DEPTH MAP SUPER-RESOLUTION USING NON-LOCAL HIGHER-ORDER REGULARIZATION WITH CLASSIFIED WEIGHTS

Hai-Tao Zhang[⋆†] *Jun Yu*[⋆] *Zeng-Fu Wang*[⋆]

[⋆] University of Science and Technology of China, Hefei, China.
[†] Southwest University of Science and Technology, Mianyang, China

ABSTRACT

High-order regularization in depth map super-resolution (SR) contributes to producing smoother depth map. However, assigning appropriate weights within regularization term is also important for preserving more detail information. In this paper, a novel and more adaptive depth SR model is proposed by using non-local total generalized variation (NLTGV) with classified weights. A random forest based classifier is trained to classify the pixels of depth map into four categories on the basis of several local structure features, such as gradient magnitude and texture energy extracted from color image, and then the weights within NLTGV are assigned with four groups of parameters corresponding to the four kinds of pixels. Evaluation results demonstrate that the local features make pixels have a good separability, and the classified weights can obviously improve the accuracy of depth map SR.

Index Terms— Depth map super-resolution, non-local generalized total variation, classified weights.

1. INTRODUCTION

The task of depth SR is to recover a high-resolution (HR) depth map given corresponding low-resolution (LR) map. It is an inherently ill-posed problem since a LR depth map can be mapped to several HR depth maps. One alternative way for tackling this problem is to incorporate some explicit smoothness assumptions and determine the HR depth map by energy minimization techniques, such as Markov random field (MRF) methods [1–3], variational methods [4–6] and filtering based methods [7–11]. Generally, energy function consists of two parts: data fidelity term and regularization term, and a weight function $\alpha(\Phi, w)$ is incorporated into regularization term for tuning its effect. Φ represents some guidance information such as the color image captured together with LR depth map, and w denotes a tunable parameter set, such as the variances of filter kernel in joint-filtering method [8].

In previous literatures, there exists several depth SR methods using the color image as guidance information to complete the up-sampling of depth map. Diebel et al. [1] performed depth SR based on a MRF framework, where the regularization term was weighted according to texture derivatives. Further, Liu et al. [2] combined the depth map own property with MRF model and improved the depth SR results. Yang et al. [7] constructed a depth cost volume with LR depth image and filtered it through joint bilateral filtering (JBF) [8], and then the HR map is obtained by using winner-takes-all approach. Chan et al. [9] combined bilateral filtering (BF) with JBF for depth SR. Yang et al. [12] used auto-regressive model to construct SR model by using non-local means regularization. A more complex approach was proposed by Park et al.[3] with non-local structure regularization. Ham et al. [13] solved the depth SR problem by using an undirected graph model. Liu et al. [14] proposed robust image-guided depth SR method using adaptive weights. The work [4] based on local total generalized variation showed the superiority of high-order regularization. Zhang et al. [6] proposed non-local total generalized variation based SR model and obtained a promising depth SR result. There also exist other depth SR methods such as sparse representation based methods [15] and deep learning based methods [16]. In this paper we mainly focus on the image-guided depth SR methods. While much progress has been made in image-guided depth SR, most of methods assign the tuning parameters with same values for all pixels. The fixed parameters are difficult to balance the smoothness of homogenous region and sharpness of depth edges. It frequently results in either over-smoothness of some depth discontinuities or texture copy artifacts in some texture-rich regions. To better overcome above shortage, we propose a local structure classification scheme for classifiedly assigning the tuning parameters to obtain more accurate depth SR result.

The contribution of this paper is two-fold. (i) A novel depth map SR approach is proposed by using NLTGV with more flexible weights. The first- and second-order term of NLTGV in our model have not only common support weights, but also the explicit independent coefficients. The common weights are assigned by fusing the structure information from the companion color image and interpolated depth map, and the independent coefficients are flexibly set to balance the effect of the first- and second-order terms. (ii) To enhance the adaptability of model for the local structure change of depth map, the tuning parameters within weights function are pixel-wisely and classifiedly assigned by considering the local structure features around pixels. Firstly, a random forest

based classifier is trained to classify the pixels of depth map into four categories by using several local features, including the gradient magnitude and texture energy from companion color image, and the multi-scale gradient magnitude from interpolated depth map. Then, the parameters corresponding to the same category pixels are assigned by a common group of values, which are determined in an experiment means. Moreover, some rules for parameter manual fine-tuning are also summarized as a reference for similar work in the future.

2. METHODOLOGY

2.1. Depth SR Model

To construct more adaptive depth SR model, we first redefine the second-order NLTGV [17] with more flexible weights as

$$\mathcal{T}(u) = \sum_{i\in\Omega}\sum_{j\in B_r(i)} \alpha_{ij}\left\{\rho|\tilde{u}_j - u_j| + \kappa\|\omega_i - \omega_j\|_1\right\}, \quad (1)$$

where Ω is the image domain, $B_r(i)$ represents the neighbourhood with radius size r around pixel i, u_j is the value of pixel j on the map u, $\tilde{u}_j := u_i + \langle\omega_i, j-i\rangle$ is a local plane through the point (i, u_i), $\omega_i \in \mathcal{R}^2$ is the normal of plane, $\alpha_{ij} \in \mathcal{R}^+$ denotes the support weight satisfying $\sum_j \alpha_{ij} = 1$, and ρ, κ are free parameters for balancing the effect of the corresponding terms. Refer to [17] [18] for details about NLTGV. The depth-map SR model based on NLTGV is defined as:

$$\hat{u} := \arg\min_u \mathcal{T}(u) + \sum\nolimits_i \lambda_i |u_i - d_i|_\varepsilon, \quad (2)$$

where u is HR depth map and d is the depth map generated by projecting LR depth map to HR color image g. $|x|_\varepsilon$ is the Huber norm for data fidelity. If $|x| \le \varepsilon$, $|x|_\varepsilon = |x|^2/(2\varepsilon)$, otherwise, $|x|_\varepsilon = |x| - \frac{\varepsilon}{2}$, where ε is a constant. The weight function is defined as $\alpha_{ij} = z_{ij}^{-1}\alpha_c(i,j)\cdot\alpha_d(i,j)$, where z_{ij} is a normalization constant. The factor α_c measures the color similarity and spatial proximity between pixels i and j as

$$\alpha_c(i,j) = \exp\left(-\|g(i) - g(j)\|^2/(2\sigma_c^2) - \|i-j\|^2/(2\sigma_p^2)\right), \quad (3)$$

where σ_c and σ_p are free parameters. The factor α_d measures the local structure similarity by using the interpolated depth map d_c, as follows:

$$\alpha_d(i,j) = |\langle d_c(i), d_c(j)\rangle| \;/\; (\|d_c(i)\|_2\|d_c(j)\|_2), \quad (4)$$

where $d_c(i)$ denotes the depth patch centered at pixel i and the mean value has been subtracted from the patch. The dot product operation $\langle\cdot,\cdot\rangle$ is used to compute the coherence between two patches. After converting the l_1 norm to its corresponding dual form by the Legendre Fenchel transform, Eq. (2) can be solved in an alternative iteration manner. Refer to [18] for more details.

2.2. Classified Parameter Assignment

In Eq. (2), the weights of NLTGV can be viewed as a parametric function $\alpha(\Phi, w)$, where $\Phi = \{g, d_c\}$ is additional information and $w = \{\rho, \kappa, \sigma_c, \sigma_p\}$ is a tunable parameter set. We note that larger σ_c, σ_p tend to obtain more smoother result and mitigate texture copy artifacts, but it may result in blurry depth edges. Conversely, smaller σ_c, σ_p are advantageous to preserve the sharpness of depth boundaries, while it may lead to more serious texture transfer problem in texture-rich regions. Intuitively, assigning parameters pixel-wisely by considering local structure changes around pixels can get better balance between smoothness of flat region and the sharpness of depth boundaries. Therefore, we first explicitly classify the pixels of depth map into four categories (Fig. 1 (a),(b)): pixels near depth discontinuities (C_1), pixels in flat regions with rich color-textures (C_2), pixels in flat region without salient textures (C_3), and other pixels are classified as the fourth class (C_4). Then, considering above local features, the parameters are pixel-wisely set by category. The pixels from same category share a group of parameters $w_i = \{\rho_i, \kappa_i, (\sigma_c)_i, (\sigma_p)_i\}, i \in \{1,2,3,4\}$. To determine appropriate parameters for pixels in C_i region, a number of small LR depth patches are extracted from a depth map dataset and the patches should contain C_i category of pixels as much as possible. Then these patches are respectively up-sampled to a HR patch by traversing parameter space w with fixed step to set parameters w_i while setting other parameters $w_j, j \ne i$ with typical values. In this way, for each patch, a group of optimum parameters corresponding to the smallest depth SR error are determined, and then we take the average of all optimum parameters of patches as the experimental parameters for C_i pixels. Still, the depth SR output is already fairly good with the experimental parameters, and additional fine-tuning give a further boost as following rules. The parameters ρ and σ_c are assigned with relatively small values in the region C_1 in order to preserve the depth edges. Generally, two local approximate planes at the different sides of edge-tangent have different normals. Hence, we set κ as small as possible for the pixels in class C_1. The rich color textures in region C_2 may result in the texture copy artifacts. Relatively larger parameters σ_c and σ_p are advantageous to mitigate this problem. On the contrary, the region C_3 requires relatively smaller parameters ρ, κ and σ_c to void over-smoothness of depth map.

2.3. Design of Classifier

The remaining problem is to determine the category of local structure for each pixels. A classifier is trained to solve this problem by using three kinds of features, including gradient magnitude (Fig. 1(f)) and Laws' texture energy (Fig. 1(j)) both from color image, and the gradient magnitude of depth at the direction of color gradient. We histogram the depth covered by two windows at the both sides of edge-tangent such as w_1 and w_2 in Fig. 1(e) and calculate the gradient magnitude of depth by χ^2 distance [21] between this two histograms. Given several triples of sample images, including color image, groundtruth of depth map and interpolated LR depth map, the procedures of classifier design are described as follows.

(a) The features are extracted from sample images, including the magnitude of gradient f_{cg} (Fig. 1(f)) and texture en-

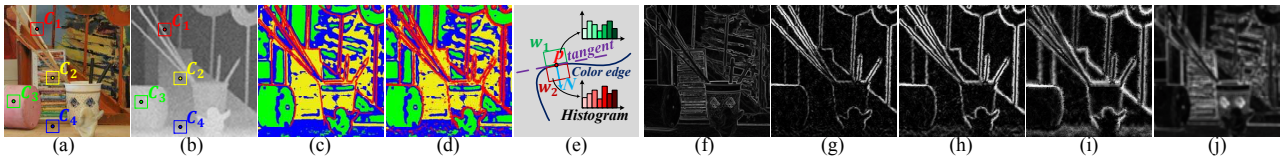

Fig. 1. The pixels are classified into four categories as shown in (a),(b),(c),(d), where (a) is color image, (b) is interpolated depth map, (c) is the groundtruth of pixel labels, and (d) is the classification result using interpolated depth map at the scale $\times 8$. The schematic diagram about the calculation fo depth gradient magnitude is shown in (e), where N is the gradient direction at pixels p. The feature maps $f_{cg}, f_{s1}, f_{s2}, f_{s3}, f_{ct}$ are shown in column (f)-(j) respectively.

Table 1. Quantitative comparison on synthetic datasets. The best result for each dataset and upscaling factor is highlighted.

	Art				*Books*				*Moebius*				*Reindeer*				*Laundry*				*Dolls*				**Average**			
	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$
MRF [1]	1.34	1.64	2.45	4.01	0.98	1.14	1.61	2.39	0.99	1.17	1.75	2.48	1.09	1.26	1.86	2.84	1.07	1.27	1.77	2.79	0.99	1.18	1.59	2.27	1.08	1.28	1.84	2.80
JBFap [9]	1.06	1.49	2.29	3.78	0.78	1.02	1.39	2.03	0.80	1.05	1.45	2.15	0.85	1.07	1.46	2.42	0.89	1.23	1.69	2.49	0.81	1.08	1.48	2.20	0.87	1.16	1.63	2.51
JBFcv [7]	0.98	1.40	2.23	4.52	0.78	1.09	1.79	3.02	0.76	1.08	1.78	3.00	0.78	1.09	1.83	3.37	0.80	1.15	1.88	3.46	0.79	1.11	1.77	2.91	0.82	1.15	1.88	3.38
TGV [4]	0.79	1.46	3.37	5.83	0.45	0.72	1.35	2.26	0.49	0.96	1.97	3.17	0.54	0.88	1.89	3.19	0.52	0.98	1.96	3.47	0.65	1.18	2.17	3.13	0.57	1.03	2.12	3.51
MRFop [2]	0.68	1.11	1.86	3.34	0.58	0.81	1.17	1.76	0.56	0.81	1.20	1.90	0.58	0.83	1.27	2.08	0.56	0.81	1.21	2.15	0.65	0.90	1.25	1.81	0.60	0.88	1.33	2.17
ARap [12]	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	2.06	0.55	0.83	1.32	2.18
WLS [19]	1.34	1.90	2.95	4.63	1.25	1.70	2.39	3.29	1.34	1.92	2.66	3.56	1.47	2.05	2.82	4.09	1.11	1.55	2.24	3.49	1.34	1.85	2.55	3.50	1.31	1.83	2.60	3.76
FGI [20]	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	0.65	0.92	1.41	2.31
RCG [14]	0.71	1.06	1.72	3.13	0.57	0.78	1.13	1.68	0.55	0.76	1.15	1.71	0.57	0.80	**1.14**	**1.88**	0.54	0.77	1.12	**1.98**	0.64	0.87	1.21	**1.73**	0.60	0.84	1.25	2.02
NLTGV [6]	0.72	1.12	2.00	3.71	0.51	0.73	1.20	1.95	0.48	0.75	1.15	1.90	0.55	0.79	1.29	2.27	0.52	0.80	1.31	2.46	0.60	0.89	1.25	1.89	0.56	0.85	1.37	2.36
Ours	**0.68**	**0.97**	**1.60**	**3.00**	**0.42**	**0.62**	**1.00**	**1.54**	**0.42**	**0.62**	**1.00**	**1.60**	0.50	**0.72**	1.16	2.00	**0.47**	**0.74**	**1.12**	2.13	**0.58**	**0.82**	**1.19**	1.77	**0.51**	**0.75**	**1.18**	**2.01**

ergy f_{ct} (Fig. 1(j)) from HR color image, and the gradient magnitude f_{dg} from the groundtruth of depth map. In addition, the multi-scale gradient magnitude of interpolated depth map is calculated with three window sizes, corresponding to three feature maps f_{s1}, f_{s2}, f_{s3} (Fig. 1(g),(h),(i)).

(b) We manually assign thresholds $t_{cg}, (t^l_{dg}, t^u_{dg}), (t^l_{ct}, t^u_{ct})$ respectively on the maps f_{cg}, f_{dg}, f_{ct}, where $t^l_{dg} < t^u_{dg}$ and $t^l_{ct} < t^u_{ct}$. The labels of pixels in sample images (Fig. 1(c)) are orderly determined as Eq. (5), where l represents the label of pixel. Namely, the pixels that have large f_{cg} and also large f_{dg} are labelled as the C_1 class, then the pixels whose texture energy f_{ct} is large constitute the second class C_2. Next, the pixels whose f_{dg} and f_{ct} are very small constitute the class C_3. The other pixels are classified as the fourth class C_4. The concrete thresholds are given in the Sect. 3.

$$l = \begin{cases} C_1 & f_{cg} > t^u_{cg} \ \& \ f_{dg} > t^u_{dg} \\ C_2 & f_{ct} > t^u_{ct} \ \& \ f_{cg} \le t^u_{cg} \ \& \ f_{dg} \le t^u_{dg} \\ C_3 & f_{dg} < t^l_{dg} \ \& \ f_{ct} < t^l_{ct} \\ C_4 & otherwise \end{cases}, \tag{5}$$

(c) Then, same number samples are randomly sampled from each category to train a random forest [22] based classifier by using the labels l and the features of $f_{cg}, f_{s1}, f_{s2}, f_{s3}$ and f_{ct}.

3. EXPERIMENTS

In this section, we first evaluate the classification scheme and depth SR method on the Middlebury datasets [23]. Twenty triples of synthetic images are selected to experiment. The LR depth map are generated by down-sampling the groundtruth of depth maps with four factors ($\times 2, \times 4, \times 8, \times 16$), and also some conditional Gaussian noises are added to the maps. Then our depth SR method is evaluated on the real sensor data [4] too.

Classifier Evaluation . Four classifiers corresponding to above four down-sampling factors are respectively trained to evaluate our classification scheme. For each classifier, eighty thousand samples are randomly sampled from six triples of sample images and each category has the same number of samples. All the features are normalized into interval $[0, 1]$ and the thresholds $t_{cg}, t^l_{dg}, t^u_{dg}, t^l_{ct}, t^u_{ct}$ for determining the labels of samples are set as 0.1/0.02/0.045/0.05/0.22. The random forest is composed of 150 trees and the maximum depth of each tree is 25. After training of classifiers, the out-of-bag errors of the four classifiers are 0.027/0.051/0.082/0.12. This shows the features make pixels in depth map have good separability. The classifiers are tested on the all twenty triples of images, and the classification error rates of the four classifiers are 0.020/0.033/0.050/0.072. Overall, the error rates are very low, but the classifier for an upper down-sampling factor obtains a larger error rate, since the depth boundaries in the interpolated depth map become more unrecognizable with the increasing of down-sampling factor. From the visual of classification results in Fig. 1(d) , we can clearly see that some pixels are erroneously classified as the class C_1. However, incorrect labels of some sporadic pixels in the C_2, C_3, C_4 regions have relatively little influence on the depth SR result than to label the pixel in class C_1 as wrong labels.

Depth SR Evaluation . Firstly, the depth SR method is evaluated on the synthetic data in terms of quantitative and qualitative comparison. The labels of pixels in depth map has been determined by using the trained classifier. The half size of support window is set $r = 2$. The parameters $\rho, \kappa, \sigma_c, \sigma_p$ for different category pixels are assigned as C_1:/2.8/0.0/1.5/2.0/, C_2:/2.8/60.0/16.0/20.0/, C_3:/2.0/20.0/2.5/2.0/, C_4:/2.8/40.0/4.5/2.0/ and unchanged on all up-sampling factors. If pixels have initial depth val-

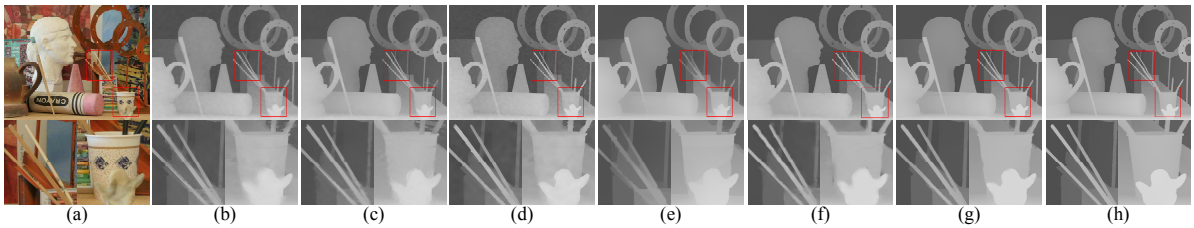

Fig. 2. Visual comparison of $\times 8$ up-sampling on the Art dataset [23], the column (a) show color image, column (b)-(g) respectively show the results by (b) MRF [1], (c) JBFap[9], (d) JBFcv [7], (e) TGV [4], (f) ARap [12] and (g) Our method. The groundtruth of depth map is shown in the column (h).

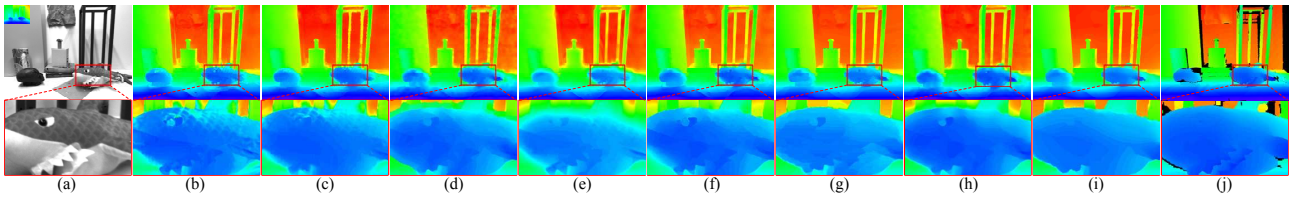

Fig. 3. Visual quality of depth SR results on the $Shark$ datasets from [4]. In column (a) the intensity image together with the LR TOF image are shown, the column (b) - (i) respectively show the results by (b) JBF [8], (c) JBFap [9], (d) JBFcv [7], (e) Guided [10], (f) MRF [1], (g) TGV [4], (h) ARap [12] and (i) our method. The groundtruth is shown in the column (j).

ues, the parameter λ is assigned as 0.55/1.05/1.6/3.0 for $\times 2, \times 4, \times 8, \times 16$ up-sampling factors respectively, otherwise set $\lambda = 0$. The accuracy of depth SR is measured by mean absolute error (MAE). Our method is compared with ten state-of-the-art methods. The numerical results for the experiments in terms of the MAE are shown in Tab. 1. It is clear that the results of our method have a great improvement over the filtering based methods [7, 9, 19, 20], MRF based methods [1][2] and TGV method [4]. Even when compared with the most recent proposed method in [12, 14] based on adaptive weights, our method has smaller MAE in almost all the cases. In particular, our method shows significant improvement for the high up-sampling factors of $\times 8$ and $\times 16$, when the associated color image contains rich textures similar to the depth edges, such as $Books$, $Moebius$ and $Dolls$. We note that the MAE of our method is decreased more than 10% for this three images at high up-sampling factors, compared to the method [12]. The obtained error of latest method [14] is smaller in a few cases, but our method gets the smallest average error in all cases. The proposed method obtains promising quality of depth map thanks to the novel classified weights which are set by considering the local structure of pixels. Compared to NLTGV method [6] with fixed weights, the accuracy of depth SR has been improved significantly.

A visual comparisons for the different methods are given in Fig. 2. The figure shows that the results of MRF method [1] and filtering methods [7, 9, 10] suffer from texture copy artifacts in the flat regions containing complexity color textures, e.g., the brush pot of Art. Our method and TGV method obtain smoother results in the flat regions thanks to the high-order regularization term, but the TGV method is prone to over-smooth out more useful signal components, e.g., the sticks of Art. The method [12] obtains low MAE similar to our results as we both use adaptive weights in depth SR model. However, upon careful comparisons in the highlighted regions, our method produces sharper depth boundaries.

Table 2. Quantitative comparison on real data from [4]. Error is calculated as MAE to the measured groundtruth in mm.

	MRF[1]	JBF[8]	JBFap[9]	JBFcv[7]	Guided[10]	TGV[4]	MRFop[2]	ARap[12]	FGI[20]	RCG[14]	Ours
$Books$	13.45	13.21	12.72	13.06	14.72	12.36	12.84	12.25	13.03	12.82	**11.36**
$Shark$	14.62	16.00	15.10	14.41	16.75	15.29	15.29	14.71	15.82	14.92	**13.38**
$Devil$	14.17	15.03	14.36	14.10	15.35	14.68	14.51	13.83	15.09	14.50	**13.23**

Finally, we evaluate our method on the TOF dataset from [4]. Quantitative comparison in Tab. 1 shows our method obtains the lowest MAE for all cases. The visual comparison of the $Shark$ up-sampling results in Fig. 3 shows our method obtains smoother result over the other methods and also performs much better in suppressing the texture copy artifacts, such as the shark model in the highlighted regions.

4. CONCLUSION

In this paper, the LR depth map has been up-sampled to a HR one by using a novel variational model based on NLTGV with classified weights. We find that the pixels of depth map have a good separability according to some appropriate local features around pixels. The experiment results show that the classified weights within the NLTGV can improve the accuracy of depth SR result and our depth SR method outperforms the state-of-the-art methods in accuracy.

References

[1] James Diebel and Sebastian Thrun, "An application of markov random fields to range sensing," in *Advances in Neural Information Processing Systems. (NIPS)*, 2005, vol. 5, pp. 291–298.

[2] Wei Liu, Shaoyong Jia, Penglin Li, Xiaogang Chen, Jie Yang, and Qiang Wu, "An mrf-based depth upsampling: upsample the depth map with its own property," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1708–1712, 2015.

[3] Jaesik Park, Hyeongwoo Kim, Yu-Wing Tai, Michael S Brown, and In So Kweon, "High-quality depth map upsampling and completion for rgb-d cameras," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5559–5572, 2014.

[4] David Ferstl, Christian Reinbacher, Rene Ranftl, Matthias Rüther, and Horst Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 993–1000.

[5] David Ferstl, Matthias Ruther, and Horst Bischof, "Variational depth superresolution using example-based edge representations," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 513–521.

[6] Hai-Tao Zhang, Kai Kang, and Zeng-Fu Wang, "Image guided depth map superresolution using non-local total generalized variation," in *Visual Communications and Image Processing (VCIP), 2016*. IEEE, 2016, pp. 1–4.

[7] Qingxiong Yang, Ruigang Yang, James Davis, and David Nistér, "Spatial-depth super resolution for range images," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[8] Johannes Kopf, Michael F Cohen, Dani Lischinski, and Matt Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 96, 2007.

[9] Derek Chan, Hylke Buisman, Christian Theobalt, and Sebastian Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. ECCV Workshop Multi-Camera Multi-Modal Sensor Fus. Algorithms Appl.*, 2008.

[10] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[11] Jiangbo Lu, Keyang Shi, Dongbo Min, Liang Lin, and Minh N Do, "Cross-based local multipoint filtering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 430–437.

[12] Jingyu Yang, Xinchen Ye, Kun Li, Chunping Hou, and Yao 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.

[13] Bumsub Ham, Dongbo Min, and Kwanghoon Sohn, "Depth superresolution by transduction," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1524–1535, 2015.

[14] Wei Liu, Xiaogang Chen, Jie Yang, and Qiang Wu, "Robust color guided depth map restoration," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 315–327, 2017.

[15] Xiaowei Deng and Xiaolin Wu, "Sparsity-based depth image restoration using surface priors and rgb-d correlations," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3881–3885.

[16] Tak-Wai Hui, Chen Change Loy, and Xiaoou Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2016, pp. 353–369.

[17] René Ranftl, Kristian Bredies, and Thomas Pock, "Non-local total generalized variation for optical flow estimation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 439–454. 2014.

[18] Hai-Tao Zhang, Jun Yu, and Zeng-Fu Wang, "Probability contour guided depth map inpainting and superresolution using non-local total generalized variation," *Multimedia Tools and Applications*, pp. 1–18, 2017.

[19] Dongbo Min, Sunghwan Choi, Jiangbo Lu, Bumsub Ham, Kwanghoon Sohn, and Minh N Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5638–5653, 2014.

[20] Yu Li, Dongbo Min, Minh N Do, and Jiangbo Lu, "Fast guided global interpolation for depth and motion," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2016, pp. 717–733.

[21] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.

[22] Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] Daniel Scharstein and Chris Pal, "Learning conditional random fields for stereo," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.