MULTISCALE DIRECTIONAL FUSION FOR DEPTH MAP SUPER RESOLUTION WITH DENOISING

Dan Xu[⋆ ▽] *Xiaopeng Fan*[⋆▽] *Shibo Zhang* [§] *Yang Wang* [⋆▽] *Debin Zhao*[⋆▽] *Wen Gao*[†⋆▽]

[⋆] Harbin Institute of Technology, School of Computer Science & Technology
[§] Northwestern University, Department of Electrical Engineering and Computer Science, Evanston, IL, USA
[†] Peking University, School of Electrical Engineering & Computer Science
[▽] Peng Cheng Laboratory, Shenzhen, China

ABSTRACT

To tackle three main problems in depth map super resolution (SR) process, which are texture copy artifacts, blurred edge artifacts and jagged edge artifacts, we propose a depth map super resolution with denoising method based on multiscale directional fusion via nonsubsampled contourlet transform (NCST). We first transform low resolution depth maps of multiple views via NSCT. Then NSCT coefficients are denoised by a BayesShrink threshold in nonsubsampled directional filter banks (NSDFB) domain and fused by the max coefficient absolute value (mCAV) rule respectively within each scale and direction. Finally the fused coefficients are synthesized and upscaled to a high resolution depth map utilizing a modified edge-guided joint bilateral filter. Experimental results demonstrate that our method significantly outperforms the state-of-the-art super resolution algorithms quantitatively and visually while mitigating the corrupted noise.

Index Terms— nonsubsampled contourlet transform, BayesShrink threshold, fusion, depth map super resolution

1. INTRODUCTION

Recent years, the rapid advancement of 3D applications, such as freeview TV, robot navigation and 3D object tracking have been witnessed. Many 3D applications require that depth map has the same resolution as the corresponding color map. Due to the limited capacity of the existing range sensors such as Microsoft Kinect and Time of Flight (ToF) cameras, generated depth maps are usually with low resolution (LR).

To enhance the resolution of depth maps, many depth map super resolution algorithms are proposed, among which two main categories are utilized to upsample the depth maps generally. One category is single depth map SR methods et al. [1–6]. They obtain the high resolution (HR) depth map with the initial LR depth map or the corresponding guided high resolution color map of the same view. For example, Ham et al. [3] propose a SR method which jointly leverages structural information of guidance and input depth maps. Utilizing the edges of the low-resolution depth image through a Markov Random Fields optimization in a patch synthesis based manner, Xie et al. [6] construct a high resolution edge map to guide the upscaling of the depth map.

The other category is multiple depth map SR methods [7–10], which enhance the resolution of the target depth map utilizing the fusion results of multiple low resolution depth maps of different views at the same time or same view at different time. Choi et al. [9] propose a novel depth map SR framework by taking interview coherence into account. Recently, Lei et al. [10] proposed a credibility based multiview depth maps fusion strategy, which takes the view synthesis quality and interview correlation into account, and gain a remarkable performance.

For single depth map SR methods, they may produce texture copy artifacts and blurred edge artifacts when the color discontinuities and the depth discontinuities at the corresponding location are not consistent. To tackle these problem, multiple depth map SR methods are applied. By fusing the depth maps of different views or times, more contour and occlusion information can be obtained which will be used subsequently in the SR process. Considering the characteristics of clear edges with no much texture in depth map, preserving more contour information is the key to the depth map super resolution. Multiple depth map SR methods are performed in spatial domain, which only take local pixels into account rather than global contour structure. Inspired by the success of the image fusion in transform domain of multi-view [11] and multi-focus [12, 13] which can obtain more high frequency details, we propose to fuse the LR depth maps in transform domain via multiscale nonsubsampled contourlet transform (NSCT) for depth map SR.

In addition, depth maps captured from the existing range sensors are usually degraded by both internal and external noise. The internal noise is caused by some factors of sensor itself, such as photon shot noise and thermal noise. The external noise is caused by some environment factors, such as strong illumination, occlusion, and high reflectivity surfaces. Among these noises, photon shot noise is the most significant factor which impacts the intensity of noise. In ac-

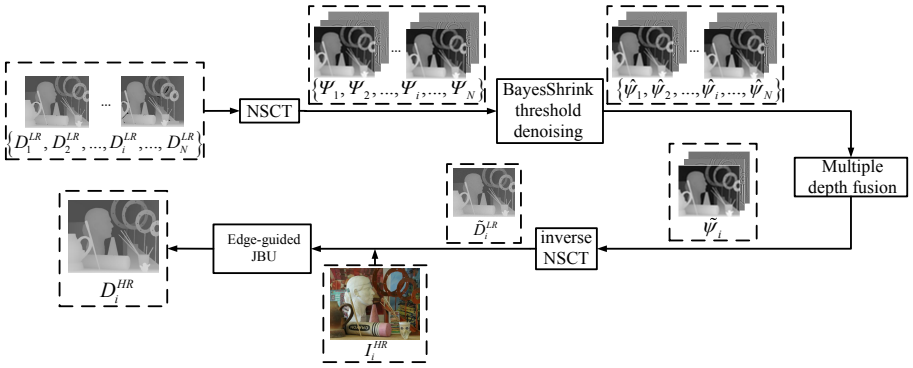

Fig. 1: The flowchart of the multiscale directional fusion for depth map super resolution with denoising model.

tual capture scene, photon shot noise is theoretically Poisson distributed [14], but it can be sufficiently approximated to a Gaussian distribution proved by Frank et al. [15]. With the corrupted noise, depth map cannot reflect the accurate depth value of the scene. Especially in 3D reconstruction procedure, noise along surfaces in 3D scenario is more obvious to cause observers' attention. Unfortunately, depth map noise does not cause much attention. Only a few literatures [16–18] take bilateral filter as a pre-processing treatment to simply refine the incorrect values of the depth map. However, bilateral filter is not a satisfying denoising filter specific to actual noise distribution. Our method uses a BayesShrink threshold model with the assumption that the corrupted noise conform to a Gaussian distribution, markedly alleviating the depth noise.

In this paper, we propose a depth map super resolution with denoising method based on multiscale directional fusion via NSCT. We first transform LR depth maps of multiple views in NSCT domain [19]. Then NSCT coefficients are denoised by a BayesShrink threshold in NSDFB domain and fused by mCAV rule respectively within each scale and direction. Finally the fused coefficients are synthesized and upscaled to the target high resolution depth map utilizing a promoted edge-guided joint bilateral upsampling (JBU).

2. FRAMEWORK OF MULTISCALE DIRECTIONAL FUSION BASED DEPTH MAP SUPER RESOLUTION WITH DENOISING

As illustrated in Fig. 1, the multiple LR depth maps $\{D_1^{LR}, D_2^{LR}, \ldots, D_i^{LR}, \ldots, D_N^{LR}\}$ of different views are input and transformed into NSCT coefficients $\{\Psi_1, \Psi_2, \ldots, \Psi_i, \ldots, \Psi_N\}$. Parameter N refers to the number of view points. Then $\{\Psi_1, \Psi_2, \ldots, \Psi_i, \ldots, \Psi_N\}$ is denoised by the BayesShrink threshold denoising method to obtain the purified coefficients $\{\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_i, \ldots, \hat{\psi}_N\}$. The fused coefficients $\tilde{\psi}_i$ of view i are get by fusing the denoised NSCT coefficients of multiple depth maps via the max Coefficient Absolute Value (mCAV) rule which will be introduced in the following part. The fused LR depth map $\tilde{D}_i^{LR}$ is acquired through the inverse NSCT of $\hat{\psi}_i$. The final HR depth map D_i^{HR} is a result upscaling $\tilde{D}_i^{LR}$ by edge-guided JBU with the guidance of I_i^{HR} which is the HR color image of view i.

3. MULTISCALE DIRECTIONAL FUSION FOR DEPTH MAP SUPER RESOLUTION WITH DENOISING

3.1. Depth map transform by NSCT

In this section we give an introduction of depth map transform by nonsubsampled contourlet transform (NSCT). NSCT consists of two parts: nonsubsampled pyramid structure (NSP) and nonsubsampled directional filter banks (NSDFB). NSP filters the image into low-frequency subbands and high-frequency subbands. NSDFB transforms the 2-D frequency plain into directional subbands. Nonsubsampled pyramids (NSP) are constructed by iterated two-channel nonsubsampled 2-D filter banks

$$H_\phi(z) = \begin{cases} H_1\left(z^{2^{\phi-1}}\right)\prod_{j=0}^{\phi-2} H_0\left(z^{\phi}\right), & 1 \le \phi < 2^J \\ \prod_{j=0}^{\phi-1} H_1\left(z^{2^{\phi}}\right), & \phi = 2^J \end{cases} \tag{1}$$

where $z = [z_1, z_2]^T$. They achieve a subband decomposition by low-pass filter H_0 and high-pass filter $H_1(z) = 1 - H_0(z)$ [20]. NSDFB is a shift-invariant extension of contourlet transform (CT) [21] which achieves a directional filtering of different frequency parts by directional filter banks (DFB) composed of fan filters as (2)

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

where U_i and U_j are the fan filters, and Q represents a rotation operator. More details can be referred in [19].

As multiple views usually lie alongside every fixed distance in the actual arrangement, there is a disparity d between the left view D_i and the right view D_j in the horizontal direction as in [22]

$$d = D_i - D_j = \frac{Bf}{Z}. \tag{3}$$

B is the baseline distance between depth sensors. f is the focal length, and Z is the actual depth of field.

We firstly find the common depth map region Ω in multiple views. Then we transform Ω in depth maps of different views by NSCT. The transform coefficients $\Psi_n^{p,q}$ of noisy depth map of view n can be obtained by (4),

$$\begin{aligned} &\Psi_n^{p,q} = U_\kappa\left(H_\phi\left(D_n^{LR}\right)\right), \kappa \in \{0,1\}, \\ &\phi \in \{1,2,\ldots,2^k\}, \text{and } n \in \{1,2,\ldots,N\}, \end{aligned} \tag{4}$$

where p is the level of NSP, and q is the direction number of NSDFB.

3.2. BayesShrink threshold based depth map denoising in NSDFB domain

The noisy NSCT coefficients $\Psi_n^{p,q}$ can be regarded as a result of original NSCT coefficients $\psi_n^{p,q}$ corrupted by noise ε as

$$\Psi_n^{p,q} = \psi_n^{p,q} + \varepsilon. \tag{5}$$

In this paper, we denoise the depth map coefficients by setting a BayesShrink threshold [23] proposed by Chang et al. For a given parameters set, the objective is to find a soft-threshold δ_n^p in pth level of view n which minimizes the Bayes risk,

$$r\left(\delta_n^p\right) = E\left(\hat{\psi}_n^p - \psi_n^p\right)^2 = E_{\psi_n^p} E_{\hat{\psi}_n^p|\psi_n^p}\left(\hat{\psi}_n^p - \psi_n^p\right)^2 \tag{6}$$

where $\hat{\psi}_n^p$ is the denoised NSCT coefficients of view n in pth level, $\hat{\psi}_n^p|\psi_n^p \sim N(\bar{\psi}_n^p, \sigma_\varepsilon^2)$, $\bar{\psi}_n^p$ is the mean value of $\hat{\psi}_n^p$, σ_ε is the standard deviation of ε. Denote the optimal threshold by $\hat{\delta}_n^p$,

$$\hat{\delta}_n^p = \arg\min_{\delta_n^p} r(\delta_n^p). \tag{7}$$

First, as the intensity of noise is unknown, a robust median estimator [24] is used from the finest scale coefficients of the first level to estimate the standard deviation σ_ε of ε from the noisy coefficients Ψ_n^p .

$$\sigma_\varepsilon = median(\Psi_n^1)/0.6745. \tag{8}$$

The standard deviation of each level σ_Ψ^p can also be estimated as (9),

$$(\sigma_\Psi^p)^2 = \frac{1}{M}\sum_q (\Psi_n^{p,q})^2 \tag{9}$$

where M indicates the number of the noisy NSCT coefficients in pth level. As $(\sigma_\Psi^p)^2 = (\sigma_\psi^p)^2 + \sigma_\varepsilon^2$, the standard deviation of original NSCT coefficients is calculated as $\sigma_\psi^p = \sqrt{\max\left((\sigma_\Psi^p)^2 - \sigma_\varepsilon^2, 0\right)}$. As there is no closed form solution for $\hat{\delta}_n^p$, BayesShrink threshold $\dot{\delta}_n^p$ is used as an approximate solution to $\hat{\delta}_n^p$,

$$\dot{\delta}_n^p = \begin{cases} \frac{\sigma_\varepsilon^2}{\sqrt{\max\left((\sigma_\Psi^p)^2 - \sigma_\varepsilon^2, 0\right)}}, & \sigma_\varepsilon^2 < (\sigma_\psi)^2, \\ \max\left(|\Psi_n^p|\right), & \sigma_\varepsilon^2 \geq (\sigma_\psi)^2. \end{cases} \tag{10}$$

Equation 11 gives a soft threshold function of the filter criterion

$$\hat{\psi}_n^p = \text{sgn}(\Psi_n^p) \cdot \max\left(|\Psi_n^p| - \dot{\delta}_n^p, 0\right). \tag{11}$$

3.3. Depth map Super resolution based on contour multiscale directional fusion via NSCT

This section gives an introduction about the process of depth map super resolution based on multiscale NSCT fusion. Firstly, the denoised transform coefficients $\{\hat{\psi}_n^{p,q} | n = 1,\ldots,i,\ldots,N\}$ of N views are fused together.Since NSCT is a multiscale transform, depth map can be transformed into various grained coefficients, such as from coarse grained $\hat{\psi}_i^c$ to fine grained $\hat{\psi}_i^f$ of view i. We fuse the coefficients of multiple views with same grain such as $\hat{\psi}_i^c$ and $\hat{\psi}_j^c$, $\hat{\psi}_i^f$ and $\hat{\psi}_j^f$, according to the max Coefficient Absolute Value (mCAV) rule,

$$\tilde{\psi}_i^{p,q}(x,y) = \max_{n\in\{1,2,\ldots,N\}}\left(|\hat{\psi}_n^{p,q}(x,y)|\right), \tag{12}$$

where $\tilde{\psi}_i^{p,q}$ indicates the fused coefficients in pth level with qth direction of view i, (x,y) refers to the coordinate of the depth pixel.

Utilizing inverse NSCT with the obtained fused coefficients from coarse grained to fine grained, we can get a fused depth map with more contour information of the size as region Ω. The fused depth map $\tilde{D}_i^{LR}$ is gained by (13),

$$\tilde{D}_i^{LR} = \sum_{\kappa\in\{0,1\},\phi\in\{0,1\}}\left(G_\phi\left(U_\kappa^{-1}(\tilde{\psi}_i)\right)\right). \tag{13}$$

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 (14),

$$\breve{D}_i^{LR}(x,y) = \begin{cases} \tilde{D}_{i_f}^{LR}(x,y), & \text{if } (x,y) \in \Omega \\ D_i^{LR}(x,y), & \text{otherwise} \end{cases} \tag{14}$$

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

Finally, the refined LR depth map $\breve{D}_i^{LR}$ is upscaled to a high resolution depth map D_i^{LR} utilizing a edge-guided JBU as (15), which effectively avoid the texture copy artifacts and blurred edge artifacts,

$$\begin{aligned} D_i^{HR}(x_2',y_2') = &\frac{1}{\theta_{x_2',y_2'}}\sum_{(x_1,y_1)\in\Theta}\breve{D}_i^{LR}(x_1,y_1)\mathcal{S}(x_1,y_1,x_2,y_2) \\ &\cdot\left(\alpha\Re\left(\|I_{x_2',y_2'} - I_{x_1',y_1'}\|\right) + (1-\alpha)\Im\left(\|I_{x_2,y_2} - I_{x_1,y_1}\|\right)\right) \end{aligned} \tag{15}$$

where $\theta_{x_2',y_2'}$ is a normalization factor, Θ is a neighborhood centered at (x_2,y_2). (x_1',y_1') and (x_2',y_2') refer to the corresponding HR pixel coordinate of the LR pixel coordinate (x_1,y_1) and (x_2,y_2) respectively. $\mathcal{S}$, $\Re$ and $\Im$ are Gaussian kernel functions about distance, illumination and depth difference respectively. α is a discriminating factor which ranges from 0 to 1 depending on the distance between its location in LR depth map to adjacent edge.

Algorithm	TGV [25]		JABDU [26]		R [9]		SDF [3]		MDMF+VSQ-TDU [10]		Ours	
	PEP	PEPD	PEP	PEPD	PEP	PEPD	PEP	PEPD	PEP	PEPD	PEP	PEPD
Art	5.79%	5.04%	11.61%	10.53%	16.41%	15.66%	4.36%	3.77%	11.60%	10.12%	**3.26%**	**2.58%**
Books	4.32%	3.66%	3.70%	3.12%	10.20%	9.35%	3.57%	2.56%	2.88%	2.54%	**1.79%**	**1.16%**
Dolls	3.53%	3.24%	4.24%	3.76%	10.59%	9.94%	3.08%	2.32%	3.18%	2.97%	**2.64%**	**2.03%**
Laundry	3.74%	3.15%	6.03%	4.28%	9.00%	8.18%	2.93%	2.16%	5.96%	4.13%	**2.77%**	**1.96%**
Midd1	7.18%	6.02%	3.54%	3.19%	13.28%	11.99%	5.93%	5.06%	2.65%	2.07%	**2.13%**	**1.52%**
Moebius	4.21%	3.42%	4.74%	4.01%	11.34%	10.65%	3.06%	2.58%	3.83%	3.40%	**2.32%**	**1.47%**
Monopoly	4.54%	3.39%	3.70%	3.22%	6.42%	5.21%	3.23%	2.89%	2.91%	2.56%	**1.98%**	**1.35%**
Reindeer	3.79%	3.14%	5.00%	4.58%	9.76%	8.35%	2.66%	2.08%	4.49%	3.41%	**2.17%**	**1.39%**
average	4.64%	3.88%	5.32%	4.59%	10.88%	9.92%	3.60%	2.93%	4.69%	3.90%	**2.38%**	**1.68%**

Table 1: Quantitative depth upsampling results on Middlebury datasets

4. EXPERIMENTAL RESULTS AND ANALYSIS

We implement our experiments on Middlebury Dataset [27]. We use eight sets which are Art, Books, Dolls, Laundry, Midd1, Moebius, Monopoly, and Reindeer. We downsample the corresponding HR depth map with the scaling factor 4 to get the LR depth map. The level of Laplacian pyramids is set as 3. The numbers of directions at each level are 2, 4 and 8 from higher to lower level respectively. Considering actual condition that depth maps are corrupted by photon shot noise, we conduct Gaussian Noise on the depth map with standard deviations of 4 as [10] to compare the performance of super resolution without denoising and with denoising.

4.1. Experimental comparison on super resolution without and with denoising

To verify the performance of the proposed method, we compare our method with five state-of-the-art methods, i.e., total generalized variation(TGV) [25], joint adaptive bilateral depth map upsampling (JABDU) [26], R-method [9], joint static and dynamic filter (SDF) [3], and MDMF+VSQ-TDU [10]. We utilize PEP (the percentage of error pixels whose disparity error is larger than 6 pixels in the up-sampled depth maps) and PEPD(PEP after denoising) as the assessment criterias simultaneously. Table 1 shows the quantitative results of super resolution without denoising and with denoising. It can be seen that we gain the lowest PEP and PEPD compared with other state-of-the-arts. As some depth map SR methods take a bilateral filter as denoising filter, in this paper, we take this treatment as the denoising method for other methods. Fig. 2 shows visually comparison on depth map super resolution results with denoising for images Books. It can be seen that result of our proposed method successfully avoids texture copy artifacts, blurred edge artifacts and jagged edge artifacts beyond the other methods while processing noise.

5. CONCLUSION

We propose a multiscale directional fusion method for depth map super resolution with denoising via NSCT to enhance the resolution of the depth map. On one hand, multiscale directional fusion based depth map SR method upscales the depth map in different scales according to the contour orientation of the depth map, which is extremely proper to retain the contours in depth map, effectively averting texture copy artifacts, blurred and jagged edge artifacts. On the other hand, BayesShrink threshold based denoising in NSDFB domain markedly mitigates the noise corrupted in the depth map. Experimental results demonstrate that our method significantly outperforms the state-of-the-arts quantitatively and visually while mitigating the corrupted noise.

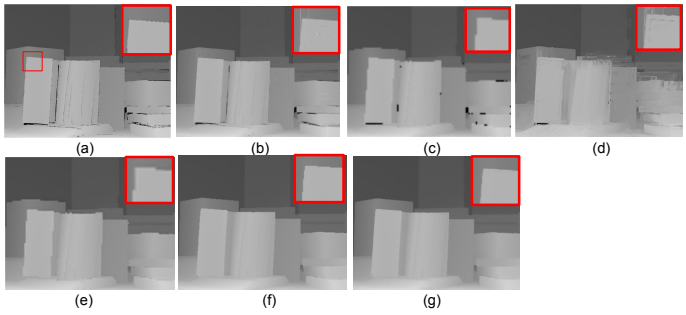

Fig. 2: Visually comparison of depth map super resolution results with denoising for images Books and Reindeer, (a) ground truth, (b–g) depth map uspampled by TGV, JABDU, R-method, MDMF+VSQ-TDU, SDF, ours.

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

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

[3] Bumsub Ham, Minsu Cho, and Jean Ponce, "Robust image filtering using joint static and dynamic guidance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, 2015.

[4] 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.

[5] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM Transactions on Graphics*, 01 2007, vol. 26.

[6] J. Xie, R. S. Feris, and M. T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, Jan 2016.

[7] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for tof 3d shape scanning," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009, pp. 343–350.

[8] J. Zhu, L. Wang, R. Yang, J. E. Davis, and Z. Pan, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1400–1414, July 2011.

[9] J. Choi, D. Min, and K. Sohn, "Reliability-based multiview depth enhancement considering interview coherence," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 603–616, April 2014.

[10] J. Lei, L. Li, H. Yue, F. Wu, N. Ling, and C. Hou, "Depth map super-resolution considering view synthesis quality," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1732–1745, April 2017.

[11] Z. Zhang, H. Bian, and Z. Song, "A multi-view sonar image fusion method based on the morphological wavelet and directional filters," in *IEEE/OES China Ocean Acoustics (COA)*, Jan 2016, pp. 1–6.

[12] S. Liu and J. Chen, "A fast multi-focus image fusion algorithm by dwt and focused region decision map," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2016, pp. 1–7.

[13] L. Lei, S. L. Zhou, and J. L. Liu, "Multi-focus image fusion based on scale vector norm of nonsubsampled contourlet transform," in *IECON 42nd Annual Conference of the IEEE Industrial Electronics Society*, Oct 2016, pp. 981–985.

[14] B. Buettgen, T. Oggier, M. Lehmann, R. Kaufmann, and F. Lustenberger, "Ccd/cmos lock-in pixel for range imaging: Challenges, limitations and state-of-the-art," in *Proceedings of Range Imaging Research Day*, 01 2005.

[15] H. Rapp U. Koethe B. Johne F. A. Hamprecht M. Frank, M. Plaue, "Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras," *Optical Engineering*, vol. 48, pp. 48 – 48 – 16, 2009.

[16] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.

[17] L. Chen, H. Lin, and S. Li, "Depth image enhancement for kinect using region growing and bilateral filter," in *International Conference on Pattern Recognition (ICPR)*, Nov 2012, pp. 3070–3073.

[18] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," pp. 71–84, 2012.

[19] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, Oct 2006.

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

[21] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec 2005.

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

[23] S. G. Chang, Bin Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sep 2000.

[24] G. Y. Chen, T. D. Bui, and A. Krzyzak, "Image denoising using neighbouring wavelet coefficients," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 2, pp. ii–917–20 vol.2.

[25] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," pp. 993–1000, 2013.

[26] Joohyeok Kim, Gwanggil Jeon, and Jechang Jeong, "Joint-adaptive bilateral depth map upsampling," *Signal Processing Image Communication*, vol. 29, no. 4, pp. 506–513, 2014.

[27] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007, pp. 1–8.