DEPTH SUPER-RESOLUTION USING JOINT ADAPTIVE WEIGHTED LEAST SQUARES AND PATCHING GRADIENT

Yuyuan Li[1], Jiarui Sun[1], Bingshu Wang[2], Yong Zhao[1,]*

[1]Shenzhen Graduate School, Peking University, Shenzhen, China
[2]Faculty of Science and Technology, University of Macau, Macau, China
1501213870@sz.pku.edu.cn, 1701120005@pku.edu.cn, yb77408@umac.mo, yongzhao@pkusz.edu.cn

ABSTRACT

This paper presents a flexible framework for the challenging task of color-guided depth upsampling. Some state-of-the-art approaches apply an aligned RGB image for depth recovery. Unfortunately, these kinds of methods may result in texture copying artifacts and edge blurring artifacts. To address these difficulties, we propose an adaptive weighted least squares framework of choosing different guidance weight for variant conditions flexibly. First of all, in the framework, we propose a joint adaptive color weighting scheme in which the depth maps and color images jointly choose a proper weight term for diverse cases. Then, a patch-based smoothness measuring approach called patching-gradient method (PGM) is proposed to distinguish the discontinuities and smooth areas. Our PGM is robust to dense noise and preserve weak edges effectively. Quantitative and qualitative experiments on noisy ToF-like datasets demonstrate our frameworks effectiveness on suppressing both texture copying artifacts and edge blurring artifacts.

Index Terms— Depth map super-resolution, ToF, WLS, Patching-gradient method, De-noising

1. INTRODUCTION

Depth map super-resolution (DSR) is a challenging task in the field of computer vision, such as gesture recognition, intelligent transportation system, 3D reconstruction, etc. A depth map captured by 3D Time-of-Flight (3D-ToF) camera is usually low-resolution and polluted by large amounts of noise. Based on these challenges, plenty of DSR methods have been exploited to improve the image resolution and get rid of the noise simultaneously.

The mainstream DSR methods can be classified as multiple maps merging DSR [1, 2], learning-based DSR [3-5, 25] and color-guided DSR [6-17]. Multiple maps merging DSR algorithms fuse multiple images of low resolution (LR) depth maps to restore HR depth maps. Most of the methods in this category [1, 2] assume that the object is static, which might fail to handle dynamic scenes. Learning-based DSR methods recover an LR depth map with a high resolution (HR) training datasets [3, 4] or with an image dictionary [5].

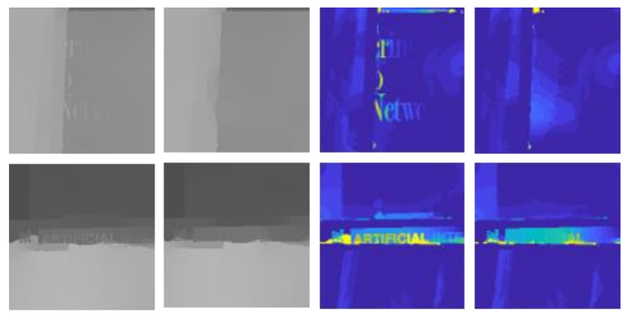

Fig. 1. The comparison between [17] and our method on $16\times$ upsampling of noisy Book dataset. (a) Result of [17]. (b) Result of our method. (c) Error map of [17]. (d) Error map of our method. It shows that our method suppresses texture copying artifacts significantly and preserve the edges well.

However, these methods have many limitations on large scale factor upsampling. Moreover, deep learning methods [25] have been gradually introduced in DSR however they have many limitations on the datasets.

Based on the assumption that the discontinuities in depth maps are consistent with the edges of color images, Color-guided depth upsampling methods exploit an aligned HR color image to guide the depth restoration. However, texture copying artifacts and edge blurring artifacts will occur when the depth discontinuities are inconsistent with the color edges. Plenty of local filtering approaches [7-12] and global optimization models [6, 13-16, 17] have been proposed to address these artifacts. Local filtering methods usually design a filter to convolve an image such as joint bilateral filter (JBF) [7-11] and joint geodesic filter (JGF) [12] in which a geodesic distance, instead of the Euclidean distance, was computed to evaluate the dissimilarities between two pixels. Global optimization methods usually construct an optimization function to obtain optimized results iteratively. Park et al. [6] presented a complex weighted least squares (WLS) optimization with a non-local means regularization. Ferstl et al. [14] proposed an anisotropic diffusion tensor, obtained from a HR intensity image, to guide the DSR and combine it with the MRF model. Yang et al. [15, 16] presented an autoregressive model (AR) which evaluated an AR minimization of prediction errors. Inspired by the AR and a mathematic penalty term [18], Liu et al. [17] proposed a robust WLS-based model computing the depth weight iteratively. These methods achieved good performance for suppressing one of the two artifacts but not both.

To address the two artifacts in color-guided DSR above and de-noise at the same time, we present an adaptive WLS model and iteratively evaluate the weight term. The main contributions include two parts. Firstly, a novel selective color-guided weighting scheme by adopting different weight in diverse cases is proposed, which is proven to be robust to noise. The depth map and color image are constrained mutually and then jointly to choose guidance weights for the proposed framework. Secondly, we propose a patch-based smoothness measuring approach called patching-gradient method (PGM) which can evaluate the smoothness of both color images and depth maps with its robustness to noise and good sensitiveness to weak edges. As shown in Fig. 1, our framework is effective for removing texture copying artifacts, especially for the artifacts close to edges which are hard to remove, and simultaneously preserving edges prominently. Experimental results on several Middlebury datasets validate that our proposed framework outperforms the existing state-of-the-art methods.

2. THE PROPOSED METHOD

Color images can be exploited to guide depth super-resolution mainly due to two points: the color image is high resolution and all of the edges in a depth map can be found in the corresponding color image at the same position [19]. However, a color image contains structural boundaries and internal texture while the depth map is sparse, structural and only describes the boundaries of the objects. Actually, only when the depth maps and color images are consistent with each other, the color guidance is meaningful to the recovery.

2.1. Joint adaptive color weighting scheme

In a global optimization, the guidance of color image is usually controlled by a color weight term. The ideal condition is that for diverse cases, the color weight have effect on the depth restoration to different extent. Therefore, according to the consistency between the depth edges and color edges, we classify the pixels into three categories and propose a selective color weighting scheme for the three cases: (1) When a pixel locates in smooth areas of depth map but in edges of color images, which always causes texture copying artifacts, we set the weight as constant 1 to avoid the impact of color edges on depth smooth regions. (2) When a pixel locates in smooth areas of color images but in edges of depth maps, which leads to edge blurring artifacts, we enhance the weak edges of color images to provide a stronger guidance. (3) When the edges in depth map are consistent with those in color image, we only need to keep the original guidance of color image. We adopt a piecewise function to formulate this selective color weight. i and j are the pixel indexes. Pixel j is the neighbor of pixel i. The color weight term $w_{C,ij}$ is defined as follows:

$$w_{C,ij} = \begin{cases} 1 & (\varepsilon_{C,j} \le T_{C1}) \cap (\varepsilon_{D,j} \ge T_{D1}) \\ w_C 1 & (\varepsilon_{C,j} > T_{C2}) \cap (\varepsilon_{D,j} < T_{D2}) \\ w_C 2 & otherwise \end{cases} \tag{1}$$

where ε_j is a parameter measuring the smooth degree of a small patch centered at pixel j. We will give ε a further discussion in Sec. 2.3. C and D denote color image and depth map respectively. T is the threshold which distinguishes discontinuities and smooth regions. When $\varepsilon_j \le T_j$, the pixel j locates at an edge. When $\varepsilon_j \ge T_j$, pixel j locates in a smooth region. The expressions $w_C 1$ and $w_C 2$ possess the same form. For $w_C 1$, we preprocess image with edge enhancement using guided filter [20]. $w_C 1$ and $w_C 2$ can be generally formulated as w_C:

$$w_C = \exp\left(-\frac{|i-j|^2}{2\sigma_S^2}\right)\exp\left(-\frac{\sum_{k\in\{R,G,B\}}\left|I_i^k - I_j^k\right|^2}{3\times 2\sigma_C^2}\right) \tag{2}$$

where σ_C and σ_S are the constant parameters corresponding to the space domain and range domain. Eq. (2) is a common bilateral filter.

2.2. Adaptive weighted WLS framework

The WLS model is widely used in DSR. But for the original WLS, it tends to suffer from the two kinds of artifacts above. To address these problems, an adaptive WLS framework with a novel weighting scheme is presented. The framework can be formulated as follows:

$$D^{n+1} = \underset{D}{\arg\min}\{\sum_{i=\Omega_0}(D_i^{n+1} - D_i^0)^2 + \beta\sum_{i=\Omega}\sum_{j=\omega_k(i)} W_{ij}(D_i^{n+1} - D_j^n)^2\} \tag{3}$$

where D^0 is the bicubic interpolation of LR depth map. The β denotes a coefficient that balances the data term and the smoothness term, Ω and Ω_0 are the coordinate space of D and D^0, respectively. The $\omega_k(i)$ is a window centered at pixel i. In the optimization, the output D^n in the *nth* iteration is used as the input of the $(n+1)th$ iteration.

The framework consists of a data term and a smoothness term. For the weight W_{ij} in the smoothness term, a novel selective weighting scheme which contains a color weight term and a depth weight term is designed to reduce the two kinds of artifacts and de-noise. The two components contribute different effect to the restoration. We define W_{ij} as:

$$W_{ij} = w_{C,ij}^{\ n} w_{D,ij}^{\ n} \tag{4}$$

Fig. 2. Examples of edge-detection maps on $8\times$interpolation of noisy LR depth map. (a) the interpolation of ToF-like degradation. (b) Sobel method. (c) Liu et al. [17]. ($r=9$. $r_\lambda=3$. θ=0.92). (d) Liu et al. [17]. ($r=9$. $r_\lambda=3$. θ=0.88). (e) Our PGM. ($r_\varepsilon=3$. T_D=0.057). (f) Our PGM. ($r_\varepsilon=3$. T_D=0.065). Two regions are highlighted by rectangles and enlarged in the second row.

For the color weight term $w_{C,ij}^{\ n}$, we utilize the joint adaptive color weight in Eq. (1). It can address the two artifacts effectively.

To depth weight term $w_{D,ij}$. Inspired by Liu et al. [17], we choose a newly updated depth map in the last iteration to evaluate $w_{D,ij}$. We define $w_{D,ij}$ as follows:

$$w_{D,ij}=\exp\left(-\frac{\left|D_i^n-D_j^n\right|^2}{2\sigma_D^2}\right) \tag{5}$$

where σ_D is the constant parameter corresponding to the depth domain on Gaussian function. $w_{D,ij}$ is computed from D^n instead of D^0 since the edges in D^0 are blurred and severely polluted by noise.

We look for the derivative of Eq. (3) with respect to D and then let the derivative equal to 0. The solution for Eq. (3) is formulated for a fixed point:

$$D^{n+1}=\frac{D_i^0+2\beta\sum_{j=\omega_k(i)}W_{ij}D_j^n}{1+2\beta\sum_{j=\omega_k(i)}W_{ij}},\qquad i\in\Omega \tag{6}$$

The newly updated depth map is applied to evaluate three variables: $w_{D,ij}^n$, ε_D^n and D^n .

Generally, in our method the depth map and color image constrain each other and jointly determinate the three conditions in which we apply diverse color weights. The result will approach to a higher accuracy with the increasing number of iterations. The proposed approach achieves good performance on two aspects: enhancing the resolution and getting rid of noise.

2.3. Patch-gradient method

To measure the local smoothness of image and compute the parameter ε in Eq. (1), we propose a general patch-based smoothness measuring method: patch-gradient method (PGM). The patch-gradient is defined as:

$$\varepsilon_j^n=\sqrt{\left(\sum_{p\in\omega_\varepsilon(j)}\partial_x S_p\right)^2+\left(\sum_{p\in\omega_\varepsilon(j)}\partial_y S_p\right)^2}\qquad j\in\omega_k(i) \tag{7}$$

where $\omega_\varepsilon(j)$ is a window centered at the pixel j with the radius r_ε . S_p represents the pixel in the image S . $\partial_x S_p$ and $\partial_y S_p$ are the partial derivative from x and y dimension respectively. We use a sliding window ω_ε to calculate the sum of partial derivatives in two dimensions respectively.

Along with the upsampling of image, noise is upsampled as well, which severely interrupts the edges in depth maps and increases difficulties of measuring smooth degree. Compared with common pixel-based smoothness measuring algorithms, PGM mainly possesses two advantages: 1) PGM is sensitive enough to weak edges and blurring edges. This is because that PGM combines all the differences within a patch and can capture the slow change even if an edge is weak in a window span; 2) The patch-based statistic can reduce the noise interruption and is robust to a variety of noisy situations because of the high stability of the gradients statistics within a patch. Moreover, our PGM has a stronger anti-noise ability than other patch-based methods such as [21]. Figure 2. Illustrates that our method performs well on de-noising and edge-preserving.

3. EXPERIMENTS

To validate the proposed approach, the proposed approach and the compared methods are tested on six ToF-like degraded depth datasets provided by Yang et al. [16]. Gaussian noise is added into the original Middlebury datasets [22] with a variance of 25. Mean-absolute-error (MAE) is employed as the performance metric and three scale factors (2x, 4x, 8x) are tested in the paper. Parameters of our approach in the experiments are set as follows: β=0.95 , $r_{\varepsilon,C}$=1 , $r_{\varepsilon,D}$=3 , T_{D1}=0.002 , T_{D2}=0.011 , T_{C1}=T_{C2}=0.05 , $r_k=9$ for most situations. The parameters are finetuned for a few cases. Qualitative and quantitative comparisons with several state-of-the-art methods [6,9,12, 14,16,17,20,23,24] are displayed in Fig. 3 and Table 1.

For the edge blurring artifacts, our method and the RGDR model [17] can preserve edges better than the other

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Visual comparison of $8\times$ upsampling on ToF-like Art dataset (with Gaussian noise). (a) Color image. (b) Ground truth. (c) NLM-MRF [6]. (d) ATGV [14]. (e) AR [16]. (f) RGDR [17]. (g) Our methods. Two regions are highlighted by rectangles and enlarged in the second row. Our method performs well on both smooth areas and depth edges.

Table 1. Quantitative comparison on noisy ToF-like datasets in terms of MAE. The best evaluation results are in bold.

MAE	Art			Book			Dolls			Laundry			Moebius			Reindeer			Average
	2x	4x	8x	2x	4x	8x	2x	4x	8x	2x	4x	8x	2x	4x	8x	2x	4x	8x	
JBF[9]	1.59	2.06	3.11	0.82	1.24	2.04	0.81	1.20	1.98	0.94	1.38	2.25	0.89	1.28	2.05	0.95	1.36	2.24	1.57
JGF[12]	1.33	1.81	2.90	0.79	1.24	2.05	0.80	1.23	2.01	0.88	1.36	2.23	0.82	1.25	2.03	0.91	1.37	2.26	1.52
NLMMRF[6]	1.66	2.47	3.44	1.19	1.47	2.06	1.19	1.56	2.15	1.34	1.73	2.41	1.20	1.50	2.13	1.26	1.65	2.46	1.83
Guided[20]	1.91	2.23	3.08	0.84	1.12	1.73	0.84	1.11	1.69	1.01	1.31	2.00	0.92	1.19	1.78	1.06	1.32	1.98	1.51
JIDCA[23]	1.69	2.98	3.68	1.53	2.71	3.04	1.54	2.71	2.94	1.45	2.72	3.16	1.55	2.72	2.94	1.65	2.80	3.13	2.50
AR[16]	1.17	1.70	2.93	0.98	1.22	1.74	0.97	1.21	1.71	1.00	1.31	1.97	0.95	1.20	1.79	1.07	1.30	2.03	1.46
WLS[24]	1.25	1.73	2.59	0.74	1.10	1.45	0.85	1.21	1.68	0.83	1.17	1.65	0.80	1.18	1.67	0.84	1.15	1.58	1.30
ATGV[14]	0.80	1.21	2.01	0.61	0.88	1.21	0.66	0.96	1.38	0.61	0.87	1.36	0.57	0.77	1.23	0.61	0.85	1.30	0.99
RGDR[17]	0.71	1.06	1.72	0.57	0.78	1.13	0.64	0.87	1.21	0.54	0.77	**1.12**	0.55	0.76	1.15	0.57	0.80	1.14	0.89
Ours	**0.57**	**0.92**	**1.55**	**0.47**	**0.69**	**1.05**	**0.56**	**0.81**	**1.17**	**0.46**	**0.74**	1.19	**0.45**	**0.69**	**1.12**	**0.48**	**0.74**	**1.10**	**0.82**

models. However, the RGDR model still suffers from the texture copying artifacts as shown in Fig. 3f. This is because that the color weight in this model has too much impact on the homogeneous areas in depth maps. In contrast, as shown in Fig. 3e, the AR [16] can achieve excellent results on suppressing texture copying artifacts but the depth edges are still blurred. On the contrary, Fig. 3g illustrates that our approach obtains outstanding performance on resolving both two problems. Especially when the texture copying artifacts occur nearby the edges, they are hard to be removed without blurring edges. From Fig. 3f and Fig. 3g, we can observe our proposed method effectively gets rid of the texture copying artifacts around the edges as well as preserve clear edges.

Table 1 displays the results of nine state-of-the-art methods and our method in terms of MAE. For the selected evaluation metric MAE, the lower is the better. The results of the compared methods in Table 1 are quoted from Liu et al. [17]. Among all of the compared methods, the RGDR model performs best performance with the lowest MAE values on all datasets. On $8\times$ Laundry dataset, the MAE value of RGDR is 0.07 lower than that of the proposed method, which shows RGDR's advantage. But our method performs the lowest MAE values with good performance on the rest of the datasets except the $8\times$ Laundry. Notably, it can be seen from the table that the minimum average MAE of the compared methods is 0.89 belonging to RGDR. However, the proposed method obtains the average MAE with 0.82, 0.07 less than the previous minimum value. Therefore, results on various datasets demonstrate that our method outperforms the state-of-the-art methods.

4. CONCLUSION

In this paper, we have proposed an adaptive weighted least squares framework which suppress the edge blurring and texture copying artifacts in a better way. There are two innovations. Firstly, we propose an adaptive weighting scheme in which the depth maps and the color images constrain each other and jointly choose a suitable guidance weight for variant conditions flexibly. Moreover, a novel patching-based smoothness measuring model called patching-gradient method (PGM) is proposed to evaluate the smooth degree of an image. In the experiment, the average MAE of our method is 0.07 lower than RGDR on noisy ToF-like datasets. It is worth mentioning that our method has successfully suppressed the artifacts nearby the edges which is hard to remove. Both the qualitative and quantitative comparisons results demonstrate our method's effectiveness and robustness.

5. ACKNOWLEDGEMENTS

This research is supported by National Sci-Tech Support Plan: 2015BAK01B04 and Shenzhen Technology Program: JCYJ20160506172651253.

6. REFERENCES

[1] S. Schuon, C. Theobalt, J. Davis, S. Thrun, "Lidarboost: Depth superresolution for tof 3d shape scanning," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 343-350, June, 2009.

[2] A. N. Rajagopalan, A. Bhavsar, F. Wallhoff, G. Rigoll, "Resolution enhancement of pmd range maps," *In Joint Pattern Recognition Symposium*, Springer, Berlin, Heidelberg, pp. 304-313, June, 2008.

[3] O. Mac Aodha, N. D. Campbell, A. Nair, G. J. Brostow, "Patch based synthesis for single depth image super-resolution," *In European Conference on Computer Vision*, Springer, Berlin, Heidelberg, pp. 71-84, October, 2012.

[4] M. Hornácek, C. Rhemann, M. Gelautz, C. Rother, "Depth super resolution by rigid body self-similarity in 3d," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1123-1130, 2013.

[5] J. Xie, R. S. Feris, M. T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, Vol. 25, No. 1, pp. 428-438, 2016.

[6] D. Chan, H. Buisman, C. Theobalt, S. Thrun, "A noise aware filter for real-time depth upsampling," *In ECCV Workshop on Multicamera and Multimodal Sensor Fusion Algorithms and Applications*, 2008.

[7] J. Park, H. Kim, Y. W. Tai, M. S. Brown, I. Kweon, "High quality depth map upsampling for 3d-tof cameras," *In Proceedings of the IEEE International Conference on Computer Vision,* pp. 1623-1630, November, 2011.

[8] E. Eisemann, F. Durand, "Flash photography enhancement via intrinsic relighting," *In ACM Transactions on Graphics*, Vol. 23, No. 3, pp. 673-678, 2004.

[9] J. Kopf, M. F. Cohen, D. Lischinski, M. Uyttendaele, "Joint bilateral upsampling," *In ACM Transactions on Graphics,* Vol. 26, No. 3, pp. 96, August, 2007.

[10] J. Zhu, L. Wang, R. Yang, J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-8, 2008.

[11] D. Chan, H. Buisman, C. Theobalt, S. Thrun, "A noise aware filter for real-time depth upsampling," *In ECCV Workshop on Multicamera and Multimodal Sensor Fusion Algorithms and Applications*, 2008.

[12] M.-Y. Liu, O. Tuzel, Y. Taguchi, "Joint geodesic upsampling of depth images," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 169-176, 2013.

[13] J. Diebel, S. Thrun, "An application of markov random fields to range sensing," *In Advances in Neural Information Processing Systems*, pp. 291-298, 2006.

[14] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 993-1000, 2013.

[15] J. Yang, X. Ye, K. Li, C. Hou, "Depth recovery using an adaptive color-guided auto-regressive model," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 158–171, October. 2012.

[16] J. Yang, X. Ye, K. Li, C. Hou, Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Transactions on Image Processing*, Vol. 23*, No.* 8, pp. 3443-3458, 2014.

[17] W. Liu, X. Chen, J. Yang, Q. Wu, "Robust color guided depth map restoration," *IEEE Transactions on Image Processing*, Vol. 26*,* No. 1, pp. 315-327, 2017.

[18] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, J. Weickert, "Generalised nonlocal image smoothing," *International Journal of Computer Vision*, Vol. 90, No. 1, pp. 62-87, 2010.

[19] C. Jung, S. Yu, J. Kim, "Intensity-guided edge-preserving depth upsampling through weighted L0 gradient minimization," *Journal of Visual Communication and Image Representation*, Vol. 42, pp.132-144, 2017.

[20] K. He, J. Sun, X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 6, pp. 1397-1409, 2013.

[21] W. Liu, X. Chen, J. Yang, Q. Wu, "Variable Bandwidth Weighting for Texture Copy Artifact Suppression in Guided Depth Upsampling," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 27, No. 10, pp. 2072-2085, 2017.

[22] Middlebury Datasets:http://vision.middlebury.edu/stereo/data/

[23] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 1545–1552, 2013.

[24] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Transactions on Image Processing*, Vol. *23*, No.12, pp. 5638-5653, 2014.

[25] W. Zhou, X. Li, D. Reynolds, "Guided deep network for depth map super-resolution: How much can color help?," *IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 1467-1461, 2017.