M - \min(\Omega))(\max(\Omega) - M)} \tag{7}$$

where $\Omega$ represent the spatial neighborhood around $p$. $M$ indicates the mean in the $\Omega$ and min, max represents the minimum depth value and maximum depth value in the $\Omega$. The standard deviation distribution is normalized in 0 to 1 by consider maximum standard deviation. Also, it can represent how much edge information is included in each kernel. If standard deviation is very small in some region then the x have almost zero. Otherwise, x is the close to one when the standard deviation is almost same with maximum standard deviation. So, we can determine the edge in the filter kernel by using the x value (Fig. 2).


(a) Blending function (b) Ground truth depth (c) Edge region (x≥0.5)

**Fig. 2.** Blending function and Edge region

### 3.3 Adaptive Multilateral Upsampling

Assume that there are input low resolution depth map $I$, output high resolution depth map $\tilde{S}$, high resolution color image $\tilde{I}$ and median filtered depth map $M$ (in case of input depth map captured by ToF camera). After blending function is determined, upsampled depth value $\tilde{S}_p$ is defined by Eq. (8).

$$\tilde{S}_p = \frac{1}{k_p} \sum\nolimits_{q_\downarrow \in \Omega} I_{q_\downarrow} f\left(\left\|p_\downarrow - q_\downarrow\right\|\right) \left[\alpha(x) g\left(\left\|\tilde{I}_p - \tilde{I}_q\right\|\right) + (1 - \alpha(x)) h\left(\left\|M_{p_\downarrow} - M_{q_\downarrow}\right\|\right)\right] \tag{8}$$

Here $p$ and $q$ denote the pixel coordinate in a high resolution color image and $p_\downarrow$ and $q_\downarrow$ are denote the corresponding coordinates in the low resolution depth map. $\Omega$ is a spatial neighborhood around target pixel. $f$, $g$ *and* $h$ are all Gaussian functions, and $\alpha(x)$ is a blending function.

If the x is increased, $\alpha$ is close to 1 and proposed method upsample of low resolution depth map while protecting the edge information. Otherwise, If the x is decreased, $\alpha$ is close to 0 and proposed method use the median filtered depth map information. Noise in ToF cameras is due to systematic errors and other sources impacting the measurements [9]. For example, surfaces with low infrared reflectance properties result in error region of the depth map. So, input depth map is median filtered to reduce the amount of noise.

## 4 Experimental Results

To evaluate the performance of the proposed upsampler, we used four test image sets Cones, Teddy, Tsukuba and Venus provided by Middlebury stereo were used [10].

These data sets are omposed of color images and ground-truth depth maps as hown in Fig. 3. So, we use the original depth map information in the range filter *h*. The ground truth depth maps were downsampled by factors of 2, 4 and 8 using the nearest

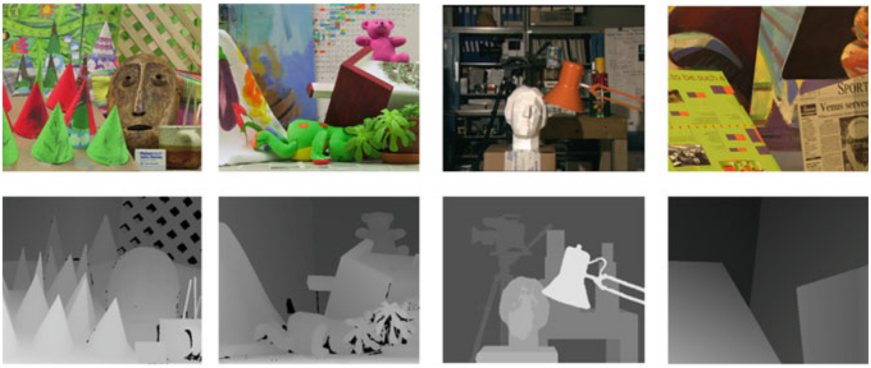
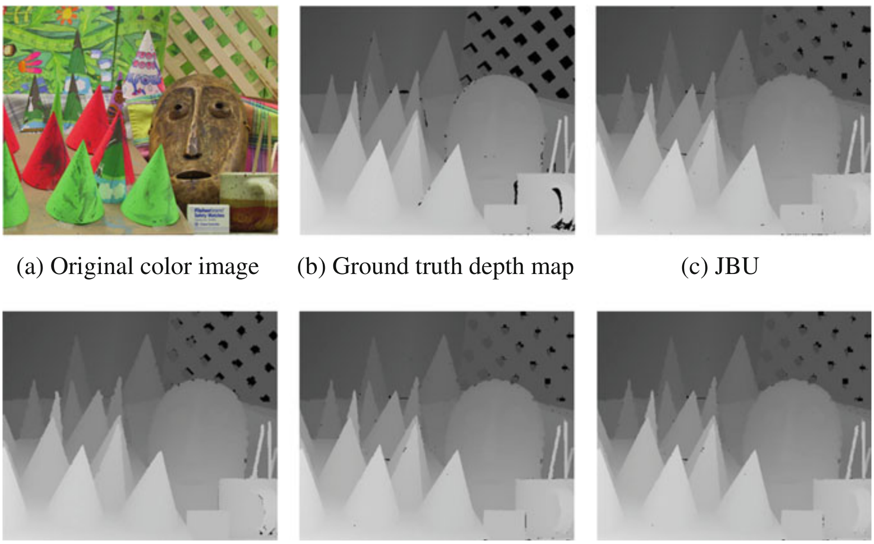


**Fig. 3.** Image sets for experimet



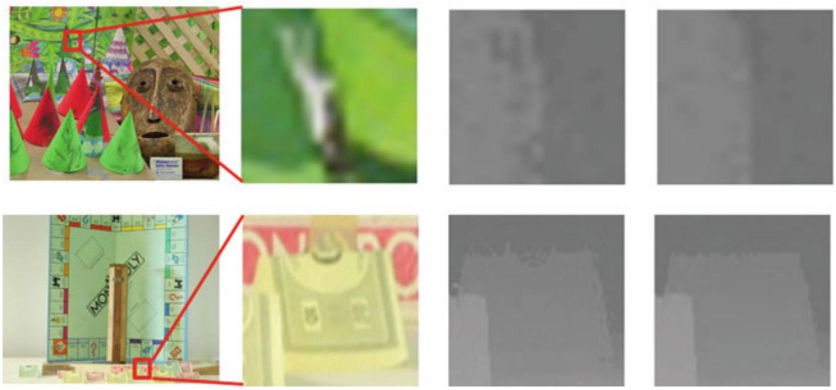
(a) Original color image (b) Ground truth depth map (c) JBU

(d) MRF (e) NAFDU (f) Proposed method

**Fig. 4.** Upsampled depth maps

**Table 1.** Performance comparison

| Dataset | Scale | JBU | MRF | NAFDU | Proposed |
|---|---|---|---|---|---|
| Cones | 2× | 2.50 | 2.84 | 2.49 | 1.92 |
| | 4× | 4.31 | 4.12 | 4.42 | 3.84 |
| | 8× | 10.01 | 10.02 | 10.02 | 10.01 |
| Teddy | 2× | 4.41 | 4.81 | 4.44 | 4.09 |
| | 4× | 6.12 | 6.48 | 6.09 | 6.01 |
| | 8× | 11.12 | 11.21 | 11.19 | **11**.09 |
| Tsukuba | 2× | 3.41 | 4.71 | 3.44 | 2.92 |
| | 4× | 5.31 | 5.39 | 5.31 | 5.12 |
| | 8× | 8.84 | 8.69 | 8.94 | 8.21 |
| Venus | 2× | 0.83 | 1.12 | 0.93 | 0.26 |
| | 4× | 0.92 | 1.2 | 0.91 | 0.76 |
| | 8× | 3.81 | 4.05 | 3.80 | 3.73 |



(a) Original color image (b) JBU (c) Proposed method

**Fig. 5.** Enlarged texture copying area

neighbor method. In the experiment, the length of one side in the local window for the blending function, the length of one side in the local window for the filtering, variance of the spatial filter $\sigma_s$, variance of the range filter $\sigma_r$ were set to factor 2: 7, 7, 1, 3, factor 4: 9, 7, 3, 8 and factor 8: 17, 11, 5, 10.

Also, ε and τ of the blending function were set to 50 and 0.3 in this test. The proposed method is compared with the joint bilateral upsampling (JBU) [4], the noise-aware filter for depth upsampling (NAFDU) [5] and the MRF-based depth up-sampling [6]. Figure 4 shows the upsampled depth maps of Cones of the scaling factor 4.

Although the JBU can reconstruct depth edges, it has texture copying problem. So, proposed method use the adaptive weighting functions based on edge information (Fig. 5).

For an objective evaluation of the depth maps, the bad pixel rate (BPR), whose absolute difference is greater than 1, was used.

Table 1. show the result of the BPR(%) comparison. According to the results, the proposed method outperforms the conventional algorithms in terms of the BPR for all the scaling factors and all the tested images.

## 5 Conclusion

In this paper, we have proposed a multilateral depth map upsampling method for low resolution depth maps. The proposed upsampler is based on edge information, and we define the edge region that considers standard deviation in the image. After that, our proposed method use the adaptive weighting functions based on blending function. It is an efficient method because it increases depth map resolution while protecting the edge region. In addition this method can solve the texture copying problem. From the experimental results, the proposed method outperformed compared to the other depth upsampling approaches in terms of bad pixel rate.

**Acknowledgements.** This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea(ROK). [GK15C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

## References

1. Okutomi, M., Kanade, T.: A multiple-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **15**(4), 353–363 (1993)
2. C.S. Swiss Ranger SR-2. The Swiss center for electronics and microtechnology. http://www.csem.ch/fs/imaging.htm
3. Fehn, C., Barre, R., Pastoor, S.: Interactive 3-D TV concepts and key technologies. Proc. IEEE **94**(3), 524–538 (2006)
4. Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. ACM Trans. Graph. **26**(3), 1–6 (2007)
5. Chan, D., Buisman, H., Theobalt, C., Thrun, S.: A noise-aware filter for real-time depth upsampling. In: ECCV Workshop on Multi-camera and Multi- modal Sensor Fusion Algorithms and Applications, pp. 1–12 (2008)
6. Diebel, J., Thrun, S.: An application of markov random fields to range sensing. In: Proceedings of Advances in Neural Information Processing Systems, pp. 291–298 (2006)
7. http://en.wikipedia.org/wiki/Standard_deviation
8. Baek, E., Ho, Y.: Gaussian alpha blending for natural image synthesis. In: Proceedings of Korean Institute Smart Media, vol. 4(1), pp. 138-141(2015)
9. Schwarz, L.A., Mkhistaryan, A., Mateus, D., Navab, N.: Human skeleton tracking from depth data using geodesic distances and optical flow. Image Vis. Comput. **20**(1), 217–226 (2012)
10. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(13), 7–42 (2002)