SINGLE DEPTH IMAGE SUPER RESOLUTION VIA A DUAL SPARSITY MODEL

Yulun Zhang[1]*, Yongbing Zhang*[1] *and Qionghai Dai*[1,2]

[1]Graduate School at Shenzhen, Tsinghua University, China
[2]Dept. of Automation, Tsinghua University, China

ABSTRACT

Depth images play an important role and are popularly used in many computer vision tasks recently. However, the limited resolution of the depth image has been hindering its further applications. To address this problem, we propose a novel dual sparsity model based single depth image super resolution algorithm, with a single low-resolution depth image as input. We formulate this problem by combining the recently developed analysis model and synthesis model exploiting the sparsity of analyzed vectors and the sparse coefficients respectively. The analysis operator and dictionaries are trained over extensive samples separately. We show that our model clearly outperforms state-of-the-art methods on the widely used Middlebury 2007 datasets both quantitatively and visually.

***Index Terms*—** Analysis model, depth image, dual sparsity model, super resolution, synthesis model

1. INTRODUCTION

Recently, the highly development of acquisition devices have been promoting the wide utilization of depth images in computer vision applications, such as image classification, object recognition, and interactive games. However, the huge unbalance between the resolution of depth image and that of its corresponding intensity image has been hindering its further applications. For example, the resolution of PMD CamCube 3.0 and SwissRange SR400 depth camera are only as low as about 200×200 [1]. The resolution of the well-known Microsoft Kinect is 640×480, which is much lower than that of its corresponding intensity image. To address this problem, we develop a sophisticated single depth image super resolution (SR) algorithm employing a single low-resolution (LR) depth image as input.

Depth image SR is a procedure where the resolution of a LR depth image is enhanced to reconstruct its high-resolution (HR) counterpart, while minimizing visual artifacts. The main drawback of depth image SR is that it is an ill-posed inverse problem, where the input LR depth image is usually corrupted by noise or bad pixels and lacks texture information due to the inherent smooth and plain property of depth image.

It is well known that the synthesis model [2] is a well-established tool to design the image reconstruction algorithms. Another interesting alternative is the analysis model [2], where an analysis operator multiplies the input signal and the output is expected to be sparse. Our proposed model is mainly based on these two fundamental models for the depth image SR, which has not witnessed the usage of the combination of synthesis and analysis models to the best of our knowledge.

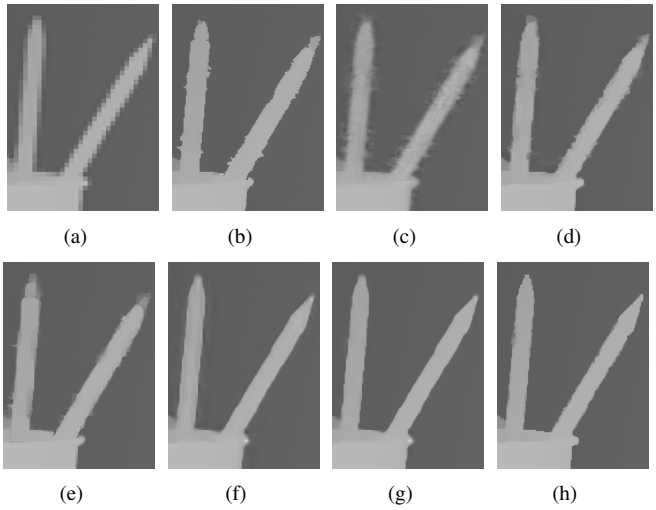

Fig. 1. Visual comparisons of $\times 4$ upsampling on a snippet of the Middlebury *Art*. (a)LR input depth image (enlarged using nearest neighbor interpolation). (b)JBFcv [3]. (c)Guided [4]. (d)Edge [5]. (e)TGV [6]. (f)ScSR [7]. (g)Ours. (h)Ground Truth. (Zoom in for better view)

The main contribution in this paper is that we conduct a novel dual sparsity model (DSM) for the depth image SR by combining both the sparse representation synthesis and analysis models. More specifically, we consider the sparsity of analysis operator and coefficients simultaneously, which can reconstruct the expected information better. In addition, we learned two classes of dictionaries: analytical dictionaries and learned dictionaries. Fig. 1 shows the depth image SR result by our method comparing with other state-of-the-art approaches.

The remainder of the paper is organized as follows: In Section 2, we give an overview of the related work. Section

3 presents our dual sparsity model based SR algorithm. In Section 4, the details of our overall algorithm are explained. We report the experimental results and conclude the paper in Section 5 and 6 successively.

2. RELATED WORK

There are numerous ways to increase the resolution of depth measurements. In general, they can be grouped in four main classes: (1) fusion of multiple depth sensors, (2) temporal and spatial fusion, (3) combination of depth and intensity sensors, (4) single depth image SR.

The typical works of fusing multiple depth sensors are proposed by Gudmundsson *et al.* [8] and Zhu *et al.* [9]. Gudmundsson *et al.* [8] presented a method for stereo and Time of Flight (ToF) depth image fusion in a dynamic programming approach. Zhu *et al.* [9] also conducted similar work by using an accurate depth calibration and then fusing the measurements under a Markov Random Field (MRF) framework. Due to its complexity and high-cost, consumers can hardly use this method to upsampling depth images directly.

Another fusion way of depth image SR is to fuse multiple depth measurements into one depth image. Schuon *et al.* [10] proposed a method to fuse ToF acquisitions from many slightly moved viewpoints. This work is further developed by Cui *et al.* [11], where a set of fused depth images with larger displacements was used. These works rely on the assumption that multiple range images are available with camera movement, which, in practical applicatons, may not always be true.

Additional intensity information can also be employed as a useful cue for depth image SR, which was first introduced by Diebel and Thrun [12], who performed the depth image SR using a MRF formulation, with a smoothness term according to the texture derivative in the corresponding intensity image. Yang *et al.* [3] used bilateral filtering of a depth cost volume and a RGB image in an iterative refinement way to perform depth image SR. Park *et al.* [5] proposed a more complex approach by using a combination of different weighting terms of a least squares optimization. In these works, it was claimed that the edges in RGB images can be helpful cues utilized for depth image SR, however, in common cases, it is very hard to obtain HR color images that is well registered with depth images.

Since the intensity information is not employed in single depth image SR, we can refer to the SR methods proposed in natural image SR, whose predominate methods are sparse representation ones. Yang *et al.* [7] reconstructed HR patches as a sparse linear combination of the learned dictionary atoms based on the assumption that low- and high-resolution patches share the same reconstruction coefficients. Aodha *et al.* [13] formulated the depth images SR problem as a multi-class MRF model with each hidden node representing the label of HR patches. Because the mapping relation from HR to LR patches is many to one, the classical learning based methods like sparse coding might suffer from the over-fitting problem.

Very recently, Nam *et al.* [14] proposed a cosparse analysis model and demonstrated its effectiveness in image reconstruction. Kiechle *et al.* [15] applied the analysis model to depth image SR, where a bimodal cosparse analysis operator is learned offline and then applied to reconstruct HR depth images.

Our work takes advantage of the sparse representation synthesis and analysis models simultaneously. More specifically, our DSM explores the sparsity of analysis vectors performed from analysis operator and the reconstructed HR depth images, together with sparse reconstruction. We demonstrate that our approach leads to better results compared to previously reported state-of-the-art methods.

3. PROPOSED APPROACH

In this section, we first introduce the traditional synthesis model, followed by representing the recently developed analysis model. Finally, DSM combining the advantage of the former two models is given in detail.

In the following discussion, we denote $\boldsymbol{I}_H$ and $\boldsymbol{I}_L$ as the HR and LR depth images, respectively, and $\boldsymbol{y}_H$ and $\boldsymbol{y}_L$ as the feature vectors of the HR and LR depth image features. $\boldsymbol{D}_H$ and $\boldsymbol{D}_L$ denote the dictionaries for HR and LR depth image patches in the synthesis model. Bold lowercase letters denote vectors. Bold uppercase letters denote regular matrices, i.e. **S** is used as a downsampling operation in matrix form. Plain lowercase letters are used as scalars.

3.1. Synthesis Model and Coupled Dictionary Learning

To derive the mapping function between paired low- and high-resolution depth image features, we learn a coupled dictionary by minimizing the objective function with a sparsity constraint and forcing the HR and LR features to share the same reconstruction coefficients:

$$\begin{aligned} &\min_{\boldsymbol{D},\boldsymbol{X}} \sum_i (\left\|\boldsymbol{y}^i - \boldsymbol{D}\cdot\boldsymbol{x}_i\right\|_2^2 + \lambda \left|\left|\boldsymbol{x}_i\right|\right|_1) \\ &s.t. \left\|\boldsymbol{D}(:,j)\right\|_2^2 \leq 1, j = 1, ..., d, \end{aligned} \tag{1}$$

where $\boldsymbol{y}^i = \begin{bmatrix} \boldsymbol{y}_H^i \\ \boldsymbol{y}_L^i \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_H \\ \boldsymbol{D}_L \end{bmatrix}$ and $\boldsymbol{x}_i$ are the features , dictionaries and coefficient vector of the *i*-th high- and low- resolution feature pairs, respectively. The coefficient matrix $\boldsymbol{X}$ contains the coefficient vectors of all feature pairs, λ is a weighting constant. In the formula above, the first term is a reconstruction constraint to promise the accuracy of reconstruction. The second term is a sparsity constraint.

Traditionally, provided two dictionaries $\boldsymbol{D}_H$ and $\boldsymbol{D}_L$ are learned, we can obtain an estimation of the HR image feature

$\boldsymbol{y}_H$ from the LR image feature $\boldsymbol{y}_L$ by first solving:

$$\min_{\boldsymbol{x}} \left\|\boldsymbol{y}_L - \boldsymbol{D}_L \cdot \boldsymbol{x}\right\|_2^2 + \lambda \left|\left|\boldsymbol{x}\right|\right|_1, \tag{2}$$

and afterwards synthesizing $\boldsymbol{y}_H$ from the computed sparse coefficient via:

$$\boldsymbol{y}_H = \boldsymbol{D}_H \cdot \boldsymbol{x}. \tag{3}$$

3.2. Analysis Model

An alternative to the synthesis model described above for reconstructing $\boldsymbol{y}_H$, is to solve:

$$\begin{aligned} &\min_{\boldsymbol{y}_H} g(\boldsymbol{\Omega}\boldsymbol{y}_H) \\ &s.t.\ \left\|\boldsymbol{y}_L - S \cdot \boldsymbol{y}_H\right\|_2^2 \leq \epsilon, \end{aligned} \tag{4}$$

which is known as the *analysis model* [2]. where, $\mathbf{S} \in \mathbb{R}^{m \times n}$ is the measurement matrix modeling the sampling process. $\boldsymbol{\Omega} \in \mathbb{R}^{k \times n}$ with $k \geq n$ is called the analysis operator, and the analyzed vector $\boldsymbol{\Omega}\boldsymbol{y}_H \in \mathbb{R}^k$ is expected to be sparse, whose sparsity can be measured via an appropriate function g. Unlike the synthesis model, where $\boldsymbol{y}_H$ is fully described by the non-zero elements of $\boldsymbol{x}$, in the analysis model the zero elements of the analyzed vector $\boldsymbol{\Omega}\boldsymbol{y}_H$ describe the subspace containing $\boldsymbol{y}_H$.

In depth image SR problems, different analysis operators are used in literatures such as the Wavelet transforms, the finite difference operators, and the learned operators. In our work, we will learn an analysis operator offline.

3.3. Dual Sparsity Model (DSM)

We propose an improved reconstruction model by simultaneously considering the sparsity of sparse coefficient $\boldsymbol{x}$ and analyzed vector $\boldsymbol{\Omega}\boldsymbol{y}_H \in \mathbb{R}^k$ after we learned the coupled dictionary $\begin{bmatrix}\boldsymbol{D}_H \\ \boldsymbol{D}_L\end{bmatrix}$ and analysis operator $\boldsymbol{\Omega}$. The proposed method named as dual sparsity model (DSM) is formulated as:

$$\min_{\boldsymbol{x}} \left\|\boldsymbol{y}_L - \boldsymbol{D}_L \cdot \boldsymbol{x}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\Omega} \cdot \boldsymbol{D}_H \cdot \boldsymbol{x}\right\|_2^2 + \lambda_2 \left\|\boldsymbol{x}\right\|_1, \tag{5}$$

where λ_1 and λ_2 are weighting parameters. Then the HR feature $\boldsymbol{y}_H$ can be recovered according to equation (3).

In our DSM, the coefficient $\boldsymbol{x}$ is expected to be sparse, the sparsity of which is measured by L_1 norm. As discussed in Section 3.2, analysis vector $\boldsymbol{\Omega} \cdot \boldsymbol{D}_H \cdot \boldsymbol{x}$ will also be sparse and we measure its sparsity by L_2 norm. As a consequence, the coefficient $\boldsymbol{x}$ solved by formulation (5) will be better than that solved by formulation (1) by considering the sparsity of both analysis vectors and coefficients.

4. IMPLEMENTATION DETAILS

To evaluate our approach we train fixed dictionaries and analysis operator from Middlebury 2006 stereo datasets [16].

4.1. Learning Dictionary and analysis operator

In the dictionary learning stage, we first downsample the HR depth images $\boldsymbol{I}_H$ by a factor 2 to get the LR depth images $\boldsymbol{I}_L$, and then upsample them back to the original resolution using bicubic interpolation. We randomly select 30000 patch pairs from $\boldsymbol{I}_H$ and $\boldsymbol{I}_L$, where we generate HR depth image features $\boldsymbol{y}_H$ by removing the mean of each HR patches and extract LR depth image features $\boldsymbol{y}_L$ by computing the 1st and 2nd-order derivatives from the LR patches along both horizontal and vertical directions, and then we stack the features of HR and LR patches into vectors $\begin{bmatrix}\boldsymbol{y}_H \\ \boldsymbol{y}_L\end{bmatrix}$. In the experiment, we set the patch size to be 5, and consequently the dimension of $\boldsymbol{y}_H$ and $\boldsymbol{y}_L$ is 25 and 100 respectively. The above dictionary learning formulation (1) is solved by an iterative optimization method [17], and the learned HR and LR dictionaries contain 1024 atoms. A smaller dictionary containing 512 atoms is displayed in Fig. 2 as an example.

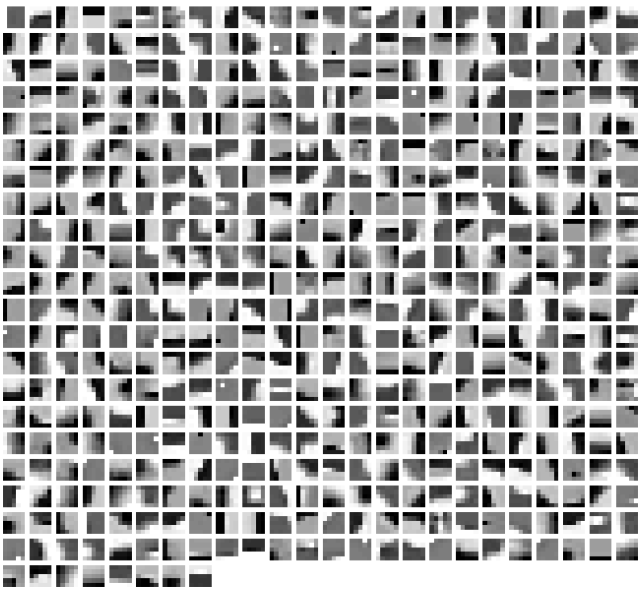

Fig. 2. HR depth image features dictionary using 30000 HR and LR depth patch pairs randomly sampled from the Middlebury 2006 stereo datasets. Totally 512 atoms are learned with each atom of size 5×5.

We use the same training set above to learn the analysis operator $\boldsymbol{\Omega}$ by GOAL [14] with the default parameters in the released code. In general, the larger value of k we choose, the better reconstruction quality can be obtained with the analysis operator. Twofold redundancy provides a good balance between reconstruction quality and computation time. Therefore, we set $k = 2n$ throughout this paper. As an example, a learned analysis operator $\boldsymbol{\Omega} \in \mathbb{R}^{100 \times 25}$ is shown in Fig. 3.

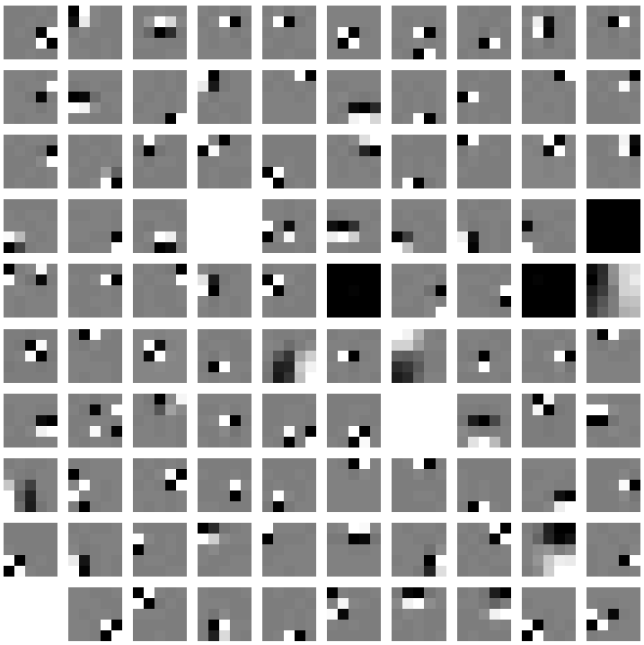

Fig. 3. Learned atoms of an analysis operator $\mathbf{\Omega} \in \mathbb{R}^{100\times 25}$. Each of the 100 atoms is represented as a 5×5 square, where black corresponds to the smallest negative entry, gray is a zero entry, and white corresponds to the largest positive entry.

4.2. Depth Image SR via DSM

After obtaining the dictionaries and analysis operator, we first get the sparse coefficients from the formulation (5) and then recover the HR depth image features by equation (3). To remove the blocking effect, we divide the HR images into overlapping patches and reconstruct the HR depth image patches by adding it with the mean of corresponding LR depth image patches and averaging the depth values over the overlapping regions. Because we recover the HR depth features patch by patch, ringing effect will arise along the edges, we use a bilateral filter (BF) as a post-processing to remove the ringing effect in the reconstructed depth images.

5. EXPERIMENTAL RESULTS

To demonstrate the superiority of the proposed DSM, the clean Milldlebury 2007 datasets [16] provided by Ferstl [6] are selected to conduct extensive experiments by DSM and numerous state-of-the-art competing methods. We compare our results with learning based method [7], and previous methods [12], [3], [4], [5], [6] and [18] quantitatively and visually. It should be noted that we compare to [18] by using the results from the project website of Yang *et al.* [18], and its resolution is different from that provided by Ferstl [6]. As a result, when we compare our results with that of [18], we would not provide the visual comparisons, but only the quantitative comparisons.

We first downsample the test HR depth images by a scaling factor of 2, 4, 8 and 16, and then restore them to have the original resolution with DSM. We set $\lambda_1 = 0.09$ and $\lambda_2 = 0.2$ throughout our experiment.

5.1. Quantitative Results

In this subsection, we compare our RMSE results with learning based method [7], and previous methods [12], [3], [4], [5], [6] and [18] from the clean dataset provided by Ferstl [6] using a scaling factor of 2, 4, 8 and 16 in Table 1. In addition, we provide our DSM results without using the bilateral filter for a fair comparison with ScSR [7].

The best result for each dataset and SR factor is highlighted and the second best is underlined. What can be clearly seen from the numerical results is that our approach is superior compared to other state-of-the-art methods in terms of RMSE. It should be noted that both ScSR [7] and the proposed DSM take only single depth image as input, whereas all the other competing methods [12], [3], [4], [5], [6] and [18] employ registered intensity information as the guided cue for depth image SR. Although [12], [3], [4], [5], [6] and [18] take more information into account, the performance is defeated by our DSM, except for *Art* with scaling factor of 8 and 16 , where AP [18] has the lowest RMSE value. This means that considering extra intensity information may not promise better performance compared with single depth image SR methods. Using the synthesis model, Yang *et al.* [7] exploited the sparsity of the coefficients while minimizing the reconstruction error. ScSR [7] was able to achieve lower RMSE than that of the majority of [12], [3], [4], [5], [6] and [18] with the scaling factor of 2, but it was not better than our DSM whether we add bilateral filter or not when the scaling factor is larger than 2. The newly developed depth image SR method AP [18] achieved the best performance when upsampling the art depth image by a scaling factor of 8 and 16, however, AP [18] had to adjust many parameters in depth SR with high complexity. By combining analysis and synthesis models, DSM shows superiors quantitative performance under the most cases whether bilateral filter is used or not, which demonstrates the robustness, effectiveness and accuracy of our DSM.

Since the depth image lacks textures and its edges contain the key information for depth SR problem, it is well-founded for us to further improve the quality of the edges by moving the ringing effects with the bilateral filter as a post-processing procedure. Furthermore, with the proper use of bilateral filter, our results can actually be further improved both in quantitatively and visually.

5.2. Visual Results

To better demonstrate the effectiveness of our DSM, we compare our visual results with other competing methods with the scaling factor set to be 4 in Fig. 1 and 8 in Fig. 4. Obvious spurs or blurring artifacts can be observed in Fig. 1 (b)-(e).

Table 1. RMSE comparisons on Middlebury 2007 datasets without noise. RMSE is measured for four different magnification factors ($\times 2$, $\times 4$, $\times 8$, $\times 16$). The best result for each dataset and SR factor is highlighted and the second best is underlined.

	Art				*Books*				*Moebius*			
	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$	$\times 2$	$\times 4$	$\times 8$	$\times 16$
Nearest	3.122	4.795	6.920	9.906	1.205	1.861	2.639	4.150	1.102	1.744	2.644	3.788
Bilinear	2.834	4.147	5.995	8.928	1.119	1.673	2.394	3.525	1.016	1.499	2.198	3.179
MRF [12]	3.119	3.794	5.503	8.657	1.205	1.546	2.209	3.400	1.187	1.439	2.054	3.078
JBFcv [3]	4.066	4.056	4.712	8.268	1.615	1.701	1.949	3.325	1.069	1.386	1.820	2.494
Guided [4]	2.934	3.788	4.974	7.876	1.162	1.572	2.097	3.186	1.095	1.434	1.878	2.851
Edge [5]	2.833	3.498	4.174	6.262	1.195	1.495	1.978	2.947	1.064	1.349	1.803	2.377
TGV [6]	3.032	3.785	4.787	7.102	1.290	1.603	1.991	2.941	1.129	1.459	1.914	2.630
ScSR [7]	1.452	2.517	4.318	7.527	0.604	1.193	1.851	2.717	0.603	1.040	1.606	2.595
AP [18]	1.791	2.851	**3.694**	**5.911**	1.328	1.530	1.839	2.919	0.870	1.031	1.550	2.573
Ours (w/o BF)	1.492	2.345	3.935	7.095	0.618	1.113	1.737	2.558	0.611	0.974	1.531	2.453
Ours (with BF)	**1.064**	**2.117**	3.851	6.884	**0.544**	**1.037**	**1.686**	**2.468**	**0.536**	**0.902**	**1.469**	**2.311**

This is mainly because these methods utilize registered intensity as guided information. For the regions with different depth value while their intensity value are very similar, the results of methods [3], [4], [5] and [6] are prone to be influenced by the property of the intensity values, which lead to serious spurs. In Fig. 1 (f), strong blurring artifacts can be perceived, since ScSR [7] takes the sparsity of the coefficients into consideration only. However, in Fig. 1 (g), sharp edges can be found and exhibits best visual quality compared with the ground truth in Fig. 1 (h). This is because the blurring effects can be greatly removed by considering both synthesis and analysis models in our DSM.

In Fig. 4, the same problem of spurs or blurring effects happened according to the results of all of the other competing methods [12], [3], [4], [5] , [6] and [7]. Furthermore, as the scaling factor increases from 4 to 8, the results of TGV [6] suffer from the texture-copying problem, especially in the regions where the depth or intensity value changes a lot. While, as a single depth SR method, our DSM is free of texture in intensity image and achieves sharper edges and less artifacts as well.

In our experiments, when the scaling factor is set to be 2 and 16, visual comparisons are not presented here due to space limitation. For all scaling factors, our DSM behaves very well and outperforms other competing methods. All of these visual comparisons can further demonstrate the excellent performance of our DSM with sharper edges, clearer structures and less artifacts.

6. CONCLUSION

In this paper, we proposed a novel Dual Sparsity Model (DSM) based single depth image SR algorithm, which takes the sparse representation synthesis and analysis models into consideration simultaneously. Experimental results on the Middlebury datasets show that our new model outperforms state-of-art algorithms both quantitatively and visually.

In the future, we would like to further improve the speed, robustness and arruracy of our DSM by modifying the formulation (1) and (5). We also believe that additional information like registered intensity would further improve the performance of depth image SR in a proper way. In addition to depth image SR, our DSM can be extended to many computer vision and image processing tasks, such as face recognition, image super resolution, denoising, deblurring, inpainting and image classification.

7. ACKNOWLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China under Grant 61170195, U1201255&U1301257.

8. REFERENCES

[1] J. Xie, C.C. Chou, R. S. Feris, and M.T. Sun, "Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering," in *ICME*, 2014.

[2] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, pp. 947, 2007.

[3] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *CVPR*, 2007.

[4] K. He, J. Sun, and X. Tang, "Guided image filtering," in *ECCV*, 2010.

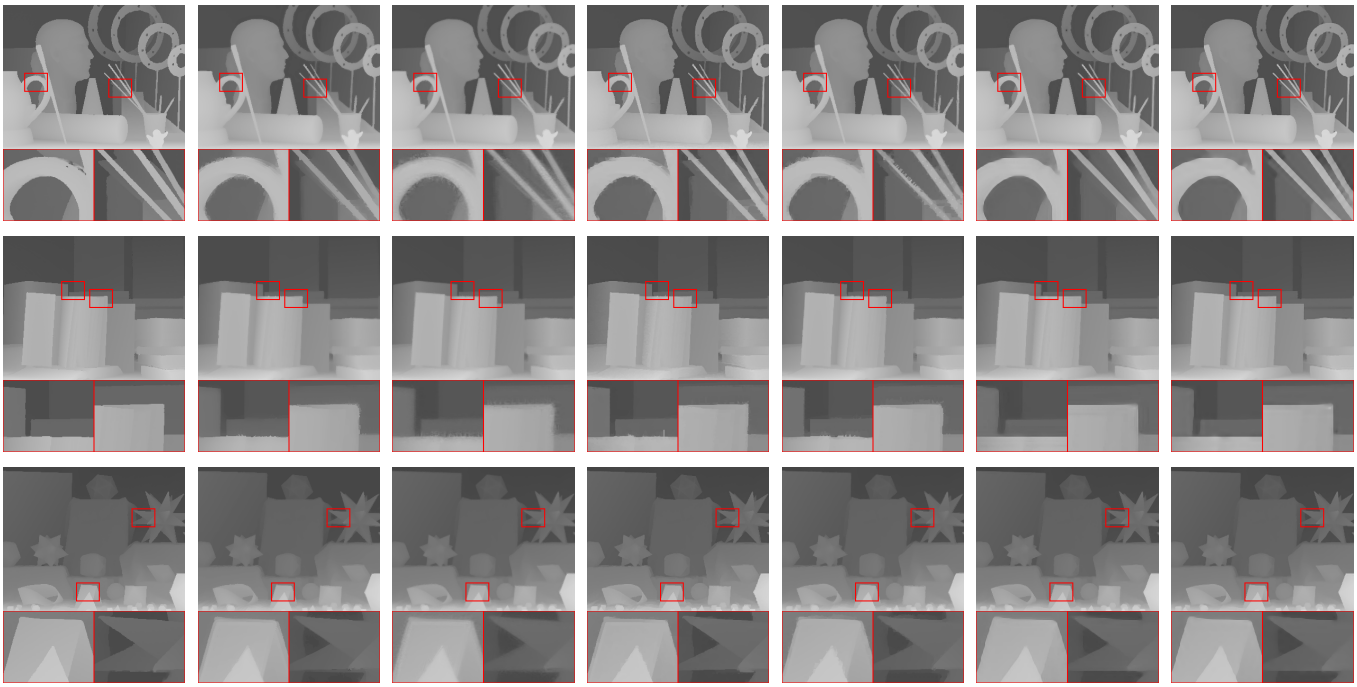

Fig. 4. Visual quality comparisons of $\times 8$ upsampling on the Middlebury *Art* , *Books* and *Moebius* datasets. Column 1: Ground Truth. Column 2: JBFcv [3]. Column 3: Guided [4]. Column 4: Edge [5]. Column 5: TGV [6]. Column 6: ScSR [7]. Column 7: Ours. (Zoom in for better view.)

[5] J. Park, H. Kim, Y.W. Tai, M. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *ICCV*, 2011.

[6] D. Ferstl, C. Reinbacher, and R. Ranftl, "Image guided depth upsampling using anisotropic total generalized variation," in *ICCV*, 2013.

[7] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *TIP*, vol. 19, no. 11, pp. 2861–2873, 2010.

[8] S. A. Gudmundsson, H. Aanaes, and R. Larsen, "Fusion of stereo vision and time-of-flight imaging for improved 3d estimation," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3, pp. 425–433, 2008.

[9] J. Zhu, L. Wang, R. Yang, J. Davis, and Z. Pan, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *PAMI*, vol. 33, no. 7, pp. 1400–1414, 2011.

[10] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for tof 3d shape scanning," in *CVPR*, 2009.

[11] Y. Cui, S. Schuon, S. Thrun D. Chan, and C. Theobalt, "3d shape scanning with a time-of-flight camera," in *CVPR*, 2010.

[12] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *NIPS*, 2005.

[13] O. M. Aodha, N. D. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *ECCV*, 2012.

[14] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[15] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," in *ICCV*, 2013.

[16] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *CVPR*, 2007.

[17] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in *NIPS*, 2006.

[18] J.Y. Yang, X.C. Ye, K. Li, C.P. Hou, and Y. Wang, "Color-guided depth recovery from rgb-d data using an adaptive autoregressive model," *TIP*, vol. 23, no. 8, pp. 3443–3458, 2014.