DEPTH MAP INPAINTING AND SUPER-RESOLUTION BASED ON INTERNAL STATISTICS OF GEOMETRY AND APPEARANCE

Satoshi Ikehata†, Ji-Ho Cho‡ and Kiyoharu Aizawa†

†The University of Tokyo, ‡Vienna University of Technology

ABSTRACT

Depth maps captured by multiple sensors often suffer from poor resolution and missing pixels caused by low reflectivity and occlusions in the scene. To address these problems, we propose a combined framework of patch-based inpainting and super-resolution. Unlike previous works, which relied solely on depth information, we explicitly take advantage of the internal statistics of a depth map and a registered high-resolution texture image that capture the same scene. We account these statistics to locate non-local patches for hole filling and constrain the sparse coding-based super-resolution problem. Extensive evaluations are performed and show the state-of-the-art performance when using real-world datasets.

Index Terms— depth-map super-resolution, depth-map inpainting, ToF sensor, sparse Bayesian learning

1. INTRODUCTION

Dynamic 3D scene geometry acquisition methods are a major research topic in image processing and computer vision. Various depth acquisition methods have been proposed to date; these can generally be categorized into two approaches: passive and active. While passive approaches generate a disparity map of the scene by multiple image correspondences and triangulation [1, 2], active sensors, *e.g.*, time-of-flight (ToF) cameras, measure the distance from the camera to the objects directly using active infrared illumination. In reality, however, depth maps acquired by these sensors often contain unreliable areas where the scene contains occluded regions or objects with low reflectivity. Also, a depth map captured by a ToF camera has very low-resolution (*e.g.*, 176×144 in Swiss Ranger SR4000). Some extra image processing steps are therefore required: depth map inpainting and super-resolution.

While image inpainting methods have been studied [3, 4, 5], there is very little available literature on depth-map inpainting [6]. To the best of our knowledge, there is no work which uses a registered high-resolution texture image for depth-map inpainting. In contrast, many works on depth-map super-resolution exist, and they are mainly categorized into two classes. One class determines depth values at interpolated coordinates of the input domain in the manner of local filtering [7, 8, 9] or Markov random field (MRF) modeling [10, 11], which uses a registered high-resolution texture image to preserve sharp boundaries. While effective, this approach is critically sensitive to the texture image quality. The second class is motivated by image super-resolution literature, which explicitly considers the low-resolution image generation process. Some algorithms in this class reconstruct a high-resolution depth map of a static scene by fusing multiple low-resolution depth maps that were observed together [12, 13]. More recently, learning-based single image super-resolution techniques were integrated into depth-map super-resolution to handle dynamic scenes. Aodha *et al*. [14] learned the statistics of low-resolution and high-resolution depth patches from external training sets and used them to upsample a new low-resolution depth map in the MRF formation. In contrast, Li *et al*. [15] derived a mapping function from a low-resolution depth patch and its corresponding color patch to a high-resolution depth patch via sparse coding. However, they require a large number of external clean training sets and time-consuming dictionary learning for each magnification factor. Also, the external statistics merely capture the co-occurring structures of depth and texture because each scene in the training set is completely different.

In this paper, we propose a new method for depth map inpainting and super-resolution which can produce a dense high-resolution depth map from a corrupted low-resolution depth map and its corresponding high-resolution texture image. Both steps use the internal statistics of the input geometry and appearance (*i.e.*, recurrence of the co-occurring structures of depth/texture patches), which are learned directly from the input depth/texture images. We account for the registered texture image to locate non-local patches for hole filling and constrain the sparse coding problem with patch-based super-resolution.

2. PROPOSED METHOD

In this section, we formulate a combined framework for depth-map inpainting and super-resolution. Let D_{l_0} and I_{l_1} be a low-resolution depth map and a registered high-resolution texture image, and let D_{l_1} be a high-resolution

This work was partially supported by the Grants-in-Aid for JSPS Fellows (No. 248615) and the Austrian Science Fund (FWF, M1383-N23).

depth map. We denote the hole and source regions in D_{l_0} as H and $S(\triangleq D_{l_0} \setminus H)$. The problem is how to reconstruct D_{l_1} from D_{l_0} and I_{l_1} while filling H in D_{l_0}. We solve these problems separately under these assumptions: (1) The relative position between the color and depth cameras is known; (2) The missing pixel region in the input depth map is known.

2.1. Patch-based depth map inpainting using internal statistics of geometry and appearance

We extract two different sets, V, W, of the overlapped $p \times p$ depth patches $\boldsymbol{x}_{Si} \in \mathbb{R}^{p\times p}, \boldsymbol{x}_{Hi} \in \mathbb{R}^{p\times p}$ from D_{l_0} as

$$V \triangleq \{\boldsymbol{x}_{Si} | i = 1, \cdots, n, \boldsymbol{x}_{Si} \in S\}, \quad (1)$$
$$W \triangleq \{\boldsymbol{x}_{Hi} | i = 1, \cdots, m, \{\boldsymbol{x}_{Hi} \cap S\} \neq \emptyset, \{\boldsymbol{x}_{Hi} \cap H\} \neq \emptyset\}.$$

In general patch-based image inpainting methods [3], hole regions in $\boldsymbol{x}_{Hi}$ are restored iteratively by example-based synthesis where a missing structure is propagated from a source patch ($\boldsymbol{x}_{Si}$) with a partially similar structure. Once one patch is inpainted, it is used as a new source, and all inpainted patches are finally merged into a completely reconstructed depth map. While effective, this propagation scheme is sensitive to the filling order, which is inevitable when the missing region is larger than a patch size. To overcome this difficulty, our method utilizes another source, *i.e*. the texture image.

We divide the low-resolution texture image (I_{l_0}), which is downsampled from I_{l_1}, into overlapped patches $\boldsymbol{y}_{Si} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{y}_{Hi} \in \mathbb{R}^{p\times p}$ with spatial positions that are the same as $\boldsymbol{x}_{Si}$ and $\boldsymbol{x}_{Hi}$. $\boldsymbol{x}_i \in \{\boldsymbol{x}_{Si}, \boldsymbol{x}_{Hi}\}$ and $\boldsymbol{y}_i \in \{\boldsymbol{y}_{Si}, \boldsymbol{y}_{Hi}\}$ represent the same scene in the different domains (*i.e*., in geometry and appearance), and thus their structures are likely to be correlated with each other. We observe that this relationship tends to recur again in the close vicinity of the same images, which implies that if the structures of two sufficiently close depth patches are similar, then corresponding texture patches are also similar and vice versa (*i.e*., $\boldsymbol{y}_i \simeq \boldsymbol{y}_j \leftrightarrow \boldsymbol{x}_i \simeq \boldsymbol{x}_j$). We confirmed these statistics by simple inpainting experiments shown in Fig. 1, where missing regions in a depth patch were synthesized from the nearest neighbor depth patch, where the distance was defined by depth or corresponding texture information. We found that appropriate depth patches could be retrieved even in depth-less areas using texture information; however, texture information did not contribute when the patch was extracted from a distant region.

Following observations above, we find a pair of $\boldsymbol{x}_{Hi} \in W$ and $\boldsymbol{x}_{Sj} \in V$ with the smallest depth/texture combined distance defined as follow,

$$f(\boldsymbol{x}_{Hi}, \boldsymbol{x}_{Sj}) = w_D d_D(\boldsymbol{x}_{Hi}, \boldsymbol{x}_{Sj}) + w_I d_I(\boldsymbol{y}_{Hi}, \boldsymbol{y}_{Sj}), \quad (2)$$

where $\boldsymbol{y}_{Hi}$ and $\boldsymbol{y}_{Sj}$ are extracted at same positions of $\boldsymbol{x}_{Hi}$ and $\boldsymbol{x}_{Sj}$, and d_D and d_I are defined as,

$$d_D(\boldsymbol{x}_{Hi}, \boldsymbol{x}_{Sj}) \triangleq \frac{1}{p^2 - N_H} \|M(\boldsymbol{x}_{Hi} - \boldsymbol{x}_{Sj})\|_2^2, \quad (3)$$
$$d_I(\boldsymbol{y}_{Hi}, \boldsymbol{y}_{Sj}) \triangleq \frac{1}{p^2} \|\boldsymbol{y}_{Hi} - \boldsymbol{y}_{Sj}\|_2^2. \quad (4)$$

Here, N_H is the number of missing pixels in $\boldsymbol{x}_{Hi}$, and M is a sampling matrix which extracts $\boldsymbol{x}_{Hi} \cap S$. The weighting factors w_D and w_I reasonably quantify the trustworthiness of the depth/texture information as follows:

$$w_D = \left(1 - \frac{N_H}{p^2}\right)\left(1 - \frac{N_C}{N_{C_{max}}}\right), \quad (5)$$
$$w_I = e^{-\|\boldsymbol{y}_{hc} - \boldsymbol{y}_{sc}\|_2^2 / 2\sigma_y}. \quad (6)$$

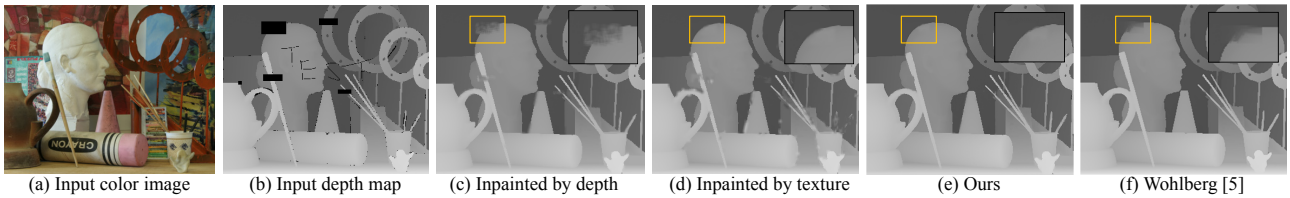
Here σ_y is a variance in the Gaussian function (4.0 in our implementation), N_c counts the connections of holes among four neighbors and $N_{C_{max}}$ is the maximum possible number of connections. w_D is designed from our observations such that the depth patch is informative when the number of holes is small and the holes are sparsely distributed rather than concentrated locally. $\boldsymbol{y}_{hc}$ and $\boldsymbol{y}_{sc}$ are the center positions of $\boldsymbol{y}_{Hi}$ and $\boldsymbol{y}_{Sj}$. w_I reflects our observation that the correlation between depth and texture decreases as the spatial distance between $\boldsymbol{y}_{Hi}$ and $\boldsymbol{y}_{Sj}$ increases. Once every $\boldsymbol{y}_{Hi}$ find their corresponding $\boldsymbol{y}_{Sj}$, hole regions are synthesized by overlapped nearest neighbor source patches via a voting-based approach [16] to acquire a final purified depth map. After the low-resolution depth map is restored, super-resolution is applied to it, and is discussed in the next section.

2.2. Patch-based super-resolution using internal statistics of geometry and appearance

A recent study of the internal statistics of a natural image [17] stated that most small image patches in a natural image tend to recur redundantly inside the image across different scales. This observation forms the basis of the internal example-based single image super-resolution approach [18, 17], and we also learn a set of high-resolution/low-resolution pairs of depth patches implicitly by using patch repetitions across multiple image scales,

$$X \triangleq \{(\boldsymbol{x}_i, \boldsymbol{y}_i) | i = 1, \cdots, n, \boldsymbol{x}_i \in D_{l_{-1}}, \boldsymbol{y}_i \in D_{l_0}\}. \quad (7)$$

Here $\boldsymbol{y}_j \in \mathbb{R}^{h^2\times 1}$ and $\boldsymbol{x}_i \in \mathbb{R}^{l^2\times 1}$ are vector representations of all possible $h \times h$ and $l \times l$ overlapped depth patches that are extracted from the same position of D_{l_0} and $D_{l_{-1}}$, which is a depth map downsampled from D_{l_0} by the magnification factor r ($= h/l$). Let $\boldsymbol{p}$ denote a $l \times l$ depth patch in D_{l_0} to be upsampled. We can search for similar patches from the learned patches ($\boldsymbol{x}_i, i = 1, \cdots, k$) in $D_{l_{-1}}$ (*e.g*., using a k-approximate nearest neighbor search [19]) and their corresponding high-resolution patches ($\boldsymbol{y}_i, i = 1, \cdots, k$) provide a strong prior indication of the appearance of a high-resolution unknown patch ($\boldsymbol{y}^*$) in D_{l_1}. However, simply averaging these high-resolution patches ($\boldsymbol{y}_i$) to recover an overall estimate of ($\boldsymbol{y}^*$) like [17] is problematic for depth-map super-resolution because of its blurring effect (which stems from multiple inconsistent high-resolution interpretations). However, only using the best patch reconstructs a false high-resolution edge



(a) Input color image (b) Input depth map (c) Inpainted by depth (d) Inpainted by texture (e) Ours (f) Wohlberg [5]

Fig. 1. Experimental results of patch-based depth map inpainting. Hole regions were synthesized by the nearest neighbor patches in the same image, which were found by using (c) depth information only, (d) texture information only, (e) both depth and texture information (proposed method). And we show the inpainted result by the method of Wohlverg [5] in (f).

when there is no ideally similar low-resolution patch in X. To overcome these issues, we instead solve the following problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{l^2}\|\boldsymbol{p}-\phi_L\boldsymbol{\alpha}\|_2^2+\frac{\mu}{h^2}\|\boldsymbol{q}-\psi_H\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_0. \quad (8)$$

Here, $\phi_L \in \mathbb{R}^{l^2\times k}$ is a matrix which consists of $\boldsymbol{p}$'s k-nearest neighbors (*i.e.*, $\phi_L \triangleq [\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k]$) and $\psi_H \in \mathbb{R}^{h^2\times k}$ is a matrix consisting of vectorized $h\times h$ texture patches in I_{l_1} with the same location of each $\boldsymbol{x}_i (i=1,\cdots,k)$. $\boldsymbol{q}\in\mathbb{R}^{h^2\times 1}$ is a high-resolution texture patch in I_{l_1} with the same location as $\boldsymbol{p}$. $\boldsymbol{\alpha}\in\mathbb{R}^{k\times 1}$ is a coefficient vector and μ and λ are trade-off parameters which are fixed as 0.05 and 10^{-6} in our implementation, respectively. $\|\cdot\|_0$ is an ℓ_0-norm penalty, which counts the number of non-zero entries in the matrix.

The process contains two essential elements. First, we introduce a sparsity penalty which is applied to $\boldsymbol{\alpha}$ to reconstruct a high-resolution depth patch with the smallest possible combination of learned patches. This penalty avoids the visual artifacts caused by multiple inconsistent high-resolution interpretations, especially in the case of large k. Second, we introduce a *texture constraint* which quantifies the observed internal statistics that the geometry and appearance of the same scene should be correlated with each other. Intuitively, the coefficients ($\boldsymbol{\alpha}$) are recovered such that the target depth patch ($\boldsymbol{p}$) is synthesized from the sparsest possible combination of learned depth patches ($\phi_L\boldsymbol{\alpha}$) and the reconstructed high-resolution texture patch ($\psi_H\boldsymbol{\alpha}$) is consistent with the registered high-resolution texture image (I_{l_1}). Then, Eq. (8) can be transformed into the standard description of sparse coding as,

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{b}-A\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_0, \quad (9)$$

where $\boldsymbol{b}\triangleq[\boldsymbol{p}/l,\sqrt{\mu}\boldsymbol{q}/h]^T$ and $A\triangleq[\phi_L/l,\sqrt{\mu}\psi_H/h]^T$. Eq. (9) entails a difficult combinatorial optimization problem that must be solved efficiently at every patch. In most super-resolution literature, it is common to replace the discontinuous non-convex ℓ_0 norm with the convex surrogate ℓ_1 norm. However, we instead apply a hierarchical Bayesian approximation called sparse Bayesian learning (SBL) to estimate $\boldsymbol{\alpha}$ which has been proved to give a better approximation of ℓ_0 norm [20].

2.3. Recovery of α and merging of high-res patches

SBL [21] assumes a standard Gaussian likelihood function for the first-level diffuse errors giving $p(\boldsymbol{b}|\boldsymbol{\alpha})=N(\boldsymbol{b};A\boldsymbol{\alpha},\lambda I)$, where $I\in\mathbb{R}^{(h^2+l^2)\times(h^2+l^2)}$ is the identity matrix. We then apply independent zero-mean Gaussian prior distributions to $\boldsymbol{\alpha}$ as $p(\boldsymbol{\alpha})=N(\boldsymbol{\alpha};\mathbf{0},\Gamma)$. $\Gamma\triangleq[\gamma_1,\cdots,\gamma_{n+1}]^T$ is a non-negative vector of the variances in one-to-one correspondence with elements of $\boldsymbol{\alpha}$. A small or zero-valued variance of γ_i implies that the associated α_i is constrained to be near zero. Combining the likelihood and the prior using Bayes' rule leads to the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{b})\propto p(\boldsymbol{b}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$. The main issue for estimation of $\boldsymbol{\alpha}$ is the value of the unknown parameter Γ. Without prior knowledge of the locations of the zero coefficients, the empirical Bayesian approach to learning Γ is to maximize the resulting likelihood function with respect to Γ [21]. Equivalently, we minimize,

$$\begin{aligned} L(\Gamma) &\triangleq -\log\int p(\boldsymbol{b}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})d\boldsymbol{\alpha} \quad (10)\\ &\equiv \log|\Sigma_b|+\boldsymbol{b}^T\Sigma_b^{-1}\boldsymbol{b} \quad \text{with} \quad \Sigma_b\triangleq A\Gamma A^T+\lambda I, \end{aligned}$$

with respect to Γ. While $L(\Gamma)$ is non-convex, optimization can be accomplished by adapting the majorization-minimization approach from [20]. This technique essentially involves the construction of rigorous upper bounds on two terms in Eq. (11). While space precludes a detailed treatment in this paper, the resulting update rules for the $(k+1)$-th iteration are given by

$$\begin{aligned} \gamma_i^{(k+1)} &\rightarrow \left(\alpha_i^{(k)}\right)^2+z_i^{(k)},\forall i, \quad \Gamma^{(k+1)}=\text{diag}[\boldsymbol{\gamma}^{(k+1)}],\\ {z_i}^{(k+1)} &\rightarrow \gamma_i^{(k)}-{\gamma_i^{(k)}}^2A_i^T(\lambda I+A\Gamma^{(k)}A^T)^{-1}A_i, \quad (11)\\ \boldsymbol{\alpha}^{(k+1)} &\rightarrow \Gamma^{(k)}A^T\left(\lambda I+A\Gamma^{(k)}A^T\right)^{-1}\boldsymbol{b}. \end{aligned}$$

Note that each γ_i is initialized with a random value. These expressions only require $O(k)$ computations, and are guaranteed to reduce $L(\Gamma)$ until a fixed point Γ^* is reached.

When a collection of overlapped patches $\{\boldsymbol{y}_i^*=\phi_H\boldsymbol{\alpha}_i^*|i=1,\cdots,n\}$, where ϕ_H is the high-resolution pairs of ϕ_L, is reconstructed, we merge them into a single high-resolution depth map. Because simple averaging of the depth values over the overlapping regions is sensitive to outliers, we introduce a

weighting factor calculated from the photo-consistency of the reconstruction. The value of j-th pixel in the high-resolution depth map $D^j_{l_1}$ is calculated as:

$$D^j_{l_1} = \frac{1}{N_j} \sum_{\forall \boldsymbol{y}^*_i | D^j_{l_1} \cap \boldsymbol{y}^*_i \neq \emptyset} e^{-\frac{\sum_{j=1}^{h^2} (I_{H^*_{i,j}} - I_{H\,i,j})^2}{2\sigma_I}} S\boldsymbol{y}^*_i, \quad (12)$$

where $N_j \triangleq \sum e^{-\frac{\sum_{j=1}^{h^2} (I_{H^*_{i,j}} - I_{H\,i,j})^2}{2\sigma_I}}$, S is a sampling operator which extracts the corresponding element in $\boldsymbol{y}^*_i$. $H_{i,j}$ is the j-th element in the reconstructed texture patch that corresponds to a depth patch ($\boldsymbol{y}^*_i$) and σ_I is the variance of the Gaussian function (5.0 in our implementation). Intuitively, we have high confidence in the value when the distance between the reconstructed texture and the original one is small.

3. RESULTS

In this section, we quantitatively evaluate our method using real data. All experiments were performed on an Intel Core i7-2640M (2.80 GHz, single thread) machine with 8 GB of RAM and were implemented in MATLAB.

First, we evaluated our inpainting algorithm from visual inspection of the inpainted results of a manually scratched *art* dataset (See Fig. 1). We observed that our method works for large missing regions, while the state-of-the-art image inpainting method [5] has considerable trouble with them.

Second, we compared our method with three state-of-the-art SR algorithms: sparse coding based image super-resolution (ScSR) [22], patch-based synthesis (PbsSR) [14], and weighted mode filtering (WMF) [9], and also compared it with nearest neighbor interpolation (NN). We also applied our method without the texture/sparsity constraints to confirm their contributions to our method. Note that only our method and WMF use a registered texture image, and external dictionary learning is required in ScSR and PbsSR. In our implementation, we use $l = 3$ and $k = 20$, and use the efficient PatchMatch algorithm [19] to find the k-NN patches. For ScSR and PbsSR, we used the authors' implementations with their default parameter settings and dictionaries, which are available from their websites. For WMF, we used the parameters described in the original paper, except that we set $\sigma_I = \sigma_S = 15$, which worked better than the original values ($\sigma_I = 6$ and $\sigma_S = 7$). Each algorithm is evaluated in terms of root mean square error(RMSE) and elapsed time for four datasets, comprising *art* and *mebius* from the Middlebury stereo dataset (which are downsampled to 347× 277), and the *fountainP11* (using only the 6th image) and *herzjesuP8* (using only the 5th image) datasets from [23] (which are downsampled to 345× 230) with a magnifying factor of 4. To make fair comparisons, inpainting is skipped in this experiment and the missing pixels in the ground truth depth map are neglected when calculating the RMSE. The results are illustrated in Table 1 and Fig. 3. We show that our method outperforms all of the state-of-the-art algorithms on the *art* and *mebius* datasets and shows the 2nd best performance on the *fountain* and *herzjesu* datasets while ScSR suffers from blurring effects, WMF appears to be sensitive to texture edges and PbsSR has problems around the object boundaries.

Table 1. Performance comparison with four real datasets.

	Root mean square error (RMSE)				Ave. running time
	art	*mebius*	*fountain*	*herzjesu*	
Ours (full configuration), k=20	**1.19**	**0.61**	1.34	1.51	1162 (sec)
Ours (w/o constraints), k=20	1.75	1.30	1.69	1.56	
ScSR [22]	2.82	3.10	8.70	6.58	1960 (sec)
WMF [9]	1.46	1.04	1.62	1.74	**275 (sec)**
PbsSR [14]	13.9	14.4	**0.50**	**0.78**	429 (sec)
NN	14.2	14.8	1.60	1.69	

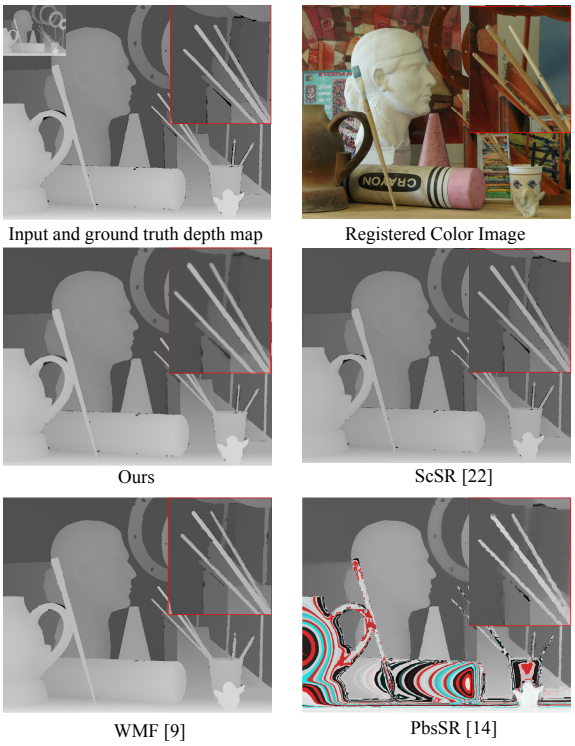


Fig. 2. Reconstructed 4× depth maps of *art* dataset.

4. CONCLUSION

In this paper, we have proposed depth-map inpainting and super-resolution algorithms which explicitly capture the internal statistics of a depth-map and its registered texture image and have demonstrated their state-of-the-art performance. The current limitation is that we have assumed the accurate registration of the texture image and have not assumed the presence of sensor noise. In future work, we will evaluate our method's robustness to these problems to assess its handling of more practical situations.

5. REFERENCES

[1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *proc. of Workshop on Stereo and Multi-Baseline Vision*. IEEE, 2001.

[2] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *proc. of Computer Vision and Pattern Recognition*. IEEE, 2006.

[3] P. Perz and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[4] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[5] B. Wohlberg, "Inpainting by joint optimization of linear combinations of exemplars," *Signal Processing Letters*, vol. 18, no. 1, pp. 75–78, 2011.

[6] A. V. Bhavsar and A. N. Rajagopalan, "Range map superresolution-inpainting, and reconstruction from sparse data," *Computer Vision and Image Understanding*, vol. 116, no. 4, pp. 572–591, 2012.

[7] A. K. Riemens, "Multistep joint bilateral depth upsampling," in *proc. of Visual Communications and Image Processing*. SPIE, 2009.

[8] K. He, J. Sun, and X. Tang, "Guided image filtering," in *proc. of 11th European Conference on Computer Vision*. IEEE, 2010.

[9] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *Transactions on Image Processing*, vol. 21, no. 3, pp. 1176–1190, 2012.

[10] J. Lu, D. Min, R. S. Pahwa, and M. N. Do, "A revisit to mrf-based depth map super-resolution and enhancement," in *proc. of International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2011.

[11] J. Park, H. Kim, Y-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *proc. of International Conference on Computer Vision*, 2011.

[12] C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," in *Computer Society Conference on Computer Vision and Pattern REcognition Workshops*. IEEE, 2008.

[13] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for tof 3d shape scanning," in *proc. of Computer Vison and Pattern Recognition*. IEEE, 2009.

[14] O. M. Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *proc. of European Conference on Computer Vision*, 2012.

[15] Y. Li, T. Xue, L. Sun, and J. Liu, "Joint example-based depth map super-resolution," in *proc. of International Conference on Multimedia and Expo*. IEEE, 2012.

[16] C. Barnes, E. Ahechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *Transactions on Graphics (Proc. SIGGRAPH)*, vol. 28, no. 3, 2009.

[17] M. Zontak, "Internal statistics of a single natural image," in *proc. of Computer Vision and Pattern Recognition*. IEEE, 2011.

[18] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *proc. of International Conference onf Computer Vision*. IEEE, 2009.

[19] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized patchmatch correspondence algorithm," in *proc. of European Conference on Computer Vision*, 2010.

[20] D. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *Transactions on Information Theory*, vol. 57, no. 9, 2011.

[21] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learns. Res.*, vol. 1, pp. 211–244, 2001.

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

[23] C. Strecha, W. V. Hansen, L. V. Gool, P. Fua, and U. Thoennessen, "On bench marking camera calibration and multi-view stereo for high resolution imagery," in *proc. of Computer Vision and Pattern Recognition*, 2008.