A GRAPH-BASED JOINT BILATERAL APPROACH FOR DEPTH ENHANCEMENT

Yongzhe Wang, Antonio Ortega*

Signal and Image Processing Institute
University of Southern California
Los Angeles, California USA

Dong Tian, and Anthony Vetro

Multimedia Group
Mitsubishi Electric Research Labs (MERL)
Cambridge, Massachusetts USA

ABSTRACT

Depth images are often presented at a lower spatial resolution, either due to limitations in the acquisition of the depth or to increase compression efficiency. As a result, upsampling low-resolution depth images to a higher spatial resolution is typically required prior to depth image based rendering. In this paper, depth enhancement and upsampling techniques are proposed using a graph-based formulation. In one scheme, the depth is first upsampled using a conventional method, then followed by a graph-based joint bilateral filtering to enhance edges and reduce noise. A second scheme avoids the two-step processing and upsamples the depth directly using the proposed graph-based joint bilateral upsampling. Both filtering and interpolation problems are formulated as regularization problems and the solutions are different from conventional approaches. Further, we also studied operations on different graph structures such as star graph and 8-connected graph. Experimental results show that the proposed methods produce slightly more accurate depth at the full resolution with improved rendering quality of intermediate views.

Index Terms— 3D Video, Depth Upsampling, Graph-based Image Processing, Joint Bilateral

1. INTRODUCTION

In 3D video applications, depth images enable many advanced features, e.g., supporting a multiview autostereoscopic display with a limited set of input views, or adjusting the perceived depth from a stereo display. To support these applications, depth has become an integral component in next-generation 3D data formats as standardized extensions of both AVC and HEVC. It is assumed that depth images are captured by range cameras or computed via stereo matching algorithms. Together with the multi-view texture images, it is possible to generate a synthesized virtual view using depth image based rendering (DIBR) techniques.

Unfortunately, depth is often available at a low resolution relative to its corresponding color component. For example, state-of-art range cameras currently acquire depth at much lower resolutions compared to conventional color cameras. Also, depth is typically coded at a lower spatial resolution in order to improve overall coding efficiency. Since a depth value is required for each color pixel in conventional DIBR rendering procedures, it is necessary to upsample and enhance the low-resolution depth image.

Traditional upsampling algorithms that are used for color images are not suitable for depth since depth images are characterized by smooth areas partitioned by sharp object boundaries. As such, various methods have been considered that utilize edge information to improve the processing of depth. Some of these methods have also exploited the similarity between the collocated color image (which is often available) and the depth image. For instance, joint bilateral filter (JBF), which is widely used for depth denoising and upsampling [1], utilizes color difference and pixel distance between target pixel and its neighboring pixel to adapt filter weights to image structure. A trilateral filter was proposed in [2] that extends JBF by making use of the pixel difference in depth map. On top of that, [3] included pixel difference across different views in the cost function.

Recently, the block graph-based transform (GBT) has been successfully applied to the problems of depth coding [4], upsampling [5] and denoising [6]. [5] features a synthesized view matching (which was also used in [3]) and piecewise smoothness enforcement by keeping only the lowest L coefficients in the GBT domain. [6] groups similar image patches and force the group sparsity of the GBT coefficients.

In this paper, we formulate the problems from the the graph-based image filtering and graph interpolation point of view recently presented in [7] and [8]. Different from [5] and [6], instead of forcing sparsity or hard thresholding the coefficients which may be viewed as ideal lowpass graph filtering operations, we enable the use of an arbitrary graph filter to regularize the solution, which includes but not limited to an ideal lowpass filter. Different from [5] where the texture maps of the same and neighboring view points are used, and different from [6] where the availability of the texture map is not assumed, we use the texture map of the same view point in calculating the link weights. The rest of the paper is organized as follows: after a brief review of existing work on graph-based image processing in Section 2, we present graph-based joint bilateral filtering and upsampling techniques applied to depth images utilizing the corresponding textures in Section 3. Experimental results are presented in Section 4 and concluding remarks are given in Section 5.

2. OVERVIEW OF GRAPH-BASED IMAGE PROCESSING

The field of graph-based signal processing is generating significant interest in recent years [9]. We start by introducing some basic definitions and notation. An undirected graph $G = (V, E)$ consists of a collection of nodes $V = \{1, 2, ..., N\}$ connected by a set of links $E = \{(i, j, w_{ij})\}$,$i, j \in V$ where (i, j, w_{ij}) denotes the link between nodes i and j having weight w_{ij}. The adjacency matrix $\mathbf{W}$ of the graph is an $N \times N$ matrix, the degree d_i of a node i is the sum of link weights connected to node i. The degree matrix $\mathbf{D} := diag\{d_1, d_2, ..., d_N\}$ is a diagonal matrix, and the combinatorial Laplacian matrix is $\mathbf{L} := \mathbf{D} - \mathbf{W}$. The normalized Laplacian matrix $\mathcal{L} := \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ is a symmetric positive semi-definite matrix. Therefore, it has eigendecomposition $\mathcal{L} = \mathbf{U}\Lambda\mathbf{U}^t$, where $\mathbf{U} = \{\mathbf{u}_1, ..., \mathbf{u}_N\}$ is an orthogonal set of eigenvectors and

*Part of this work was performed during Yongzhe Wang's internship at MERL in 2013.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

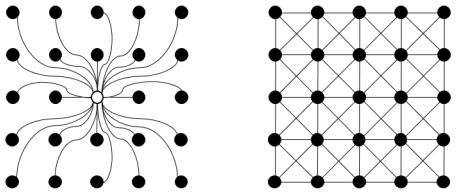


Fig. 1. An illustration of the graph structures used in a bilateral filter (left) and the proposed graph-based joint bilateral filter (right).

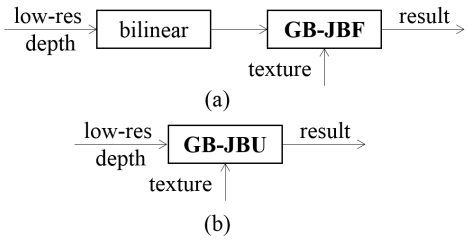


Fig. 2. Application of GB-JBF for depth denoising (a) and GB-JBU for depth upsampling (b).

$\mathbf{\Lambda} = diag\{\lambda_1, ..., \lambda_N\}$ is its corresponding eigenvalue matrix.

The eigenvectors and eigenvalues of the Laplacian matrix provide a spectral interpretation of the graph signals. The eigenvalues $\{\lambda_1, ..., \lambda_N\}$ can be treated as graph frequencies and are always situated in the interval $[0, 2]$ on the real line for the normalized Laplacian. The *Graph Fourier Transform* (GFT) is defined as a projection of a signal $\mathbf{x}$ onto the eigenvectors $\mathbf{U}$ of the graph Laplacian:

$$\tilde{\mathbf{x}} := \mathbf{U}^t \mathbf{x}. \tag{1}$$

Since a graph spectral domain is introduced with GFT, *Graph Spectral Filtering* (GSF) is defined as:

$$\tilde{\mathbf{x}}_{out} := \mathbf{H_1}(\Lambda)\tilde{\mathbf{x}}_{in}, \tag{2}$$

where $\tilde{\mathbf{x}}_{in}$ and $\tilde{\mathbf{x}}_{out}$ are the spectral representations of the input and output signals, respectively, and $\mathbf{H_1}\left(\Lambda\right) = diag\left(h_1\left(\lambda_i\right)\right)$. $h_1\left(\lambda_i\right)$ is the spectral response of the filter. If $h_1\left(\lambda_i\right)$ is a polynomial, the graph filtering can be directly applied on the graph Laplacian in the vertex domain:

$$\mathbf{x}_{out} = \mathbf{U}\mathbf{H_1}(\Lambda)\mathbf{U}^t\mathbf{x}_{in} = \mathbf{H}(\mathcal{L})\mathbf{x}_{in}. \tag{3}$$

It should be noted that when normalized Laplacian matrix is used to define the GFT, the input signal should be properly normalized and a normalized output signal is produced as will be described in the following subsection.

2.1. Graph spectral interpretation of conventional image filters

A graph spectral interpretation of conventional image filters has been discussed in [7] and will be briefly reviewed here. For any image filter and an input image $\hat{\mathbf{x}}_{in}$, the pixels in the output image $\hat{\mathbf{x}}_{out}$ can be written as a weighted average of the pixels in $\hat{\mathbf{x}}_{in}$:

$$\hat{\mathbf{x}}_{out}\left[j\right] = \sum_i \frac{w_{ij}}{\Sigma_i w_{ij}} \hat{\mathbf{x}}_{in}\left[i\right], \tag{4}$$

where w_{ij} is the weighting coefficients between pixel i and j. Note that with this general notation we can accommodate adaptive filtering operators and we do not need to consider only shift invariant operators. As a special case, the weights of the bilateral filter [10] are defined by

$$w_{ij} = \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_s^2}\right)\exp\left(-\frac{\left(\mathbf{x}_{in}\left[i\right] - \mathbf{x}_{in}\left[j\right]\right)^2}{2\sigma_r^2}\right). \tag{5}$$

Now a graph is constructed as shown in Fig. 1 (left), where pixel j at the center is to be filtered. It is connected to all pixels within a specific window. The weights of the links in the graph are defined by the image filter weights w_{ij}, where i is the index of one of the pixels connected to j. Then the filtering operation can be rewritten as:

$$\hat{\mathbf{x}}_{out} = \mathbf{D}^{-1}\mathbf{W}\hat{\mathbf{x}}_{in}, \tag{6}$$

where $[\mathbf{W}]_{ij} = w_{ij}$ and $\mathbf{D}$ is the corresponding degree matrix, so that $\mathbf{D}^{-1}$ is the normalization term in (4).

Given the graph filtering framework, we are now able to analyze the graph spectral response of a conventional image filter by rewriting (6) and left multiplying $\mathbf{D}^{1/2}$ on both sides:

$$\begin{aligned} \mathbf{D}^{1/2}\hat{\mathbf{x}}_{out} &= (\mathbf{I} - \mathcal{L})\mathbf{D}^{1/2}\hat{\mathbf{x}}_{in} \\ \mathbf{x}_{out} &= (\mathbf{I} - \mathcal{L})\mathbf{x}_{in}, \end{aligned} \tag{7}$$

where the last step is obtained by defining the normalized input $\mathbf{x}_{in} = \mathbf{D}^{1/2}\hat{\mathbf{x}}_{in}$ and output $\mathbf{x}_{out} = \mathbf{D}^{1/2}\hat{\mathbf{x}}_{out}$. This normalization enables us to have a spectral interpretation of the conventional filter and it also ensures that a normalized constant signal is an eigenvector of $\mathcal{L}$ associated with zero eigenvalue [7]. A graph vertex domain filtering interpretation of a conventional image filter is thus obtained by comparing (3) and (7):

$$\mathbf{H}_c = \mathbf{I} - \mathcal{L}. \tag{8}$$

In the graph spectral domain, the response of the graph spectral filter $\mathbf{H}_{1,c}$ is:

$$\mathbf{H}_{1,c} = \mathbf{U}^t\mathbf{H}_c\mathbf{U} = \mathbf{I} - \mathbf{\Lambda}. \tag{9}$$

This shows that the conventional image filter penalizes the higher frequencies while allowing lower frequencies to pass in the graph spectral domain, as illustrated in Fig. 3.

An extension of the conventional image filters using the graph spectral filtering has been proposed in [7]. After defining a graph from the conventional filter weights, the signal defined on the vertices can be processed using a graph filtering operation. It is demonstrated that a general design of the graph spectral filter is beneficial and an approximation method using the Chebyshev approximation of the desired spectral response is used in order to avoid expensive diagonalization of the Laplacian matrix. An additional extension of [7] is proposed in this paper with graph specified from a guidance image.

3. GRAPH-BASED JOINT BILATERAL FILTERING AND UPSAMPLING

In this paper, the graph-based image processing approaches presented in the previous section are applied to depth processing problems, especially when depth is present at low resolution and denoising or upsampling is required.

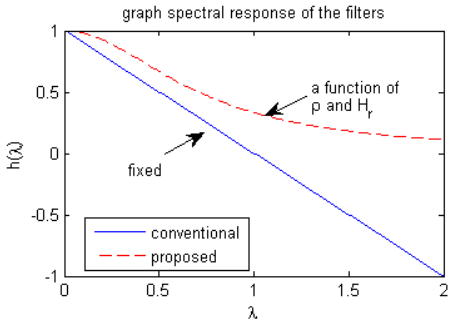


Fig. 3. Comparison of the graph spectral response between the conventional image filter (9) and the proposed one (13).

It is known that directly applying conventional image processing methods to depth images may result in annoying artifacts along object boundaries, so that the rendered object may have artifacts along the contours. Viewers are very sensitive to such artifacts, so the processing of edges in depth images requires special attention. As done in prior work, e.g. [1] [3], we aim to exploit the correlation between color images and depth images along the object boundaries. To achieve this, we consider graph constructions where the weights between the pixels are defined as joint bilateral weights, which, as proposed in [1], take the form:

$$w_{ij} = \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_s^2}\right)\exp\left(-\frac{\left(\tilde{I}[i] - \tilde{I}[j]\right)^2}{2\sigma_r^2}\right). \quad (10)$$

Note that in contrast to the bilateral filter, the weights of the joint bilateral filter are calculated using a guidance image $\tilde{I}$ rather than the input values $\mathbf{x}_{in}$. This is especially effective when the input image is noisy (as in the depth denoising problem) or unavailable (as in the depth upsampling problem).

3.1. Graph-based joint bilateral filtering (GB-JBF)

In a first scheme, we assume that a conventional upsampling algorithm, e.g., a bilinear filter, is used to produce the depth image at full resolution, as shown in Fig. 2(a). Since the resulting image will be smooth along depth edges or contain erroneous sample values, we further process the depth image using a proposed graph-based joint bilateral filter (GB-JBF) to denoise the upsampled image.

A regular grid graph is first constructed with nodes only connected to their immediate 8-connected neighbors as shown in Fig. 1 (right) with joint bilateral weights. The color image contains a superset of edges of the depth image, and thus the graph filtering operation will not filter across the depth edges. After the joint bilateral graph is defined, the denoising problem is formulated as a regularization problem as in [7]:

$$\dot{\mathbf{x}} = \arg\min_{\mathbf{x}} \underbrace{\|\mathbf{x}_{noisy} - \mathbf{x}\|^2}_{A} + \rho \underbrace{\|\mathbf{H}_r\mathbf{x}\|^2}_{B}, \quad (11)$$

where $\mathbf{x}_{noisy}$ is the observation or noisy depth image, and $\mathbf{H}_r$ is the regularization kernel, which is designed as a high pass graph filter. Part A is a data-fitting term that computes the error between the reconstructed signal and the original signal at the known samples, while part B is the Euclidean norm of the output of a highpass graph filter $\mathbf{H}_r$. The closed form solution is given by:

$$\dot{\mathbf{x}} = \left(\mathbf{I} + \rho\mathbf{H}_r^t\mathbf{H}_r\right)^{-1}\mathbf{x}_{noisy}. \quad (12)$$

Compared with the conventional image filters, such as BF and JBF, whose graph spectral responses are fixed, the proposed scheme allows us to design the response of the filter in the graph spectral domain. An illustration of the comparison is shown in Fig. 3. In this paper specifically, we choose $h_{1,r}(\lambda) = \lambda$ as the regularization kernel. The parameter ρ in (11) is pre-determined to balance the two terms. Hence the spectral response of the whole denoising filter is

$$h_1(\lambda) = \frac{1}{1 + \rho\lambda^2}, \quad (13)$$

which is a lowpass filter.

Further, Chebyshev approximation is used to approximate the spectral response of the graph filter using a degree k polynomial, so that the graph filter is a degree k polynomial of the graph Laplacian. This can be interpreted as a k-hop operator operated on a simple regular grid graph. In contrast, from the graph filtering point of view, the conventional filter is a one-hop operator on a more complicated graph with each node connected to all pixels in the window.

3.2. Graph-based joint bilateral upsampling (GB-JBU)

In the second scheme, a graph-based joint bilateral upsampling (GB-JBU) is proposed to upsample the low-resolution depth image using the corresponding color image, as shown in Fig. 2(b). Similar to [8], the depth upsampling problem is formulated as a graph interpolation problem with regularization:

$$\dot{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{J}(\mathbf{x}_{du} - \mathbf{x})\|^2 + \rho\|\mathbf{H}_r\mathbf{x}\|^2, \quad (14)$$

where $\mathbf{J} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ denotes the downsampling operator. M is the size of the known subset of samples, and N is the size of the full set of samples. $\mathbf{J}$ can be represented in the following form by appropriate permutation: $\mathbf{J}_{M\times N} = \left(\mathbf{I}_{M\times M}|\mathbf{0}_{M\times(N-M)}\right)_{M\times N}$.

$\mathbf{x}_{du}$ is the down-upsampled signal and can be expressed in the following form with appropriate permutation $\mathbf{x}_{du} = [\mathbf{x}(S)^t, \mathbf{0}(S^c)^t]^t$, and $\mathbf{H}_r$ is a regularization graph filter which is designed as high pass. Note that the weights follow the same form as the GB-JBF in (10). The solution is then given as:

$$\dot{\mathbf{x}} = \left(\mathbf{J}^t\mathbf{J} + \rho\mathbf{H}_r^t\mathbf{H}_r\right)^{-1}\mathbf{x}_{du}. \quad (15)$$

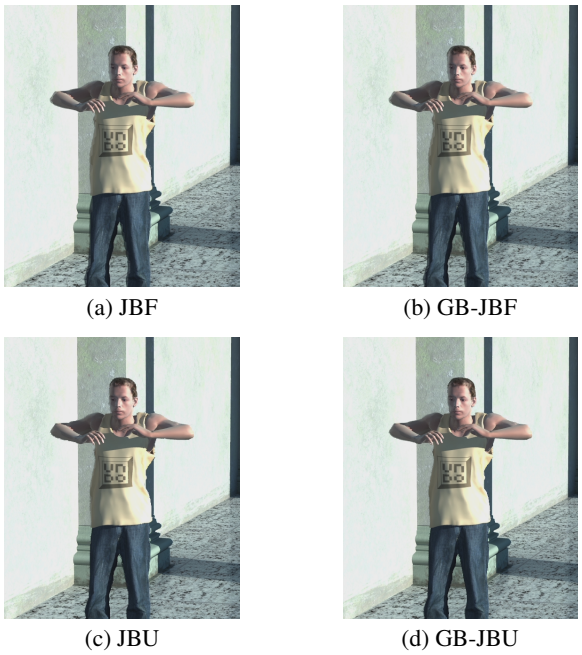
To reduce the complexity, the down-upsampled high resolution image is first partitioned into blocks with overlapped borders. We explored two approaches using two different graph structures: start graph and 8-connected graph. In the first approach, a star graph is constructed for each missing pixel in the block. The node at the center represents the missing pixel and the graph links are connected to nodes corresponding to known pixels within the block. In the second approach, an 8-connected graph is constructed on the down-upsampled image block. Then all missing pixels are interpolated in a single step using (15).

4. EXPERIMENTS AND DISCUSSIONS

Experiments are conducted to study the performance of the proposed graph-based joint bilateral approaches. They are tested in the scenario shown in Fig. 2. The inputs are uncompressed depth videos with the down-sampling factor of 4 in both horizontal and vertical

Table 1. View synthesis PSNR[dB] for depth upsampling; *UD*: *Undo_Dancer*, *GT*: *GT_Fly*; S: star-graph; 8: 8-connected graph

seq.	view	BF	JBF	GB-JBF	JBU[1]	GB-JBU(S)	GB-JBU(8)
UD	left	37.30	36.96	36.69	39.95	**39.97**	39.73
	right	35.84	36.25	37.17	40.16	**40.24**	40.12
	avg.	36.57	36.60	36.93	40.06	**40.10**	39.92
GT	left	45.77	46.02	46.07	49.22	**49.30**	47.39
	right	45.60	45.83	45.89	49.27	**49.36**	47.44
	avg.	45.69	45.93	45.98	49.24	**49.33**	47.41



(a) JBF (b) GB-JBF

(c) JBU (d) GB-JBU

Fig. 4. Example view synthesis results using different algorithms performed on down-sampled depth map with a factor of 4 in each direction.

directions. For GB-JBF, a bilinear filter is used to upsample the input depth map to the high resolution before GB-JBF is used to correct the resampling error. GB-JBU, on the other hand, is performed directly on the down-sampled depth video. The reconstructed depth videos together with their co-located texture videos are then used for view synthesis. For each test sequence, there are three input views and two synthesized views specified in the Common Test Condition (CTC) in JCT-3V [11].

Two reference methods are compared with GB-JBF: BF and JBF. A fast algorithm proposed in [12] is used for BF and JBF experiments. An implementation of the block-based JBU is compared with GB-JBU so the the supporting pixels are the same for both approaches. In both graph-based approaches, the regularization kernel is selected as a high pass filter $h_1(\lambda_i) = \lambda_i$, and $\rho = 1$. For GB-JBF, we used a polynomial of degree 5 to approximate the original graph spectral response. For filtering, $\sigma_s = 2, \sigma_r = 0.2$. For upsampling, the block size is set to 9, $\sigma_s = 7$, $\sigma_r = 0.06$. For GB-JBU, unnormalized Laplacian is used.

Table 1 shows the average PSNR over 10 frames between virtual views synthesized by the original depth and the reconstructed depth for the test sequences *Undo_Dancer* (*UD*) and *GT_Fly* (*GT*). The subjective results are consistent with the numerical ones. We present sample images in Fig. 4.

It can be observed that the one-step upsampling approaches (JBU/GB-JBU) are in general better than the two-step approaches in terms of objective quality and subjective quality. Although the filtering after the bilinear filter corrected some errors, we noticed that the object boundaries in the virtual view are blurred. This is due to the bilinear filter processes across edges in the depth map. This negative effect can be reduced by the one-step upsampling approaches.

In terms of objective quality, both proposed graph-based approaches provide a slight gain over the conventional ones. Specifically, GB-JBF is 0.05-0.3dB better than JBF on average. This shows that with a simpler graph (8-connected) and polynomial approximation one can make use of a larger bilateral filter window and get some gains (without much higher complexity). Note here we use an 11×11 window for GB-JBF instead of 5×5 for JBF, but it's 11×11 on the simpler graph. The GB-JBU with the star-graph is 0.05-0.1 better than JBU with the same settings.

Note that it is possible to replicate exactly JBU in the graph domain by using $\mathcal{L}$. Thus the main difference between JBU and GB-JBU(S) is that the approach in GB-JBU(S) is based on L. Filtering approaches based on $\mathcal{L}$ and L will differ only in vertices whose degrees are different from those of neighboring ones (since the normalization is based on degree). That is, they will operate differently near image edges, but will lead to the same results in areas with no image edges. This explains why the difference in performance is small, since the interpolation results will be identically the same in most flat regions, and will differ only where images edges are present.

The current result for GB-JBU(8) is worse than both the JBU and GB-JBU(S). This is mainly because the weak links in the sparse graph. If no weak link prevention method is used, some thin outliers can be observed in the interpolated depth map. So we manually set a hard threshold to prevent weak links, and it has effectively suppressed the outliers. However, the threshold is empirical and may have deteriorated the objective quality. This may be a research topic of our future work.

5. CONCLUSIONS

In this paper, a graph-based joint bilateral filtering (GB-JBF) and a graph-based joint bilateral upsampling (GB-JBU) are developed and evaluated for depth map filtering and upsampling in the context of 3D video rendering applications. Both the filtering and interpolation problems are formulated as a regularization problem as in [7] and [8]. This paper proposes to use joint bilateral weights to define the graph weights such that the depth image can be filtered in the resulting graph spectral domain. We have also studied the trade-off using different graph structures.

Several extensions of the work may be considered, e.g. the design of the optimal regularization kernel. A single-step depth denoising and upsampling is to be evaluated.

6. REFERENCES

[1] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[2] S. Liu, P. Lai, D. Tian, and C. W. Chen, "New depth coding techniques with utilization of corresponding video," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 551 –561, June 2011.

[3] Y. Wang, D. Tian, and A. Vetro, "Local depth image enhancement scheme for view synthesis," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2725–2728.

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

[5] 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)*, 2012.

[6] W. Hu, X. Li, G. Cheung, and O. Au, "Depth map denoising using graph-based transform and group sparsity," in *IEEE International Workshop on Multimedia Signal Processing*, 2013.

[7] A. Gadde, S. K. Narang, and A. Ortega, "Bilateral filter: Graph spectral interpretation and extensions," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept. 2013.

[8] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Signal and Information Processing (GlobalSip), 1st IEEE Global Conference on*, Dec. 2013.

[9] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[11] D. Rusanovskyy, K. Mueller, and A. Vetro, "Common test conditions of 3dv core experiments," *JCT3V-E1100, Joint Collaborative Team on 3D Video between MPEG and ITU-T VCEG*, Jul. 2013.

[12] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Computer Vision–ECCV 2006*, pp. 568–580. Springer, 2006.