SAMPLING-BASED ROBUST MULTI-LATERAL FILTER FOR DEPTH ENHANCEMENT

Kyoung-Rok Lee, Ramsin Khoshabeh, Truong Nguyen

University of California, San Diego
Department of Electrical and Computer Engineering
{*krl006, ramsin, tqn001*}*@ucsd.edu*

ABSTRACT

Depth maps are an integral component of 3D video processing. They have a number of uses, including view synthesis for multi-view video, human computer interaction, augmented reality, and 3D scene reconstruction. However, depth maps are often captured at low quality or low resolution due to sensor hardware limitations or estimation errors. In this paper, we propose a new method to enhance noisy or low-resolution depth maps using high-resolution color images. Our method is based on sample selection and refinement in conjunction with multi-lateral filtering, a method derived from joint bilateral filtering using a new weighting metric. Our experimental results verify that the proposed method performs very well in comparison to existing methods.

Index Terms— Depth map, range camera, depth estimation, depth upsampling, depth superresolution

1. INTRODUCTION

Depth maps are commonly used in many three-dimensional (3D) applications. For these applications, the depth maps should be of high geometric quality and resolution since a minor error may result in distortions. Recent advances have shown us various types of sensors to obtain depth maps, such as Time-of-Flight cameras (ToF), real-time infrared projectors and cameras (e.g. Microsoft Kinect), or stereo vision systems. Unfortunately, most of the time, the quality and resolution of the acquired depth is not up to par with the analogous color images obtained from standard cameras.

Due to this limitation, the subject of depth map upsampling/refinement has been extensively studied. As depth sensors have been widely used in the field of computer vision, the problem of depth upsampling has received ever-increasing attention. A seminal work in the study of the depth map upsampling problem is the work by Diebel et al. [1]. They assumed that discontinuities in range and color tend to co-align. In their work, the posterior probability of the high-resolution reconstruction is designed as a Markov Random Field (MRF) and it is optimized with the Conjugate Gradient (CG) algorithm.

This work is supported in part by NSF grant CCF-1065305.

Following with a similar depth upsampling method, Kopf et al. proposed Joint Bilateral Upsampling (JBU) [2]. This approach leverages a modified bilateral filter. They upsample a low-resolution depth by applying a spatial filter to it, while jointly applying a similar range filter on the high-resolution color image. Research in depth map upsampling has recently experienced significant progress by the initial introduction of JBU.

Additionally, Yang et al. presented an upsampling method based on bilateral filtering the cost volume with sub-pixel estimation [3]. They build a cost volume of depth probability and then iteratively apply a standard bilateral filter to it. The final output depth map is generated by taking the winner-takes-all approach on the weighted cost volume. Finally, a sub-pixel estimation algorithm is applied to reduce discontinuities.

It is worth noting that all of the methods mentioned above may suffer from artifacts, such as texture copying and edge blurring. Texture copying occurs in smooth areas with noisy depth data and textures in the color image, while edge blurring occurs in transition areas if different objects (located in different depth layers) have similar color.

To overcome these problems, the Noise Aware Filter for Depth Upsampling (NAFDU) was proposed by Derek et al. [4]. The method switches between two different filters; a standard bilateral upsampling filter on smooth regions and a joint bilateral upsampling filter on transition areas in the depth map. A blending function is used for a gradual intermixing between the two filter outputs.

The approach closest to ours, the Pixel Weighted Average Strategy (PWAS) by Frederic et al., also proposes to resolve artifacts [5]. They build multi-lateral upsampling filters, which are an extended joint bilateral filter with an added credibility factor. The factor takes into account the low reliability of depth measurements along depth edges and the inherent noisy nature of real-time depth data. As a further improvement upon PWAS, Adaptive Multi-lateral Filtering (AMF) has been proposed to improve accuracy within smooth regions [6].

These approaches may solve the texture copying and edge blurring problems, but their performances are unfortunately very sensitive to the window size of their filter, making the window size the most critical parameter. If the window size

is too large, it might cause boundary blurring and lose details of complex objects. If the window size is too small, it may fail to collect significant information from its neighborhood.

In light of these problems, we present a new depth sampling method and multi-lateral filtering technique. Our approach is based on selecting reliable depth samples from a neighborhood of pixels and applying multi-lateral filtering. We first define unreliable regions by calculating a measure of reliability for each pixel in the depth map. The reliability is determined by calculating the sum of gradients for each pixel's neighborhood. Then every pixel in the region of low reliability collects samples from the region of high reliability and selects the best sample with the highest fidelity. Each pixel's selected depth sample is refined by sharing its information with its neighbors' selected samples in the sample refinement stage. Finally, a robust multi-lateral filter, which is an extended joint bilateral filtering technique with an additional factor for robustness weights, is applied to reduce noise while preserving sharpness along edges. We evaluate our approach on the Middlebury datasets [7] and show that our method provides performance gains over existing methods. To our knowledge, our work is the first of its kind that presents a sampling-based depth refinement.

The remainder of this paper is organized as follows. Section 2 presents the proposed method. A visual and quantitative comparison of a number of key methods and improvement results are reported in Section 3. Finally, a conclusion is given in Section 4.

2. PROPOSED METHOD

Errors in range images generated by real-time depth sensors or stereo vision systems can be roughly categorized into two broad categories:

- Errors in transition areas: Inadequate calibration, occlusion area, or motion artifacts often lead to wrong distance values at object boundaries when we fuse the depth maps with color images.
- Random noise on geometrically flat or smooth surfaces: Properties of the object surface, lighting conditions, or systematic errors may generate noise on the surface.

In our work, we investigate a method that is able to fix both errors. Our method takes a color image I and a depth map D as inputs. The process consists of sample selection, sample refinement, and robust multi-lateral filtering. Before refining the depth map, we first measure depth reliabilities and define unreliable regions in it. In the sample selection stage, for every pixel in the unreliable region we collect samples from reliable regions and select the best sample giving the highest fidelity. Then the selected depth samples are refined by sharing their information with their neighbors' selected samples. Finally, a robust multi-lateral filter is applied to reduce noise in smooth areas, while preserving sharpness along the edges.

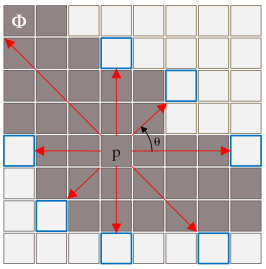


Fig. 1. Pixel p shoots several rays (red lines) toward reliable (white) region and collects the closest samples (blue squares).

2.1. Unreliable Region Detection

We use the gradient of the depth map as an important key for measuring reliability of depth values based on our assumption that depth values with high variance in their neighborhood or depth values along edges are not reliable. We first take the derivative on the depth map and calculate its magnitude. Then for each pixel, we compute an average function of Gaussian of the gradient magnitude in a small window. Thus, the reliability for each pixel is determined by the following equation.

$$K_p = \sum_{q \in \Omega_p} f_t(|\nabla D_q|)/|\Omega p| \tag{1}$$

where f_t is the Gaussian function with variance σ_t , Ω_p is the window centered at pixel p , and ∇ is the gradient operation. The unreliable region Φ is the set of pixels whose reliability values are less than a certain threshold.

$$\Phi \leftarrow \{p | K_p < \tau\} \tag{2}$$

The threshold τ controls the width of the unreliable region. The remaining pixels in the depth map are identified as the reliable region. Every depth value in this unreliable region Φ will be refined by the following sample selection and refinement steps.

2.2. Sample Selection

To recover the pixels in the unreliable region, we collect depth samples from a nearby reliable region, inspired by [8]. In the alpha matting problem, color samples are extracted from the reliable regions to estimate the unknown pixels' alpha values. Similar to this, we collect depth samples from the reliable regions to refine the unreliable depth values.

We collect depth samples from a neighborhood by shooting several rays toward the known region. The slope of each ray is given by Θ, $\Theta \leftarrow \{\frac{j\pi}{\eta} | j = 0, 1, ..., 2\eta - 1\}$. When the ray from p meets the known region, the closest depth sample

is saved to Ψ_p. Fig. 1 illustrates rays (red lines) and collected depth samples (blue squares) for the pixel location p.

The best depth sample for pixel $p \in \Phi$ is obtained by calculating the fidelity for all collected depth samples Ψ_p. For every $i \in \Psi_p$, the fidelity function is computed based on the criteria that pixels with similar color tend to share similar depth values and that they are likely to have the same depth value if they are spatially close. Therefore, at a given pixel p, we compute the fidelity function as

$$g^{SS}(p,i) = \sum_{q \in \Omega_p} f_r(|I_q - I_i|) f_s(||q - i||)/|\Omega p| \quad (3)$$

where f_r and f_s are taken to be Gaussian functions with standard deviations σ_r and σ_s respectively. The chromatic similarity in the RGB color space is computed by the Euclidean distance metric. To suppress errors from image noise caused by low lighting, high ISO settings, or chromatic distortion, we take into account the average of all pixels in a 3×3 window centered at pixel p.

Eq. (3) will have a large response for the depth value d_i which has a similar color and spatial proximity, but a value close to zero for the rest. We select the i^* giving the largest fidelity value among depth samples:

$$i^* = \arg\max_{i \in \Psi_p} g^{SS}(p,i) \quad (4)$$

Then we save the selected depth and fidelity for each pixel $p \in \Phi$.

$$\begin{aligned} D_p^{SS} &= D_{i^*}, \\ E_p^{SS} &= g^{SS}(p, i^*) \end{aligned} \quad (5)$$

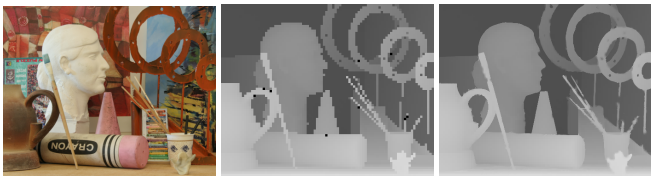
We use the selected disparity map D^{SS} and the fidelity map E^{SS} in the following sample refinement stage.

2.3. Sample Refinement

Since we collect depth samples with the ray searching scheme, where several rays spread out like the spokes of a wheel, the method occasionally fails to collect appropriate depths in some situations. For example, the desirable depth sample may be located between the rays or it may happen that the color affected by noise with a false depth sample is accidentally very similar to the target's color. Therefore, an additional refinement stage is required.

In the sample refinement stage, the samples are refined by comparing their own choice of best depth sample with the choices of their neighborhood. The design of the fidelity function for sample refinement is based on color fitness and spatial distance between pixel p and its neighbor q as well as q's fidelity value from the sample selection stage. The fidelity for sample refinement is determined by

$$g^{SR}(p,q) = E_q^{SS} f_r(|I_p - I_q|) f_s(||p - q||) \quad (6)$$



(a) Color image (b) Input disparity (c) Proposed method

Fig. 2. Experimental result of our method. The input disparity map is generated by downsampling ground truth with factor of 5.

where E_q^{SS} is the fidelity value of the pixel q from the previous stage. q^* is chosen to give the largest fidelity value among p's neighborhood:

$$q^* = \arg\max_{q \in \Omega_p} g^{SR}(p,q) \quad (7)$$

Similar to the sample selection stage, both refined depth value and fidelity are saved.

$$\begin{aligned} D_p^{SR} &= D_{q^*}, \\ E_p^{SR} &= g^{SR}(p, q^*) \end{aligned} \quad (8)$$

Up to this point, we have estimated new depth values for pixels in the unreliable region. In the next step, the refined depth data will be used as an initial depth estimate and the fidelity values will be used to determine the robustness factor.

2.4. Robust Multi-lateral Filtering

After the unreliable region is refined by the sample selection and refinement stages, the depth map still needs to be processed to reduce discontinuities in the final depth map. Therefore, we apply a robust multi-lateral filter. It is an extended joint bilateral filtering technique, but the robustness factor is added to reduce blurring along edges as well as to refine edges. The robustness value for pixel p is determined by choosing the minimum value between K_p and E_p^{SR} because we want to disregard depth values with low plausibility as much as we can.

$$\tilde{K}_p = \min\{K_p, E_p^{SR}\} \quad (9)$$

With this robustness factor, our final depth is determined by

$$D_p^{MF} = \frac{\sum_{q \in \Omega_p} f_r(|I_p - I_q|) f_s(||p - q||) \tilde{K}_q D_q^{SR}}{\sum_{q \in \Omega_p} f_r(|I_p - I_q|) f_s(||p - q||) \tilde{K}_q} \quad (10)$$

The spatial weighing term f_s is based on pixel position and the range weight f_r is based on color data. Thus, this filter adjusts the edges in the input depth map D^{SR} to the edges in the guidance color image I and the robustness factor $\tilde{K}$

gives low weight to depth values with low fidelity, preventing artifacts such as texture copying and edge blurring. Fig. 2 shows the result of our algorithm as will be discussed in the next section.

3. RESULTS

In this section we discuss the experiments to evaluate the algorithm's performance. The proposed method has been implemented with GPU programming and tested on computer with In Intel Corei7 CPU 2.93GHz Processor and an NVIDIA GeForce GTX 460 graphics card. The performance speed of our method is on average 26 fps on a video with 640×480 resolution.

We provide both qualitative and quantitative comparisons with other existing methods. Also, we show an improvement benchmark by applying our refinement method to disparity estimation results from all of the 109 methods on the Middlebury stereo evaluation website.

3.1. Visual and Quantitative Comparison

For a quantitative comparison, we utilize the *Moebius*, *Books*, and *Art* scenes from the Middlebury datasets. We have evaluated the performance of the proposed method against the state-of-the-art methods presented by [6]. Downsampled disparity maps are generated by downsampling the ground truth by a factor of $3\times$, $5\times$, and $9\times$. In [6], they used the structural similarity (SSIM) measure as a quantitative comparison. However, this measure is not appropriate for depth map evaluation because it does not function properly with disparities in unknown or occluded regions. Middlebury's ground truth maps contain regions of unknown disparity and depth upsampling algorithms do not produce meaningful results in those regions. As a fair comparison, we calculate the average percentage of bad pixels with an error threshold of 1 for all known regions; pixels whose disparity error is greater than threshold are regarded as the bad pixels. This is the same scoring scheme employed in the Middlebury evaluation. Fig. 3 and Table 1 show that our method performs better than all of the other methods.

3.2. Improvements

Also, we apply our refinement method to the disparity estimation results of all methods submitted to the Middlebury stereo evaluation. Fig. 4 shows the improvement in terms of the percentage of bad pixels. Note that the proposed method improves the results of most methods. One limitation of the proposed algorithm is that its performance drops with small and complex images or poorly estimated initial disparity maps.

More experiments and results of our method can be found at: `http://videoprocessing.ucsd.edu/~ultralkl/projects/SRMF/`

Table 1. Quantitative comparisons (average percent of bad pixels)

Dataset		JBU	PWAS	AMF	Proposed
Moebius	3x	7.43	4.68	4.5	**3.62**
	5x	12.22	7.49	7.37	**4.87**
	9x	21.02	12.86	12.75	**9.02**
Books	3x	5.4	3.59	3.48	**2.38**
	5x	9.11	6.39	6.28	**3.58**
	9x	15.85	12.39	12.24	**7.11**
Art	3x	15.15	7.05	6.79	**5.07**
	5x	23.46	10.35	9.86	**6.91**
	9x	38.41	16.87	16.87	**11.7**

4. CONCLUSION

In this paper, we have proposed a new depth map enhancement method based on sample selection, refinement, and robust multi-lateral filtering. Specifically, we have introduced a new sampling method to get better accuracy on boundaries. Also we have investigated multi-lateral filtering with a new robustness factor. Experiments clearly show that the proposed algorithm significantly outperforms other state-of-the-art methods in the depth upsampling problem.

Our future work will be focused on improving the model to determine unreliable regions adaptively.

5. REFERENCES

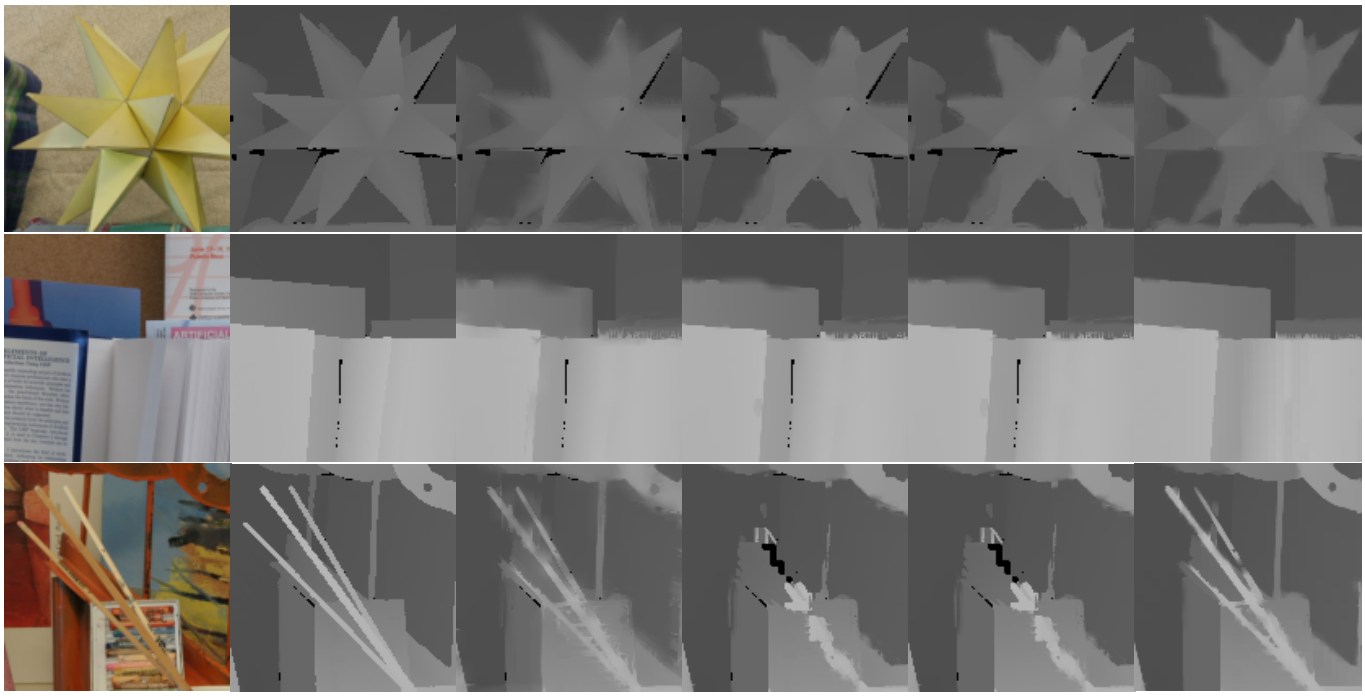
[1] James Diebel and Sebastian Thrun, "An application of markov random fields to range sensing," in *In NIPS*. 2005, pp. 291–298, MIT Press.

[2] Johannes Kopf, Michael F. Cohen, Dani Lischinski, and Matt Uyttendaele, "Joint bilateral upsampling," in *ACM SIGGRAPH 2007 papers*, New York, NY, USA, 2007, SIGGRAPH '07, ACM.

[3] Qingxiong Yang, Ruigang Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, june 2007, pp. 1 –8.

[4] Derek Chan, Hylke Buisman, Christian Theobalt, and Sebastian Thrun, "A Noise-Aware Filter for Real-Time Depth Upsampling," in *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008*, Marseille, France, 2008, Andrea Cavallaro and Hamid Aghajan.

[5] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier, and A. Cuesta, "Pixel weighted average strategy for depth sensor data fusion," in *Image Processing (ICIP), 2010*



(a) Color image (b) Ground truth (c) JBU (d) PWAS (e) UML (f) Proposed method

Fig. 3. Visual comparison on the Middlebury datasets. The upsampling methods include: (c) JBU, (d) PWAS, (e) UML, (f) proposed method.

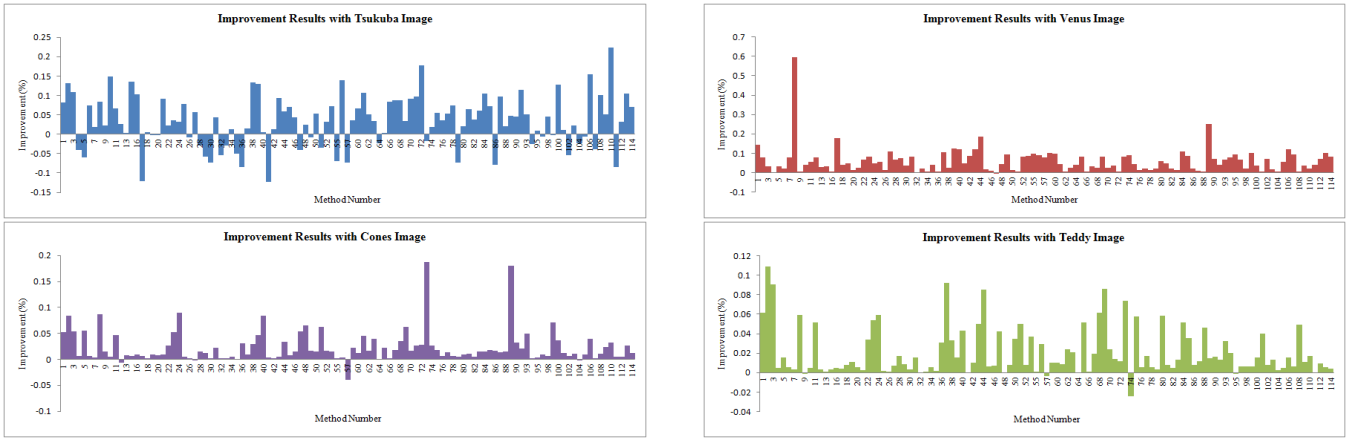


Fig. 4. Percentage improvement in terms of number of bad pixels after applying the proposed algorithm to all the 109 methods on the Middlebury stereo evaluation.

17th IEEE International Conference on, sept. 2010, pp. 2805 –2808.

[6] F. Garcia, D. Aouada, B. Mirbach, T. Solignac, and B. Ottersten, "A new multi-lateral filter for real-time depth enhancement," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 30 2011-sept. 2 2011, pp. 42 –47.

[7] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, june 2007, pp. 1 –8.

[8] Eduardo S. L. Gastal and Manuel M. Oliveira, "Shared sampling for real-time alpha matting," *Computer Graphics Forum*, vol. 29, no. 2, pp. 575–584, May 2010, Proceedings of Eurographics.