Depth-Assisted Stereo Video Enhancement Using Graph-Based Approaches

Tian, D.; Mansour, H.; Vetro, A.; Wang, Y.; Ortega, A.

TR2014-091 October 2014

Abstract

In stereo video applications, the quality of the two views may vary based on different camera capturing conditions and setup, compression/transmission, and sensor noise. Although some studies show that the perceived video quality may not be significantly affected by the lower quality view, maintaining a similar video quality is still desired in order to prevent eye strain during extended viewing sessions. In this paper, we study a graph-based approach to enhance the lower quality views by referring to the high quality view in addition to an accompanying depth map. We construct a graphical signal model with joint bilateral edge weights and show that graph-based joint bilateral filtering can better suppress several types of noises, e.g., Gaussian, motion as well as quantization noise.

IEEE International Conference on Image Processing (ICIP)

DEPTH-ASSISTED STEREO VIDEO ENHANCEMENT USING GRAPH-BASED APPROACHES

Dong Tian, Hassan Mansour, Anthony Vetro

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

Yongzhe Wang, Antonio Ortega

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

ABSTRACT

In stereo video applications, the quality of the two views may vary based on different camera capturing conditions and setup, compression/transmission, and sensor noise. Although some studies show that the perceived video quality may not be significantly affected by the lower quality view, maintaining a similar video quality is still desired in order to prevent eye strain during extended viewing sessions. In this paper, we study a graph-based approach to enhance the lower quality views by referring to the high quality view in addition to an accompanying depth map. We construct a graphical signal model with joint bilateral edge weights and show that graph-based joint bilateral filtering can better suppress several types of noises, e.g., Gaussian, motion as well as quantization noise.

Index Terms— 3D video, graph-based filtering, graph-based joint bilateral filtering (GB-JBF)

1. INTRODUCTION

In stereo video applications, the two views are often presented at different quality levels due to variation in the source signals. For example, changes of brightness or color may be produced by the imaging sensor and circuitry of the stereo camera or even from shot noise. Another example is asymmetric coding of stereo video to reduce the bitrate required for storage/transmission, where one of the views is compressed at a lower resolution and later upsampled at the receiver side before display. The down/upsampling process applied on one of the views would typically lead to unequal quality in the stereo video.

Though studies show that little deterioration in one view would not degrade the perceived 3D quality, equal quality is still desired in order to avoid visual strain from long time viewing [1]. Traditional image enhancement techniques may be used to improve the quality of the degraded view. For example, if the low quality view is captured at a lower resolution, multi-image or example-based super resolution (SR) techniques may be used to match the resolutions of both views. Furthermore, with emerging 3D video formats, e.g., 3D-AVC and 3D-HEVC, the depth signal is included as part of the data format. Hence, it is desirable to further exploit the depth information to enhance a low quality view.

In [2], Garcia et al. proposed to use depth information for super resolution of multiview images. The approach uses the depth map to generate high frequency content from a neighboring full resolution view that is warped to the viewpoint of the low resolution image. The low-resolution image is then enhanced by exploiting the warped high frequency content. Depth information is required in the warping procedure and in handling occlusions.

In recent years, the emerging graph signal processing tools have been applied to classical image processing tasks [3]. For example, a typical interpolation problem was studied using spectral graph theory in [4], where the upsampling problem is formulated as a regularized least squares problem. In our previous work, we extended this approach to depth image upsampling [5] and demonstrated the benefits from graph spectral domain processing.

In this paper, we further study applying the graph-based approach to enhance the texture view in a stereo video with more emphasis on denoising than interpolation. It is demonstrated that the proposed approach can effectively suppress several types of noises including Gaussian, motion as well as quantization noise. The presented method is also an extension of the depth assisted framework of Garcia et al. [2]. In particular, a graph-based image enhancement approach is employed to remove different sensor noises from the low quality views; furthermore, preliminary studies to remove the quantization artifacts from compression are conducted also. We formulate the image enhancement problem as a filtering procedure applied in the graph spectral domain. We abstract the image as a graph signal, where the graph vertices are the image pixels and the graph edges connect pixels that lie in close spatial neighborhood. Moreover, we use joint-bilateral weights derived from the warped high-quality view to specify the edge weights.

The remainder of the paper is organized as follows. Section 2 provides background on graph-based image processing. Section 3 constructs a joint bilateral graph for stereo image interpolation/denoising and then uses a graph-based joint bilateral filter (GB-JBF) formulation for graph spectral denoising problems in stereo images. In Section 4, we performed experiments and verified the types of noise that could be efficiently suppressed by GB-JBF approaches.

2. GRAPH-BASED IMAGE PROCESSING

2.1. Basics of Images on Graphs

In general graph signal processing [3], 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 weights w_{ij}. For image processing applications, a pixel may be treated as a node in a graph. 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 $\mathcal{L} := \mathbf{D} - \mathbf{W}$.

It is noted that every graph design can be associated to an underlying conventional image filtering. More specifically, for an input image $\hat{x}_{in}$, and the filtered output image $\hat{x}_{out}$, can be written as,

$$\hat{x}_{out}[i] = \sum_j \frac{w_{ij}}{\Sigma_k w_{ik}} \hat{x}_{in}[j], \quad (1)$$

or in graph notation,

$$\hat{x}_{out} = \mathbf{D}^{-1}\mathbf{W}\hat{x}_{in}. \quad (2)$$

Furthermore, the normalized Laplacian matrix is defined as $\mathbf{L} := \mathbf{D}^{-1/2}\mathcal{L}\mathbf{D}^{-1/2}$, which is a symmetric positive semi-definite matrix. Hence, it admits an eigendecomposition $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^t$, where $\mathbf{U} = \{\mathbf{u}_1, ..., \mathbf{u}_N\}$ is an orthogonal set of eigenvectors and $\mathbf{\Lambda} = \text{diag}\{\lambda_1, ..., \lambda_N\}$ is its corresponding eigenvalue matrix. The eigenvectors and eigenvalues of the Laplacian matrix provide a spectral interpretation of the graph signals. Note that eigenvalues $\{\lambda_1, ..., \lambda_N\}$ can be treated as graph frequencies and are always situated in the interval $[0, 2]$ on the real line.

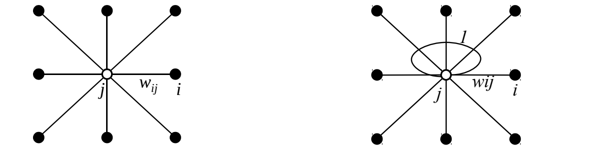
Hence in the graph spectral domain, a *Graph Fourier Filtering* (GSF) $\mathcal{H}$ can be designed for image processing purposes, where $\mathcal{H}$ is a diagonal matrix. The corresponding graph filter $\mathbf{H}$ in the vertex domain can be expressed as,

$$\mathbf{H} = \mathbf{U}\mathcal{H}\mathbf{U}^t. \quad (3)$$

Note that the spatial domain signal needs to be normalized $x = \mathbf{D}^{\frac{1}{2}}\hat{x}$ before applying a GSF that is designed on top of a normalized Laplacian matrix.

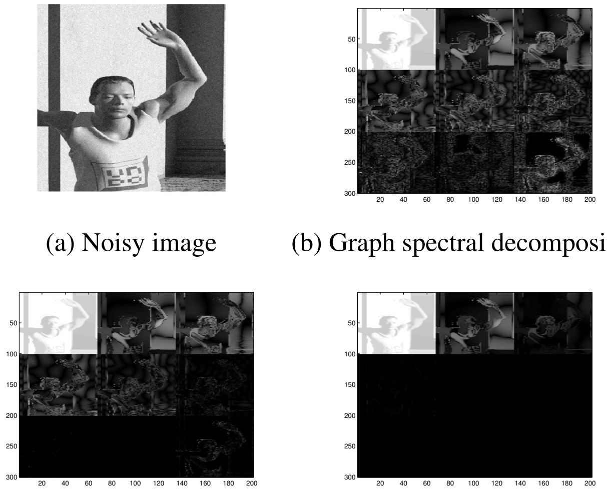
2.2. Graph-based Image Enhancement

Consider the example graph shown in Fig. 1. The center circle pixel is a target pixel to be filtered or interpolated. We adopt bilateral [6] and a joint bilateral [5] filtering kernels where the weights w_{ij} are defined by,



(a) Graph w/o self looping (b) Graph w/ self looping

Fig. 1. Graph structure (a) used in [6] and (b) used for stereo image enhancements in this paper.



(a) Noisy image (b) Graph spectral decomposition

(c) JBF filtering (d) Proposed GB-JBF filtering

Fig. 2. Graph spectral decomposition and GSF filtering.

$$w_{ij} = \exp(-\frac{\|p_i - p_j\|^2}{2\sigma_s^2})\exp(-\frac{(x_{in}[i] - x_{in}[j])^2}{2\sigma_r^2}). \quad (4)$$

The first exponential term is a spatial distance penalty (p_i refers to the pixel's spatial location) and the second exponential term is an intensity distance penalty (x_{in} refers to the intensity value from the guide image). In next section, we will propose to use a warped image as guide image in stereo image enhancement.

Fig. 2 illustrates the concept of graph spectral decompositions and the effects from different graph spectral filtering. Fig. 2(a) shows the original image. Fig. 2(b) depicts how the original image is projected into the eigen-spaces spanned by the eigen vectors of a graph Laplacian matrix $\mathbf{L}$. In order to simplify the illustration, the eigen vectors are first sorted based on the corresponding eigen values in an ascending order. The top-left subband image shown in Fig. 2(b) is the projected image into the eigen-space corresponding to the first eigen vector. Similarly, the other subband images are projections into the eigen-spaces spanned by the remaining eigen vectors. For the other 8 subband images, each of them is projected to an eigen-space spanned by a combination of several eigen vectors so as to limit the total number of subbands to be shown is 9.

3. GRAPH SPECTRAL FILTERING ON STEREO IMAGES

3.1. Joint Bilateral Weights for Stereo Video

We reuse the (joint) bilateral structure as given in (4), but it is proposed to assign the warped pixels from the other view at higher quality to define the intensity domain. The motivations are that the most salient features have already been preserved in the low quality image, and the details that are missing cannot be recovered by simply warping the other view. Instead, with the accompanying depth signal, the other view is warped and used to as a guidance image. A standard depth image based rendering (DIBR) process is invoked to generate the guidance image as described below.

For each pixel location $i = [u, v]$ in the current view, a corresponding location $i' = [u', v']$ can be determined based on the camera parameters and the pin hole camera model. Hence, $x_{in}(i)$ in the warped image is obtained from $x'_{in}(i')$ in the other view at high quality. Note that though $[u, v]$ is always at a pixel at integer location, $[u', v']$ may point to a subpixel location. As a result, subpel interpolation in the high quality view needs be performed to maintain the accuracy. In this paper, we used the 8- or 7- tap interpolation as defined in H.265/HEVC video coding standard [7].

Dis-occlusions may happen near to the foreground objects during the warping process, which are marked as hole areas. A typical hole filling algorithm would propagate the background pixels into the holes using inpainting approaches. Several ways may be used to handle the disoccluded area. One way is to use the filled pixels or the original noisy pixels to determine the JBF weights. However, pixels filled in such a way or the noisy pixels are assumed to be too unreliable to be used as guidance image values. In our experiments, we used an alterative traditional filtering method (bilinear filter) to handle the disoccluded area.

By defining the joint bilateral weights, a conventional joint bilateral filter (JBF) can be designed by referring to (2), that is, the JBF is the underlying filter to construct the graph signal. We can see that a graph filter $\mathbf{H}$, derived from the graph spectral filter $\mathcal{H} = \mathbf{I} - \Lambda$ as in (5), is quivalent to the underlying JBF filter $\hat{x}_{out} = \mathbf{D}^{-1}\mathbf{W}\hat{x}_{in}$.

$$\begin{aligned} x_{out} &= \mathbf{U}(\mathbf{I} - \Lambda)\mathbf{U}^t x_{in} \\ &= (\mathbf{I} - \mathbf{L})x_{in}. \end{aligned} \tag{5}$$

So $\mathcal{H}_{JBF} = \mathbf{I} - \Lambda$ is an illustration of a conventional filter in graph spectral domain, which is actually a band pass filter in graph spectral domain and can suppress the frequency noises in the center of the graph spectral domain. As demonstrated in Fig. 2(c), the lowest frequency band and highest frequency band images would be mostly maintained; while the middle frequency band images are removed. JBF will be used as a benchmark to study the additional gains from the proposed graph based filtering. Our next target is to design a strict low pass GSF filter $\mathcal{H}_{GB-JBF}$ to replace $\mathcal{H}_{JBF}$.

3.2. Graph Spectral Denoising

In this section, we present how to design a graph-based joint bilateral filter (GB-JBF) $\mathcal{H}_{GB-JBF}$ to filter stereo images with accompanying depth information available. The graph G is constructed as shown in Fig. 1(b) using the joint bilateral weights defined in section 3.1. Note that there is a self loop link with weight 1 in the graph, but it will not be reflected in the Laplacian matrix as it is canceled when constructing the combinatorial Laplacian matrix $\mathcal{L} = \mathbf{D} - \mathbf{W}$.

The graph spectral denoising problem is formulated as a regularization problem as in [6],

$$\dot{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - f\|^2 + \frac{\rho}{2}\|\mathbf{H}x\|^2, \tag{6}$$

where f is the noising image, ρ is a weighting factor to balance the regularization problem. The above formulation provides a way to find out the optimal GB-JBF filter design in terms of least square errors. There is a closed form solution given by,

$$\begin{aligned} \dot{x} &= \mathbf{U}(\mathbf{I} + \rho\mathcal{H}^2)^{-1}\mathbf{U}^t f \\ &= (\mathbf{I} + \rho\mathbf{H}^t\mathbf{H})^{-1} f. \end{aligned} \tag{7}$$

Provided the penalty function $\mathbf{H}$ being a high pass filter, GB-JBF filter $\mathcal{H}_{GB-JBF} = (\mathbf{I} + \rho\mathcal{H}^2)^{-1}$ is a low pass filter.

4. SIMULATIONS

In this section, the denoising capability of the proposed GB-JBF filter is compared against its underlying JBF filter with $\mathcal{H}_{JBF} = \mathbf{I} - \Lambda$. The penalty function in (6) is chosen to be $\mathcal{H} = \Lambda$, and hence the GB-JBF filter is $\mathcal{H}_{GB-JBF} = (\mathbf{I} + \rho\Lambda^2)^{-1}$. The regularization parameter in (6) is selected as $\rho = 1000$ in our experiments.

Two stereo video are used to demonstrate the cases when a graph spectral filter may bring some benefits. The first video is Undo_Dancer with ground truth depth, and the second video is Kendo with depth being estimated from texture video [8]. Both sequences have multiple views available and a stereo pair is selected for the tests. The selected left view is kept intact, denoting a view with higher quality, while the selected right view is degraded with a few types of noises, including Gaussian white noise, Gaussian blurring, motion blurring and compression artifacts.

Fig. 3 shows the PSNR values obtained for JBF and GB-JBF filtering results.

Noise Types	Undo_Dancer		Kendo	
	JBF	GB-JBF	JBF	GB-JBF
Gaussian Noise	34.78	34.11	36.10	36.37
Gaussian Blurring	32.71	33.09	36.75	37.68
Motion Blurring	33.61	34.33	38.35	39.79
Coding Artifacts	33.20	33.72	38.20	39.54

Fig. 3. Filtering results in PSNR (dB) with various noise, GB-JBF vs. JBF.

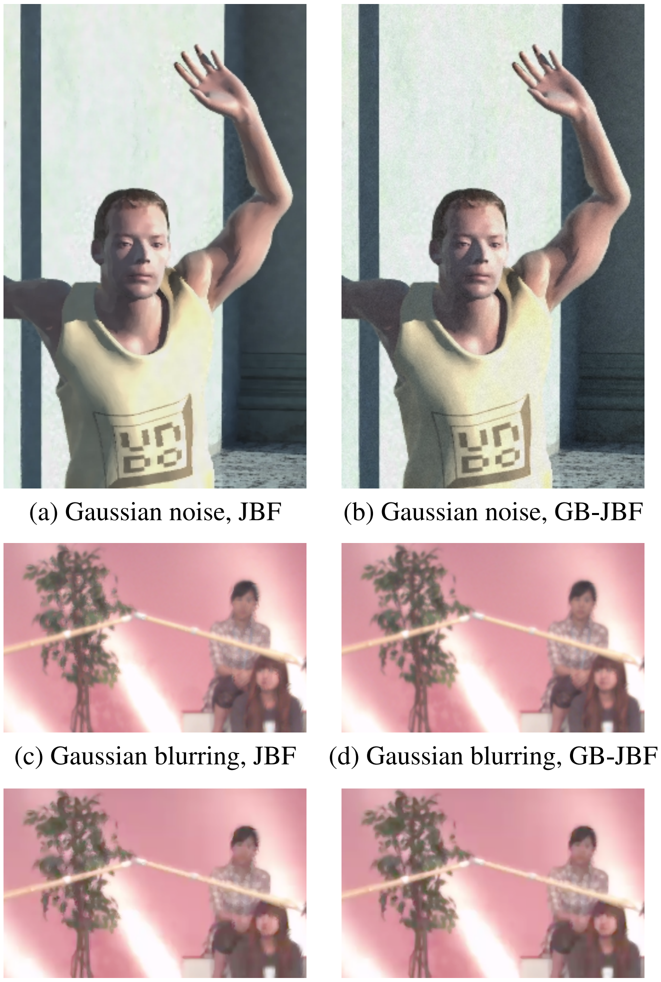
For Gaussian white noise with a constant mean and variance, GB-JBF is clearly better than JBF in terms of both objective and subjective measures for Kendo. With Undo_Dancer, it is noted that JBF shows higher PSNR than GB-JBF. Perceptually, however, GB-JBF maintained more details in the image while keeping the object boundaries sharper. See the wall in the background and arm boundary in Figs. 4(a) and 4(b). With JBF, there is a fake contour around the hand raised aloft and many details in the image are removed. In addition, Fig. 2(d) shows the effects of the proposed GB-JBF on the subband images. Compared to JBF as in Fig. 2(c), it could be noticed that GB-JBF could suppress more high frequency noises while still maintaining the low frequencies well.

The Gaussian blurring noise are added with a 3 × 3 window and $\sigma = 1$. It is noticed that both PSNR and subjective quality of GB-JBF are better than JBF. Figs. 4(c) and 4(d) shows the JBF and GB-JBF results, and GB-JBF has much less annoying noises compared to JBF.

Similarly, with motion blurring within a range of 3 pixels, GB-JBF is more favorable than JBF in terms of both PSNR and subjective quality.

Finally, in a compression system for stereo video, one view may be compressed more severely and have more coding artifacts, e.g., coded at a lower resolution by asymmetrical compression techniques or coded using a higher QP. In one example experiment, one view is degraded more from a high QP 34, and the PSNR of the filtered images are reported in Fig. 3. GB-JBF showed higher PSNR, and more enhanced visual quality. In Figs. 4(e) and 4(f), the snapshots from Kendo sequence show that GB-JBF demonstrates much sharper edges than JBF along the shoulder of the lady in the front and around the leaves of the plant. It is subject to future work to study how robust GB-JBF performs with compressed depth.

As the graph based approach involves eigen decomposition, GB-JBF would incur higher complexity than JBF. Fortunately, with the GB-JBF implemented in this paper, we employed a local graph for each pixel and the graph size is limited by the range of the neighborhood pixels, which is selected as 7 × 7 window (instead of 3 × 3 as in Fig. 1). With a MATLAB implementation and the selected parameters in the above tests, GB-JBF takes about 20% more run time than JBF.



(a) Gaussian noise, JBF (b) Gaussian noise, GB-JBF

(c) Gaussian blurring, JBF (d) Gaussian blurring, GB-JBF

(e) Coding artifacts, JBF (f) Coding artifacts, GB-JBF

Fig. 4. Denoising snapshots, GB-JBF vs. JBF.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we studied the use of graph spectral filtering to enhance a stereo image pair, when one view is more degraded than the other view. The higher quality view and the accompanying depth image are used to enhance the lower quality view. A graph based joint bilateral filter (GB-JBF) was presented. For applications like denoising, deblurring, where all existing pixels are to be altered, it was found that GB-JBF provides better quality in both objective and subjective quality while dealing with several types of noise.

In a 3D coding system, one of the views may be degraded from asymmetrical coding. One way to exploit the proposed GB-JBF is to upsample the low-resolution view with a prior-art method, and then have it enhanced by the GB-JBF approach. It is subject to our future work to study efficient interpolation directly from the low-resolution image using graph based approaches.

6. REFERENCES

[1] L. Stelmach, Wa James Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 2, pp. 188–193, 2000.

[2] D.C. Garcia, C. Dorea, and R.L. De Queiroz, "Super resolution for multiview images using depth information," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 9, pp. 1249–1256, 2012.

[3] 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.

[4] 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.

[5] Y. Wang, A. Ortega, D. Tian, and A. Vetro, "A graph-based joint bilateral approach for depth enhancements," in *to appear in ICASSP 2014*, 2014.

[6] Akshay Gadde, Sunil K Narang, and Antonio Ortega, "Bilateral filter: Graph spectral interpretation and extensions," *ICIP 2013*, October 2013.

[7] G.J. Sullivan, J. Ohm, Woo-Jin Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[8] D. Rusanovskyy, K. Muller, and A. Vetro, "Common test conditions of 3dv core experiments," in *JCT3V meeting, JCT3V-F1100*, October 2013.