ICIC Express Letters
Volume **10**, Number **10**, October **2016**

ICIC International ©2016 ISSN 1881-803X
pp. **2479–2484**

DEPTH RECOVERY BASED ON IMAGE GUIDED MULTI-POINT FILTERING

Li Li, Youen Zhao, Xiaohong Shen and Caiming Zhang

School of Computer Science and Technology
Shandong University of Finance and Economics
No. 7366, Erhuan East Road, Jinan 250014, P. R. China
{ lilisd; yzh; xhsh }@sdufe.edu.cn

Received March 2016; accepted June 2016

Abstract. *The depth maps captured by consumer RGB-D cameras are usually degraded by noise, low resolution, missing values, etc. This paper presents a novel depth recovery method to improve the spatial resolution and quality of initial depth map. Our solution is based on an image guided multi-point filtering framework. Different from the conventional point-wise filtering, the estimates are calculated for all observation pixels in the multi-point filtering. Firstly, we calculate the depth estimates of pixels in an adaptive support region with a piecewise constant model. The adaptive support region is determined based on the distance preserving domain transform technology guided by the color image. Then a number of such estimates are aggregated together by a weighted averaging strategy to acquire final depth estimates. By quantitative and qualitative experiments on publicly available test sequences, we demonstrate the capabilities of our method on the depth recovery task.*

Keywords: Depth recovery, Guidance image, Multi-point filtering

1. Introduction.

A depth map represents the distance of each point to a reference camera and can be used in many applications such as 3DTV, new view rendering, and robot vision. The high quality and resolution depth maps are required in these applications. However, there exists at the moment no depth map generation technique that is able to produce a perfect depth map. For instance, the depth maps obtained by a ToF (time of flight) range sensor, 'Mesa Imaging SR4000', are of low resolution (176 × 144) and noisy. Due to this limitation, the subject of depth recovery has been extensively studied. Usually there are different recovery methods to address different aspects of depth map corruption such as low spatial resolution, noise, blur edge, holes, and low accuracy. In this paper, we focus on the improvement of the spatial resolution and accuracy of non-ideal low resolution and noisy depth map.

In filtering-based methods, joint bilateral filtering (JBF) is widely used and may work well, because it can preserve the discontinuity of up-sampled depth map with the help of the accompanied color image. However, one challenging problem of the JBF is its high computational complexity. In recent years, several methods enable joint bilateral filtering to be computed at constant time or even video rate by modifying the model or using GPU implementation [1, 2, 3]. The guided image filtering (GF) was proposed recently [4] and has demonstrated its unique advantage over JBF in some applications such as stereo matching and HDR (high dynamic range) compression [4, 5, 6]. Due to the linear model and using integral image technique, the guided filtering runs much faster than JBF method. According to [7], GF is essentially a multi-point estimator which calculates the estimates of all observation pixels, compared with the point-wise estimator JBF which calculates the estimate of a single pixel only. However, GF has a fixed-sized square filter window and simply averages for multi-estimate which may generate fuzzy object boundaries in the depth recovery task. An example is shown in Figure 1.

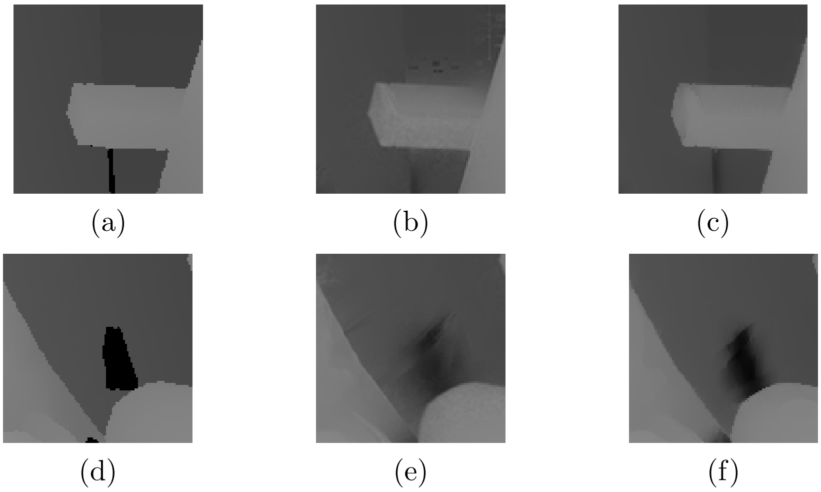

Figure 1. An example: Detail crop of recovered Teddy (the square windows in Figure 2), (a) and (d) Truth depth map, (b) and (e) results of GF, (c) and (f) results of our method

Motivated from the above researches, we proposed a new depth recovery method based on a image guided multi-point filtering model, as the expansion of the guided filtering. Different from GF, our multi-point filtering has an adaptive window and a weighted average for multiple estimates, which makes our method potential to give better results. To decide the adaptive support region, the domain transforming technique [8] is used here. The transform enables the 2D filtering to be performed using a sequence of 1D filters, which lowers computational and memory costs compared with conventional 2D filters. The key contribution of this paper is to integrate domain transforming technique into multi-point filtering and finish depth map recovery efficiently and effectively. This paper is the revised version of our previous research [9] with no public release.

The remainder of this paper is organized as follows. Section 2 describes our proposed depth map recovery method. Then Section 3 presents objective and subjective experimental results. Finally, Section 4 gives our conclusions.

2. Depth Map Recovery.

Given a depth map D, our goal is to improve its spatial resolution and quality using aligned high resolution color image $\mathbf{I}$. The resulting recovered depth map at the same resolution of $\mathbf{I}$ is denoted as J.

2.1. Algorithm overview.

First we upscale the raw depth map to the same size as color image by bilinear interpolation. The interpolated depth map D_{ini} is not reliable. Then an adaptive multi-point filtering is performed to improve its quality. The multi-point filtering includes the following steps. Firstly, for each pixel p, a set of four varying support arm lengths is decided, which is based on the image guided domain transforming technique. Once such four arm lengths of each pixel p are decided, an adaptive support region Ω_p is available. Secondly, we use the piecewise constant model to compute multi-point estimates J_s^k for a set of points $s \in \Omega_k$. For a pixel p in the support region Ω_k, it then has an estimate J_p^k. And a pixel p is generally covered by multiple support regions and has a number of multi-point estimates $\left\{J_p^k | p \in \Omega_k\right\}$. Finally, the multi-point estimates are fused by a weighted averaging strategy to obtain the output J_p for each pixel p.

2.2. Adaptive support region.

With the help of aligned color image, we use the domain transforming technique to decide for each direction an appropriate arm length of each pixel p in the initial depth map D_{ini}. So they jointly delineate an adaptive support region Ω_p. The domain transform recently presented [8] is a dimensionality reduction technique, which is a distance-preserving transform and exists for a 1D domain. For a 1D signal

I embedded in 2D $(x, I(x))$ space, a transformation $t : \mathrm{R}^2 \rightarrow \mathrm{R}$ preserves L_1 distance between two neighboring pixels in the original domain R^2 and in the new dimensionality reduction domain R, which is expressed as

$$gt(x+h) - gt(x) = h + |I(x+h) - I(x)| \tag{1}$$

where $gt(x) = t(x, I(x))$ represents the transformation operator at pixel x, and h is the sampling interval. After the derived strategy presented in [8], the values at any pixel u in the new domain can be computed by

$$gt(u) = \int_0^u 1 + \frac{\sigma_s}{\sigma_r} |I'(x)| dx \tag{2}$$

where σ_s and σ_r control the influence of spatial and intensity range information respectively similar to the bilateral filtering. For multichannel signal such as a RGB color image I embedded in 4D (x, I_R, I_G, I_B) space, the transformation $t : \mathrm{R}^{C+1} \rightarrow \mathrm{R}$ can be expressed as

$$gt(u) = \int_0^u 1 + \frac{\sigma_s}{\sigma_r} \sum_{k=1}^{C} |I'_k(x)| dx \tag{3}$$

where I_k is the k-th channel of signal I, and C is the number of channels.

Unfortunately, for a 2D image signal, there exists no such transformation in general as described in [8]. In this work, we use 1D transform to perform 2D filtering. That is, for a 2D color image $\mathbf{I}$, the horizontal and vertical passes are conducted for each row and column using Equation (3) respectively. The constant radius r is adopted based on the transformed values to decide which pixels are included in the support region, and then horizontal and vertical arm lengths of each pixel p in initial depth map D_{ini} are derived. It is worth noting that the radius is constant in new domain, but a space-varying and non-symmetric radius in original domain, in which its size changes according to the similarity between two neighboring pixels in 4D $(x, \mathbf{I}_R, \mathbf{I}_G, \mathbf{I}_B)$ space. Once the four arm lengths are decided for each pixel p, an adaptive filter support region Ω_p can be defined as an area integral of multiple horizontal segments $H(q)$. That is expressed by $\Omega_p = \bigcup_{q \in V(p)} H(q)$, where q is a pixel located on the vertical segment $V(p)$ of pixel p. Obviously the adaptive support region Ω_p only includes pixels belonging to the same population as p to support multi-point filtering described below.

2.3. **Multi-point filtering.** After a pixel-wise adaptive support region Ω_p for each pixel p is given, we use a multi-point filtering similar to the method presented in the paper [10] to perform depth map recovery. For the multi-point filtering, the output depth is supposed to be a linear transformation of the guidance color image. For a pixel p in an adaptive support region Ω_k centered at a pixel k, its output depth J_p^k can be expressed as

$$J_p^k = \mathrm{a}_k^T \mathbf{I}_p + b_k \tag{4}$$

where $p \in \Omega_k$, $\mathbf{I}_p$ is a 3×1 RGB components vector of pixel p, a_k is a 3×1 coefficient vector, and a_k and b_k are constant parameters corresponding to the support region Ω_k. The parameters a_k and b_k can be computed by minimizing the difference between the output value J_i^k and input D_{ini}^i of each pixel i in the support region Ω_k. [10] has proven that a lower-order fitting model can help depth recovery task without causing blurry boundaries as the guided image filtering does. As the extension of GF, the piecewise constant model is used here to fit data between guidance color image and initial depth map. Then the parameter a_k is set to zero and the above equation is reduced to

$$J_p^k = b_k \tag{5}$$

Similar to GF, the window parameter b_k can be determined by minimizing differences between input image D_{ini} and output image J. It has been proven that b_k can be expressed as

$$b_k = \frac{1}{|\Omega_k|} \sum_{s \in \Omega_k} D_{ini}^s \tag{6}$$

where $|\Omega_k|$ is the number of pixels in Ω_k. It is worth noting that the multi-point filtering is used here to calculate an estimate J_s^k for all pixels in support region, i.e., $s \in \Omega_k$, which is contrast to the pixel-wise filtering that gives the central pixel estimate only.

Then we can apply the linear transformation model to all the support regions in the entire image. However, a pixel p is often involved in different support regions that contain p and have different window parameters. It is hence a number of different estimates $\left\{J_p^k | p \in \Omega_k\right\}$ for each pixel p. Taking the confidence of each estimate into account, we use the weighted averaging to compute the final output by

$$J_p = \frac{\sum_{k:p \in \Omega_k} w_k J_p^k}{\sum_{k:p \in \Omega_k} w_k} \tag{7}$$

where w_k is the relative weight for each estimate J_p^k. As the adaptive support region is intended to involve similar pixels with the central pixel, we set the number of pixels in support region as its corresponding weight. So the fusion equation is rewritten as

$$J_p = \frac{\sum_{k \in \Omega_p} |\Omega_k| J_p^k}{\sum_{k \in \Omega_p} |\Omega_k|} \tag{8}$$

Note that, for more easily data process and computation, we have modified the summations of J_p^k for $k : p \in \Omega_k$ in Equation (7) into those for $k \in \Omega_p$ in Equation (8). This is an approximate transformation which may not always hold. With this modification, about four O(1) time multiple estimates fusion over 2D adaptive support region are needed in the constant model by using the integral image technique. By combining the domain transform with multi-point filtering, initial depth map can be enhanced exactly and efficiently by our proposed method.

3. Experimental Results.

To validate the effectiveness and efficiency of the proposed method, we evaluate our method through various experiments. The performance was compared with the JBF-based method and the GF-based method which represent the top performances. For the three algorithms, we use our own Matlab implementation using the Intel Core i3 CPU, 2.3GHZ PC. We perform experiments using ground truth depth maps provided by the Middlebury test bed [11, 12].

We evaluate the performance of up-sampling low resolution depth maps firstly. The low resolution depth maps are generated by down-sampling the ground truth disparity maps. The down-sampling ratio is set to 8. The low resolution depth maps are firstly up-sampled to the same size as high resolution color image by bilinear interpolation. The interpolated depth maps are called initial depth maps and then recovered by our proposed method. For optimal results, our parameters are set: $\sigma_s = 10$, $\sigma_r = 0.2$, and window radius $r = \sqrt{3}\sigma_s$. The results of our method compared with the JBF-based and the GF-based method are given in Figure 2 and Figure 1. The JBF-based and the GF-based methods perform the joint bilateral filtering and the guided filtering on initial depth map respectively. In order to fairly compare performances of filtering-based three methods, we adopt the same initial depth maps. The other parameters of two compared methods are adjusted to acquire the optimal results. From the figures, we can see that the proposed method yields superior results over the two compared methods, especially in discontinuity and occluded areas. The objective evaluation of these methods is shown in Table 1. The accuracy is evaluated by measuring the percent of bad pixels (where the absolute disparity error is greater than 1) for $Vis.$ (visible pixels in the image) and $Dis.$ (near discontinuity area) pixels.

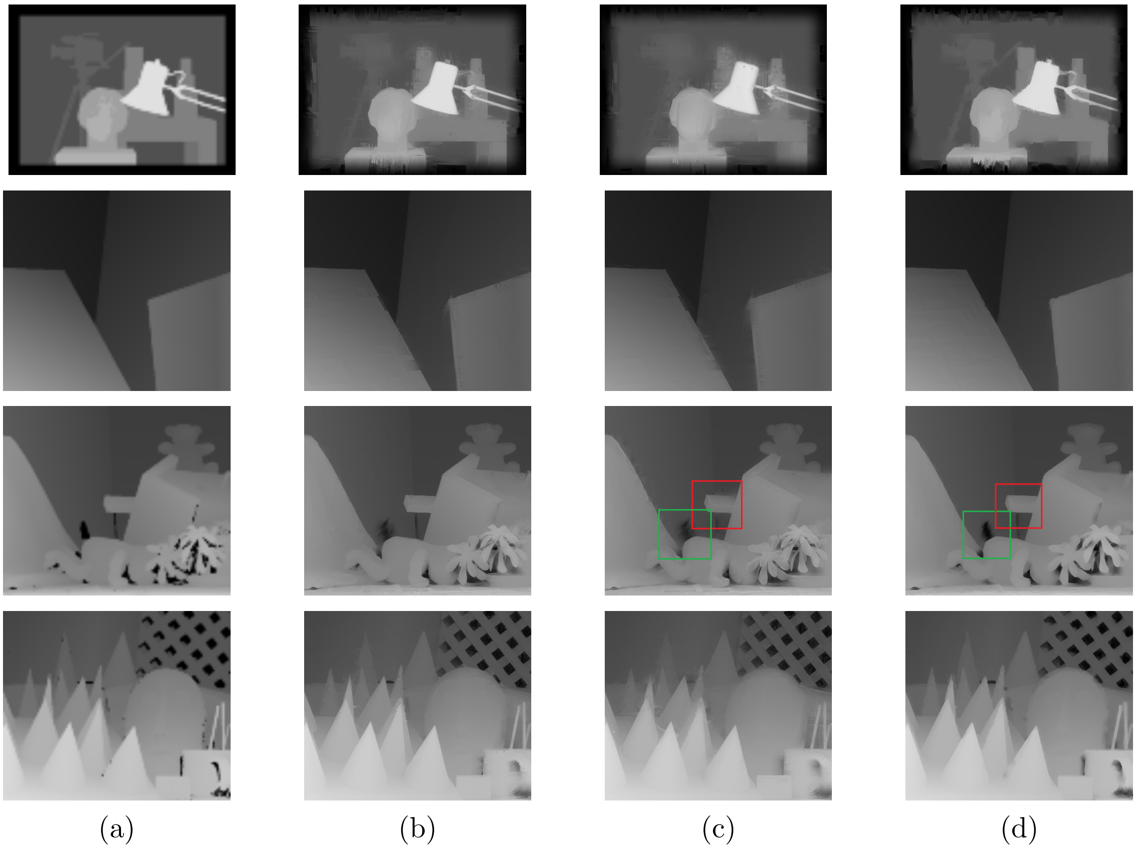

Figure 2. Depth up-sampling results for "Tsukuba, Venus, Teddy and Cones": (a) The initial depth maps, (b) JBF-based method results, (c) GF-based method results, (d) Our method results

Table 1. Objective evaluation for recovered results

Methods	*Tsukuba*		*Venus*		*Teddy*		*Cones*	
	Vis.	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*	*Vis.*	*Dis.*
INITIAL	14.50	48.74	2.38	41.49	18.10	57.02	18.34	57.04
JBF-BASED	16.25	32.73	0.98	12.85	14.35	40.92	13.48	44.53
GF-BASED	18.88	42.92	1.79	23.78	20.86	53.09	18.27	55.49
PROPOSED	13.02	27.00	0.46	8.26	11.51	37.68	10.65	38.66

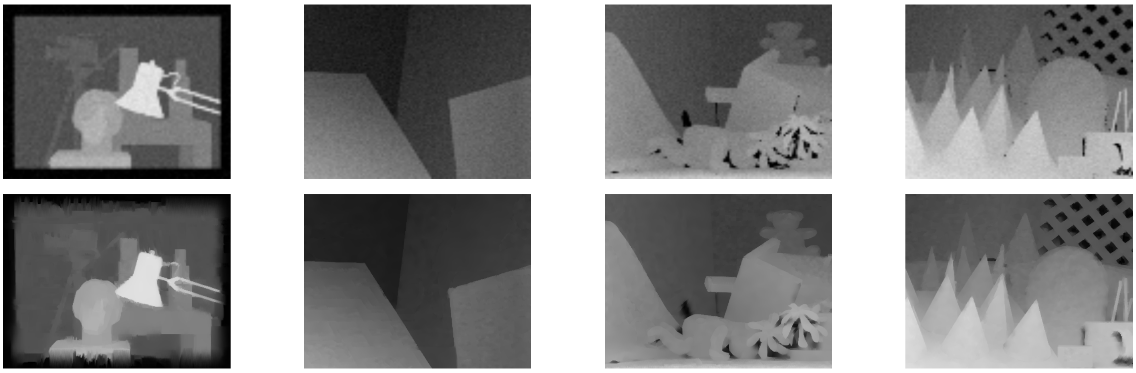

Figure 3. Depth recovered results in a noisy environment for "Tsukuba, Venus, Teddy and Cones": the first row is the noisy initial depth maps and the last row is recovered results by our method.

It is worth noting that the process time of our method is close to that of the GF-based method, which runs much faster than the JBF-based method. The execution time of our method is independent of the window radius r, which makes the algorithm scalable for higher resolution images in future application. Furthermore, our method gains better results than the GF method due to adopting an adaptive support region, weighted averaging process and a piecewise constant fitting model.

Lastly denoising performance of our proposed method is evaluated. The down-sampled depth maps are added additive white Gaussian noise with a mean of 0 and variation 20. Then the proposed method is performed on the noisy initial depth maps and results are shown in Figure 3. We found that the proposed method may provide accurate high resolution depth maps even in a noisy environment.

4. **Conclusions.** In this paper, we have presented a novel approach for low resolution and noisy depth map recovery. As an extension to the guided filtering, our method adopts an adaptive support region and weighted averaging for multi-estimate fusion. The color image guided domain transformation is used to set up the adaptive support region and can be efficiently computed by a series of 1D filters. So the computational complexity of our method does not depend on the filtering size, which fits for processing high resolution images. The proposed method is efficient and effective for depth recovery confirmed by experimental results. However, the same as the other filtering-based methods, our proposed method only uses local information for depth recovery task and may blur discontinuity when the noise of initial depth map is high, such as outdoor images. In future work, we will implement the proposed method with GPU for real-time performance. And we are also interested in combining it with the global optimal algorithm to improve its performance further.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 61373078, 61373080, 61303088 and 61402261 for funding.

REFERENCES

[1] Q. Yang, H.-H. Tan and N. Ahuja, Real-time o(1) bilateral filtering, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.557-564, 2009.

[2] J. Yang, X. Ye, K. Li, C. Hou and Y. Wang, Color-guided depth recovery from RGB-D data using an adaptive auto-regressive model, *IEEE Trans. Image Processing*, vol.23, no.8, pp.3443-3458, 2014.

[3] X. Shen, C. Zhou, L. Xu and J. Jia, Mutual-structure for joint filtering, *Proc. of IEEE International Conference on Computer Vision*, 2015.

[4] K. He, J. Sun and X. Tang, Guided image filtering, *Proc. of IEEE European Conference on Computer Vision*, pp.1-14, 2010.

[5] C. Rhemann, A. Hosni, M. Bleyer, C. Rother and M. Gelautz, Fast cost-volume filtering for visual correspondence and beyond, *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011.

[6] L. De-Maeztu, S. Mattoccia, A. Villanueva and R. Cabeza, Linear stereo matching, *Proc. of IEEE International Conference Computer Vision*, 2011.

[7] V. Katkovnik, A. Foi, K. Egiazarian and J. Astola, From local kernel to nonlocal multiple-model image denoising, *International Journal of Computer Vision*, vol.86, pp.1-32, 2010.

[8] E. Gastal and M. Oliveira, Domain transform for edge-aware image and video processing, *ACM Trans. Graphics*, vol.30, 2011.

[9] L. Li and C. Zhang, Depth enhancement with domain transform-based multipoint filter, *Proc. of Asian Conference on Design and Digital Engineering*, 2013.

[10] J. Lu, K. Shi, D. Min, L. Lin and M. Do, Cross-based local multipoint filtering, *Proc. of IEEE Computer Vision and Pattern Recognition*, 2012.

[11] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision*, vol.47, no.1, 2002.

[12] D. Scharstein and R. Szeliski, High-accuracy stereo depth maps using structured light, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp.195-202, 2003.