UNIVERSITY OF BATH

Citation for published version:
Qiao, Y, Jiao, L & Hou, B 2018, 'High-quality depth up-sampling based on multi-scale SLIC', *Electronics Letters*, vol. 54, no. 8, 8345115, pp. 494-496. https://doi.org/10.1049/el.2017.4393

DOI:
10.1049/el.2017.4393

Publication date:
2018

Document Version
Peer reviewed version

High-quality depth up-sampling based on multi-scale SLIC

Yiguo Qiao, Licheng Jiao and Biao Hou

A multi-scale SLIC based depth up-sampling method is proposed in order to obtain high-quality depth maps, especially in the case of high up-sampling rate. The proposed method is implemented hierarchically, where the high-resolution image is segmented from coarse to fine by using multi-scale simple linear iterative clustering (SLIC) superpixels. A depth guided discriminant function is defined to distinguish the validity of the segmented superpixels, and only the valid ones will be interpolated each layer. The experimental results show that the proposed method solves the depth missing and the depth confusion problems largely.

Introduction: High-resolution (HR) depth map plays an important role in a variety of 3D applications such as human computer interaction, 3D reconstruction, 3DTV, FTV and so on. Both passive methods and active methods are studied for generating the HR depth maps. The passive methods acquire the depth by calculating it from numbers of multi-view images, however, occlusions might be produced. In the active methods, depth maps are obtained by using depth sensors with a time-of-flight technique, but the low-resolution (LR) of the captured depth map is a serious drawback. Thus for up-sampling the LR depth maps, both interpolation-based methods and optimization-based methods are studied. Interpolation-based methods, such as the classical JBU[1], are most likely cause depth confusions at the depth boundary regions. Optimization-based methods such as the MRF-based method[2], may lead over-smoothing problems. Moreover, if the up-sampling rate is relatively large, both of these two kinds of methods may result in depth missing artifacts. To overcome the problems presented above, the multi-scale SLIC based up-sampling method is proposed in this letter.

Depth up-sampling using multi-scale SLIC: Local feature is commonly used in the traditional up-sampling methods, however, pixels that have similar features may sometimes far away from each other. That is why artifacts like depth confusion or even depth missing arises in the up-sampled results. To avoid this from happening, a hierarchical model is proposed as Fig.1 shows. The inputs are the HR color image I and the to be interpolated depth map D_I. The intermediate depth map D_I is generated by mapping the LR depth map to the HR grid, in which pixels directly from the LR depth map are called seeds, the other blank ones are called targets. The color image is roughly segmented in the first layer for producing relatively large superpixels, and that makes the most similar but distant seeds available for interpolating the targets. Then a discriminant function is introduced to select the valid superpixels that can be interpolated in the current layer. With the purpose of obtaining more and more valid regions, the color image will be segmented thin and thin each layer based on a multi-scale SLIC.

SLIC is an image segmentation method by clustering similar pixels into a superpixel[3]. Firstly, the HR color image will be converted into a 5-D vector space $[l, a, b, x, y]$, where $[l, a, b]$ represents the pixel color intensity in CIELAB color space, and $[x, y]$ represents the pixel coordinate. Secondly, with a desired initial superpixels number k, the interval between two neighboring superpixel centers can be obtained as $L = \sqrt{N/k}$, where N denotes the total number of pixels. Thus the superpixel centers are distributed uniformly. Then, a search range of $2L * 2L$ is used to label the neighboring pixels to the above superpixel centers. The geodesic distance d_g is measured by Eq.(1) $\sim$ Eq.(3),

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \tag{1}$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{2}$$

$$d_g = \sqrt{(\frac{d_c}{m})^2 + (\frac{d_s}{L})^2} \tag{3}$$

where d_c, d_s and d_g denote the color distance, the spatial distance and the geodesic distance, respectively; pixel i and j denotes the superpixel center and its neighbors, respectively; m is a scale constant and is relevant to the image size. In this way, all pixels will be labeled to its nearest center point. Next, the center point vector of each superpixel will be updated by averaging the vectors of all pixels within the active superpixel. The above clustering and updating processes will be performed alternately, until the clustering result converges. At last, the clustered superpixels will be divided into 8-connected sub-superpixels. For segmenting the color image from coarse to thin, we set the superpixels numbers $\{k_1, k_2, \cdots, k_N\}$ of the multi-scale SLIC in ascending order.

A depth guided discriminant function as shown in Eq.(4) is used for identifying the valid ones among the segmented superpixels,

$$J_{\Psi}^{(n)} = \begin{cases} 1, & \max|D_n(i) - D_n(j)| < th_D \\ 0, & else. \end{cases}$$
$$s.t. \quad i, j \in \Psi \cap S \tag{4}$$

where $J_{\Psi}^{(n)}$ reflects whether a current superpixel Ψ in the nth layer is valid or not; S denotes the seeds set; th_D is a given depth threshold, which is closely related to the depth gaps between different objects in a scene. If $J_{\Psi}^{(n)} = 1$, it indicates that the maximum depth difference of the current region Ψ is not that large. In this case, the current region will be regarded as depth-continuous and valid; otherwise if $J_{\Psi}^{(n)} = 0$, we consider the current region Ψ is depth-discontinuous and invalid. In particular, a region that has no seed pixel will be skipped. With this depth guided discriminant function, only the target pixels in those valid regions will be interpolated in each layer. The invalid regions with large depth spans are most likely across the depth borders, and will be count as wrong-segmented. Therefore, those regions are omitted during the interpolation of each layer.

Target pixels in each valid region will be interpolated with all seeds in the same region. A region based joint bilateral filter (R-JBF) as shown in Eq.(5) is adopted[1],

$$D_{n+1}(t) = \frac{\sum\limits_{s \in \Psi \cap S} D_n(s) \cdot g_d(s,t) \cdot g_c(s,t)}{\sum\limits_{s \in \Psi \cap S} g_d(s,t) \cdot g_c(s,t)}$$
$$s.t. \quad J_{\Psi}^{(n)} = 1\ ,\ t \in \Psi \cap T \tag{5}$$

where T denotes the targets set; g_d and g_c are two gaussian kernels with respect to the spatial distance and the color difference, respectively. Eq.(6) and Eq.(7) present the details,

$$g_d(s,t) = exp(-\frac{(x_s - x_t)^2 + (y_s - y_t)^2}{2d_\sigma^2}) \tag{6}$$

$$g_c(s,t) = exp(-\frac{(I_s - I_t)^2}{2c_\sigma^2}) \tag{7}$$

where x and y represent the horizontal and vertical coordinates, respectively; d_σ and c_σ are two scale parameters. Target pixels will be updated after the interpolation of each layer.

The proposed R-JBF ensures that a target will be interpolated with the seeds satisfying both the two criteria as follows. The first one is that the seeds should be in the same superpixel with the target; another one is that the maximum depth difference of this superpixel should be small enough. Through this way, all targets will be interpolated with the most closest corresponding seeds, so that clear boundary will be produced even if under a fairly large up-sampling rate.

After the N layers loop, target pixels still can be found in the intermediate up-sampling result. The remaining targets will be interpolated with its neighboring non-zero pixels based on the JBF.

Experimental results: We test our method on 6 data sets from Middlebury. They are Art, $Laundry$, $Dolls$, $Book$, $Moebius$, and $Reindeer$, respectively. To obtain the LR depth maps, we down-sample the highest resolution depth map in the database with the down-sampling rates of 2, 4, 8, and 16, respectively.

Fig.2 and Fig.3 present the 8 times up-sampling results on $Laundry$ and $Reindeer$ data sets, respectively. In which, we also compare our method with the JBU method[1], the JBLM method[4], the MRF based method[2], the JGU method[5], the TGV method[6], the AR Model method[7] and the JTF method[8]. Comparisons show that the proposed method avoids the depth confusion artifacts and results in clear depth boundary.

We also evaluate the performance of the proposed method according to **BP** and **MSE**, respectively. **BP** denotes the bad pixel rate and **MSE** stands for the root-mean-square error. Bad pixels are those whose values in the up-sampled image deviate from the ground truth by more than one disparity. Table1 shows the quantitative evaluation result on $Dolls$ and the comparisons with the other 7 methods above. The stability of our method is confirmed by the superior results under different up-sampling rates.

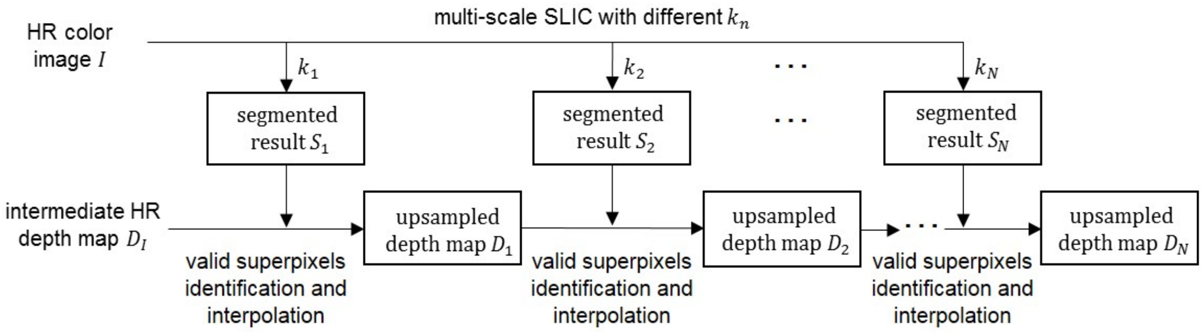

Fig. 1. Framework of the proposed method.

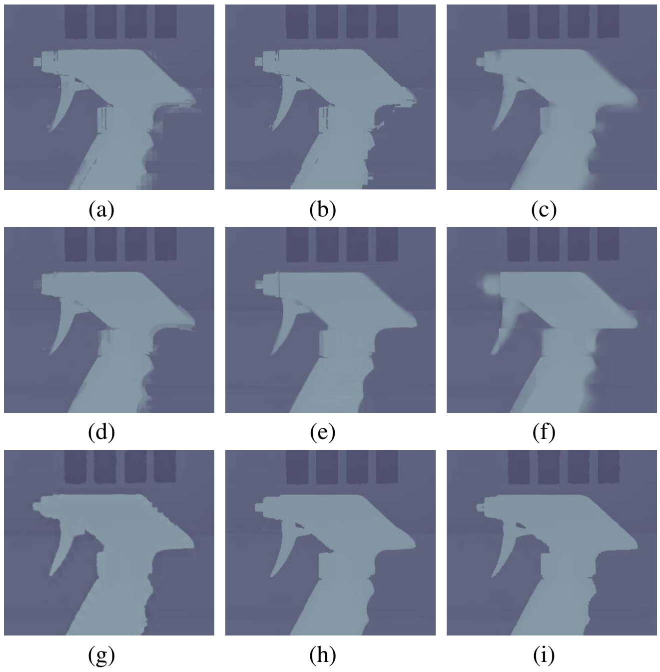

Fig. 2 *8 times up-sampling results on Laundry. Enlarged details with different methods of: (a) JBU, (b) JBLM, (c) MRF,(d) JGU, (e) TGV, (f) AR Model, (g)JTF, (h) the proposed method and (i) the ground truth.*

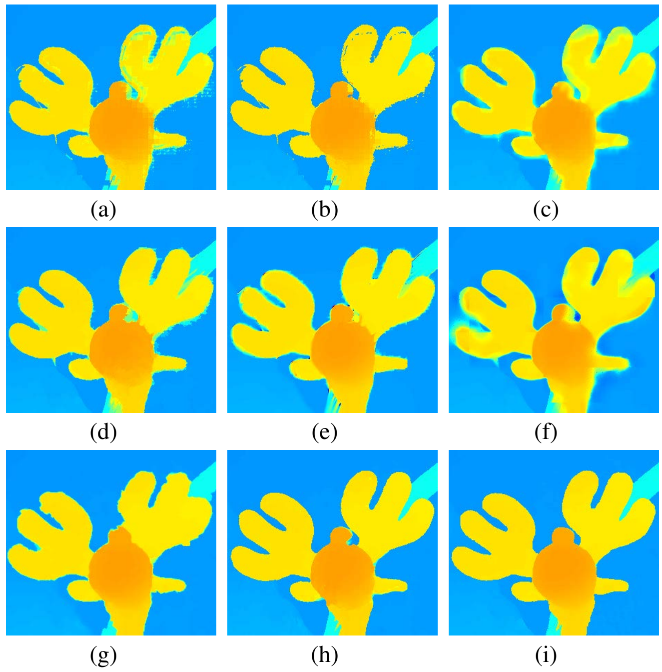

Fig. 3 *8 times up-sampling results on Reindeer. Enlarged details with different methods of: (a) JBU, (b) JBLM, (c) MRF,(d) JGU, (e) TGV, (f) AR Model, (g)JTF, (h) the proposed method and (i) the ground truth.*

Table 1: Quantitative evaluations of *Dolls* with up-sampling rates of 2, 4, 8 and 16. For each image, the best performance is shown in bold.

Methods	BP(%)				MSE			
	2	4	8	16	2	4	8	16
JBU	1.54	3.90	10.07	23.24	1.14	1.66	2.39	3.39
JBLM	1.24	3.23	8.97	20.07	1.21	1.79	2.55	3.60
MRF	2.85	4.11	9.35	22.65	1.11	**1.32**	**1.82**	2.90
JGU	2.29	4.29	9.02	21.12	1.29	1.71	2.37	3.11
TGV	1.52	3.94	6.65	15.23	**1.02**	1.55	1.96	**2.62**
AR	2.80	6.53	10.14	18.21	1.20	1.56	1.90	2.86
JTF	1.21	3.89	7.72	18.45	1.20	1.67	2.48	3.27
Ours	**0.86**	**1.86**	**5.89**	**15.11**	**1.02**	1.45	1.94	2.79

Conclusion: We present a multi-scale SLIC based depth up-sampling method in this letter. Based on a coarse-to-fine segmentation strategy, we interpolate the target pixels layer by layer. In each layer, we firstly use a depth threshold to judge whether a superpixel is reasonably segmented or not. Then only the targets in the valid ones will be interpolate by the proposed R-JBF. The proposed method effectively solves the depth confusion and the depth missing artifacts. Experimental results indicate that the proposed method performs well in both visual effect and quantitative evaluation.

Acknowledgment: This work was supported by the Major Research Plan of the National Natural Science Foundation of China (No.91438201).

Y. Qiao, L. Jiao and H. Biao (*Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation Xidian University, Xiąŕan, Shaanxi Province 710071, China*)

E-mail: qiaoyiguo123@126.com

References

1 Kopf, J., Cohen, M. F. and Lischinski, D. et al.:'Joint bilateral upsampling', *Acm Transactions on Graphics*, 2007, **26(3)**:pp. 1-5.

2 Kang, Y. S., Lee, S. B. and Ho, Y. S.:'Depth map upsampling using depth local features', *Electronics Letters*, 2014, **50(3)**, pp. 5255-5262.

3 Achanta, R., Shaji, A. and Smith, K. et al.:'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34(11)**, pp. 2274-2282.

4 Lee, S. B., Kwon, S. and Ho, Y. S.:'Discontinuity adaptive depth upsampling for 3D video acquisition', *Electronics Letters*, 2013, **49(25)**, pp. 1612-1614.

5 Liu, M. Y., Tuzel, O. and Taguchi, Y.:'Joint Geodesic Upsampling of Depth Images', *Computer Vision and Pattern Recognition*, 2013, pp. 169-176.

6 Ferstl, D., Reinbacher, C. and Ranftl, R. et al.:'Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation', *IEEE International Conference on Computer Vision*, 2014, pp. 993-1000.

7 Yang, J., Ye, X. and Li, K. et al.:'Color-Guided Depth Recovery From RGB-D Data Using an Adaptive Autoregressive Model', *IEEE Transactions on Image Processing*, 2014, **23(8)**, pp. 3443-3458.

8 Lo, K. H., Wang, Y. F. and Hua, K. L.:'Edge-Preserving Depth Map Upsampling by Joint Trilateral Filter', *IEEE Transactions on Cybernetics*, 2017, **99**,PP: 1-14.