Depth Map Upsampling via Progressive Manner Based on Probability Maximization

Rongqun Lin[1](✉), Yongbing Zhang[1], Haoqian Wang[1], Xingzheng Wang[1], and Qionghai Dai[1,2]

[1] Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
linrq14@mails.tsinghua.edu.cn
[2] Department of Automation, Tsinghua University, Beijing 100084, China

Abstract. Depth maps generated by modern depth cameras, such as Kinect or Time of Flight cameras, usually have lower resolution and polluted by noises. To address this problem, a novel depth upsampling method via progressive manner is proposed in this paper. Based on the assumption that HR depth value can be generated from a distribution determined by the ones in its neighborhood, we formulate the depth upsampling as a probability maximization problem. Accordingly, we give a progressive solution, where the result in current iteration is fed into the next to further refine the upsampled depth map. Taking advantage of both local probability distribution assumption and generated result in previous iteration, the proposed method is able to improve the quality of upsampled depth while eliminating noises. We have conducted various experiments, which show an impressive improvement both in subjective and objective evaluations compared with state-of-art methods.

Keywords: Progressive manner · Denoising · Depth map · Upsampling · Probability Maximization

1 Introduction

As an indication of true position in 3D space, depth maps play a more and more important role in a variety of different applications including object reconstruction, medical, 3D television and entertainment. Although there exist several range measuring approaches to capture depth map, it is difficult to acquire depth information accurately and sufficiently. For instance, time of flight (ToF) cameras can use active sensing to capture depth map per-pixel at video frame-rate, and they become easily accessed and popular. However, the main disadvantage of such cameras is that the resolution of generated depth maps is relatively low compared with their associated color image. This is due to chip size limitations and the captured depth maps always contain amounts of acquisition noise. These defects limit the practical applications based on depth information.

This work was partially supported by the National Natural Science Foundation of China under Grant 61170195, U1201255, and U1301257.

Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 84–93, 2015.
DOI: 10.1007/978-3-319-24078-7_9

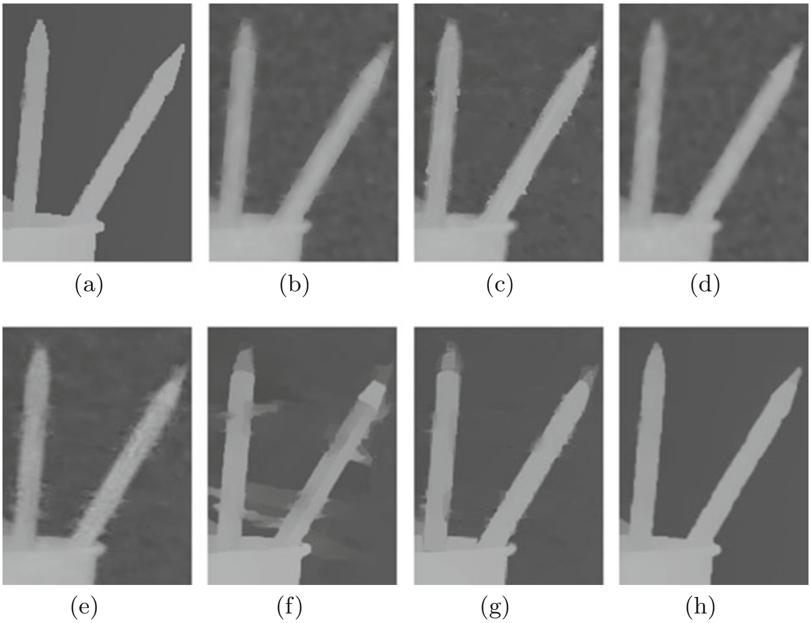


Fig. 1. Depth map upsampling results of art compared in details (upscaled ×4) with noise. (a) Ground truth. (b) Diebel et al. [2]. (c) Yang et al. [11]. (d) Chan et al. [1]. (e) He et al. [5]. (f) Park et al. [7]. (g) Ferstl et al. [3]. (h) Ours. (Zoom in for better view.)

Therefore, depth upsampling and denoising becomes a vital problem to the development of 3D applications [4,8].

To address the above problem, several approaches are proposed to upsample the depth maps via using additional corresponding color or intensity image as cues. Diebel et al. [2] performed upsampling using a MRF formulation exploiting the fact that discontinuities in depth and intensity image tend to co-align and weights of the smoothness term were computed according to texture derivatives. Yang et al. [11] used a joint bilateral filtering of a depth cost volume and a RGB image in an iterative process. Chan et al. [1] used a noise-aware bilateral filter to eliminate the noise during upsampling the depth map. Park et al. [7] proposed a more complicated approach, which is based on a least-square optimization that combines several weighting factors with nonlocal means filtering, segmentation, image gradients and edge saliency for depth upsamling. Frestl et al. [3] formulated the depth map upsampling as a global energy optimization problem using Total Generalized Variation (TGV) regularization.

While these methods achieve good quality in smooth regions, the major drawback is that their upsampling results contain blurring, over smoothing or texture-copying around thin structures or sharp discontinuities. Moreover, when upsampling a noised LR depth map, these methods produce worse results, since the noise can be propagated to the upsampled regions, especially around sharp edges or thin objects. All of the above methods cannot generate accurate HR depth maps with sharp edges and less noise.

To generate accurate HR depth maps with sharp edges while depressing noise effects, in this paper we propose a depth upsampling algorithm based on probability maximization in a progressive manner. The main contributions of our work are two-fold: (1) Inspired by the work in [10], we provide the mathematic derivations and build our model based on the assumption that the depth of a pixel in HR depth map can be generated from a distribution determined by the depth of pixels in its neighborhood in the same HR depth map. (2) To preserve the sharp edges of the upsampled depth map and remove the noise, we exploit the progressive framework via accumulating the influence of the initial input and remove noise progressively. Compared to state of the art methods, our method is superior in terms of both subjective and objective evaluations. Figure 1 shows that our results can simultaneously remove the noise and achieve sharper edges with less artifacts .

The paper is organized as follows: Sect. 2 details about our approach including our model and the progressive framework. The experimental results are reported in Sect. 3, followed by a conclusion of our work in Sect. 4.

2 Proposed Method

In this section, we will detail the proposed method to solve the depth upsampling and denoising problem simultaneously. Firstly, we derive a probability maximization problem to build our model. After necessary derivations, we analyze our model and show that our model describes the intrinsic properties of the depth map. Secondly, to preserve the sharp edges of the upsampled depth map and remove the noise, we introduce the progressive framework. In each iteration, we use the output of previous iteration to update the input accumulating the influence of the initial input until the output converges to a stable result, which can remove some noise progressively.

2.1 Our Model

Inspired by [10], we assume the depth of a pixel in HR depth map can be generated from a distribution determined by those in its neighborhood in the same HR depth map. It yields the following model for each pixel i.

$$p(d_i^H|d_\Omega^H) \propto \sum_{j\in\Omega} \eta_{ij} exp\{-(d_i^H - d_j^H)^2/2\sigma_d^2\}, \tag{1}$$

where Ω is a spatial pixel block centered at pixel i, d_i^H is the value of pixel i in HR depth map. d_Ω^H is the depth of pixels in its neighborhood in the same HR depth map. η_{ij} is the mixture coefficient corresponding statistically to color image and spatial position. And $\eta_{ij} = \alpha_{ij}/\sum_{j\in\Omega}\alpha_{ij}$, where

$$\alpha_{ij} = exp\{-\frac{\sum\limits_{k=1}^{3}(I_i^k - I_j^k)^2}{2\sigma_I^2}\}exp\{-\frac{(p-q)^2}{2\sigma_s^2}\}, \tag{2}$$

with I_i^k being the k_{th} color channel of pixel i in color image. p is the position of pixel i and q is the position of pixel j.

Then we maximize the probability and the optimal d_i^H for each pixel i is yielded as

$$d_i^H = \underset{d_i^H}{\operatorname{argmax}}\, p(d_i^H|d_{\Omega}^H) = \underset{d_i^H}{\operatorname{argmin}}\,(-ln\ p(d_i^H|d_{\Omega}^H)). \tag{3}$$

And we get

$$\begin{aligned}\frac{\partial(-ln\ p(d_i^H|d_{\Omega}^H))}{\partial d_i^H} &= \frac{\sum\limits_{j\in\Omega}\eta_{ij}exp\{\frac{-(d_i^H-d_j^H)^2}{2\sigma_d^2}\}\frac{(d_i^H-d_j^H)}{\sigma_d^2}}{\sum\limits_{j\in\Omega}\eta_{ij}exp(\frac{-(d_i^H-d_j^H)^2}{2\sigma_d^2})},\\ &\propto d_i^H - \frac{\sum\limits_{j\in\Omega}\alpha_{ij}exp\{\frac{-(d_i^H-d_j^H)^2}{2\sigma_d^2}\}d_j^H}{\sum\limits_{j\in\Omega}\alpha_{ij}exp\{\frac{-(d_i^H-d_j^H)^2}{2\sigma_d^2}\}}.\end{aligned} \tag{4}$$

Setting Eq. 4 to zero yields a per-pixel constraint for each d_i^H.

However, we meet a **chicken-egg** dilemma. Our goal is to obtain the HR depth map, while d_i^H can be reliably acquired only when d_{Ω}^H is available. Thus we introduce a roughly estimated HR depth map (eg. Bicubic upsampling result) to break this dilemma. In addition, it can be used to avoid incorrect depth prediction due to depth color inconsistency (some pixels with similar intensity may have different depth, vice versa).

Therefore, we rewrite our model as follow. We write the whole objective function in a matrix form. We denote by d^{in} the input vector, d^{out} the output vector, $\boldsymbol{W}$ the weight matrix. Each matrix element is $\boldsymbol{W}_{ij} = w_{ij}/\sum_{j\in\Omega} w_{ij}$, where w_{ij} is the mixture coefficient corresponding statistically to color image, spatial position and the roughly estimated HR depth map.

$$w_{ij} = exp\{-\frac{\sum\limits_{k=1}^{3}(I_i^k - I_j^k)^2}{2\sigma_I^2}\}exp\{(-\frac{(p-q)^2}{2\sigma_s^2}\}exp\{\frac{-(d_i^{est} - d_j^{est})^2}{2\sigma_d^2}\}, \tag{5}$$

σ_I , σ_s and σ_d adjust the importance of the spatial distance , intensity difference and estimated depth changes.

The objective function with respect to d^{out} is therefore formulated as

$$min(E(d^{out})) = min\{(d^{out} - \boldsymbol{W}d^{in})^T(d^{out} - \boldsymbol{W}d^{in})\}. \tag{6}$$

The global minima can be obtained by solving $d^{out} = \boldsymbol{W}d^{in}$, which is our final practical model.

Our model describes the intrinsic properties of the depth map. Although our model is similar to the combination of the neighborhood smoothness term and NLM regularization in [7], their method takes segmentation, image gradient, edge saliency and non-local means into consideration, which is more complicated. Moreover, the difference between our method and that of [7] is the idea how we get the final HR depth map, which will be detailed in the following part.

2.2 Progressive Framework

In our model mentioned above, we set the parameters appropriately and consequently the w_{ij} of the pixels in the neighborhood almost has slight difference in their values. From a different perspective, pixels in the nearby area often have similar depth.

Algorithm A1. PROPOSED METHOD

```
Input: low resolution depth map $d^L$ .
initialization: Map $d^L$ to high-resolution image as initial input $d^{ini}$, compute $\boldsymbol{W}$.
for:  k = 1 : max-step  do
    update the $k_{th}$ input $d^{in}(k)$ according to 7 .
    compute $d^{out}(k) = \boldsymbol{W} d^{in}(k)$.
    if  ($k$ = max-step)
        $d^H = d^{out}(k)$.
        break.
    end if
    if  $sum(d^{out}(k) - d^{in}(k)) \leq \varepsilon$, ($\varepsilon = 0.0001$)
        $d^H = d^{out}(k)$.
        break.
    end if
end for:
Output: high resolution depth map $d^H$.
```

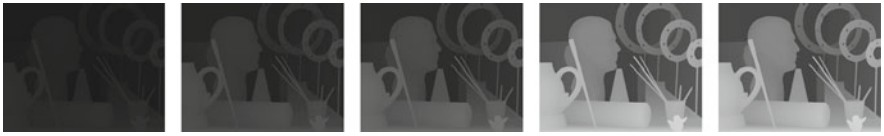
If the input contains more information, our model can produce the result once. However, the initial input is the one with high resoltion but has fewer non-zero values in the position we map the low one to high one. It means that in initial depth map there exist many zero values, which is not enough to get the final HR depth map once.

The idea of our method is that the depth of every pixel in HR depth map can be acquired via **accumulating the influence of the initial input** . During the progressive process, **the updated result gets improved and contains more information for the next iteration**. Therefore we propose a progressive framework.

In each step, we update the input as shown in the following.

$$d_i^{in}(k) = \begin{cases} d_i^{ini}, & \text{if } i \in \Psi \\ d_i^{out}(k-1), & \text{otherwise} \end{cases} \tag{7}$$

where $d_i^{in}(k)$ is the depth of pixel i in k_{th} input, $d_i^{out}(k-1)$ is the depth of pixel i in $(k-1)_{th}$ output. We map d^L to high-resolution image as initial input d^{ini},



(a) s1 (b) s2 (c) s3 (d) s66 (e) s100

Fig. 2. Progressive process displayed and intermediate results with specified step.

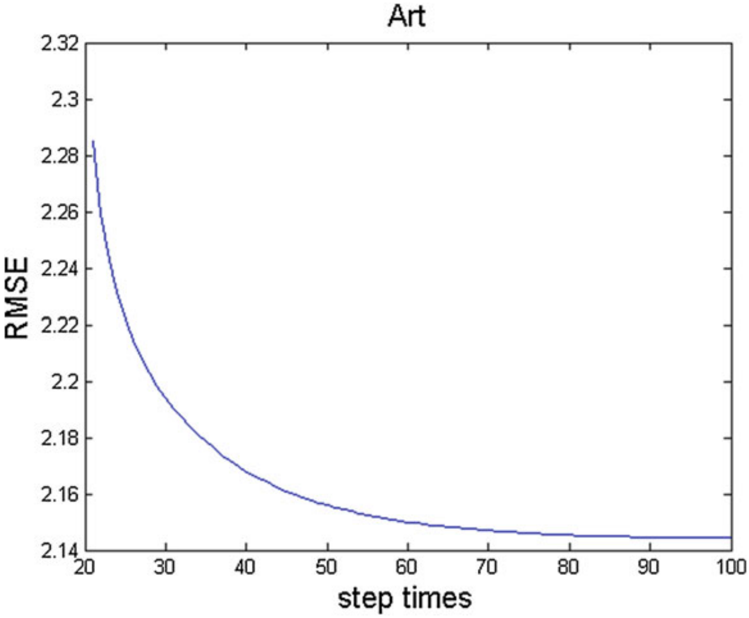


Fig. 3. RMSE in different step(up-scaled ×2)without noise

and the d_i^{ini} is the depth of pixel i in the initial input. Ψ is the non-zero position in the initial input. The replacement makes k_{th} input have the same value as the initial input on the Ψ, which was called **anchor points**. But in the last step, we don't do the replacement, in order to cope with noisy input.

Pixels in the output of first step usually have a small value, since there exist fewer non-zero values in initial input. Thus in Fig. 2, result in the first step seems dark and it gets brighter and brighter during the progressive process. As is shown in Fig. 3, our method is so powerful that the RMSE results converge so quickly. The reason we can deal with the noised situation well is that we can remove some noise in every step and also keep the thin structures like sharp edges. In the last step the output except the anchor points has noiseless and accurate values. We regenerate the values on the anchor points, thus the noise on the anchor points is removed.

3 Experimental Results

We test our method using synthetic examples from Middlebury 2007 datasets [6,9] and the dataset of Frestl et al. [3] for quantitative and qualitative comparisons with the state of the art methods. In our experiments, we normalize all the values of pixels and spatial coordinates and the roughly estimated depth map is obtained by bicubic interpolation from the low resolution depth map.

During our experiments, σ_I influences the importance of guided color image. When it fixes at a very small digit, the results will contain serious artifacts. It means that the color image has excessive influence. σ_s influences the importance of spatial difference. When it fixes at a very small digit, the results will be over-smoothing. σ_d influences the importance of estimated depth changes. When it fixes at a very small digit, the results will have blurings around the edges. Consequenlty, in this paper we set the values of parameters as follows: $\sigma_I = 0.12, \sigma_s = 0.02, \sigma_d = 0.04$. And the size of neighborhood region Ω is 9×9. We use the same setting when upsampling the noised input LR depth map.

To demonstrate the effectiveness of our proposed method, we show our results in two parts: upsample the clean LR depth map and upsample the noised LR depth map.

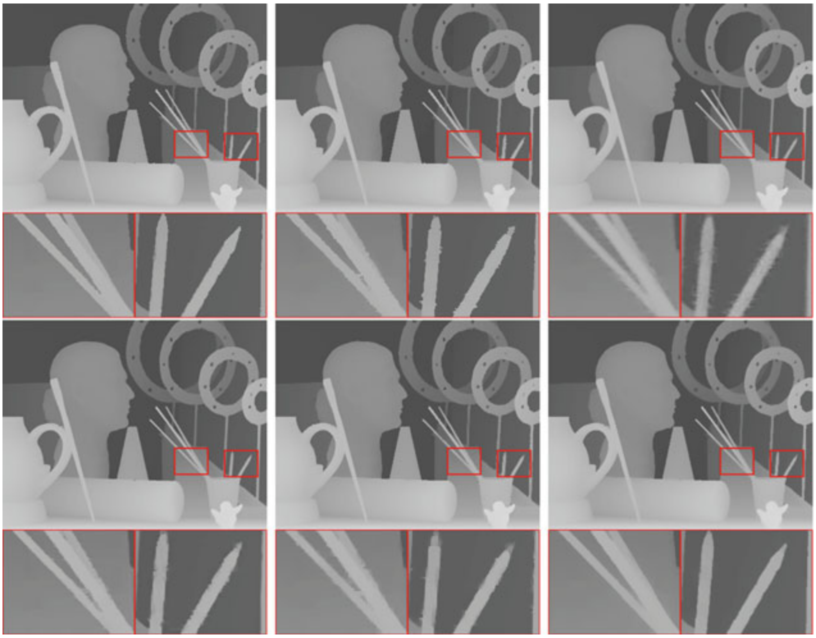


Fig. 4. Visual quality comparisons of ×4 upsampling on the Middlebury *Art* datasets ***without noise***. Row 1 Column 1: Ground Truth. Row 1 Column 2: Yang et al. [11]. Row 1 Column 3: He et al. [5]. Row 2 Column 1: Park et al. [7]. Row 2 Column 2: Ferstl et al. [3]. Row 2 Column 3: Ours. (Zoom in for better view.)

3.1 Upsample the Clean LR Depth Map

We downscale the Middlebury 2007 datasets [6,9] by bicubic interpolation method as the clean LR depth map. Especially, we choose three of them to form our test dataset: Art, Book and Moebius, which have clutter depth values and their corresponding color images have complicated textures. We conduct quantitative evaluations on our results and the results provided by Frestl et al. [3].

The quantitative results in terms of the Root Mean Square Error (RMSE) against the ground-truth depth maps are shown in Table 1. Beside the bilinear interpolation, the proposed method is compared with five recent methods: Diebel et al. [2], Yang et al. [11], He et al. [5], Park et al. [7], and Ferstl et al. [3]. The best result for each dataset is highlighted. What can be clearly seen from the numerical results is that our approach is the best compared to other state of the art methods.

A visual comparison for the different methods is given in Fig. 4. For clean inputs, Yang et al. [11], He et al. [5], Park et al. [7], and Ferstl et al. [3] methods introduce some jaggy artifacts along edges because they depend too much on the guide color image. Our method can generate results with sharper edges and less artifacts.

Table 1. RMSE comparisons on Middlebury 2007 datasets **without noise**(upscaled ×2, ×4).

	Art		*Books*		*Moebius*	
	×2	×4	×2	×4	×2	×4
Bilinear	2.834	4.147	1.119	1.673	1.016	1.499
Diebel et al. [2]	3.119	3.794	1.205	1.546	1.187	1.439
Yang et al. [11]	4.066	4.056	1.615	1.701	1.069	1.386
He et al. [5]	2.934	3.788	1.162	1.572	1.095	1.434
Park et al. [7]	2.833	3.498	1.195	1.495	1.064	1.349
Ferstl et al. [3]	3.032	3.785	1.290	1.603	1.129	1.459
Ours	**2.144**	**3.457**	**1.025**	**1.230**	**0.993**	**1.334**

3.2 Upsample the Noised LR Depth Map

In reality, captured depth maps always have plenty of noise. For fair comparison, we employ the **noisy input** dataset used in Frestl et al. [3].

Our method has a great advantage both in quantitative and qualitative evaluations compared with previous methods. Experiments show that our method can remove lots of noise and preserve sharp edges.

The quantitative results in terms of the Root Mean Square Error (RMSE) are shown in Table 2. Beside the bilinear interpolation, the proposed method

Table 2. RMSE comparisons on Middlebury 2007 datasets **with noise**(upscaled ×2, ×4).

	Art		*Books*		*Moebius*	
	×2	×4	×2	×4	×2	×4
Bilinear	4.580	5.621	3.948	4.309	4.200	4.565
Diebel et al. [2]	3.489	4.514	2.064	3.002	2.127	3.105
Yang et al. [11]	3.005	4.021	1.874	2.383	1.917	2.418
Chan et al. [1]	3.437	4.464	2.091	2.773	2.076	2.759
He et al. [5]	3.546	4.412	2.375	2.737	2.481	2.831
Park et al. [7]	3.759	4.564	1.946	2.607	1.956	2.508
Ferstl et al. [3]	3.188	4.063	1.522	2.213	1.475	2.030
Ours	**2.305**	**3.747**	**1.495**	**2.098**	**1.455**	**1.821**

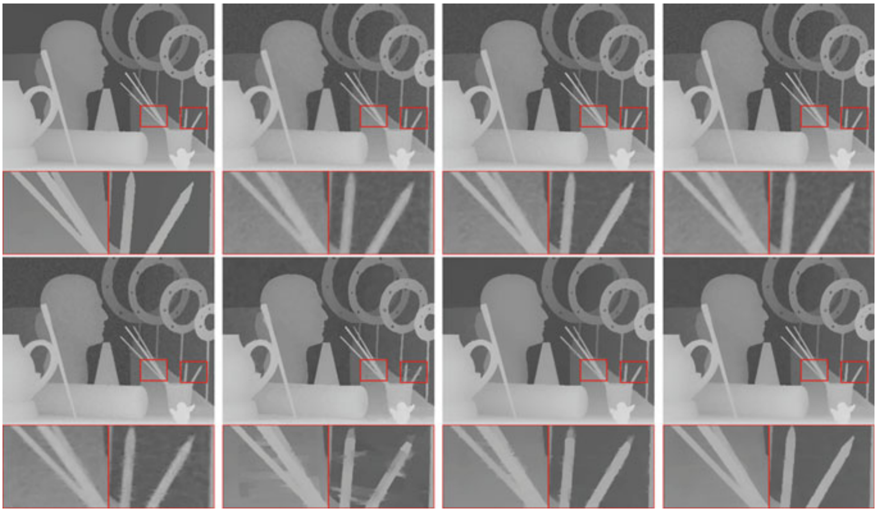


Fig. 5. Visual quality comparisons of ×4 upsampling on the Middlebury *Art* datasets ***with noise***. Row 1 Column 1: Ground Truth. Row 1 Column 2: Diebel et al. [2]. Row 1 Column 3: Yang et al. [11]. Row 1 Column 4: Chan et al. [1]. Row 2 Column 1: He et al. [5]. Row 2 Column 2: Park et al. [7]. Row 2 Column 3: Ferstl et al. [3]. Row 2 Column 4: Ours. (Zoom in for better view.)

is compared with six recent methods: Diebel et al. [2], Yang et al. [11], Chan et al. [1], He et al. [5], Park et al. [7], and Ferstl et al. [3]. Our results always rank first in terms of RMSE.

A visual comparison for the different methods is given in Fig. 5. For noised inputs, Diebel et al. [2], Yang et al. [11], Chan et al. [1], He et al. [5] methods generate the noised results with blurrings. Although Park et al. [7], and Ferstl et al. [3] methods can remove some noise, they still introduce some jaggy artifacts along edges because they depend too much on the guided color image. As can be seen, our proposed method can produce high resolution depth maps with sharper edges, clearer structures and fewer artifacts.

4 Conclusions

In this paper, we propose a novel method for depth map upsampling via progressive manner based on probability maximization. The formulated model is able to reveal and employ the intrinsic properties of the depth map. To preserve the sharp edge of the upsampled depth map and remove the noise, we exploit the progressive framework through accumulating the influence of the initial input and remove noise progressively. Experiments show that our method outperforms the state-of-art methods in terms of quantitative and qualitative comparisons. Our method can produce the HR depth map with sharp edges, more accurate values and less noise.

As future work, we would like to improve our work to meet the need of real-time reconstructions and extend our method to more applications.

References

1. Chan, D., Buisman, H., Theobalt, C., Thrun, S., et al.: A noise-aware filter for real-time depth upsampling. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008 (2008)
2. Diebel, J., Thrun, S.: An application of markov random fields to range sensing. In: NIPS (2005)
3. Ferstl, D., Reinbacher, C., Ranftl, R.: Image guided depth upsampling using anisotropic total generalized variation. In: ICCV (2013)
4. Guomundsson, S.A., Larsen, R., Aanæs, H., Pardas, M., Casas, J.R.: Tof imaging in smart room environments towards improved people tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–6. IEEE (2008)
5. He, K., Sun, J., Tang, X.: Guided image filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 1–14. Springer, Heidelberg (2010)
6. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: CVPR (2007)
7. Park, J., Kim, H., Tai, Y., Brown, M., Kweon, I.: High quality depth map upsampling for 3D-tof cameras. In: ICCV (2011)
8. Prasad, T., Hartmann, K., Weihs, W., Ghobadi, S.E., Sluiter, A.: First steps in enhancing 3D vision technique using 2D/3D sensors. In: Computer Vision Winter Workshop, pp. 82–86 (2006)
9. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: IEEE Conference onComputer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
10. Xu, L., Yan, Q., Jia, J.: A sparse control model for image and video editing. ACM Trans. Graph. (TOG) **32**(6), 197 (2013)
11. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: CVPR (2007)