SINGLE DEPTH IMAGE SUPER-RESOLUTION AND DENOISING BASED ON SPARSE GRAPHS VIA STRUCTURE TENSOR

Yihui Feng[⋆,♯] *Xianming Liu*[†] *Yongbing Zhang*[♯] *Qionghai Dai*[⋆]

[⋆] Department of Automation, Tsinghua University, Beijing 100084, China
[†] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[♯] Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

ABSTRACT

The existing single depth image super-resolution (SR) methods suppose that the image to be interpolated is noise free. However, the supposition is invalid in practice because noise will be inevitably introduced in the depth image acquisition process. In this paper, we address the problem of image denoising and SR jointly based on designing sparse graphs that are useful for describing the geometric structures of data domains. In our method, we first cluster similar patches in a noisy depth image and compute an average patch. Different from the majority of the graph Fourier transform (GFT) that assumed an underlying 4-connected graph structure with vertical and horizontal edges only, we select more general sparse graph structures and edges weights based on the difference of the blocks' structure tensors. For the average patch, a graph template with edges orthogonal to the principal gradient is designed. Finally, the graph based transform (GBT) dictionary is learned from the derived correlation graph for signal representation. As shown in our experimental results, the proposed method obtains a lot of improvement in performance.

Index Terms— Denoising, depth image, dictionary learning, graph signal processing, super-resolution.

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 the depth image and that of its corresponding intensity image has been hindering its further applications. Moreover, due to the limitations of the current depth sensing technologies, acquired depth images are often corrupted by non-negligible acquisition noise. But, most of the existing single depth image super-resolution (SR) schemes assume that the original depth image is noise free. This assumption is not valid in practice because noise will be inevitably introduced in the image acquisition process. Nowadays, the problems of single depth image SR can be handled ignoring the noise which may not be able to yield satisfying result because the SR process may mistakenly enlarge the noise. Hence, new SR schemes for noisy depth images need to be developed for better restraining the noise-caused artifacts and preserving the edge structures.

(a) Texture map (b) Depth map

Fig. 1. The depth map of Books and its color counterpart.

Although the depth maps and color images have the close connection, they have amazingly different characteristics. Depth maps do not contain rich texture information reflecting the physical attributes of a surface. In other words, a depth image is often piecewise smooth: it contains sharp edges. See Fig. 1 for an illustration where the depth map exhibits this piecewise smooth characteristic, and how it is different from its color counterpart.

In this paper, in order to fully exploit the smoothness and noise factors of the deep images, we propose to solve the denoising and SR problem jointly based on the graphs that are generic data representation forms that are useful for describing the geometric structures of data domains in numerous applications. The important point is that a depth image implies a simple-enough graph representations in the pixel domain. In a graph, an edge connecting two vertices contains a weight that represents the similarity between the two vertices [1]. So that, the graph-based filtering techniques [2] can be readily applied. Specially, we first cluster similar patches

This work was supported in part by the National Natural Science Foundation of China under Grant 61571254, Guangdong Special Support program(2015TQ01X161), and the Research and Development Program of Shenzhen under grant No. JCYJ20160513103916577.

via K-nearest-neighbors (kNN) in a depth image and compute an average patch. Then we cluster the average patches via the Lloyd-Max algorithm based on their principal gradients. When designing the graph for each cluster, we select more general sparse graph structures and edge weights which differs from the previous graph Fourier transform (GFT) assuming an underlying 4-connected graph structure with vertical and horizontal edges only. Then, the graph based transform (GBT) dictionary is learned from the derived correlation graph for signal representation. Experimental results show that our proposed algorithm receives better result in both subjective and objective quality.

The outline of the paper is as follows. We first discuss the related work in Section 2. Then we review the necessary mathematical tools and discuss the graph template design methodology in Section 3 and overview our proposed joint denoising / SR algorithm in Section 4. Finally, experimentation and conclusion are presented in Section 5 and 6, respectively.

2. RELATED WORK

Single depth image SR is a very attractive and important area in computer vision, because it offers the promise of generating a high-resolution (HR) image from its degraded low-resolution (LR) measurement. However, most of the existing image SR schemes assume that the depth image to be interpolated is noise free and sometimes the two problems are typically studied independently. Instead, [2] performed denoising and color demosaicking (a special case of SR) under a unified framework. [3] solved denoising and SR under the same linear minimum mean square-error estimation (LMMSE) framework to estimate both the noiseless and missing samples from the noisy and LR image. Different from [3], our joint denoising / SR method is based on the local smoothness and nonlocal self-similarity of the depth maps.

Recently, it has been shown [4, 5, 6] that using GFT for transform coding, where each GFT is computed from a weighted undirected graph, we can get significant compression gain over discrete cosine transform (DCT). Most of the recent approaches use a simple 4-connected graph, where each pixel can be connected to its nearest vertical and horizontal pixels neighbors. In order to make the method more efficient, we generalize these schemes by considering diagonal edges as well when designing sparse graphs.

3. PRELIMINARIES

3.1. Graph Fourier Transform

In this section, we briefly overview a few basic definitions for signal on graph. A more complete description of the graph signal processing framework can be found in [7]. We define a weighted and undirected graph $\mathcal{G} = \{\mathcal{V}, \varepsilon, \mathbf{W}\}$ consists of a finite set of vertices $\mathcal{V}$ with cardinality $|\mathcal{V}| = N$, a set of edges ε connecting vertices, and a weighted adjacency matrix $\mathbf{W}$. $\mathbf{W}$ represents the matrix of edge weights, with $\mathbf{W}_{ij} = \mathbf{W}_{ji}$ denoting the positive weight of an edge connecting vertices *i* and *j*; otherwise $\mathbf{W}_{ij} = 0$. We assume that the graph is connected. The unnormalized combinatorial graph Laplacian operator is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is the diagonal degree matrix whose i^{th} diagonal element is equal to the sum of the weights of all the edges incident to vertex *i*. The normalized graph Laplacian is defined as $\mathcal{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$. Both operators are real symmetric and positive semidefinite matrices and have a complete set of real orthonormal eigenvectors with corresponding nonnegative eigenvalues. Note that zero appears as an eigenvalue with multiplicity equal to the number of connected components in the graph [8].

3.2. Computing Optimal Graph

The majority of the GFTs assumed an underlying 4-connected graph structure with vertical and horizontal edges only. Recently, Rotondo et al. [9] proposed a design methodology to select more general sparse graph structures and edge weights, on which GFTs are defined for block-based coding. In our experiment, we also utilize the design methodology to learn more grneral sparse graphs. In order to achieve the goal, the first step is to identify blocks with dominant principal gradients. We accomplish this by examining the two eigenvalues of the computed structure tensor matrix:

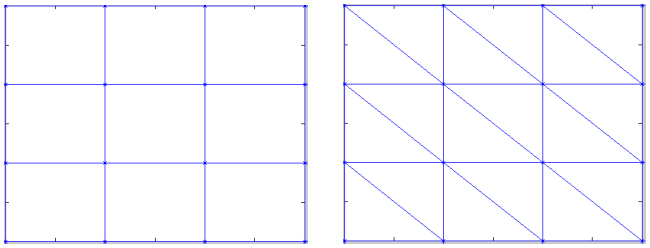
$$\begin{bmatrix} \sum_{\mathbf{r}} w(\mathbf{r})(I_x(\mathbf{p}-\mathbf{r}))^2 & \sum_{\mathbf{r}} w(\mathbf{r})I_x(\mathbf{p}-\mathbf{r})I_y(\mathbf{p}-\mathbf{r}) \\ \sum_{\mathbf{r}} w(\mathbf{r})I_x(\mathbf{p}-\mathbf{r})I_y(\mathbf{p}-\mathbf{r}) & \sum_{\mathbf{r}} w(\mathbf{r})(I_y(\mathbf{p}-\mathbf{r}))^2 \end{bmatrix}, \tag{1}$$

where $w(\mathbf{r})$ is a weight parameter, I_x and I_y are the partial derivatives with respect to the *x*- and *y*- axis, respectively. By performing eigen-decomposition on the 2D structure tensor $S_w(\mathbf{p})$, we can obtain eigenvalues λ_1 and λ_2, where $\lambda_1 \geq \lambda_1 \geq 0$, and the corresponding eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that describe the gradient $\nabla = (\boldsymbol{I}_x, \boldsymbol{I}_y)$ of the patch. $\boldsymbol{v}_1$ corresponding to the larger λ_1 is the principal gradient. For example, if λ_1 and λ_2 are both large, the patch has complex structure and there is no one dominant gradient direction. If the patch is mostly flat with little or no detectable gradient in any direction, the $\lambda_1 \approx \lambda_2 \approx 0$. Finally, $\lambda_1 \gg \lambda_2 \approx 0$ would indicate a dominant principal gradient in the patch in direction $\boldsymbol{v}_1$.

When confirming the blocks with dominant principal gradients, each block is associated with a principal gradient of angle θ, where $0^\circ \leq \theta < 180^\circ$. Then we can select different graph tempalte for signal representation according to the block's principal gradient of angle θ. Fig. 2 shows the examples of the graph template. Experimental results show that the various graphs can obtain better quality.

4. PROPOSED METHOD

After discussing the construction of a graph, we formulate the depth map joint denoising / SR problem in this section. We



(a) template1 (b) template2

Fig. 2. Examples of the graph template.

combine the graph conducting a GBT dictionary and sparse representation [10] to achieve the image reconstruction. We branch this section into two parts, the first one is the GBT dictionary learning and then we combine sparse representation to develop our method in order to have good performance.

4.1. GBT Dictionary Learning

In order to process the noisy and LR depth image more efficiently, we use the image patch for our experiment. For a given patch, we first search for its K-nearest-neighbors (kNN) and compute an average patch, from which we produce a graph describing correlations among adjacent pixels.

For the average patch, we use the structure tensor matrix calculated by Eq. (1) to judge that if the patch has dominant principal gradients. By performing eigen-decomposition on the 2D structure tensor, one can calculate eigenvalues λ_1 and λ_2, and corresponding eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that describe the gradient $\nabla = (\boldsymbol{I}_x, \boldsymbol{I}_y)$ of the patch. In the experiment, we state the patch to have a dominant principal gradient iff λ_1 - $\lambda_2 > \delta$, where δ is a predefined threshold. Blocks with dominant principal gradients can select the graph template to achieve the signal representation. In the constrast, the blocks without dominant principal gradients are encoded using conventional DCT to learn a fixed graph structures. The patch with dominant principal gradient has a principal gradient of angle θ. We take advantage of the range of θ to select the graph template as shown in the Fig. 2. When the graph is confirmed, we also can obtain a weighted adjacency matrix $\mathbf{W}$ where $W_{i,j} = w_{i,j}$. Denote by $\mathbf{D}$ the degree matrix, where $d_{i,i} = \sum_j W_{i,j}$. So the graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$. We denote $\mathbf{U}$ as the GBT dictionary and it is a matrix with eigenvectors of $\mathbf{L}$ as rows. By the above algorithm, we can learn a GBT dictionary for the average patch to achieve image denoising and SR.

4.2. Image Reconstruction

Having discussed the construction of the GBT dictionary, we start the denoising / SR algorithm using the conventional sparse representation. Sparse representation stands for that a patch can be represented as a sparse linear combination in a GBT dictionary $\mathbf{U}$. In other words, in order to represent a patch, we can find a dictionary $\mathbf{U}$ and weight vector α and the patch is well approximated by $\mathbf{U}\alpha$. The constrained optimization can be reformulated as:

$$\min_{\boldsymbol{U},\alpha} \|y - \mathbf{U}\alpha\|_2 + \tau \|\alpha\|_0, \tag{2}$$

where τ is Lagrange multiplier trading off approximation error and sparisity.

In our experiment, we group a set of the similar patches together and optimize the joint sparity of the group using the same GBT dictionary $\mathbf{U}$:

$$\min \sum_{i=1}^{N} \|y_i - \mathbf{U}\alpha_i\|_2 + \tau \min \sum_{i=1}^{N} \|\alpha_i\|_0. \tag{3}$$

This idea is based on the observation that similar edge structures naturally appear nonlocally throughout a depth image.

However, the Eq. (3) is NP-hard because the regularizer $\|\alpha\|_0$ is associated with the l_0 pseudo norm. Although we can solve the l_0 -norm by replacing it with l_1-norm, the computational complexity of the convex optimization is a hurdle to a realtime implementation. So we draw lessons from [11] and [12] where they proposed a effective approach - transform spectrum shrinkage- to obtain the best possible solution. Transform spectrum shrinkage means that we can represent all the similar patches in the derived GBT domain and sparisify the transform representations by hard-thresholding the transform coefficients.

After calculating the GBT dictionaty $\mathbf{U}$ and the sparse coefficient α for the average patch of the noisy and LR depth image, we can reconstruct the noiseless and HR patch by the following equation:

$$\boldsymbol{x} = \mathbf{U}\alpha. \tag{4}$$

Finally, all the patches are reconstructed from the Eq. (4) and they are noiseless and HR patches. In our experiment, we use the overlap patch so that every pixel admits several estimates in overlapped patches and the depth map is updated by weighted averaging over overlapped patches. In order to make the experimental results more accurate, we empirically set the weight as c_j = 1-(c_j/n), where r_j denotes the rank of the sparsified coefficient matrix and n is the patch size. Because our input depth image has noise, so we borrow the interative regularization technique in [13] to add filtered noise back to the image at each iteration to iteratively improve the quality based on the previous reconstruction:

$$\bar{y}^{(k+1)} = \bar{y}^{(k)} + \beta(y - \bar{y}^{(k)}), \tag{5}$$

where y is the previous depth image, $\bar{y}^{(k)}$ is the processed depth at the *k-th* interation and β is a relaxation parameter.

5. EXPERIMENTAL RESULTS

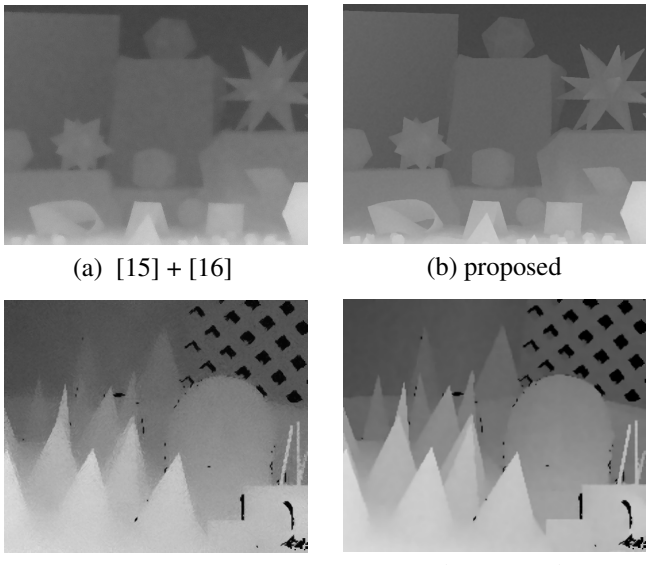
In this section, we evaluate our method based on designing sparse graphs via structure for single depth image denoising

Fig. 3. 8 test images from left to right: **Cones**, **Moebius**, **Aloe**, **Flowerpots**, **Lampshade**, **Midd**, **Rocks**, and **Teddy**.

and SR. First, we will outline our experimental parameters including details in datasets, evaluation metrics, and then we show the performance comparisons with other competitive methods.

5.1. Experimental Parameters

In this sub-section, we analyze the main parameters of our proposed method. For a fair comparison, we use 8 commonly used depth images shown in Fig. 3 including animals, people and objects. In the experiment, the original images are first downsampled and then added additive white Gaussian noise (AWGN), with standard deviation σ ranging from 5 to 25. We employ the structural similarity (SSIM) [14] to evaluate the performance of denoising / SR results by different methods: one performs denoising via Bilateral Filtering (BF) first [15] and then SR via New Edge-Directed Interpolation [16], another is linear minimum mean square-error estimation (LMMSE) [3]. The parameter δ is empirically set to 12 and the upscaling factor is $2\times$.



(a) [15] + [16] (b) proposed

(c) [15] + [16] (d) proposed

Fig. 4. The denoised and super-resolved images of (a)(b) Moebius at noise level 5; (c)(d) Cones at noise level 10.

5.2. Experimental Performance

Table I shows numerical evaluations in terms of SSIM with standard deviation σ ranging from 5 to 25. As shown in the table, we can see that our method has the best results on average for the evaluation metric. The table indicates that our algorithm outperforms LMMSE [3] at all noise levles and performs better than [15] + [16] in most test images. Moreover, the denoising / SR algorithm based on the sparse graph can reduce the calculation and save time.

Table 1. Quantitative Performance Comparisons In SSIM.

Images	method	5	10	15	20	25
Cones	[15] + [16]	0.8820	0.8735	0.8722	0.8684	**0.8693**
	lmmse	0.9477	0.9278	0.9048	0.8769	0.8554
	proposed	**0.9518**	**0.9330**	**0.9127**	**0.8890**	0.8678
Moebius	[15] + [16]	0.9417	0.9396	0.9383	0.9385	**0.9377**
	lmmse	**0.9892**	0.9739	0.9519	0.9271	0.8962
	proposed	**0.9892**	**0.9794**	**0.9664**	**0.9510**	0.9335
Aloe	[15] + [16]	0.8984	0.8939	0.8915	0.8912	0.8909
	lmmse	0.9640	0.9480	0.9249	0.8980	0.8674
	proposed	**0.9676**	**0.9545**	**0.9385**	**0.9191**	**0.8975**
Flowerpots	[15] + [16]	0.8456	0.8443	0.8439	0.8447	0.8444
	lmmse	0.9705	0.9500	0.9212	0.8883	0.8557
	proposed	**0.9803**	**0.9727**	**0.9613**	**0.9478**	**0.9304**
Lampshade	[15] + [16]	0.9116	0.9085	0.9080	0.9075	0.9071
	lmmse	0.9734	0.9590	0.9368	0.9115	0.8798
	proposed	**0.9775**	**0.9676**	**0.9547**	**0.9380**	**0.9162**
Midd	[15] + [16]	0.9050	0.9025	0.9017	0.9019	0.9014
	lmmse	0.9711	0.9566	0.9346	0.9075	0.8763
	proposed	**0.9747**	**0.9652**	**0.9510**	**0.9345**	**0.9138**
Rocks	[15] + [16]	0.9160	0.9159	0.9153	0.9154	0.9152
	lmmse	0.9722	0.9576	0.9361	0.9088	0.8808
	proposed	**0.9868**	**0.9792**	**0.9672**	**0.9520**	**0.9328**
Teddy	[15] + [16]	0.8811	0.8670	0.8669	0.8653	0.8662
	lmmse	0.9574	0.9378	0.9107	0.8828	0.8466
	proposed	**0.9619**	**0.9395**	**0.9160**	**0.8910**	**0.8681**

5.3. Experimental Visual Performance

In order to have a visual comparison and demonstrate the effectiveness of our proposed method, we show the results in Figs. 4 with different test images with two standard deviation. We compare our visual results with [15] + [16]. From Figs. 4(a) and 4(b), we can see that our method can receive the sharper results and the surface of the depth image more cleaner. So our proposed algorithm would achieve high-quality results more faithful to the original depth images with sharper edges and better details.

6. CONCLUSION

In this paper, we propose a novel method for single depth image denoising / SR based on more general sparse graph structures that capture principal gradients in blocks. Unlike the conventional schemes that perform denoising first and iterpolation later, the proposed method jointly solve the denoising / SR problem. More specially, different from the majority of the GFTs that assumed an underlying 4-connected graph structure, we select more general sparse graph based on the structure tensors. After designing the sparse graph, we can obtain the GBT dictionary for the single noisy and LR depth image reconstruction. The results of the experiment have shown that our method obtains the best results and is superior to the state-of-art algorithms both quantitatively and visually.

7. REFERENCES

[1] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, "The emerging field of signal processing on graphs," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[2] Peyman Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013.

[3] Lei Zhang, Xin Li, and David Zhang, "Image denoising and zooming under the lmmse framework," *Iet Image Processing*, 2010.

[4] G Shen, W. S Kim, S. K Narang, A Ortega, Jaejoon Lee, and Hocheon Wey, "Edge-adaptive transforms for efficient depth map coding," in *Picture Coding Symposium*, 2010, pp. 566–569.

[5] Wei Hu, Gene Cheung, Xin Li, and O Au, "Depth map compression using multi-resolution graph-based transform for depth-image-based rendering," in *IEEE International Conference on Image Processing*, 2012, pp. 1297–1300.

[6] W Hu, G Cheung, A Ortega, and O. C Au, "Multiresolution graph fourier transform for compression of piecewise smooth images," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 419–33, 2015.

[7] D. I Shuman, S. K Narang, P Frossard, and A Ortega, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine IEEE*, vol. 29, no. 5, pp. 633–634, 2013.

[8] R. K. Chung Fan, *Spectral graph theory*, American Mathematical Society, 1997.

[9] I. Rotondo, G. Cheung, A. Ortega, and H. E. Egilmez, "Designing sparse graphs via structure tensor for block transform coding of images," in *Asia-Pacific Signal and Information Processing Association Summit and Conference*, 2015, pp. 571–574.

[10] J. Yang, J Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation.," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 19, no. 11, pp. 2861, 2010.

[11] David L. Donoho and Iain M. Johnstone, "Ideal spatial adaption by wavelet shrinkage, biometrika 81," *Biometrika*, vol. 81, pp. 425–455, 1994.

[12] Wei Hu, Xin Li, Gene Cheung, and Oscar Au, "Depth map denoising using graph-based transform and group sparsity," in *IEEE International Workshop on Multimedia Signal Processing*, 2013, pp. 001–006.

[13] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, "An iterative regularization method for total variation-based image restoration," *Siam Journal on Multiscale Modeling Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[14] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *ICCV*, pp. 839 – 846, 1998.

[16] Xin Li and M. T Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 10, no. 10, pp. 1521–7, 2000.