ALTERNATELY GUIDED DEPTH SUPER-RESOLUTION USING WEIGHTED LEAST SQUARES AND ZERO-ORDER REVERSE FILTERING

Kailong Zhou, Shengtao Yu and Cheolkon Jung

School of Electronic Engineering, Xidian University, Xian, Shaanxi 710071, China
zhengzk@xidian.edu.cn

ABSTRACT

Due to the structural inconsistency between color and depth, texture copying and edge blurring artifacts appear in color-guided depth super-resolution. In this paper, we propose alternately guided depth super-resolution using weighted least squares (WLS) and zero-order reverse filtering. We adopt WLS for alternating guidance, and alternately use color and depth as guidance in WLS to suppress texture copying artifacts. Since color guidance causes edge blurs in depth due to the mismatch between color and depth, we apply zero-order reverse filtering to depth images to alleviate edge blurring artifacts. Moreover, WLS is a global optimization-based filter and thus it is effective in removing depth noise. Experiments on Middlebury and real scene datasets show that the proposed method outperforms state-of-the-art methods in terms of quantitative and qualitative measurements.

Index Terms— Depth super-resolution, alternating guidance, edge blurring, texture copying, weighted least squares, zero-order reverse filtering.

1. INTRODUCTION

Acquiring depth information from real world scenes is of vital importance in many applications such as image-based rendering and virtual reality. However, depth images obtained by depth cameras have a series of problems. Time-of-flight (ToF) camera can acquire depth images with a video rate [1], however, the captured depth images are of low resolution and noisy. To address the issues in ToF cameras, many methods have been proposed by researchers and are classified into two main categories: Single depth image super-resolution (SR) and color-guided depth image SR. Single depth image SR often uses an external database to establish the relationship between low-resolution (LR) patch and high-resolution (HR) patch to generate HR depth images. Aodha *et al.* [2] used a generic database of HR local patches to upsample LR depth images. Xie *et al.* [3] proposed edge-guided single depth image SR. However, dictionary training and patch matching in these methods have high computational complexity. With the assumption of structural consistency between color and depth images, color-guided depth image SR increases the resolution of LR depth images guided by its aligned HR color image. Kopf *et al.* [4] proposed joint bilateral upsampling for depth SR. In addition to the bilateral filter, weighted mode filter [5], anisotropic diffusion [6], and joint geodesic filter [7] can also be used for depth SR. By minimizing an energy function which contained data and smoothing terms, Diebel and Thrun [8] proposed depth SR based on Markov random field (MRF). However, the assumption of structural consistency between color and depth images does not always hold. When it fails, artifacts such as texture copying and edge blurring are introduced into the reconstructed depth images [9].

In this paper, we propose alternately guided depth super-resolution using weighted least squares (WLS) and zero-order reverse filtering. We adopt WLS for alternating guidance, and alternately utilize color and depth as guidance in WLS to suppress texture copying artifacts. The upsampled depth image guided by color (the first guidance) contains much weaker textures than the original color guidance image [10]. Thus, the upsampled depth image can be also used to guide depth SR. However, directly using the upsampled depth image as the second guidance would cause serious edge blurring artifacts when adjacent pixels have similar colors but distinct depths [11]. To handle the edge blurring artifacts, we perform zero-order reverse filtering and fast median filtering sequentially on the upsampled depth map to generate the second guidance image with clearer edges [12]. Moreover, we also perform zero-order reverse filtering on the input LR depth image (after bicubic interpolation) to recover its high-frequency information. Thanks to zero-order reverse filtering, the edge blurring artifacts are significantly suppressed. Fig. 1 illustrates the entire framework of the proposed depth SR based on alternating guidance.

This work was supported by the National Natural Science Foundation of China (No. 61872280) and the International S&T Cooperation Program of China (No. 2014DFG12780).

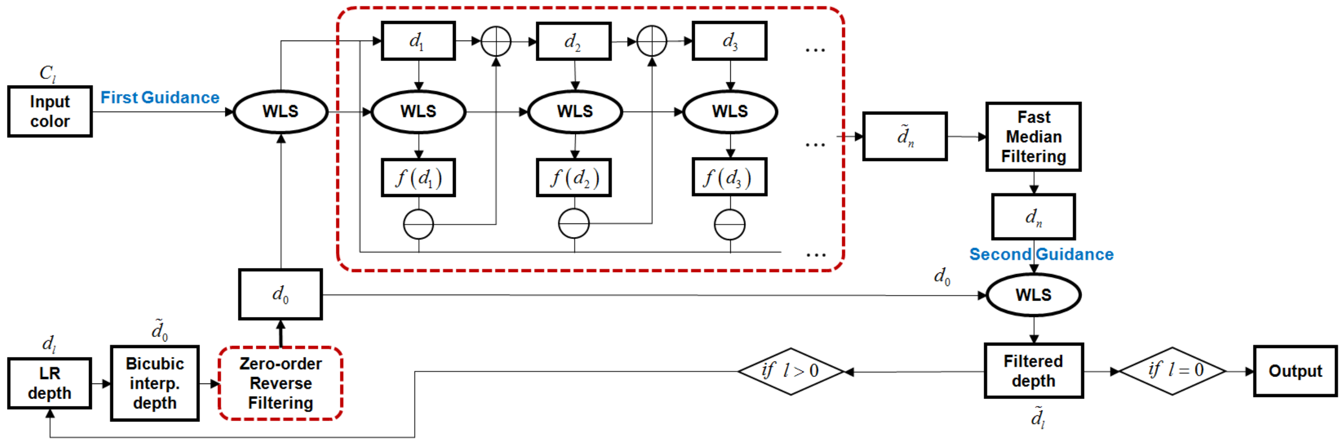


Fig. 1. Entire framework of the proposed depth SR based on alternating guidance. l indicates different processing levels. The number of levels is L, while the total upsampling factor is $M = 2^L$. l starts at $l = L - 1$ and stops at $l = 0$. C_l and d_l are the color guidance image and input depth image of level l, respectively. Operation $f(\cdot)$ represents WLS filtering under the guidance of C_l. Red dotted boxes: Zero-order reverse filtering. Blue texts: Alternating guidance.

2. PROPOSED METHOD

2.1. Weighted Least Squares

Given an input image v and a guidance image g, we obtain an output image u in WLS [13] by minimizing:

$$\varepsilon(u) = \sum_p \{(u_p - v_p)^2 + \lambda \sum_{q \in N(p)} \omega^g_{p,q}(u_p - u_q)^2\} \quad (1)$$

where $N(p)$ represents a four neighbour set of a pixel p, λ controls the balance between data and smoothing terms, increasing λ results in smoother output, $\omega^g_{p,q}$ is the weight calculated from the guidance image g and measures the similarity between pixels p and q. $\omega^g_{p,q}$ is defined as follows:

$$\omega^g_{p,q} = \exp(-\|g_p - g_q\|/\sigma) \quad (2)$$

where σ is a range parameter. $\|\cdot\|$ represents the L_2 norm. Eq. (2) is rewritten in a vector form as follows:

$$\varepsilon(\mathbf{u}) = (\mathbf{u} - \mathbf{v})^{\mathrm{T}}(\mathbf{u} - \mathbf{v}) + \lambda \mathbf{u}^{\mathrm{T}} \mathbf{A}_g \mathbf{u} \quad (3)$$

where $\mathbf{u}$ and $\mathbf{v}$ denote $S \times 1$ column vectors containing values of u and v, respectively, where S is the total number of pixels, T denotes the transposition, and $\mathbf{A}_g$ is an $S \times S$ Laplacian matrix defined in a way similar to the random walk approach [14] as follows:

$$\mathbf{A}_g(m,n) = \begin{cases} \sum\limits_{l \in N(m)} \omega^g_{m,l} & n = m \\ -\omega^g_{m,n} & n \in N(m) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Based on a large sparse matrix, this energy function is solved through a linear system as follows:

$$(\mathbf{I} + \lambda \mathbf{A}_g)\mathbf{u} = \mathbf{v} \quad (5)$$

where $\mathbf{I}$ is an $S \times S$ identity matrix.

However, in depth SR, using one-time WLS is insufficient [10] especially when upsampling factor is large. Based on the fast WLS solver [13], we divide the depth SR problem into multiple scales. In each level, we perform WLS multiple times to get better SR results.

2.2. Alternating Guidance

Color-guided depth SR often causes texture copying artifacts. The upsampled depth image guided by color contains much weaker textures than the original color image [10]. As shown in Fig. 2, the intermediate depth map d_1 generated under the guidance of color image C_l (the first guidance) has less textures than C_l. Thus, using intermediate depth image d_1 as another guidance along with guidance C_l is able to reduce texture copying artifacts. d_1 is obtained by minimizing the following energy function:

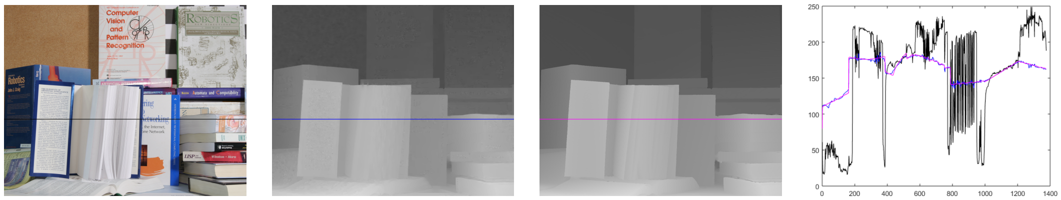
$$\varepsilon(\mathbf{d}_1) = (\mathbf{d}_1 - \mathbf{d}_0)^{\mathbf{T}}(\mathbf{d}_1 - \mathbf{d}_0) + \lambda_1 \mathbf{d}_1{}^{\mathbf{T}} \mathbf{A}_{C_l} \mathbf{d}_1 \quad (6)$$

where $\mathbf{A}_{C_l}$ denotes Laplacian matrix defined by C_l.

However, d_1 contains blurring edges because of the guidance of C_l. To alleviate edge blurring artifacts in d_1, we perform zero-order reverse filtering [12] and fast median filtering sequentially [15], and obtain the second guidance image d_n from d_1. By using d_0 as input and d_n as guidance, the filtered data $\tilde{d}_l$ is obtained by minimizing the following energy function:

$$\varepsilon(\tilde{\mathbf{d}}_l) = (\tilde{\mathbf{d}}_l - \mathbf{d_0})^{\mathbf{T}}(\tilde{\mathbf{d}}_l - \mathbf{d_0}) + \lambda_2 \tilde{\mathbf{d}}_l^{\mathbf{T}} \mathbf{A}_{d_n} \tilde{\mathbf{d}}_l \quad (7)$$

where $\mathbf{A}_{d_n}$ denotes Laplacian matrix defined by d_n.



(a) (b) (c) (d)

Fig. 2. 1D scanline comparison on *Books*. (a) Color image. (b) Intermediate depth map d_1. (c) Ground truth. (d) 1D scanline.

2.3. Zero-Order Reverse Filtering

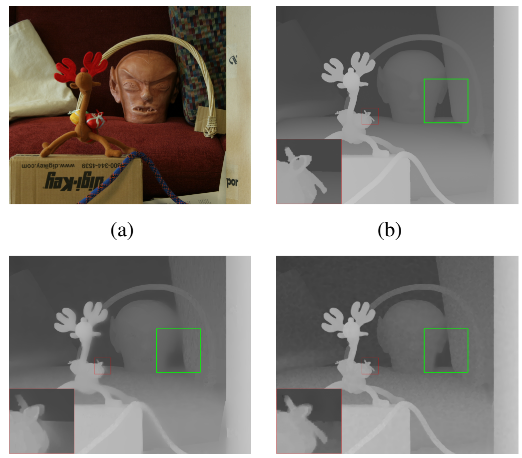
Zero-order reverse filtering, also called DeFilter, removes part of or all filtering effect without needing to know the exact filter in prior [12]. Given the filtering result J^* of image I^* and operator $F(\cdot)$, zero-order reverse filtering is formulated by an iterative scheme as follows:

$$X^{t+1} = X^t + J^* - F(X^t) \tag{8}$$

where $J^* = F(I^*)$, X^t is the current estimate of I^* in the t-th iteration.

Previous work for color-guided depth SR [10], [16] used the bicubic interpolated depth image as the input. However, bicubic interpolation usually generates blurring edges in depth images. Thus, we can improve the input depth d_0 with clearer edges by performing zero-order reverse filtering on bicubic interpolated depth image $\tilde{d}_0$.

In our method, the intermediate depth d_1 is used to reduce texture copying artifacts. However, because d_1 is generated with the guidance of color image C_l, it contains blurring edges at the location where depth has edges but color image does not. To reduce the edge blurring artifacts in d_1, we use zero-order reverse filtering for d_1 to get the enhanced depth image $\tilde{d}_n$. After that, we perform fast median filtering [15] on $\tilde{d}_n$ for denoising (the zero-order reverse filtering increases noise) and get the second guidance image d_n. Compared with the intermediate depth map d_1, the edge blurring artifacts in the second guidance image d_n are significantly reduced as shown in Fig. 3.



(a) (b) (c) (d)

Fig. 3. Performance comparison on *Reindeer*. (a) Color image. (b) Ground truth. (c) Intermediate depth image d_1. (d) Second guidance image d_n.

3. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed method, we compare it with some state-of-the-art methods such as He *et al.* (GF) [17], Park *et al.* (MRF-NLM) [18], Lu *et al.* (CLM-F0) [19], Liu *et al.* (JGF) [7], Ferstl *et al.* (TGV) [20], Li *et al.* (FGI) [10] and Yang *et al.* (AR) [21]. We perform our experiments on a PC with Intel i5-6500 CPU (3.2GHz) and 16GB RAM using Matlab 2015b and C++. In our experiments, we set the smooth weights $\lambda_1 = 10$, $\lambda_2 = 10$. Range parameters of the first and second guidances are set to $\sigma_1 = 0.02$ and $\sigma_2 = 0.003$, respectively. For zero-order reverse filtering, the number of iterations on $\tilde{d}_0$ and d_1 is set to 2 and 3, respectively.

To evaluate the proposed method, we calculate the mean absolute difference (MAD) between the upsampled depth image and ground truth on *Art*, *Book*, *Moebius*, *Reindeer*, *Laundry*, *Dolls* from Middlebury dataset whose resolution is 1376×1088 [22]. Each scene contains the ground truth depth and aligned RGB image. The initial depth is acquired from the ground truth by adding noise and downsampling. Table 1 shows MAD of these methods in the test scenes with different upsampling factors: Smaller values represent better results. In Table 1, bold and underlined numbers indicate the best and second performance, respectively. In most scenes, the proposed method achieves the minimum MAD under different upsampling factors. In addition, the speed of the proposed method is much faster than AR and TGV, i.e. the average runtime of the proposed method on Middlebury dataset is 4.74, 5.81, 6.24 and 6.31 s/image for x2, x4, x8 and x16, respectively, while those of AR and TGV are more than 3000 and 800 s/image for each scale factor, respectively. Fig. 4 shows visual comparison between different methods on *Art*. As shown in the figure, the proposed method performs well in removing noise and reducing texture copying and edge

Table 1. MAD comparison on Middlebury dataset under noisy environment. The best and second performance is highlighted in bold and underlined, respectively.

	Art				*Books*				*Moebius*				*Reindeer*				*Laundry*				*Dolls*			
	x2	x4	x8	x16	x2	x4	x8	x16	x2	x4	x8	x16	x2	x4	x8	x16	x2	x4	x8	x16	x2	x4	x8	x16
Bicubic	3.52	3.84	4.47	5.72	3.30	3.37	3.51	3.82	3.28	3.36	3.50	3.80	3.39	3.52	3.82	4.45	3.35	3.49	3.77	4.35	3.28	3.34	3.47	3.72
GF [17]	1.49	1.97	3.00	4.91	0.80	1.22	1.95	3.04	1.18	1.90	2.77	3.55	1.29	1.99	2.99	4.14	1.28	2.05	3.04	4.10	1.19	1.94	2.80	3.50
MRF-NLM [18]	1.69	2.40	3.60	5.75	1.12	1.44	1.81	2.59	1.13	1.45	1.95	2.91	1.20	1.60	2.40	3.97	1.28	1.63	2.20	3.34	1.14	1.54	2.07	3.02
CLMF0 [19]	1.19	1.77	2.95	4.91	0.90	1.48	2.38	3.36	0.87	1.44	2.32	3.30	0.96	1.56	2.54	3.85	0.94	1.55	2.50	3.81	0.96	1.54	2.37	3.25
JGF [7]	2.36	2.74	3.64	5.46	2.12	2.25	2.49	3.25	2.09	2.24	2.56	3.28	2.18	2.40	2.89	3.94	2.16	2.37	2.85	3.90	2.09	2.22	2.49	3.25
TGV [20]	0.82	1.26	2.76	6.87	0.50	0.74	1.49	2.74	0.56	0.89	1.72	3.99	0.59	0.84	1.75	4.40	0.61	1.59	1.89	4.16	0.66	1.63	1.75	3.71
FGI [10]	0.79	1.17	2.01	3.65	0.58	0.80	1.13	1.75	0.58	0.80	1.15	1.71	0.65	0.89	1.36	2.37	0.65	0.97	1.49	2.43	0.67	0.91	1.31	**1.95**
AR [21]	0.76	1.01	1.70	3.05	0.47	0.70	1.15	1.81	**0.46**	0.72	1.15	1.92	**0.48**	0.80	1.29	**2.02**	**0.51**	0.85	**1.30**	**2.24**	0.59	0.91	1.32	2.08
Proposed	**0.61**	**0.91**	**1.60**	**3.02**	**0.45**	**0.62**	**0.97**	**1.56**	**0.46**	**0.66**	**1.05**	**1.59**	0.52	**0.73**	**1.19**	2.13	0.53	**0.82**	**1.30**	2.26	**0.57**	**0.84**	**1.29**	2.01

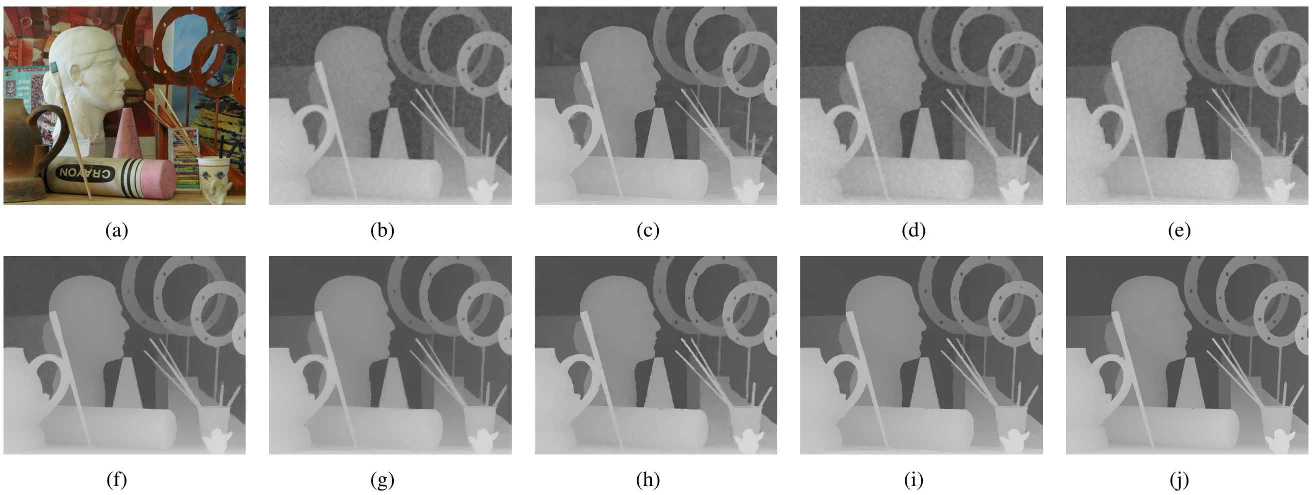


Fig. 4. Visual comparison of x8 upsampling on *Art*. (a) Color image. (b) GF [17]. (c) MRF-NLM [18]. (d) CLMF0 [19]. (e) JGF [7]. (f) TGV [20]. (g) FGI [10]. (h) AR [21]. (i) Proposed method. (j) Ground truth.

blurring artifacts.

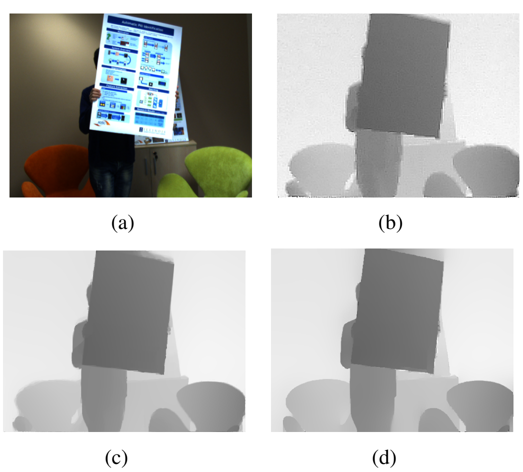


Fig. 5. Visual comparison on ADSC dataset [23]. (a) Color image. (b) Bilinear interpolation. (c) TGV [20]. (d) Proposed method.

We also perform experiments on the real-world ADSC dataset [23]. In ADSC dataset, LR depth images (with the resolution of 176×144) and aligned HR color images (with the resolution of 512×384) are given in pair. We set $\lambda_1 = 100$, $\lambda_2 = 30$, $\sigma_1 = 0.01$, $\sigma_2 = 0.007$. Fig. 5 shows the results on ADSC dataset. Compared with TGV, the proposed method achieves better performance in suppressing texture copying and edge blurring artifacts.

4. CONCLUSION

In this paper, we have proposed alternating guidance based depth SR using WLS and zero-order reverse filtering. To suppress texture copying artifacts, we have alternately used color and depth as guidance in WLS. To alleviate edge blurring artifacts, we have performed zero-order reverse filtering on the intermediate and bicubic interpolated depth images. Experimental results demonstrate that the proposed method significantly removes noise while reducing texture copying and edge blurring artifacts as well as outperforms state-of-the-art methods in terms of MAD.

5. REFERENCES

[1] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 141–159, 2010.

[2] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proc. Eur. Conf. Comput. Vis. Workshop*. Springer, 2012, pp. 71–84.

[3] J. Xie, R. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 428–438, 2016.

[4] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 96, 2007.

[5] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1176–1190, 2012.

[6] J. Liu and X. Gong, "Guided depth enhancement via anisotropic diffusion," in *Proc. PCM*. Springer, 2013, pp. 408–417.

[7] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 169–176.

[8] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Proc. NIPS*, 2005, pp. 291–298.

[9] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. Eur. Conf. Comput. Vis. Workshop*. Springer, 2008, pp. 1–12.

[10] Y. Li, D. Min, M. N. Do, and J. Lu, "Fast guided global interpolation for depth and motion," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 717–733.

[11] B. Ham, M. Cho, and J. Ponce, "Robust image filtering using joint static and dynamic guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2015, pp. 4823–4831.

[12] X. Tao, C. Zhou, X. Shen, J. Wang, and J. Jia, "Zero-order reverse filtering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 222–230.

[13] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5638–5653, 2014.

[14] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.

[15] Q. Zhang, L. Xu, and J. Jia, "100+ times faster weighted median filter (WMF)," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2830–2837.

[16] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2007, pp. 1–8.

[17] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2010, pp. 1–14.

[18] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3D-ToF cameras," in *Proc. IEEE Int. Conf. Comput. Vis.* IEEE, 2011, pp. 1623–1630.

[19] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2012, pp. 430–437.

[20] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 993–1000.

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

[22] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2007, pp. 1–8.

[23] B. Ham, D. Min, and K. Sohn, "Depth superresolution by transduction," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1524–1535, 2015.