MULTIRESOLUTION CONTOURLET TRANSFORM FUSION BASED DEPTH MAP SUPER RESOLUTION

Dan Xu[⋆] *Xiaopeng Fan*[⋆] *Debin Zhao*[⋆] *Wen Gao*[†⋆]

[⋆] Harbin Institute of Technology, School of Computer Science & Technology
[†]Peking University, School of Electrical Engineering & Computer Science

ABSTRACT

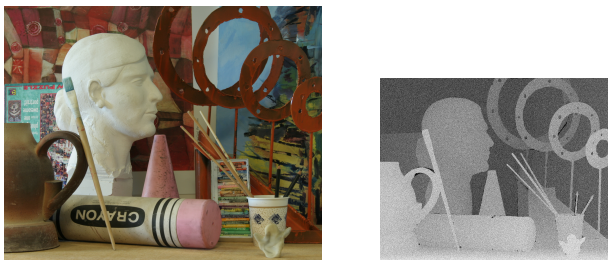
With the rapid advancement of 3D applications, depth map super resolution has been a serious problem to be solved. Many researches has proposed depth map super resolution methods focusing on spatial domain which can't produce clear edges in the super resolution results. However, different from color images, depth maps have clear edges along them with internal smoothness. In this paper, we propose a novel multiresolution contourlet transform fusion based depth map super resolution method to enhance the quality of depth maps with preserving more contour information. We first transform the depth maps of multiple views via contourlet transform (CT) into multiresolution coefficients. Then we fuse the coefficients of same resolution to get the fused coefficients. Finally the target depth map is upscaled utilizing the prevalent upscaling framework JBU or WMF. Depth map fusion in transform domain is first proposed to improve the quality of the target depth map. A CT based depth map fusion model can not only produce sharper edges, but alleviate the noise in the depth map. Experimental results demonstrate that our proposed method outperforms many state-of-the-art algorithms in both objectively and subjectively.

Index Terms— contourlet transform, fusion, depth map, super resolution

1. INTRODUCTION

With the rapid advancement of the 3D applications, such as 3D navigation, freeview TV, virtual reality, the quality of depth map has a significant effect on 3D applications. In general, methods of depth map acquisition can be divided into two categories: passive acquisition and active acquisition. Passive methods aim to acquire depth map from several color images in the process of stereo matching. On account of the poor matching performance in occluded or untextured region, the quality of acquired depth map may be degraded.

Depth maps acquired by active methods are generally sensed by depth sensor devices, such as Microsoft Kinect [1] and Time of Flight (ToF) cameras [2]. Generated depth maps are usually with low resolution, noise corruption and holes phenomenon. Fig.1(a) shows the high resolution color image and Fig.1(b) shows the corresponding low resolution(LR) depth maps with holes of the same view. However, in many 3D applications, depth map need to have same resolution with the color image. Therefore, depth map super-resolution (SR) has become a hot issue for researchers to explore.



(a) (b)

Fig. 1: (a) Color image obtained from the optical camera, and (b) the corresponding low resolution depth map with holes and with noise.

In order to enhance the quality of the depth maps, depth map SR methods are applied. Common depth map super-resolution methods can be divided into two categories. One is single depth map SR methods which upsample the depth value using the depth map itself or guidance of the corresponding color image. For example, filter based methods [3–7] aim to construct upsampling filters to enhance the depth map resolution with the guidance of the registered color image. Leveraging the HR color image and the given low resolution depth map, Kopf et al. [3] propose joint bilateral upsampling method (JBU) which combine a range filter and a spatial filter to produce very good full resolution results. Paper [5] proposes a weighted mode filter (WMF) by seeking a global mode on the histogram which uses the weight considering color similarity between reference and neighboring pixels of the color image to super-resolve the depth map. In addition to filter based methods, there are optimization based methods to upsample the depth map. A typical example is Markov Random Fields (MRF) [8–13] based depth map SR model. In [8], Diebel and Thrun first formulate the depth map SR as a multi-labeling optimization problem based on MRF model. [9] develops an extension of MRF based method by

proposing a novel data term to adaptively determine appropriate depth reference value of the target pixel.

The other category of depth map SR methods is multiple depth map SR methods [14–16] which fuse low resolution depth maps from a view point at different time or multiple views at the same time to get a high resolution depth map. Hahne et al. [14] utilize some LR depth maps of ToF sensors at different exposures to obtain a high resolution(HR) depth map. Choi et al. [15] propose a novel depth map SR framework by taking interview coherence into account. Most recently in [16], Lei et al. proposed a credibility based multiview depth maps fusion strategy, which considers the view synthesis quality and interview correlation.

For single depth map SR methods, they may produce texture copy artifacts when the color discontinuities and the depth discontinuities at the corresponding location are not consistent. To tackle the texture-transfer problem, researchers proposed multiple depth map SR methods. Considering the characteristics of clear edges with no texture of depth map, preserving more contour information is the key to the depth map super resolution. To the best of our knowledge, multiple depth map SR methods concentrate on the spatial domain, which only take local pixels into account rather than local edge structure. Inspired from the color image fusion in transform domain, we propose to fuse the LR depth maps via contourlet transform in transform domain on depth map SR for the first time. CT [17] can not only isolate the discontinuities of contours, but retain the smoothness along the contours, especially suitable for depth map. Simultaneously, multiple depth maps fusion in transform domain can mitigate the noise while retaining more high-frequency details.

In this paper, we propose a multiresolution contourlet transform fusion based depth map super resolution framework. We first transform LR depth maps of multiple views into multiresolution contourlet coefficients via contourlet transform. Then we fuse the CT coefficients of same resolution to get the fused coefficients. Finally we upscale the target depth map utilizing the prevalent upscaling framework JBU or WMF. Experimental results on benchmark depth map dataset demonstrate that our method outperforms many state-of-the-art methods in both objective and subjective performance.

The rest of this paper is organized as follows. We will give a revisit of contourlet transform in Section 2. The proposed CT based depth map super-resolution algorithm is presented in Sections 3. Experimental results are presented in Section 4 and Section 5 contains the conclusions.

2. A REVISIT TO CONTOURLET TRANSFORM

This section will give a revisit of contourlet transform. CT consists of two parts as Fig.2 described: Laplacian pyramid structure and directional filter banks (DFB). Laplacian pyramid structure filters the image into low frequency subbands and high-frequency subbands, and directional filter banks transform the 2-D frequency plain into directional subbands.

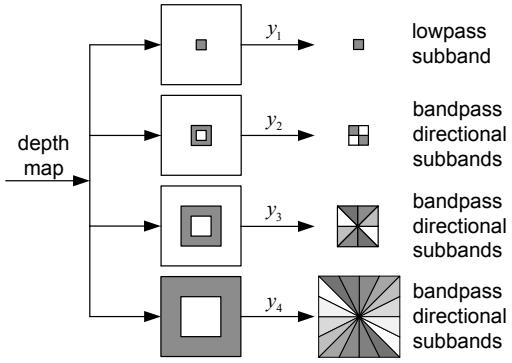


Fig. 2: The overall structure of contourlet transform: laplacian pyramid structure and directional filter banks (DFB), $y1$-$y4$ is the four chanel of Laplacian pyramids

Laplacian pyramids are constructed by iterated two-channel 2-D filter banks. They achieve a subband decomposition by low-pass filter $H_0(z)$ and high-pass filter $H_1(z)$=1-$H_0(z)$ [18]. $G_0(z)$ and $G_1(z)$ are the corresponding synthesis low-pass and high-pass filters respectively. The perfect reconstruction condition is given as

$$\begin{aligned} &H_0(z)\,G_0(z) + H_1(z)\,G_1(z) = 2 \\ &H_0(z+\pi)\,G_0(z) + H_1(z+\pi)\,G_1(z) = 0 \end{aligned} \tag{1}$$

A two-channel 2-D Laplacian pyramid decomposition with 3 level is illustrated in Fig.3. At the j-th level, the ideal subband region of the low-pass filter is $[-(\pi/2^j),(\pi/2^j)]^2$, and the ideal subband region of the high-pass filter is $[-(\pi/2^{j-1}),(\pi/2^{j-1})]^2 \setminus [-(\pi/2^j),(\pi/2^j)]^2$.

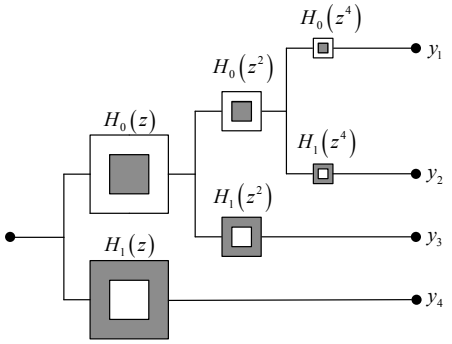
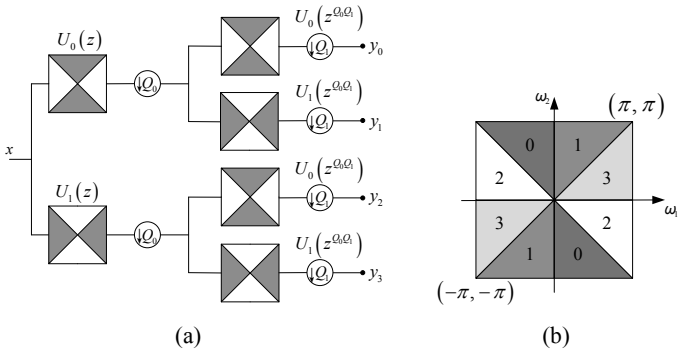


Fig. 3: Laplacian pyramid decomposition with 3 levels

Directional filter banks (DFB) composed of fan filters achieve a directional filter of different frequency parts. Fig.4 shows a four-channel DFB by cascading two level fan filters. The equivalent filter of channel k in DFB is given in (2),

$$U_k^{eq}(z) = U_i(z)\,U_j\left(z^{Q_0Q_1}\right), i,j \in \{0,1\} \tag{2}$$

where U_i and U_j are the fan filters, and Q_0 and Q_1 represent sample matrixes, referring to the rotation operators. More details can be referred in [17].



(a) (b)

Fig. 4: Four-channel directional filter bank constructed with two-channel fan filter banks. (a) Two-level fan filtering banks structure. (b) Corresponding frequency directional decomposition.

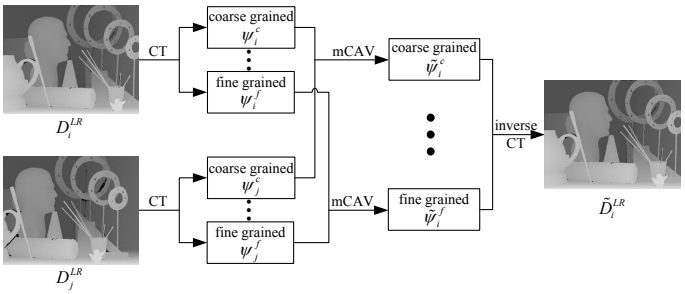


Fig. 5: The fusion process based on contourlet transform of multiple views.

3. MULTIRESOLUTION CONTOURLET TRANSFORM FUSION BASED DEPTH MAP SUPER RESOLUTION

In this section, we will give an introduction of the proposed multiresolution contourlet transform fusion based depth map super-resolution algorithm.

Due to the actual arrangement of the multiple views, they usually lie alongside every fixed distance. There's a disparity d between the left view D_i and the right view D_j in the horizontal direction. To simplify, we use two views to illustrate the formulation of disparity as in [19]

$$d = x_i - x_j = \frac{Bf}{Z}, \tag{3}$$

where d is the disparity between D_i and D_j, x_i and x_j refer to the horizontal coordinates of D_i and D_j respectively. B is the baseline distance between depth sensors. f is the focal length, and Z is the actual depth of field.

Firstly, we can find the common region Ω of the low resolution depth maps by shifting multiple views corresponding disparities to view i as (3). Then we transform Ω of multiple views by CT. The transform coefficients $\Psi_n^{p,q}$ of view n can be obtained by (4),

$$\Psi_n^{p,q} = U_\delta\left(H_\phi\left(D_n^{LR}\right)\right), \delta \in \{0,1\}\ \&\ n \in \{1,2,\ldots,N\} \tag{4}$$

where p is the level of Laplacian pyramids, and q is the direction number of DFB. Next we fuse the transform coefficients of the target depth map. The fusion process is depicted in Fig.5. Since CT is a multiresolution transform, depth map can be transformed into various grained coefficients, such as from coarse grained Ψ_i^c to fine grained Ψ_i^f of view i. The coefficients of multiple views with same grained such as Ψ_i^c and Ψ_j^c, Ψ_i^f and Ψ_j^f, are fused respectively according to the max Coefficient Absolute Value (mCAV) rule as (5).

$$\tilde{\Psi}_i^{p,q}(x,y) = \arg\max_{\Psi_n^{p,q}(x,y)}\left(\left|\Psi_n^{p,q}(x,y)\right|\right), n \in \{1,2,\ldots,N\} \tag{5}$$

where (x,y) is the pixel coordinate.Then we can get a low resolution fused depth map $\widehat{D}_i^{LR}$ with more contour information by inversely transforming the fused coefficients $\tilde{\Psi}_i$ as (6).

$$\widehat{D}_i^{LR} = \sum_{\delta\in\{0,1\},\phi\in\{0,1\}}\left(G_\phi\left(U_\delta^{-1}\left(\tilde{\Psi}_i\right)\right)\right), \tag{6}$$

The fused depth map $\widehat{D}_i^{LR}$ has the same size as region Ω. In order to gain a complete fused depth map of the target view i, we incorporate the uncommon region in view i into the fused depth region as in (7),

$$\tilde{D}_i^{LR}(x,y) = \begin{cases} \widehat{D}_i^{LR}(x,y), & if\,(x,y)\in\Omega \\ D_i^{LR}(x,y), & otherwise \end{cases} \tag{7}$$

where $\tilde{D}_i^{LR}$ is the refined depth value of view i.

4. EXPERIMENTAL RESULTS

In this section, we evaluate our method quantitatively and qualitatively comparing to several state-of-the-art methods.

4.1. PARAMETERS SETTING

We implement our experiments on computer with Intel 2.8GHz CPU, 12GB RAM, 64-bits Windows 7 and MATLAB R2014a. The raw depth maps are from Middlebury Dataset [20], which is widely used in a large number of researches in 3D applications. Among the Middlebury Stereo Dataset, we use eight sets of color images and depth maps. They are Art, Books, Dolls, Laundry, Midd1, Moebius, Monopoly, and Reindeer.

To get the LR depth map, we downsample the corresponding HR depth map with the scaling factor 4. View 1 is chosen

to be the target view, and view 5 is chosen to be the matching view. The parameters are set as follows. The number of views N is 2. The level of Laplacian pyramids is set as 3. The numbers of directions at each level are 2, 4 and 8 from lower to higher level respectively. Considering actual condition that depth maps are corrupted by external and internal noise, we conduct Gaussian noise on the depth map. For Gaussian noise, the standard deviations on the left view and the right view are 4 respectively.

Table 1: Comparison of SR experiments by R-method, MDMF and our method under JBU framework

images	JBU[3]	R-JBU [15]	MDMF +JBU[16]	proposed +JBU	gain
Art	28.37	28.42	29.18	30.38	1.20
Books	30.22	30.35	30.97	32.39	1.42
Dolls	31.60	31.65	32.69	33.76	1.07
Laundry	30.32	30.48	31.63	32.31	0.68
Midd1	31.14	31.23	32.92	33.34	0.42
Moebius	31.00	31.27	32.00	33.08	1.08
Monopoly	31.66	31.74	32.71	33.89	1.18
Reindeer	31.80	31.94	32.77	32.91	0.14
average	30.76	30.89	31.86	32.76	0.90

Table 2: Comparison of SR experiments by R-method, MDMF and our method under WMF framework

images	WMF [3]	R-WMF [15]	MDMF +WMF[16]	proposed +WMF	gain
Art	28.44	28.51	29.32	31.27	1.95
Books	30.49	30.55	31.29	32.54	1.25
Dolls	31.76	31.93	32.80	35.91	3.11
Laundry	30.57	30.71	31.85	34.39	2.54
Midd1	31.30	31.41	33.20	35.01	1.81
Moebius	31.16	32.16	32.42	35.98	3.56
Monopoly	31.80	31.84	32.90	34.60	1.70
Reindeer	31.87	31.93	32.89	33.84	0.95
average	30.92	31.13	32.08	34.19	2.11

4.2. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the performance of the proposed contourlet transform based fusion method, we compare our method with the state-of-the-art fusion method MDMF [14] and R-method[19] with the common upscaling frameworks JBU[9] and WMF [20]. From Table.1 and Table.2, we have an average gain of 0.90dB, most 1.42dB in the JBU framework, and an average gain of 2.11dB, most 3.56dB in the WMF framework over the latest MDMF methods. Meanwhile, we exceed R-method 1.87dB in the JBU framework, and 3.06dB in the WMF framework on average. Comparing with R-method and MDMF visually in Fig.6, our method successfully avoids texture copy artifacts beyond the other two methods, while alleviating the noise and completing the holes. Simultaneously, our method preserved more contour information and produced sharper edges of the target depth map without jagged artifacts.

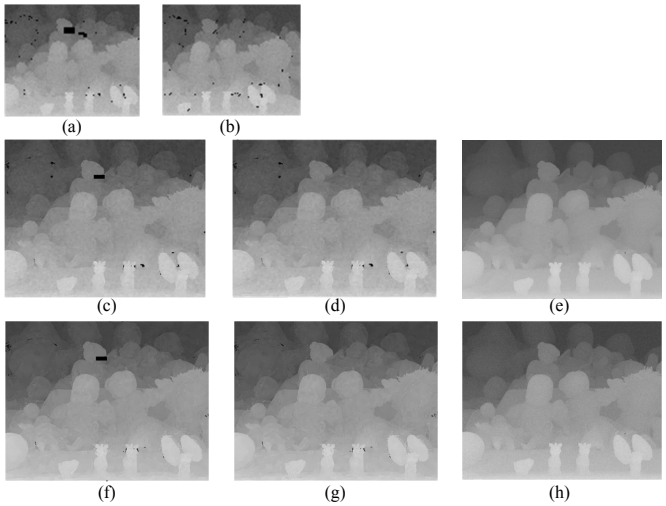


Fig. 6: Visual comparison of SR results by R-method, MDMF, and our method of image Dolls. (a-b) The LR depth maps of left view and right view, (c) R-JBU, (d) MDMF-JBU, (e) CT-JBU, (f) R-WMF, (g) MDMF-WMF and (h) CT-WMF.

5. CONCLUSION

In this paper, we have proposed a multiresolution contourlet transform fusion based depth map super resolution method. Inspired from image fusion in transform domain, we present to fuse LR depth maps of multiple views in multiresolution via contourlet transform to get a HR depth map of the target view. Contourlet transform can retain the smoothness along the contours and isolate the discontinuities of contours, especially suitable for depth map. Comparing with conventional depth map SR methods which concentrate on spatial domain, our method can preserve more contour information and avoid texture-copy artifacts, while mitigating the external and internal noise. Experimental results on Middlebury Stereo Dataset demonstrate that our method is superior to several state-of-the-art algorithms objectively and subjectively.

6. ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation of China (NSFC) under grants 61472101 and 61631017 and the Major State Basic Research Development Program of China (973 Program 2015CB351804).

7. REFERENCES

[1] S. Izadi, D. Kim, and O. Hilliges, "Kinectfusion:real-time 3d reconstruction and interaction using a moving depth camera," in *ACM Symposium on User Interface Software and Technology*. ACM, 2011, pp. 559–568.

[2] J. Park, H. Kim, Yu-Wing Tai, M. S. Brown, and I. K-weon, "High quality depth map upsampling for 3d-tof cameras," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 1623–1630.

[3] J. Kopf, M.F. Cohen, and D. Lischinski, "Joint bilateral upsampling," in *Acm Transactions on Graphics*. ACM, 2007, vol. 26(3), p. 96.

[4] J. Kim, G. Jeon, and J. Jeong, "Joint-adaptive bilateral depth map upsampling," in *Signal Processing Image Communication*. ELSEVIER, 2014, vol. 29(4), pp. 506–513.

[5] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," March 2012, vol. 21, pp. 1176–1190.

[6] M. Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 169–176.

[7] K. H. Lo, Y. C. F. Wang, and K. L. Hua, "Edge-preserving depth map upsampling by joint trilateral filter," 2017, vol. PP, pp. 1–14.

[8] J. Diebel and ThrunS., "An application of markov random fields to range sensing," in *Advances in Neural Information Processing Systems*, 2005, pp. 291–298.

[9] J. Lu, D. Min, R. S. Pahwa, and M. N. Do, "A revisit to mrf-based depth map super-resolution and enhancement," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 985–988.

[10] Y. Zuo, Q. Wu, J. Zhang, and P. An, "Explicit edge inconsistency evaluation model for color-guided depth map enhancement," 2017, vol. PP, pp. 1–1.

[11] J. Xie, R. S. Feris, and M. T. Sun, "Edge-guided single depth image super resolution," Jan 2016, vol. 25, pp. 428–438.

[12] K. H. Lo, K. L. Hua, and Y. C. F. Wang, "Depth map super-resolution via markov random fields without texture-copying artifacts," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1414–1418.

[13] J. Li, Z. Lu, G. Zeng, R. Gan, and H. Zha, "Similarity-aware patchwork assembly for depth image super-resolution," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 3374–3381.

[14] U. Hahne and M. Alexa, "Exposure fusion for time-of-flight imaging," in *Computer Graphics Forum*. Blackwell Publishing Ltd, 2011, pp. 1887–1894.

[15] J. Choi, D. Min, and K. Sohn, "Reliability-based multiview depth enhancement considering interview coherence," April 2014, vol. 24, pp. 603–616.

[16] J. Lei, L. Li, H. Yue, F. Wu, N. Ling, and C. Hou, "Depth map super-resolution considering view synthesis quality," April 2017, vol. 26, pp. 1732–1745.

[17] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," Dec 2005, vol. 14, pp. 2091–2106.

[18] J.L. Starck, F. Murtagh, and A. Bijaoui, "Image processing and data analysis," Cambridge Univ Pr, 1998.

[19] Darius Burschka, Myron Z. Brown, and Gregory D. Hager, "Advances in computational stereo," Los Alamitos, CA, USA, 2003, vol. 25, pp. 993–1008, IEEE Computer Society.

[20] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.