Depth Map Denoising using Graph-based Transform and Group Sparsity

Wei Hu [#1], Xin Li [*2], Gene Cheung [$3], Oscar Au [#4]

[#] *Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology*
Clear Water Bay, Hong Kong, China
[1] huwei@ust.hk
[4] eeau@ust.hk

[*] *West Virginia University*
PO Box 6109, Morgantown, WV 26506
[2] xin.li@ieee.org

[$] *National Institute of Informatics*
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
[3] cheung@nii.ac.jp

***Abstract*—Depth maps, characterizing per-pixel physical distance between objects in a 3D scene and a capturing camera, can now be readily acquired using inexpensive active sensors such as Microsoft Kinect. However, the acquired depth maps are often corrupted due to surface reflection or sensor noise. In this paper, we build on two previously developed works in the image denoising literature to restore single depth maps—i.e., to jointly exploit local smoothness and nonlocal self-similarity of a depth map. Specifically, we propose to first cluster similar patches in a depth image and compute an average patch, from which we deduce a graph describing correlations among adjacent pixels. Then we transform similar patches to the same graph-based transform (GBT) domain, where the GBT basis vectors are learned from the derived correlation graph. Finally, we perform an iterative thresholding procedure in the GBT domain to enforce group sparsity. Experimental results show that for single depth maps corrupted with additive white Gaussian noise (AWGN), our proposed NLGBT denoising algorithm can outperform state-of-the-art image denoising methods such as BM3D by up to 2.37dB in terms of PSNR.**

I. Introduction

Recent advances in active depth sensors such as time-of-flight cameras[1] and Microsoft Kinect® have made the acquisition of depth maps (per pixel physical distance between objects in the 3D scene and the capturing camera) widely affordable. Acquired depth maps—3D geometrical information of the scene projected as 2D images to the chosen camera viewpoints—can enable a variety of novel imaging applications, such as *Depth-Image-Based Rendering* (DIBR) [1], human gesture recognition[2], etc. However, due to the limitations of current depth sensing technologies, acquired depth maps are often corrupted by non-negligible acquisition noise and/or suffer from partially missing data. Thus a major challenge in depth map processing is the restoration of depth maps from their corrupted versions.

More generally, image denoising—or removal of *additive white Gaussian noise* (AWGN) from photographic images—is one of the classical problems in image restoration and has been extensively studied in the literature. Recent notable advances in this field include the class of nonlocal image denoising techniques [2], [3], [4], as well as dictionary learning-based approaches [5], [6]. Nonlocal image denoising [2] builds on a simple assumption that similar patterns are likely to recur throughout an image. Thus, one can first cluster similar patches in an image, in order to jointly restore them by exploiting their dependency (i.e., nonlocal self-similarity of an image). Dictionary learning [5] assumes that a signal can be represented by the linear combination of a few atoms out of a possibly over-complete dictionary. The primary challenge with dictionary learning is to simultaneously construct/learn an appropriate dictionary and identify the sparsest representation for the given signal.

(a) Depth map (b) Texture map

Fig. 1. The depth map of *Teddy* and its color counterpart.

Despite the close connection between depth maps and color images (e.g., Kinect acquires both), they have strikingly different characteristics. An important observation is that unlike typical RGB color images, depth maps do not contain

[1] http://en.wikipedia.org/wiki/Time-of-flight_camera
[2] http://code.google.com/p/kineticspace/

MMSP'13, Sept. 30 - Oct. 2, 2013, Pula (Sardinia), Italy.
978-1-4799-0125-8/13/$31.00

rich texture information reflecting the physical attributes of a surface. In other words, a depth map is often *piecewise smooth*: it contains sharp edges (e.g., boundaries between foreground objects and background), and within the edges, the surfaces are varying spatially only gradually due to the absence of textures. 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, we propose to build on the existing image denoising techniques to jointly exploit the local smoothness and nonlocal self-similarity of depth maps. More specifically, we first cluster similar patches in a depth image and compute an *average* patch, from which we can deduce a graph describing discontinuities (e.g., edges) as well as correlations among adjacent pixels. Second, we transform similar patches to a common *graph-based transform* (GBT) domain [7], where the GBT basis vectors can be derived from the deduced correlation graph. Finally, we perform an iterative thresholding procedure similar to [4] to enforce group sparsity in the GBT domain. From this perspective, the newly-developed *Nonlocal GBT* (NLGBT) denoising algorithm can be viewed as a nonlocal extension of our previous work on super-resolution reconstruction of depth images [8].

It should be noted that patch clustering and GBT have to work hand-in-hand because local transience and nonlocal invariance are the two sides of the same coin. Unlike previous works, NLGBT is both locally adaptive (through the construction of a correlation graph reflecting the edge structure embedded into the average patch) and globally consistent (via a common derived GBT transform using which all similar patches are sparsified). Experimental results have shown that for single depth maps corrupted with AWGN, our proposed denoising algorithm can outperform state-of-the-art image denoising methods such as BM3D by up to 2.37dB in terms of PSNR.

The outline of the paper is as follows. We first overview related work in Section II. We then describe the standard procedure to construct a GBT in Section III. We present our depth map denoising formulation in Section IV, and we develop our algorithm in Section V. Finally, experimentation and conclusion are presented in Section VI and VII, respectively.

II. Related Work

Denoising of photographic images have advanced rapidly in recent years especially after the publication of patch-based image denoising (including nonlocal-mean [2] and BM3D [9]). The new insight along the line of patch-based image denoising lies in that important image structures such as edges and textures can be more effectively characterized by their nonlocal self-similarity than local transience (the conventional wisdom at the heart of transform-based image models). Under the context of denoising, clustering similar patches distant from each other makes it possible to more accurately estimate the signal variance from a bilateral perspective [4]. Therefore, it is reasonable to expect that patch-based models could lend themselves to depth maps where nonlocal self-similarity is present by the abundance of depth discontinuities (counterpart of edges in photos).

Given the piecewise smooth characteristic of depth maps, GBT has been proposed in [7] for efficient transform coding; given the derived GBT basis functions do not filter across a detected edge, the GBT representation for a given depth block likely has zero high-frequency components, resulting in few non-zero coefficients that required coding. In our previous work [8], we have also used GBT for depth map super-resolution. This paper can be viewed as an extension of [8] to depth map denoising.

There are some recent work on denoising of depth maps [10], [11], where the availability of both color and texture maps are assumed, and their correlations are exploited for denoising. We do not assume the availability of color maps in our work; this is a practical consideration for depth sensors like Mesa's SwissRanger[3] that do not capture texture images from the same viewpoint, and for challenging environments like a dark room, where the lighting conditions are not reliable to capture good quality color images.

III. Graph-based Transform

We first overview the conventional three-step procedure to construct a GBT [7] from an *unweighted* graph; the method to construct a GBT from a *weighted* graph to define a set of basis functions for similar patches in Section V would be a straightforward extension.

First, prominent edges in a $\sqrt{n} \times \sqrt{n}$ target pixel patch are detected. Edge detection can be done using a number of methods; in [7], edges are detected in a target patch based on the difference between neighboring pixel values using a simple thresholding technique.

In the second step, we treat each pixel in the $\sqrt{n} \times \sqrt{n}$ block as a node in a graph $\mathcal{G}$, and connect it to its four or eight immediate neighbors in the patch, resulting in a 4- or 8-connectivity graph. If there is a detected edge between two neighboring pixels (nodes), we eliminate their connection. Given the connectivity graph, we can define an adjacency matrix $\mathbf{A}$, where $\mathbf{A}(i,j) = \mathbf{A}(j,i) = 1$ if pixel positions i and j are connected, and 0 otherwise. We can similarly compute the degree matrix $\mathbf{D}$, where $\mathbf{D}(i,i)$ is the number of connections for node i, and $\mathbf{D}(i,j) = 0$ for all $i \neq j$.

In the third step, using computed $\mathbf{A}$ and $\mathbf{D}$, we can compute the *graph Laplacian matrix* $\mathbf{L} = \mathbf{D} - \mathbf{A}$. If we now project a graph signal $\mathbf{x}$ in the graph $\mathcal{G}$ onto the eigenvectors of the Laplacian $\mathbf{L}$, it becomes the spectral decomposition of the graph signal; *i.e.*, it provides a "frequency domain" interpretation of signal $\mathbf{x}$ given graph support $\mathcal{G}$. Hence, we can construct GBT transform using eigenvectors of $\mathbf{L}$. In particular, we can stack pixels in the $\sqrt{n} \times \sqrt{n}$ patch into a length-n vector $\mathbf{x}$ and compute $\mathbf{y} = \mathbf{U} \cdot \mathbf{x}$, where $\mathbf{U}$ is a matrix with eigenvectors of $\mathbf{L}$ as rows. Fig. 2 gives an example of constructing GBT from a 2×2 pixel block.

[3] http://www.mesa-imaging.ch/prodview4k.php

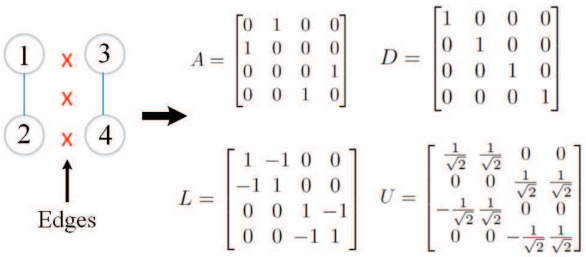


Fig. 2. An example of constructing GBT from a 2×2 pixel block. The vertical edges separate pixel 1 and 2 from pixel 3 and 4 and a graph is constructed by connecting pixels on each side of the edges. The corresponding adjacency matrix $\mathbf{A}$, degree matrix $\mathbf{D}$, Laplacian matrix $\mathbf{L}$ as well as the computed GBT $\mathbf{U}$ are shown on the right.

The above discussion of GBT is based on an unweighted graph, which can be easily extended to *weighted GBT* associated with a weighted graph. In a weighted graph the edge weight $w(i,j)$ can be any nonnegative value other than 0 and 1, offering more flexibility to describe the pairwise relationship between nodes i and j, such as the similarity in pixel intensities. The corresponding adjacency matrix $\mathbf{W}$ is defined as $W(i,j) = w(i,j)$ and the degree of a node i is defined to be $d_i = \sum_j w(i,j)$. We can then compute the graph Laplacian matrix as $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and construct weighted GBT using eigenvectors of $\mathbf{L}$.

IV. Problem Formulation

Having discussed the construction of a basic GBT, we formulate the depth map denoising problem in this section, arriving at a formal problem definition we call *Nonlocal GBT (NLGBT)*. We start by introducing the popular sparse coding in patch space, which is the foundation of our formulation. Then we elaborate on our formulation tailored for depth maps, which combines depth maps' local piecewise smooth characteristic and the nonlocal self-similarity prior by exploiting group sparsity of similar patches.

A. Sparse Coding

Sparse coding means that a patch $\mathbf{y}_i$ can be represented by a weighted combination of only a few atoms out of a learned dictionary $\mathbf{U}_i$. In other words, it means finding a dictionary $\mathbf{U}_i$ and weight vector α_i for patch $\mathbf{y}_i$ such that: i) each patch $\mathbf{y}_i$ is well approximated by $\mathbf{U}_i\alpha_i$, and ii) the sparsity of $\|\alpha_i\|_0$ is minimized. Mathematically, we can write:

$$\min_{\mathbf{U}_i,\alpha_i} \|\mathbf{y}_i - \mathbf{U}_i\alpha_i\|_2 + \tau\|\alpha_i\|_0 \tag{1}$$

where τ is a Lagrange multiplier trading off approximation error and sparsity. The l_0-norm regularizer enforces sparsity of the weight vector α_i.

Image denoising can be formulated as sparse coding of a given noise-corrupted observation $\mathbf{y}_i$. The noise, which generally produces high-frequency components, can be removed by finding a sparse representation of the signal.

B. Group Sparsity

Instead of sparsifying each patch $\mathbf{y}_i$ separately using a dictionary $\mathbf{U}_i$, one can group a set of similar patches together and optimize the joint sparsity of the group using the same learned dictionary $\mathbf{U}$:

$$\min_{\mathbf{U},\alpha} \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{U}\alpha_i\|_2 + \tau \sum_{i=1}^{N} \|\alpha_i\|_0 \tag{2}$$

The motivation is that the collection of similar patches can be treated as different noisy observations of a *principle manifold*. This is based on the observation that similar edge structures naturally appear nonlocally throughout a depth image. We thus cast the denoising problem as manifold reconstruction with noisy data.

The key to manifold reconstruction is to discover the geometry of the manifold. One can learn a dictionary capturing the common structure of a group of similar patches, and enforce sparse representation in this dictionary domain for the group—so-called *group sparsity*—to reconstruct the clean manifold geometry.

C. Nonlocal GBT

Due to the desirable characteristic of piecewise smoothness, we further exploit the group sparsity of depth maps in GBT domain in a nonlocal fashion. The main idea of *Nonlocal GBT* is to enforce sparsity of similar patches in the same GBT domain which well reflects the structure of the principle manifold. We find this particular GBT domain by first representing the common geometry of the group via an average patch by computing the average of the similar patches, which computes to the average statistics. We then derive the GBT from the similarity graph built on the average patch. The similarity graph describes the pairwise similarity between adjacent pixels in depth values. Hence the learned GBT dictionary $\mathbf{U}$ is dependent on the average patch $\bar{\mathbf{y}}$ of the group:

$$\mathbf{U} = \mathbf{U}(\bar{\mathbf{y}}). \tag{3}$$

Having constructed the GBT which is adaptive to the average statistics of the group of similar patches, we employ this GBT dictionary as the dictionary $\mathbf{U}$ in Eq. (2). In this way we can enforce the group sparsity in this common GBT domain, thus reconstructing each patch in the group well by referencing the nonlocal similar geometry. Further, by allowing different sparse representations in the same GBT domain, the texture of each individual patch is also preserved.

Another advantage of our NLGBT approach is that the GBT dictionary can be efficiently learned from the similar patches, which avoids complicated dictionary learning process.

In a nutshell, our NLGBT method is a hybrid of group sparsity and GBT representation, exploiting the group sparsity by adapting a nonlocal GBT dictionary into Eq. (2). Both the local piecewise smoothness prior and nonlocal self-similarity prior are taken into consideration via the nonlocal GBT representation.

V. Algorithm Development

In this section, we develop an iterative depth map denoising algorithm based on the NLGBT model discussed in Section IV. We describe our algorithm step-by-step and summarize it in **Algorithm 1**.

A. Patch clustering

For a given $\sqrt{n} \times \sqrt{n}$ patch (called exemplar in the vision literature), we first search for its K-nearest-neighbors (kNN). kNN together with the exemplar patch is called a *cluster* in the sequel. We then stack them as columns to create a data matrix $\mathbf{Y}$. Similar to BM3D, the K most similar patches to the exemplar patch are found using block-matching and Euclidean distance.

B. GBT dictionary learning

We learn the GBT dictionary for similar patches by exploiting their common structure as elaborated in Section IV. Specifically, we compute an average patch, from which a similarity graph is constructed modeling the local neighborhood correlations among pixels in the average patch. There are different flavors of similarity graphs, such as the ϵ-neighborhood graph, k-nearest neighbor graph and the fully connected graph [12]. We choose a four-connected graph where only pairwise adjacent pixels are connected for simplicity. In particular, we construct an undirected weighted graph by treating each pixel in the average patch as a node and connect adjacent pixels (i, j) with edge weight

$$w_{i,j} = e^{\frac{-\|y_i - y_j\|^2}{\sigma_w^2}}, \tag{4}$$

where $\|y_i - y_j\|^2$ calculates the squared intensity difference in pixel i and j as a measure of similarity. The parameter σ_w controls the sensitivity of the similarity measure to the noise and the range of the intensity difference. It is empirically set to 20% of the sum of the noise variance and maximum intensity difference in the patch.

With the weighted similarity graph, we calculate the corresponding Laplacian matrix and derive the GBT basis functions via eigen-decomposition. Note that unlike previous works on GBT [7], [8], the GBT we use here is a *weighted GBT* constructed from a weighted graph instead of an unweighted one. The motivation is that weighted GBT bases generally give sparser representation of the average patch since it better captures the subtle inter-pixel correlations in the patch.

C. Transform spectrum shrinkage

With the GBT dictionary learned, we denoise the depth map by finding a sparse solution to Eq. (2) for each cluster. The regularizer $\|\alpha_i\|_0$ is associated with the l_0 pseudo norm, which makes the optimization problem NP-hard. One can relax it to a convex optimization by replacing the l_0-norm with the l_1-norm and acquire the solution via an off-the-shelf convex optimization solver. However, the computational complexity of the convex optimization is a hurdle to a real-time implementation. We thereby adopt a simple yet effective approach—*transform spectrum shrinkage*—to yield the best possible solution.

Transform spectrum shrinkage means that we represent all the similar patches in the derived GBT domain and sparsify the transform representations by *hard-thresholding* the transform coefficients. Since the GBT representation of a clean depth map is expected to be dominated by the low-frequency components because of the piecewise-smooth characteristic, the high-frequency components are most likely to be generated by noise. Therefore, we attenuate the noise by hard-thresholding the transform coefficients. The threshold is set as $t = \sigma\sqrt{(2\log(n^2K))}$, where σ is the standard deviation of the noise, n is the patch size and K is the number of similar patches, as per the rule from [13].

D. Image update

Finally, all the patches are reconstructed from inverse GBT with the sparsified transform coefficients. As every pixel admits several estimates in overlapped patches, the depth map is updated by weighted averaging over overlapped patches. The weights should be inversely proportional to the rank of the sparsified coefficient matrix so that highly sparse patches get higher priority during weighted averaging. We empirically set the weight as $c_j = 1 - (r_j/n)$ for the j-th cluster, where r_j denotes the rank of the sparsified coefficient matrix and n is the patch size.

E. Iterative regularization

Borrowing the iterative regularization technique in [14], we add filtered noise back to the denoised image at each iteration to iteratively enhance the quality of the noisy depth map based on the previous estimate:

$$\hat{\mathbf{y}}^{(k+1)} = \hat{\mathbf{y}}^{(k)} + \delta(\mathbf{y} - \hat{\mathbf{y}}^{(k)}), \tag{5}$$

where $\mathbf{y}$ is the input noisy depth map, $\hat{\mathbf{y}}^{(k)}$ is the denoised version at the k-th iteration and δ is a relaxation parameter.

Algorithm 1 Image Denoising via NLGBT

1: **Input**: One noisy depth map $\mathbf{y}$
2: **Initialization**: $\hat{\mathbf{y}}^{(1)} = \mathbf{y}$;
3: **for** $k = 1$ to $iter$ **do**
4: Step A. Patch clustering
5: Step B. GBT dictionary learning
6: Step C. Transform spectrum shrinkage
7: Step D. Image update
8: Step E. Iterative regularization
9: **end for**
10: **Output**: The denoised depth map

VI. Experimentation

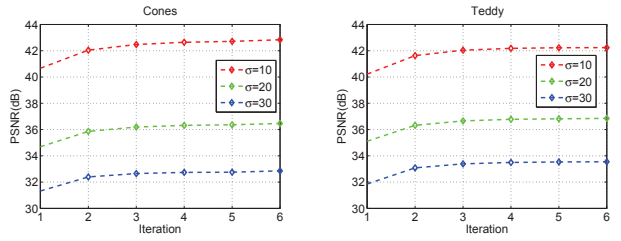
In this section we present and discuss the denoising performance of the proposed NLGBT for depth maps to support our depth map model and representation. We compare against several competing denoising methods and further show the effectiveness of NLGBT by demonstrating its benefit in Depth-Image-Based Rendering.

A. Experimental setup

We evaluate our NLGBT denoising approach with three Middlebury depth maps *Cones* (450×375), *Teddy* (450×375) and *Sawtooth* (434×380)[4]. Additive white Gaussian noise (AWGN) is added to these images, with standard deviation σ ranging from 10 to 30. We compare our approach with three other competing methods: Bilateral Filtering (BF) [15], Non-Local Means Denoising (NLM) [2] and Block-Matching 3D (BM3D) [9], which exploits the local, nonlocal and a hybrid of local and nonlocal prior respectively. Note that BM3D exhibits one of the best denoising performance in the literature.

B. Convergence of the iterative algorithm

We first demonstrate the convergence of our iterative algorithm. Fig. 3 shows that the depth map quality is greatly enhanced at the second iteration and converges fast afterwards. Hence, the computational complexity can be further reduced by controlling the number of iterations as a trade-off between complexity and performance for a fast profile.



(a) Cones (b) Teddy

Fig. 3. The PSNR curves of denoised depth maps *Cones* and *Teddy* respectively with iterative enhancement at different noise level.

C. Objective quality

We then compare our approach with BF, NLM and BM3D. Table I shows the objective quality of denoising results (measured in PSNR) by these methods at different noise level. Our scheme produces superior results, achieving up to 16.40dB gain over BF, 7.74dB gain over NLM and 2.37dB gain over BM3D.

D. Subjective quality

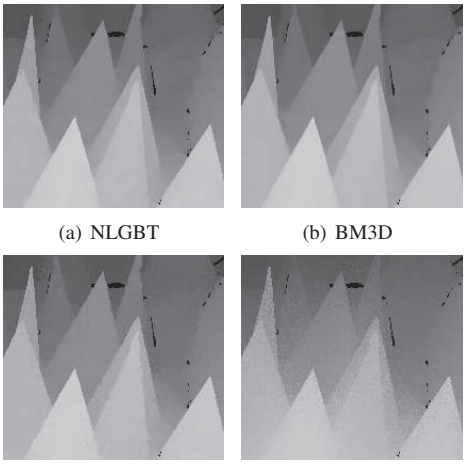
We now present the subjective quality comparison among different denoising methods. Fig. 4 and Fig. 5 demonstrate fragments of different denoised versions of depth maps for *Cones* and *Teddy* respectively. It can be observed that the depth maps denoised by our NLGBT exhibit clean sharp edges and smooth surface, while the ones produced by BM3D are blurred along the edges to some extent. Those produced by NLM and BF still look noisy all over the image.

The subjective quality of the reconstructed depth maps by NLGBT validates the superiority of exploiting the group sparsity in GBT domain for piecewise-smooth depth maps. On one hand, similar structures do recur throughout depth maps as in Fig. 4 and Fig. 5, which provides desirable inter-patch references for exploiting group sparsity; One the other hand, GBT representation is perfect to preserve the piecewise smooth characteristic of depth maps, which is why NLGBT outperforms the two nonlocal methods NLM and BM3D.

[4]http://vision.middlebury.edu/stereo/data/

TABLE I
DEPTH MAP DENOISING: PERFORMANCE COMPARISON IN PSNR (DB) WITH THREE COMPETING METHODS

		σ				
Image	Method	10	15	20	25	30
Cones	NLGBT	42.84	39.18	36.53	34.43	32.97
	BM3D	40.56	37.49	35.28	33.81	32.75
	NLM	39.42	35.84	34.64	32.95	31.62
	BF	33.34	30.53	27.96	26.03	24.21
Teddy	NLGBT	42.29	39.38	36.71	34.62	33.42
	BM3D	41.36	38.33	36.12	34.45	33.25
	NLM	39.57	36.24	35.17	33.49	32.22
	BF	34.49	31.25	28.87	26.50	23.70
Sawtooth	NLGBT	48.41	45.30	43.22	41.71	40.01
	BM3D	46.04	43.51	41.84	40.16	39.13
	NLM	41.14	37.56	38.28	36.54	35.01
	BF	36.36	30.99	27.62	25.38	23.61



(a) NLGBT (b) BM3D

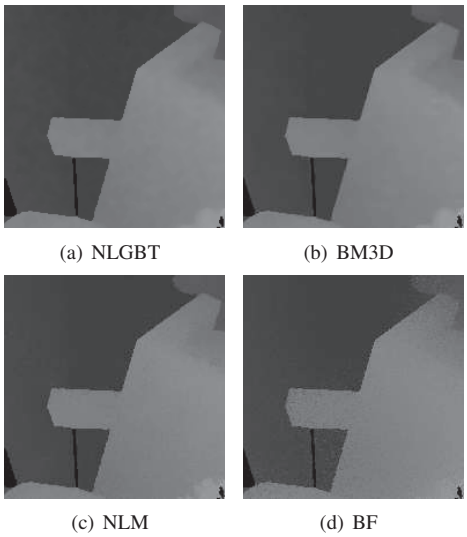
(c) NLM (d) BF

Fig. 4. Fragment of different denoised versions of the depth map *Cones* corrupted by AWGN (σ=10).

E. Application to Depth-Image-Based Rendering

As depth maps generally facilitate various end applications instead of being observed directly, we further investigate the effectiveness of our method by applying the denoised depth maps to one popular image application—Depth-Image-Based Rendering (DIBR). A simple implementation of 3D warping [1] is used to perform DIBR.

We report the objective quality of the DIBR-synthesized views facilitated with stereo depth maps (corrupted by AWGN with σ=10) denoised by different methods in Table II. Again NLGBT outperforms BF, NLM and BM3D by up to 2.03dB, 0.85dB and 0.92dB respectively. We also show the DIBR-synthesized virtual views of *Cones* with different denoised versions of the stereo depth maps in Fig. 6. It can be observed that the synthesized result with NLGBT is more pleasant than those produced with other methods, with fewer ringing



(a) NLGBT (b) BM3D

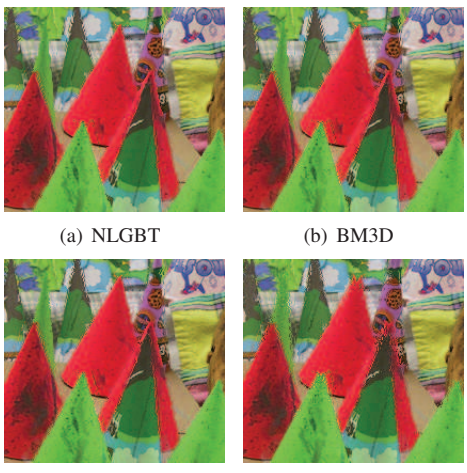
(c) NLM (d) BF

Fig. 5. Fragment of different denoised versions of the depth map *Teddy* corrupted by AWGN (σ=10).

artifacts and corrupted boundaries. The credit mainly goes to the well-preserved depth discontinuities by NLGBT, which plays a critical role in DIBR.

TABLE II
DIBR: PERFORMANCE COMPARISON IN PSNR (DB) WITH THREE COMPETING METHODS (σ=10)

	Image		
Method	Cones	Teddy	Sawtooth
NLGBT	27.30	29.96	32.06
BM3D	26.71	29.04	31.46
NLM	27.02	29.54	31.21
BF	25.27	28.39	30.48



(a) NLGBT (b) BM3D

(c) NLM (d) BF

Fig. 6. Fragment of the DIBR-synthesized images of *Cones* by different denoised versions of the corresponding stereo depth maps (σ=10).

VII. CONCLUSION

In this paper, we presented a nonlocal extension of graph-based transform and explored its application into depth map restoration. The developed nonlocal GBT (NLGBT) denoising algorithm is capable of jointly exploiting the local smoothness and nonlocal self-similarity of depth maps. When tested on standard depth images corrupted by additive white Gaussian noise, our algorithm has shown to outperform several competing approaches including BM3D. This work seems to suggest that the joint local-and-nonlocal image model underlying the proposed NLGBT algorithm is particularly effective for characterizing piecewise smooth signals such as depth maps. In the future, we plan to test the effectiveness of NLGBT on corrupted depth images acquired directly from noisy depth sensors such as Kinect.

VIII. ACKNOWLEDGMENT

This work has been supported in part by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF Project no. 610210).

REFERENCES

[1] D. Tian, P.-L. Lai, P. Lopez, and C. Gomila, "View synthesis techniques for 3D video," in *Applications of Digital Image Processing XXXII, Proceedings of the SPIE*, vol. 7443 (2009), 2009, pp. 74 430T–74 430T–11.

[2] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, June 2005.

[3] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *IEEE 12th International Conference on Computer Vision (ICCV)*, Kyoto, Japan, September 2009.

[4] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," in *IEEE Transactions on Image Processing*, vol. 22, no.2, February 2013, pp. 700–711.

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representation over learned dictionaries," in *IEEE Transactions on Image Processing*, vol. 15, no.12, December 2006.

[6] W. Dong, G. Shi, X. Li, L. Zhang, and X. Wu, "Image reconstruction with locally adaptive sparsity and nonlocal robust regularization," in *Signal Processing: Image Communication*, vol. 27, no.10, November 2012, pp. 1109–1122.

[7] G. Shen, W.-S. Kim, S. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in *IEEE Picture Coding Symposium*, Nagoya, Japan, December 2010.

[8] W. Hu, G. Cheung, X. Li, and O. Au, "Depth map super-resolution using synthesized view matching for depth-image-based rendering," in *3rd International Workshop on Hot Topics in 3D (in conjunction with ICME 2012)*, Melbourne, Australia, July 2012.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," in *IEEE Transactions on Image Processing*, vol. 16, no.8, August 2007, pp. 2080–2095.

[10] B. Huhle, T. Schairer, P. Jenke, and W. Straber, "Robust non-local denoising of colored depth data," in *IEEE CVPR Workshop on Time of Flight Camera based Computer Vision*, Anchorage, AK, June 2008.

[11] M. Tallón, S. D. Babacan, J. Mateos, M. Do, R. Molina, and A. Katsaggelos, "Upsampling and denoising of depth maps via joint-segmentation," in *20th European Signal Processing Conference (EUSIPCO 2012)*, Bucharest, Romania, August 2012.

[12] U. von Luxburg, "A tutorial on spectral clustering," in *Statistics and Computing*, vol. 17(4), 2007, pp. 395–416.

[13] D. Donoho and I. Johnstone, "Ideal spatial adaption by wavelet shrinkage," in *Biometrika*, vol. 81, 1993, pp. 425–455.

[14] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," in *Multiscale Model. Simul.*, vol. 4, no.2, 2005, p. 460489.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, 1998.