Non-local L_0 Gradient Minimization Filter and Its Applications for Depth Image Upsampling

Hang Yang[1(✉)], Xueqi Sun[2], Ming Zhu[1], and Kun Wu[1]

[1] Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Science, Changchun 130033, China
yanghang@ciomp.ac.cn, zhuming@163.com, wukun_ciomp@126.com
[2] Changchun University of Science and Technology, Changchun 130022, China
xueqiSun@163.com

Abstract. In this work, we propose a non-local L_0 gradient minimization filter. The nonlocal idea is to restore an unknown pixel using other similar pixels, and the nonlocal gradient model has been verified for feature and structure-preserving. We introduce the nonlocal idea into a L_0 gradient minimization approach, which is effective for preserving major edges while eliminating the low-amplitude structures. An optimization framework is designed for achieving this effort. Many optimized based filters do not have the property of joint filtering, so they can not be used in many problems, such as joint denoising, joint upsampling, while the proposed filter not only inherits the advantages of the L_0 gradient minimization filter, but also has the property of the joint filtering. So our filter can be applied to joint super resolution. With the guidance of the high-resolution image, we propose upsampling the low-resolution depth image with the proposed filter. Experimental results demonstrate the effectiveness of our method both qualitatively and quantitatively compared with the state-of-the-art methods.

Keywords: L_0 norm · Non-local · Joint filtering · Edge-preserving

1 Introduction

Edge-aware image processing technique is broadly studied for smoothing images without destroying different levels of structures. It is wildly applied for computer graphics community. Edge-preserving filters can be broadly divided into two broad categories: average based approaches and optimization based approaches.

The methods of first class smooth images by taking a weighted average of nearby pixels, where the weights depend on the intensity/color difference. Average based filters include bilateral filter [17], nonlocal means filter [1], guided image filter [8] and rolling guidance filter [23]. They often use guidance image to

define the similarity between pixels. The main drawback of these filters is that they will produce the halo effect near the edge.

The total variation (TV) model [15], L_0 gradient minimization filter (L_0 filter) [20], weighted least squares (WLS) [3] and curvature filter [7] belong to the optimization based methods. These approaches smooth images by optimizing objective functions containing terms defined in L_p norm ($p = 0, 1, 2$). Although the optimization based methods can avoid the halo effect along salient edges and often generate high quality results, it does not have the property of joint filtering with reference image, and this shortcoming limits their applications.

Recently, the nonlocal framework has been extensively studied by many scholars as a regularization term to overcome the staircase effect and obtain better performance. Gilboa and Osher defined a variational functional based nonlocal TV operators [6]. Zhang et al. [24] proposed a fast split Bregman iteration for this nonlocal TV minimization. Lately, the nonlocal regularizations are extended to process more general inverse problems in [12]. However, they penalize large nonlocal gradient magnitudes, and it possibly influence contrast during smoothing.

In summary, most image smooth models aim to preserve edges from noise and textures, and each of them has its limitations. In this work, we present a new edge-preserving filter based on an optimization framework, which incorporates the nonlocal strategy into the L_0 gradient minimization model and takes advantage of both variational models and spatial filters. This notion leads to an unconventional global optimization process involving discrete metrics, whose solution is able to manipulate the edges in a variety of ways depending on the saliency.

The proposed framework is general and can be used for several applications. Different from other optimization based methods, the proposed algorithm can use the reference image for joint filtering.

The depth images captured by 3D scanning devices such as ToF camera or Kinect camera may be highly degraded, which have limited resolution and low quality. As a result, it's hard to recover high quality depth maps from single depth image. Fortunately, the depth map is often coupled by high resolution (HR) color image which shows the same scene and they have strong structural similarities [4,19,22]. In recently, deep learn based depth upsampling methods [5,9,11,16] achieve well results. These methods produce the end-to-end upsampling networks, which learn high-resolution features in the intensity image and supplement the low-resolution depth structures in the depth map.

So this paper applies the proposed filter for depth image super resolution and treats the natural image as the reference image. With the guidance of the high-resolution RGB image, the proposed algorithm is well suited for upsampling the low-resolution depth image and it can not only reduce noises, but also preserve the sharp edges during super resolution. With simulations, the experimental results demonstrate that the proposed approach is promising, and it does significantly improve the visual quality of the low-resolution depth image compared with the existing upsampling methods.

2   Non-local L_0 Gradient Minimization

Different from the definition of gradient, the nonlocal gradient $\nabla_\omega S_p$ of each pixel p on the image S is defined as follows:

$$\nabla_\omega S_p = \{S_p(q), \forall q \in \Omega_p\} \tag{1}$$

where

$$S_p(q) = (S_q - S_p)\sqrt{\omega(p, q)} \tag{2}$$

and $S_p(q)$ is the vector element corresponding to q, $\omega(p, q)$ is the weight function, which is assumed to be nonnegative and symmetric, it measures the similarity features between two patches (the size is $m \times m$) centered at the pixels p and q, Ω_p is a search window centered at the pixel p (the size of Ω_p is $n \times n$) [12, 24].

The weight function $\omega(p, q)$ in Ω_p has the form:

$$\omega(p, q) = \frac{1}{C_p} \exp(-\frac{(G_a* \mid J(p + \cdot) - J(q + \cdot) \mid^2)(0)}{h^2}) \tag{3}$$

and the normalizing factor C_p is

$$C_p = \sum_{q \in \Omega_p} \exp(-\frac{(G_a* \mid J(p + \cdot) - J(q + \cdot) \mid^2)(0)}{h^2}) \tag{4}$$

where G_a is the Gaussian kernel with standard deviation a, h is a smoothing parameter, and J is a reference image which can be chosen according to different applications.

In this work, we denote the input image and filtered image as I and S, respectively. Our nonlocal gradient measure is written as

$$C(S) = \sharp\{p \mid \parallel \nabla_\omega S_p \parallel_1 \neq 0\} \tag{5}$$

It counts p whose magnitude

$$\sum_{q \in \Omega_p} \mid (S_q - S_p)\sqrt{\omega(p, q)} \mid$$

is not zero. Based on this definition, we can estimate S by solving:

$$\min_S\{\parallel S - I \parallel_2^2 + \lambda C(S)\} \tag{6}$$

The first term constrains image structure similarity.

It is a discrete counting metric involved in Eq. (6). These two terms describe the pixel-wise difference and global discontinuity respectively, it is commonly regarded as computationally intractable. In this work, we introduce an auxiliary variable based on the half-quadratic splitting method, which can expand the original terms and update them iteratively. This approach leads to an alternating optimization strategy.

Due to the discrete nature, our method contains new subproblems, and it is different from other L_0-norm regularized optimization problems. Although the proposed method can only approximate the solution of Eq. (6), but it can make the original problem easier to handle and inherit the property to maintain salient structures [20].

The auxiliary variables $\mathbf{d}_p$ are introduced, and they are corresponding to $\nabla_\omega S_p$. We can rewrite the cost function as

$$\min_{S,\mathbf{d}}\{\sum_p (S_p - I_p)^2 + \lambda C(\mathbf{d}_p) + \beta \parallel \mathbf{d}_p - \nabla_\omega S_p \parallel_2^2\} \tag{7}$$

where $C(\mathbf{d}) = \sharp\{p \mid \parallel \mathbf{d}_p \parallel_1 \neq 0\}$, and β is a an automatically adapting controlling parameter.

Our split variables approaches motivate us to propose this iterative method. In practice, a good result can be obtained by solving the following two subproblems iteratively.

Subproblem 1: computing S

$$S = \arg\min_S\{\sum_p (S_p - I_p)^2 + \beta \parallel \mathbf{d}_p - \nabla_\omega S_p \parallel_2^2\} \tag{8}$$

Now, the subproblem for S consists in solving the linear equations

$$(S - I) - \beta div_\omega(\nabla_\omega S - \mathbf{d}) = 0 \tag{9}$$

which provides

$$S = (1 - \beta\Delta_\omega)^{-1}(I - \beta div_\omega\mathbf{d}) \tag{10}$$

Here, $div_\omega\mathbf{d}$ is defined as the divergence of $\mathbf{d}$, and its discretization at p can be written as

$$div_\omega\mathbf{d}_p = \sum_{q\in\Omega_p} (\mathbf{d}_p(q) - \mathbf{d}_q(p))\sqrt{\omega(p,q)} \tag{11}$$

The non-local Laplacian Δ_ω is defined as

$$\Delta_\omega S = div_\omega\nabla_\omega S = \sum_{q\in\Omega_p} (S_q - S_p)\omega(p,q) \tag{12}$$

Since the non-local Laplacian is negative semi definite, the operator $1 - \Delta_\omega$ is diagonally dominant. Therefore we can solve S by a Gauss-Seidel algorithm.

Subproblem 2: computing d

$$\mathbf{d} = \arg\min_{\mathbf{d}}\{\sum_p \parallel \mathbf{d}_p - \nabla_\omega S_p \parallel_2^2 + \frac{\lambda}{\beta}C(\mathbf{d}_p)\} \tag{13}$$

This subproblem can be solved efficiently because the Eq. (13) can be spatially decomposed where $\mathbf{d}_p$ are estimated individually. It is the main benefit

of the proposed scheme, which makes the altered problem empirically solvable. Equation (13) is accordingly decomposed to:

$$E_p = \| \mathbf{d}_p - \nabla_\omega S_p \|_2^2 + \frac{\lambda}{\beta} H(\mathbf{d}_p) \tag{14}$$

where $H(\mathbf{d}_p)$ is a binary function returning 1 if $\| \mathbf{d}_p \|_1 \neq 0$ and 0 otherwise.

Equation (14) reaches its minimum E_p^* under the condition

$$\mathbf{d}_p = \begin{cases} \mathbf{0}, & \| \nabla_\omega S_p \|_2^2 \leq \lambda/\beta \\ \nabla_\omega S_p, & otherwise \end{cases} \tag{15}$$

Proof: (1) When $\lambda/\beta \geq \| \nabla_\omega S_p \|_2^2$, non-zero $\mathbf{d}_p$ yields

$$E(p) = \| \mathbf{d}_p - \nabla_\omega S_p \|_2^2 + \lambda/\beta \tag{16}$$
$$\geq \lambda/\beta \tag{17}$$
$$\geq \| \nabla_\omega S_p \|_2^2 \tag{18}$$

Note that $\mathbf{d}_p = \mathbf{0}$ leads to

$$Ep = \| \nabla_\omega S_p \|_2^2 \tag{19}$$

Comparing Eq. (16), the minimum energy $E_p = \| \nabla_\omega S_p \|_2^2$ is produced when $\mathbf{d}_p = \mathbf{0}$.

(2) When $\lambda/\beta < \| \nabla_\omega S_p \|_2^2$, Eq. (19) still holds. But when $\mathbf{d}_p = \nabla_\omega S_p$, E_p has its minimum value λ/β. Comparing these two values, the minimum energy E_p is produced when $\mathbf{d}_p = \nabla_\omega S_p$.

Parameter β is automatically adapted in iterations starting from a small value, it is multiplied by 2 each time. This scheme is effective to speed up convergence [20].

Continuous nonlocal gradient L_1 norm was enforced in nonlocal total variation (NLTV) [24] smoothing to suppress noise. In our method, strong smoothing inevitably curtails originally salient edges to penalize their magnitudes. In this framework, large nonlocal gradient magnitudes are allowed by nature with our discrete counting measure.

In Fig. 1, we show a natural image smoothing example compared with other competitive algorithms. One can see that L_0 filter [20] ($\lambda = 0.035$) generates a sharp but not completely smooth image which is shown in Fig. 1(b). Many details are still retained after filtering, such as flower diameter and butterfly, it is not good enough for applications. The result obtained by NLTV [24] ($\lambda = 0.05$) is shown in Fig. 1(c), in the case of overall non-local gradients with small energies, the edges are not sharp, which makes them difficult to distinguish low contrast details around. In Fig. 1(d), our result ($\lambda = 0.05$) contains the most significant structures, which are slightly sharper as the nonlocal gradient energy increases.

Fig. 1. Visual quality comparison of image smoothing. (a) original image, (b) the result of L_0 filter, (c) the result of NLTV, (d) the result of our method.

Our alternating minimization method is described in Algorithm 1.

Algorithm 1. Non-local L_0 Gradient Minimization

---

1. Input: image I, reference image J, smoothing weight λ , parameters β
2. Initialization: $S^0 = I$, compute $\omega(\cdot, \cdot)$ using Eqs. (3)–(4).
3. for $i = 0 : MaxIters$
      With S^i, solve for $\mathbf{d}^i$ in Eq. (15);
      With $\mathbf{d}^i$, solver for S^{i+1} with Eq. (12).
   end
4. Output: result image $S^{MaxIters}$.

In [14], Petschnigg *et al.* proposed to denoise a no-flash image with its flash version as the reference image. In Fig. 2, we show a comparison of using the joint bilateral filter (JBF) [14], NLTV [24] and our method. Although JBF works well, from the Fig. 2(b), one can find that the gradient inversion artifacts are significant near some edges. And NLTV does not obtain a satisfactory result. Our result is sharper and contains few noise, which is shown in Fig. 2(d).
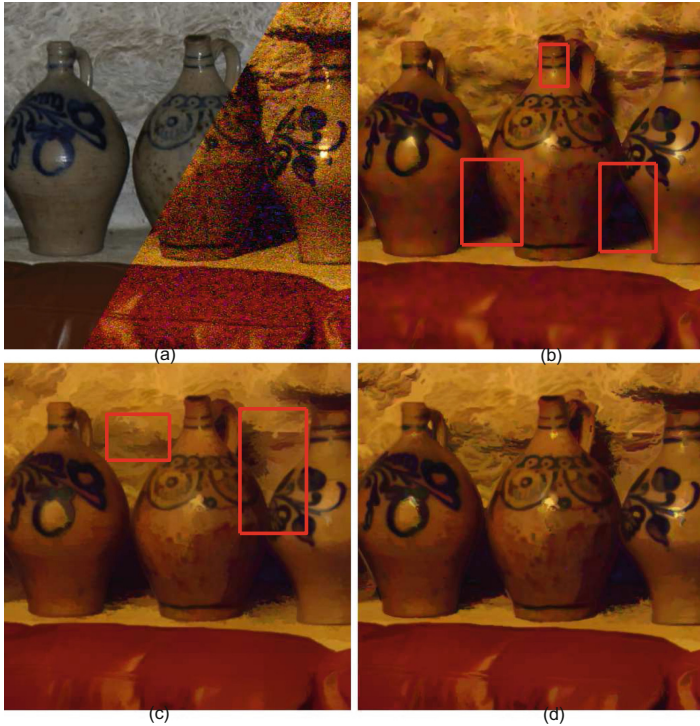
Fig. 2. Visual quality comparison of flash/no flash denoising. (a) original flash/no flash image pair, (b) the result of JBF, (c) the result of NLTV, (d) the result of our method.

3   Depth Image Upsampling

In this application, we upscale a single depth image d (size of $m \times n$) which is guided by a high-resolution natural image T (size of $M \times N$). One can see that depth images are textureless compared with natural images and have quite sparse gradients. However, according to the statistics of depth image gradient [21], the sparse gradient assumption is not accurate enough. That is to say, most gradient values of depth image are not always 0 but rather very small.

The proposed nonlocal L_0 gradient regularization can reduce the penalty for small elements, because we deal with the nonlocal gradient of the image as a whole, take into account the energy sum of the multi-directional weighted gradients, and avoid to obtain an overly smooth result.

In the first step, we upsample the depth image d to the size of $M \times N$ with nearest neighbor interpolation, and obtain an initial image D. In the second step, we compute the weights $\omega(p, q)$ with the high-resolution natural image T in Eqs. (3) and (4), that is to say, T is used as the reference image. In the last step, we use D as the input image and solve the minimization problem Eq. (7). The result of Eq. (7) is the final joint upsampling image.

Table 1. PSNR (in dB) comparison on middlebury 2007 datasets with added noise for magnification factors (×4, ×8).

	Art		Books		Moebius	
	×4	×8	×4	×8	×4	×8
He et al. [8]	29.98	**27.72**	33.94	31.99	32.23	30.22
Park et al. [13]	29.69	27.42	34.36	32.28	33.28	31.12
Chan et al. [2]	29.88	27.15	33.82	31.12	32.46	29.51
SRF [10]	29.79	26.94	33.13	31.92	31.43	29.86
Ours	**30.46**	27.53	**35.98**	**33.20**	**34.41**	**31.32**

Table 2. SSIM comparison on middlebury 2007 datasets with added noise for magnification factors (×4, ×8).

	Art		Books		Moebius	
	×4	×8	×4	×8	×4	×8
He et al. [8]	0.84	0.78	0.80	0.74	0.78	0.71
Park et al. [13]	0.92	0.88	0.89	0.84	0.90	0.83
Chan et al. [2]	0.80	0.71	0.91	0.89	0.91	0.87
SRF [10]	0.82	0.72	0.74	0.64	0.73	0.62
Ours	0.97	0.94	0.96	0.93	0.97	0.94

We show some experimental evaluations of our algorithm compared with the competitive methods for depth image upsampling. We work on 3 depth images from Middlebury 2007 datasets [4] with the scaling factors of 4 and 8, respectively. To simulate the acquisition process, these depth images are added Gaussian noise [13].

The numerical results for this experiment in terms of the PSNR are shown in Table 1. In our experiments, our method clearly outperforms the other four methods in the most cases.

The numerical results for this experiment in terms of the Peak Signal Noise Ratio (PSNR) are shown in Table 1 and Structural Similarity (SSIM) [18] in Table 2. From the Table 1, one can see that our method clearly outperforms the other four method in the most cases. In Table 2, the proposed method achieve significant SSIM improvements over other leading methods. In average, our algorithm outperforms other methods by 0.05 for the SSIM comparison.

To show the visual comparison clearly, we show some results of experiments in Fig. 3. One can find that our method can enhance edges and reduce noise better, whereas other algorithms suffer from edge blurring or noise. From Table 1 and Figs. 3, 4 and 5 one can observe that the proposed approach is effective for noisy complex scenes and can obtain clearer high resolution depth images.

In order to show the stability of the proposed deconvolution algorithm, we give the convergence curve of the alternative optimization in Fig. 6. We plot the
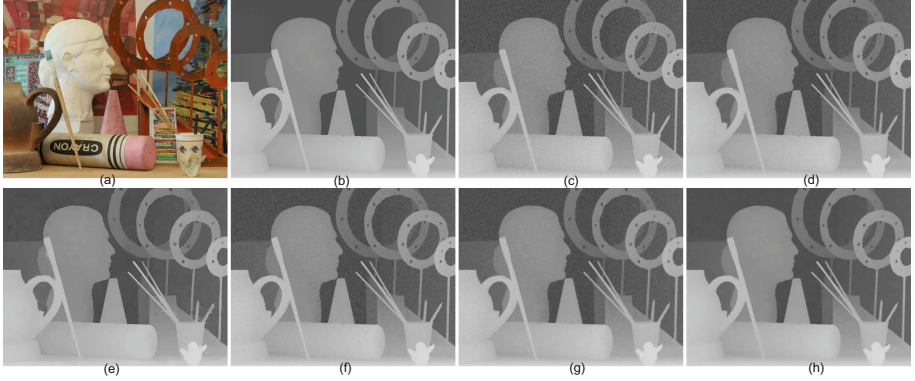
Fig. 3. Joint upsampling on "Art" image. (a) high-resolution RGB image, (b) original depth map, (c) low-resolution and noisy depth image (enlarged using nearest neighbor upsampling), (d) He et al. [8], (e) Park et al. [13], (f) Chan et al. [2], (g) SRF [10], (h) our method.
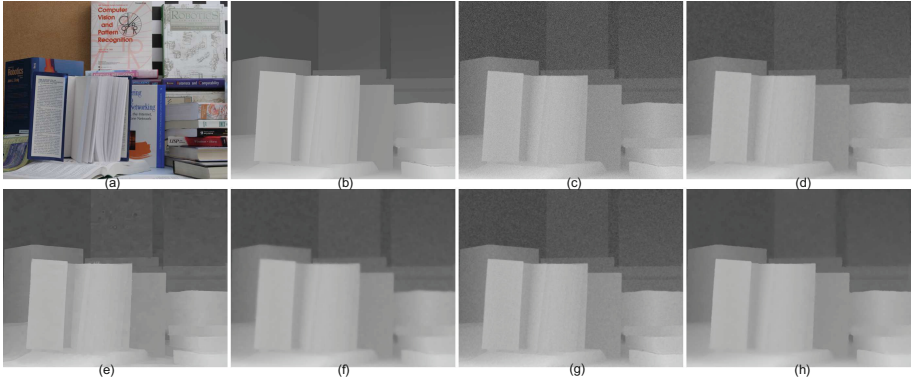

Fig. 4. Joint upsampling on "Books" image. (a) high-resolution RGB image, (b) original depth map, (c) low-resolution and noisy depth image (enlarged using nearest neighbor upsampling), (d) He et al. [8], (e) Park et al. [13], (f) Chan et al. [2], (g) SRF [10], (h) our method.

histories of the relative error $| S^{k+1} - S^k |$. Three depth images (*Art*, *Book* and *Moebius*) are used and the scaling factor is 8. It is noticeable that the proposed method is stable.

In order to improve computational time and storage efficiency, we only compute the "best" neighbors, that is, for each pixel p, we only include the 10 best neighbors in the searching window of 7×7 centered at p and the size of patch is 5×5, the parameters a and h are empirically set to 0.5 and 0.25, respectively. 7–10 iterations are generally performed in our algorithm.

For computational time, the proposed approach takes about 2.3 s for a computer which runs Windows 7 64bit version with Intel Core i5 CPU and 8 GB RAM to construct the weight function of a 256×256 image in Matlab 2010b.
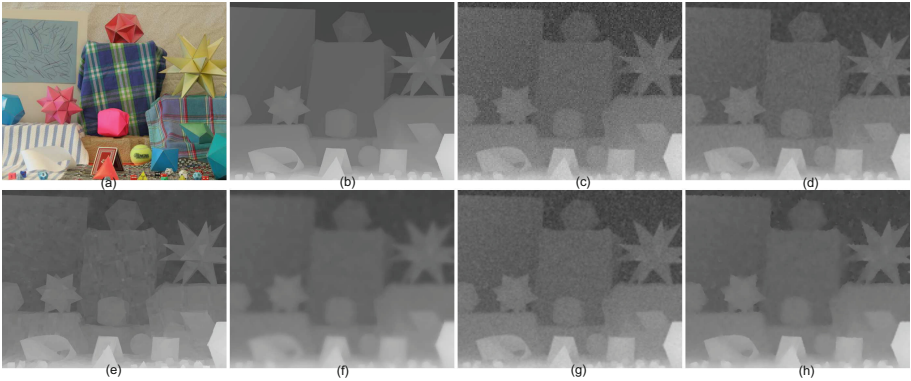
Fig. 5. Joint upsampling on "Moebius" image. (a) high-resolution RGB image, (b) original depth map, (c) low-resolution and noisy depth image (enlarged using nearest neighbor upsampling), (d) He et al. [8], (e) Park et al. [13], (f) Chan et al. [2], (g) SRF [10], (h) our method.
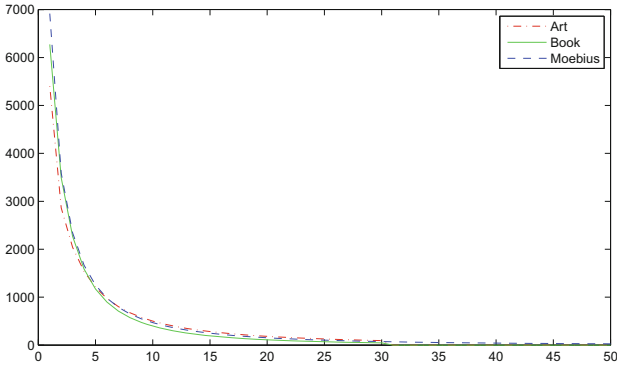

Fig. 6. Convergence curves of the alternative optimization.

Once the weight is constructed, the iteration of our method is comparable to ROF [15] in speed. The computation speed depends on the number of iterations. In general, it takes around 3.5 s for 10 iterations.

4   Conclusion

In this work, we propose a solution for nonlocal L_0 gradient minimization and show its applications for depth image upsampling. We propose an effective smoothing approach based on minimizing discretely counting nonlocal spatial changes. Different from many optimized based filters, the proposed method has the property of joint filtering, so our filter can be used for many applications. In particular, it achieves good performance in the depth image super resolution.

Treating the high-resolution RGB image as a reference image, the proposed algorithm is well suited for upsampling the low-resolution depth image. The experimental results demonstrate that the proposed approach is promising, and it has better objective performance compared to the existing upsampling methods.

References

1. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 60–65. IEEE (2005)
2. Chan, D., Buisman, H., Theobalt, C., Thrun. S.: A noise-aware filter for real-time depth upsampling. In: The Workshop on Multi-camera Multi-modal Sensor Fusion Algorithms Applications (2008)
3. Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Trans. Graph. **27**, 67 (2008)
4. Ferstl, D., Reinbacher, C., Ranftl, R., Ruether, M., Bischof, H.: Guided depth upsampling using anisotropic total generalized variation. IEEE International Conference on Computer Vision, pp. 993–1000. IEEE (2013)
5. Ferstl, D., Ruther, M., Bischof, H.: Variational depth superresolution using example-based edge representations. In: IEEE International Conference on Computer Vision (ICCV), pp. 513–521. IEEE (2015)
6. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. **7**, 1005–1028 (2008)
7. Gong, Y.: Bernstein filter: a new solver for mean curvature regularized models. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1701–1705. IEEE (2016)
8. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397–1409 (2013)
9. Hornacek, M., Rhemann, C., Gelautz, M., Rother, C.: Depth super resolution by rigid body self-similarity in 3D. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1123–1130. IEEE (2013)
10. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5197–5206. IEEE (2015)
11. Hui, T.-W., Loy, C.C., Tang, X.: Depth map super-resolution by deep multi-scale guidance. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 353–369. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_22
12. Lou, Y., Zhang, X., Osher, S.: Image recovery via nonlocal operators. J. Sci. Comput. **42**, 185–197 (2010)
13. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.: High quality depth map upsampling for 3D-TOF cameras. In: IEEE International Conference on Computer Vision (ICCV), pp. 1623–1630. IEEE (2011)
14. Petschnigg, G., Szeliski, R., Agrawala, M.: Digital photography with flash and no-flash image pairs. ACM Trans. Graph. (TOG) **23**, 664–672 (2004)
15. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenom. **60**, 259–268 (1992)

16. Song, X., Dai, Y., Qin, X.: Deep depth super-resolution: learning depth super-resolution using deep convolutional neural network. arXiv preprint arXiv:1607.01977 (2016)
17. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: International Conference on Computer Vision, pp. 839–846. IEEE (1998)
18. Wang, Z., Bovik, A.C., Sheikh, H.R.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
19. Xie, J., Feris, R.S., Sun, M.T.: Edge-guided single depth image super resolution. IEEE Trans. Image Process. **25**, 428–438 (2016)
20. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via L_0 gradient minimization. ACM Trans. Graph. **30**, 174 (2011)
21. Xue, H., Zhang, S., Cai, D.: Depth image inpainting: improving low rank matrix completion with low gradient regularization. arXiv preprint arXiv:160405817 (2016)
22. Yang, J., Ye, X., Li, K., Hou, C., Wang, Y.: Color-guided depth recovery from RGB-D data using an adaptive autoregressive model. IEEE Trans. Image Process. **23**, 3443–3458 (2014)
23. Zhang, Q., Shen, X., Xu, L., Jia, J.: Rolling guidance filter. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 815–830. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_53
24. Zhang, X., Burger, M., Bresson, X.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**, 253–276 (2010)