HIERARCHICAL EXAMPLE-BASED RANGE-IMAGE SUPER-RESOLUTION WITH EDGE-PRESERVATION

Srimanta Mandal, Arnav Bhavsar, and Anil Kumar Sao

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi, India
srimanta_mandal@students.iitmandi.ac.in, arnav@iitmandi.ac.in, anil@iitmandi.ac.in

ABSTRACT

We propose an example-based approach for enhancing resolution of range-images. Unlike most existing methods on range-image super-resolution (SR), we do not employ a colour image counterpart for the range-image. Moreover, we use only a small set of range-images to construct a dictionary of exemplars. Considering the importance of edges in range-image SR, our formulation involves an edge-based constraint to better weight appropriate patches from the dictionary in a sparse-representation framework. Moreover, realizing the need for large up-sampling factors in case of range-images, we follow a hierarchical strategy for estimating the high-resolution range-images. We demonstrate that our strategy yields considerable improvements over the state-of-the-art approaches for range-image SR.

***Index Terms*—** Range-image super-resolution, Hierarchical estimation, Edge-preservation, Sparse-representation.

1. INTRODUCTION

In recent years, range cameras and scanners are being realized as important dominant acquisition modality in computer vision and multimedia community, a reason being that a variety of technologies have been developed for range acquisition such as laser scanning [1], time-of-flight (ToF) imaging [3, 2], and structured or coded lighting [4], among which some provide high-acquisition speed, require little manual intervention, and are relatively low-cost and easily available. However, these advantages are often traded-off with limited resolution and structural accuracy in some types of range imaging devices; examples being the ToF range cameras [2], and the Kinect camera [5]. This motivates the development of computational approaches to enhance the resolution and accuracy.

Typically, range-images acquired from low-resolution (LR) range cameras are such that it is required to enhance their resolutions by large factors (unlike in optical image SR) such as 4 or 8, so as to attain a reasonable perceptual quality [6, 2]. As most of the content in range-images lacks texture details, the primary challenge in range-image SR is to achieve good localization of the inter-object edges and discontinuities when considering such large up-sampling factors, while maintaining the intra-object gradual depth variations. In this respect, naive approaches of resolution enhancement by off-the-shelf image interpolation methods result in heavy loss of localization and accuracy.

As a result, in recent years various sophisticated approaches for resolution enhancement of range-images have been reported [6, 7, 2, 8]. Considering the primary requirements of large up-sampling factors and edge preservation, a key principle upon which most of the approaches are based, is the utilization of a high-resolution (HR) colour image of the same scene. This is motivated by the observation that range discontinuities often coincide with those in colour image, and the latter can be easily captured using any off-the-shelf digital camera. Thus, colour discontinuities and local similarity of colour information derived from a HR colour image helps in localizing range discontinuities on the HR grid.

While such colour-image-based approaches do yield good quality resolution enhancement, it is possible that in some scenarios only range data is available. For instance, applications involving transmitting range information, 3D modeling, some cases of gesture analysis etc. may require the involvement of only range data. Moreover, the above mentioned colour-image-based approaches inherently require accurate registration of the range and colour image, which in turn necessitates calibration between the optical and range cameras. Using only range data can circumvent such a need for calibration/registration. Thus, one can clearly recognize a need for resolution enhancement approaches which use only range cameras.

To address this concern, we propose an example-based approach for range-image resolution enhancement, which does not employ a HR colour image. Moreover, unlike some similar methods for optical image SR, we use only a small set of range-images to construct our exemplar dictionary. Our formulation involves an edge-based constraint so as to aid in the primary task of discontinuity localization in range-image SR. Moreover, as range-image SR typically involves large up-sampling factors, we propose a hierarchical strategy for estimating the HR range-images. We demonstrate that our strategy yields considerable improvements over the state-of-the-art approaches for range-image SR.

1.1. Related work

As mentioned above, most range-image enhancement approaches employ a registered HR colour image, in order to exploit the discontinuity coincidence between the range and colour images [6, 7, 2, 8]. Our approach takes a different route which does not require HR colour image and hence does away with the necessity of an optical camera, and associated calibration/registration. Nevertheless, we do compare our approach with some of these methods, as a part of our experimentation.

Our method is motivated from the learning-based image super-resolution approaches, which uses training data to create a dictionary of local patches. More related to our work are those learning-based SR approaches which reconstruct the HR patches which involves selection of dictionary patches via sparse representation [9]. However, as compared to such image SR approaches, we require a very small set of range-images (about 3-4), to construct our dictionary. We believe that this might be because of the textureless behaviour of the

range data, which involves much less variations in the visual content than the optical images.

Among the sparsity-based colour image SR approaches, the one which is most closely related to ours is [11], whose edge-preserving strategy forms the basis of our approach. We believe that an edge-preserving approach is especially suitable for range-images, where the resolution enhancement is primarily gauged in terms of retention of object shape and inter-object discontinuities. Moreover, another key difference with this (and with other image SR-based methods), we improve the approach to work better for large up-sampling factors (e.g. 4, 8) which is a crucial requirement for range-image SR. On the other hand, the image SR methods typically demonstrate resolution enhancement by small factors (e.g. 2, 3) [6, 10, 11, 12].

Indeed, recently an example-based super-resolution approach is reported in [13], which follows an MRF-based strategy similar to [14]. However, unlike our method, this approach requires a much larger training dataset, does not use any explicit edge-based constraint, and demonstrates resolution enhancement only by factor of 4. In fact, as we demonstrate in Section 3, our method also shows significant improvements over [13].

Thus, the contributions of this work are: 1) Our approach obviates the need for registered colour image, unlike the state-of-the-art methods. 2) Our edge-preserving sparse-representation framework is particularly novel and useful for range-image SR considering the need for good discontinuity localization. 3) Our proposed hierarchical approach further helps to improve the SR efficacy for large up-sampling factors. 4) Our method requires much less training data than similar existing approaches.

2. PROPOSED APPROACH

We now discuss our approach in detail. For better clarity, the description below is divided into: 1) Edge-preserving super-resolution method of [11], in the context of range-image SR, and 2) Hierarchical strategy for range-image SR with high up-sampling factors.

2.1. Super-resolution via edge-preserving sparse-representation

We employ a patch-based super-resolution approach, where the HR range-image is constructed in a patch-wise manner. Each HR range-image patch is in turn reconstructed as a linear combination of patches in a dictionary acquired from a small set of high-resolution range-images. It often so occurs that only a few patches from the dictionary are sufficient to reconstruct the HR patch, and thus the weight vector (a.k.a coefficient vector) that weighs the dictionary patches for reconstruction, is sparse. A class of image SR methods are based on this *sparse-representation* principle, and involve solving the following problem to estimate the coefficient vector $\hat{\mathbf{c}}$

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \{\|\mathbf{y} - \mathbf{SBAc}\|_2^2 + \lambda\|\mathbf{c}\|_1\}. \quad (1)$$

The first term in the above equation computes the matching cost between a low-resolution patch $\mathbf{y}$ from the LR range observation and the corresponding weighted patches in the dictionary $\mathbf{A}$, which are blurred and down-sampled by known operators $\mathbf{B}$ and $\mathbf{S}$, respectively. The second term is an l_1-norm which enforces a sparse $\hat{\mathbf{c}}$.

The work in [11] proposes an additional gradient-based constraint in equation (1). The gradient information mentioned above is defined as 'edginess', which is computed using 1-D processing of images which is known to perform better than the conventional gradient [15]. This is computed by applying a smoothing operator along one direction and it's derivative operator along the orthogonal direction. Denoting the gradient magnitude operator as $\mathbf{E_g}$, equation (1) is modified as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \{\|\mathbf{y} - \mathbf{SBAc}\|_2^2 + \lambda\|\mathbf{c}\|_1 + \beta\|\mathbf{E_g}\{\mathbf{y}\} - \mathbf{E_g}\{\mathbf{SBAc}\}\|_2^2\}. \quad (2)$$

The additional constraint in eq. (2) minimizes the differences between gradient information of LR patch and that of the down-sampled version of the reconstructed patch. This is particularly important in the context of range data, where the discontinuity information is perceptually most significant.

The dictionary $\mathbf{A}$ is computed based on an adaptive approach proposed in [10], which involves creating multiple sub-dictionaries out of a mother-dictionary via K-means clustering of HR training patches, and Principal Component Analysis (PCA) applied to each cluster. During the coefficient estimation each sub-dictionary is selected adaptively for each LR patch, and is used as $\mathbf{A}$. The parameters λ and β are assigned weights to the edge-based and sparsity constraints, respectively. These are also computed using MAP (Maximum A Posteriori) estimation [10]. Given, the dictionary and the parameters, the equation (2) can be solved using iterative shrinkage algorithm as explained in [16] to estimate $\hat{\mathbf{c}}$. The estimated $\hat{\mathbf{c}}$ is then used to linearly combine the dictionary patches to reconstruct the HR patch, which are in turn used to reconstruct the complete image. This is done in reverse way of patch extraction with the averaged overlapping portions of patches. Due to space limitation, we encourage the reader to refer to [11, 10] for details on dictionary computation and image reconstruction.

2.2. Hierarchical strategy

The above approach has been shown to perform for optical image SR by factors of 2 or 3. However, as mentioned above, for range-image SR, the up-sampling factors involved are much higher (e.g. 4, 8). An associated concern with this, is the localization of perceptually important discontinuities in range-images. This is because the LR images in such cases are very small and do not provide enough information about the shape definitions and discontinuity localization at HR. While the edge-preserving constraint in the above approach can contribute some additional information in this respect, the matching costs can still involve ambiguities when the resolution of the input images is very low.

To mitigate this concern, we follow an hierarchical strategy. Instead of directly up-sampling by a large factor, we improve resolution in steps of 2. Clearly, an advantage of such an hierarchical structure is that at each step the information loss in the down-sampling of the estimate (involved in the matching cost), is not as high as in the direct case. This reduces the ambiguities in the matching cost at each step in both the first and second terms of equation (2). Moreover, the edge-based constraint ensures a reasonable localization at each step, which provides a well-localized 'LR' input image for the next step.

Interestingly, we have used same mother-dictionary across all the steps, which is derived from patches extracted from the training images at the highest resolution (as opposed to creating dictionary from down-sampled versions of training images). This is due to the fact that, although the training images are few, one can still have hundreds of patches from these. Given the limited amount of visual content in the range-images, such a dictionary is enough to contain the multi-scale local information required for an hierarchical approach.

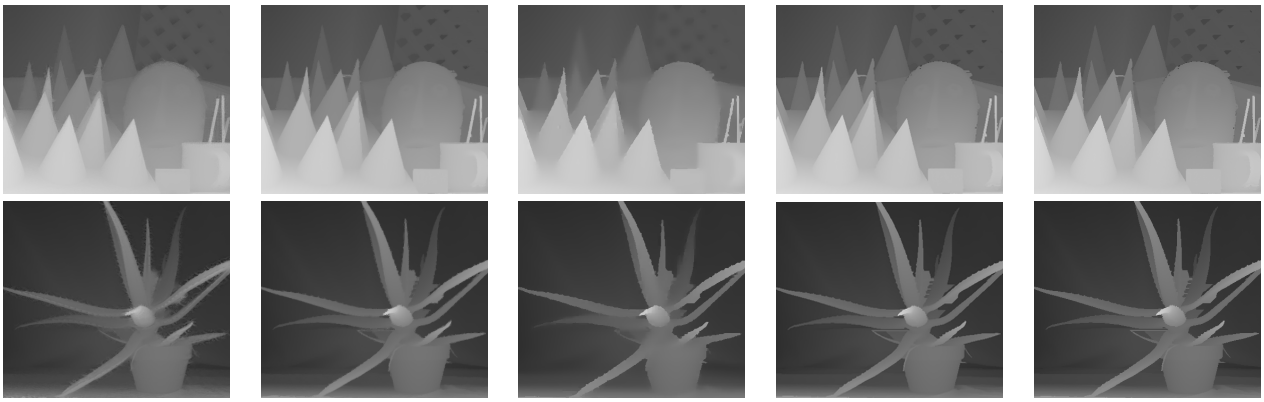

Fig. 1. SR by factor of 2: Rows 1-2 show the results for scenes Cones and Aloe, respectively. Columns 1-5 show the result on each scene for GIF [19], ATGV [2], EB [13], the proposed approach, and the original scene respectively.

3. EXPERIMENTAL RESULTS

We now provide some experimental results for our approach. We used range data from the Middlebury dataset [17, 18], which involves a rich variety of scenes. The dimensions of the HR range-images used for our training data and for ground-truth validation are of the order of 500×500. We blurred and down-sampled some HR images depending on the resolution factor in each experiment, and used the corresponding LR images as our observations. For instance, after down-sampling by factor of 2, 4, and 8, the image sizes were of the order 250×250, 125×125 and 60×60, respectively. The training data consists of 4 HR images, (viz. different from those on which our approach was validated). The patch-size in our experiments was 7×7, and the total no. of training patches were 1000.

We experimented with SR factors of 2, 4, and 8, using both the direct approach and hierarchical approach. In addition, we also compared with two recent approaches which employ the registered colour image [7, 19, 2]. These are arguably two of the best performing state-of-the-art approaches. We also compare with the recent example-based approach [13]. For all these approaches we use the publicly available codes, provided by the respective authors [20, 21, 22]. Additionally, for the method in [13], we show results obtained when using the training data provided by the authors, as this yields much better results. We provide qualitative as well as quantitative results for these experiments.

3.1. Visual results and comparisons

We first show some visual results in Figs. 1, 2, and 3, for the cases of resolution enhancement by 2, 4, and 8, respectively. In all figures, each row shows results for one example scene, over different approaches. The results from left to right in each row are for the following methods: Guided Image Filter (GIF) [19, 7], Anisotropic Total Generalized Variation (ATGV) [2], Example-based approach (EB) [13], and the proposed approach (with hierarchical estimation for factors of 4 and 8). The selected scenes are chosen as they contain significant amount of discontinuity variations, so as to better gauge the performance [1].(Please zoom the pdf soft-copy up to 300-400%, to view the images close to their actual sizes.)

Note that for the case of SR by a factor of 2 (Fig. 1), results of all approaches are relatively comparable at a first glance. Nevertheless, one can observe clear improvements in the edge-preservation using our approach, while the other methods show relatively more edge-distortions and/or bleeding at the edges.

These distortions at the discontinuities increase for higher up-sampling factors, as can be seen in the Figs. 2 and 3, for higher up-sampling factors. Particularly, the example-based approach (which, like us, does not use any colour images information) deteriorates heavily in terms of preserving object shapes. Interestingly, our approach also performs better in terms of discontinuity localization and shape preservation as compared to the more popular colour-image based methods in many cases. This clearly indicates that even without the HR colour image information, one can indeed achieve high-quality resolution enhancement given a good estimation framework.

3.2. Quantitative results

Having demonstrated some qualitative improvements over the state-of-the-art, we now validate our approach and its better performance quantitatively. We provide quantitative results over more scenes, which we could not show visually due to space constraints. Our quantitative metrics involve root mean square error (RMSE) and structural similarity (SSIM) [23], where the latter is shown to better correlate with human perception. Thus, both metrics gauge the approaches differently. The bracketed number, mentioned in the column corresponding to our method, denotes the rank of our approach in that row. The two columns for our approach in scale-4 and scale-8 cases mentioning (NH) and (H), indicate results for non-hierarchical and hierarchical approach, respectively.

The quantitative results mirror the improvements in the qualitative results shown above. We can note that we significantly outperform the example-based approach in all cases. Additionally, our approach also favourably compares with the image-based approaches in most cases. Even for these cases, the improvement is considerable in many cases.

4. CONCLUSION

We proposed an approach for example-based resolution enhancement for range-images. Unlike the popular strategy, our approach does not use any associated colour image, and requires only a range camera, thus obviating any registration/calibration. We formulate our method in an elegant sparse representation framework which also employs an edge-preserving constraint. Moreover, we also propose an hierarchical strategy to enable enhancement by high up-sampling factors. Our work indicates that given a good estimation framework, an example-based approach can outperform state-of-the-art methods including those based on using HR colour-image.

[1] Interested readers are encouraged to visit the url http://faculty.iitmandi.ac.in/∼arnav/sup_mat.pdf for more visual results at larger scale. (Please type the link manually.)

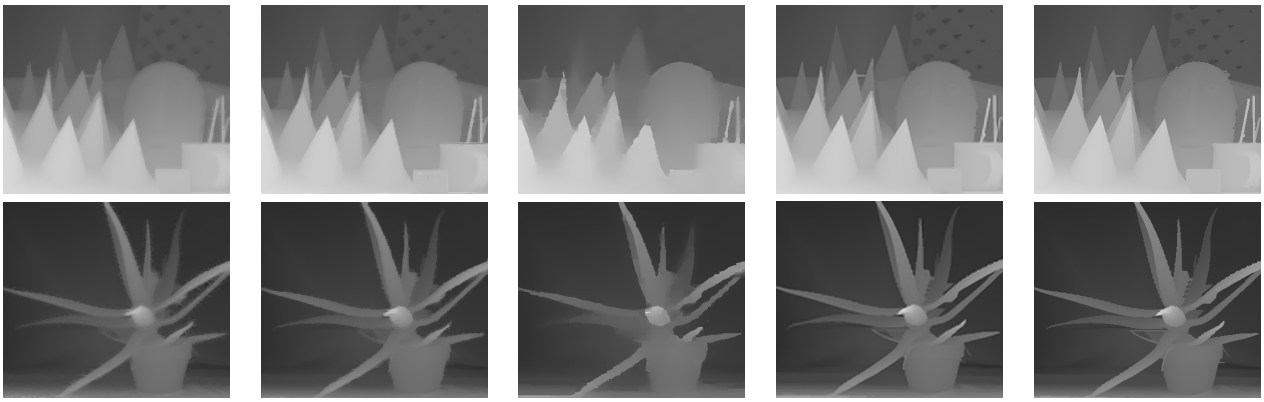

Fig. 2. SR by factor of 4: Rows 1-2 show the results for scenes Cones and Aloe, respectively. Columns 1-5 show the result on each scene for GIF [19], ATGV [2], EB [13], the proposed approach, and the original scene respectively.

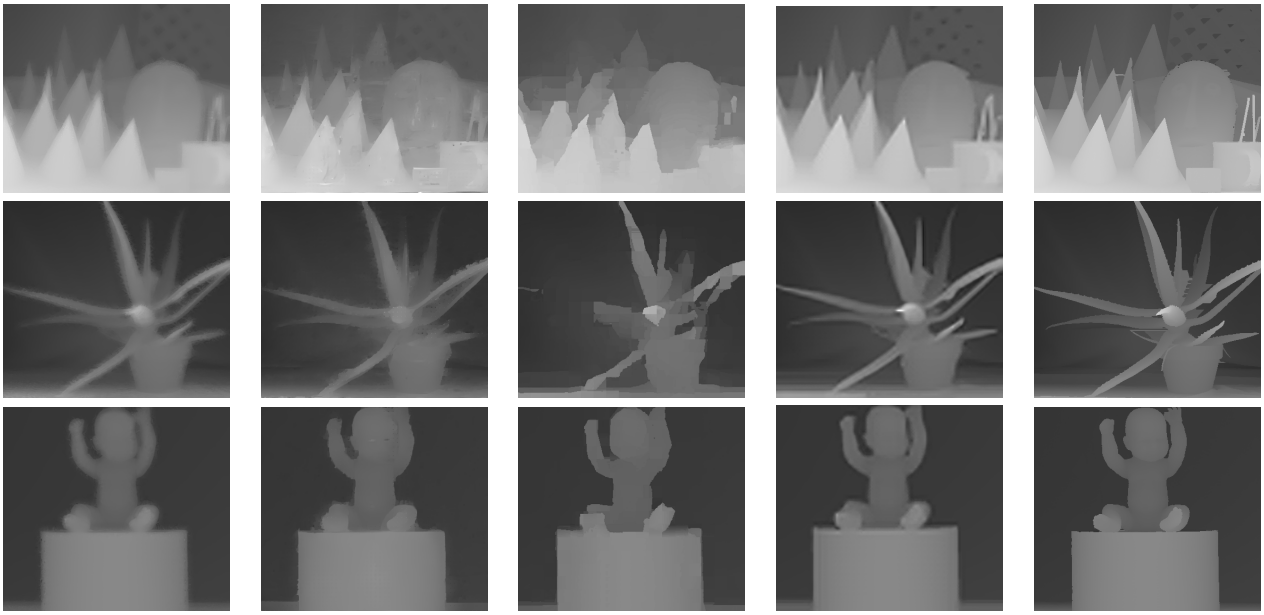

Fig. 3. SR by factor of 8: Rows 1-3 show the results for scenes Cones, Aloe and Baby, respectively. Columns 1-5 show the result on each scene for GIF [19], ATGV [2], EB [13], the proposed approach, and the original scene respectively.

Table 1. Quantitative results and comparisons

Images	Metric	Scale-2				Scale-4					Scale-8				
		GIF [19]	AGTV [2]	EB [13]	Ours	GIF [19]	AGTV [2]	EB [13]	Ours (NH)	Ours (H)	GIF [19]	AGTV [2]	EB [13]	Ours (NH)	Ours (H)
Cones	RMSE	3.62	2.82	4.08	2.13 (1)	4.28	4.16	5.88	4.20 (3)	3.73 (1)	6.32	6.99	9.39	6.45 (3)	6.23 (1)
	SSIM	0.9650	0.9788	0.9606	0.9870 (1)	0.9571	0.9580	0.9356	0.9560 (4)	0.9640 (1)	0.9353	0.9195	0.8963	0.9241 (3)	0.9252 (2)
Teddy	RMSE	2.70	2.19	3.18	1.74 (1)	2.30	2.98	4.53	3.11 (4)	2.86 (2)	4.27	4.96	7.00	4.89 (3)	4.73 (2)
	SSIM	0.9731	0.9825	0.9668	0.9890 (1)	0.9685	0.9680	0.9495	0.9657 (4)	0.9701 (1)	0.9531	0.9453	0.9196	0.9369 (4)	0.9387 (3)
Aloe	RMSE	5.43	4.15	4.93	2.89 (1)	6.09	6.00	7.29	5.68 (2)	5.12 (1)	9.04	10.94	12.84	8.67 (2)	8.57 (1)
	SSIM	0.9336	0.9670	0.9511	0.9826 (1)	0.9236	0.9350	0.9209	0.9345 (3)	0.9462 (1)	0.8906	0.8842	0.8441	0.8840 (4)	0.8846 (2)
Baby	RMSE	3.02	2.41	3.26	1.81 (1)	3.55	3.44	4.49	3.36 (2)	2.97 (1)	4.88	5.80	6.33	5.37 (3)	4.86 (1)
	SSIM	0.9768	0.9872	0.9803	0.9921 (1)	0.9713	0.9750	0.9659	0.9732 (3)	0.9786 (1)	0.9575	0.9575	0.9445	0.9465 (3)	0.9520 (2)
Venus	RMSE	2.66	1.53	1.92	0.98 (1)	2.75	2.75	1.89	1.95 (3)	1.67 (1)	3.30	4.85	3.80	2.99 (2)	2.63 (1)
	SSIM	0.9785	0.9928	0.9902	0.9967 (1)	0.9774	0.9777	0.9898	0.9875 (3)	0.9904 (1)	0.9739	0.9478	0.9712	0.9771 (2)	0.9798 (1)
Plastic	RMSE	2.12	1.64	3.16	1.81 (2)	2.42	2.39	3.31	2.68 (4)	2.63 (3)	3.76	4.16	5.33	5.00 (4)	4.55 (3)
	SSIM	0.9888	0.9928	0.9827	0.9932 (1)	0.9847	0.9843	0.9751	0.9826 (4)	0.9833 (3)	0.9738	0.9721	0.9633	0.9614 (4)	0.9613 (5)

5. REFERENCES

[1] D. Koller, M. Turitzin, and M. Levoy, "Protected interactive 3D graphics via remote rendering," *ACM SIGGRAPH 2004*, pp. 695-703, 2004.

[2] D. Ferst, C. Reinbacher, R. Ranft, M. Ruther, and H. Bischof, "Image guided depth up-sampling using anisotropic total generalized variation," *International Conference on Computer Vision (ICCV 2013)*, 2013.

[3] http://www.pmdtec.com/

[4] D. Scharstein and R. Szeliski. "High-accuracy stereo depth maps using structured light," *International Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, vol. 1, pp. 195-202, 2003

[5] J. Han, L. Shao, D. Xu, and J. Shotton "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318-1334, 2013.

[6] Q. Yang, R. Yang, J. Davis, D. Nister, "Spatial-depth super resolution for range-images," *International Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, pp. 1-8.

[7] Y. Yang, Z. Wang, "Range-image super-resolution via guided image filter," *International Conference on Internet Multimedia Computing and Service (ICIMCS 2012)*, pp. 200-203, 2012.

[8] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM SIGGRAPH 2007*, 2007

[9] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation", *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861 - 2873, 2010.

[10] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838-1857, 2011.

[11] S. Mandal and A. K. Sao, "Edge preserving single image super-resolution in sparse environment," *IEEE International Conference on Image Processing, (ICIP 2013)*, pp. 967-971, 2013.

[12] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction," *IEEE Transactions on Image Processing*, vol.18, no.1, pp.36,51, Jan. 2009.

[13] O. Mac Aodha, N. Campbell, A. Nair, and G.J. Brostow, "Patch based synthesis for single depth image super-resolution," *European Conference on Computer Vision (ECCV 2012)*, pp. 71-84, 2012.

[14] W.T. Freeman, T.R. Jones, and E.C. Pasztor, "Example-Based Super-Resolution", *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56-65, 2002.

[15] A.K. Sao, B. Yegnanarayana, and B.V.K. Vijaya Kumar, "Significance of image representation for face verification,", *Signal, Image and Video Processing*, vol. 1, pp. 225237, 2007.

[16] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 14131457, 2004.

[17] D. Scharstein and C. Pal. "Learning conditional random fields for stereo," *International Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, 2007.

[18] H. Hirschmller and D. Scharstein, "Evaluation of cost functions for stereo matching," *International Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, 2007.

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

[20] http://research.microsoft.com/en-us/um/people/kahe/eccv10/index.html

[21] http://rvlab.icg.tugraz.at/project_page/project_tofusion/project_tofsuperresolution.html

[22] http://visual.cs.ucl.ac.uk/pubs/depthSuperRes/

[23] Zhou Wang; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol.13, no.4, pp.600-612, 2004.