ADAPTIVE COMPRESSED SENSING FOR DEPTHMAP COMPRESSION USING GRAPH-BASED TRANSFORM

Sungwon Lee and Antonio Ortega

Ming Hsieh Department of Electrical Engineering
University of Southern California
sungwonl@usc.edu, ortega@sipi.usc.edu

ABSTRACT

In this paper we present an adaptive compressed sensing (CS) framework for depth map compression using a family of graph-based transforms (GBT). To improve overall performance we propose a greedy algorithm that selects for each block a GBT that minimizes a metric that takes into consideration both the edge structure of the block and the characteristics of the CS measurement matrix, using an estimate of average mutual coherence. As compared to coding using H.264/AVC, the proposed approach applied to intra-frames shows an average of 39 % bitrate savings or 3.8 dB PSNR gain for views rendered using a depth image based rendering (DIBR) technique.

Index Terms— Compressed Sensing (CS), Graph-based Transform (GBT), Depthmap Compression

1. INTRODUCTION

Standard compressed sensing (CS) theory prescribes that robust signal recovery is possible when a signal is sparse in a given sparsifying basis. Based on the signal characteristics, the sparsifying basis is often assumed to be known *a priori* at the decoder. However, for coding applications where signals are first captured and then compressed, better performance can be achieved by adaptively selecting a transform or sparsifying basis and then signaling the chosen transform to the decoder. For instance, for piecewise smooth signals, where sharp edges exist between smooth regions, edge-adaptive transforms can provide sparser representation at the cost of some overhead.

In this paper we consider block-based depth map compression as an example application. Previous work has shown that edge adaptive transforms can be more efficient than standard transforms (e.g., DCT) due to the piecewise smooth nature of these signals [1]. Moreover, correct representation of edges is important because errors in edge information lead to significant degradation of the quality of interpolated views in 3-D TV applications [2, 3]. For depth map compression, CS-based methods have been recently proposed. CS is applied by either projecting depth map on random sensing matrix (Cartesian grid sampling technique) [4] or down-sampling 2D-DCT coefficients [5]. However, performance gains achieved by these techniques are limited because the standard DCT is chosen as the sparsifying basis, which is inefficient for coding blocks containing arbitrarily shaped edges (i.e., neither vertical, nor horizontal) separating smooth regions.

To improve the efficiency of depth map coding, a graph based transform (GBT) has been proposed as it provides a sparser representation, especially when arbitrary edges (e.g., diagonal or a mixture of horizontal and vertical edges) exist in a block [1]. This transform is based on representing each block as a graph, where each vertex corresponds to a pixel, and vertices are linked only when no strong edges are present between the corresponding pixels. For any given block, different graphs can be chosen, leading to different transforms, which depend on the edge structure and therefore requiring that overhead bits be sent to the decoder. In [1] it was shown that these adaptive GBTs improved performance as compared to DCT-only methods, even when the overhead was taken into account. This work was further extended in [6], which proposed a simple cost function and a search technique to optimize the GBT selection for each block, balancing the increased sparseness achievable if more edges are considered, with the added overhead required for transmitting this information to the decoder.

In this paper, we propose a novel CS approach where the adaptive GBT is used as a block-adaptive sparsifying basis. We consider the problem of, given a specific sensing matrix (a Hadamard matrix in our work), optimizing the choice of GBT, by taking into account the quality of reconstruction and the overhead required to specify the GBT. Note that the approach in [6] aims at selecting a GBT that provides maximum sparsity for a block, without requiring excessive overhead. A key result in this paper is to show that maximum sparsity *does not* guarantee optimal performance when using CS. As studied in [7, 8], CS reconstruction depends on not only the sparsity of signal representation but also the mutual coherence between sensing matrix and sparsifying basis. Thus, a GBT providing the sparsest representation of depth map data is not necessarily maximally incoherent with a given sensing matrix. Thus, joint optimization is required to select best GBT for a given depth map, taking into account rate overhead (to specify the transform), sparsity of the representation and mutual coherence. We propose a greedy iterative algorithm that evaluates a metric for different edge configurations before selecting one. This algorithm uses a low-complexity estimate of the mutual coherence, so that explicit construction of the GBT at the encoder is only required once the edge map has been selected (i.e., it is not required in the iterative process leading to this selection). The proposed block adaptive CS approach is integrated within an H.264 codec. When evaluating its intra coding performance on three depth map sequences, we observe 3.8 dB PSNR gain in the quality of interpolated views obtained from the decoded depth map, or an average of 39 % bitrate savings.

2. PROBLEM FORMULATION

For the construction of GBT, each depth block is represented as a graph, $G(V, E)$ with nodes (pixels) and links (connections) between pixels. Note that we only use the term "edge" only to refer to image edges in order to avoid confusion. A link is present in the graph only when no edge was selected between the two corresponding pixels. In this work, we assume 4-neighbor connectivity for the pixels so

that each node, V, can have at most 4 links. From the graph, the adjacency matrix $\boldsymbol{A}$ is formed, where $\boldsymbol{A}(i,j) = \boldsymbol{A}(j,i) = 1$ if pixel positions i and j are immediate neighbors not separated by an edge. Otherwise $\boldsymbol{A}(i,j) = \boldsymbol{A}(j,i) = 0$. Then we compute the degree matrix $\boldsymbol{D}$, where $\boldsymbol{D}(i,i)$ is the number of links connected to i^{th} pixel and $\boldsymbol{D}(i,j) = 0, \ \ \forall i \neq j$. Then, the Laplacian matrix can be computed as:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} = \begin{cases} -1 & if \ \ (i,j) \in E \\ d_i & if \ \ i = j \\ 0 & otherwise \end{cases} \tag{1}$$

A spectral decomposition, defined as the projection of a signal onto the eigenvectors of $\boldsymbol{L}$, can be interpreted as providing the "frequency" contents of the graph signal [9]. Note that $\boldsymbol{L}$ is symmetric, leading to real eigenvalues and a set of orthogonal eigenvectors. Thus we define the GBT for a given graph as the eigenvector matrix, $\boldsymbol{\Psi}$, whose columns are the eigenvectors of the Laplacian $\boldsymbol{L}$ of the graph. Since $\boldsymbol{\Psi}$ is orthogonal, its inverse is $\boldsymbol{\Psi}^T$. In our depth map compression application, block-adaptive GBTs are applied to residual blocks obtained after intra/inter prediction, where the graph from which the GBT is derived is chosen based on the edges present in each residual block. For each block, these edges could be detected by applying a simple threshold to the difference between neighboring residual pixel values [1]. However, using the same threshold for all blocks is suboptimal because it does not take into account the overhead required to transmit the chosen edge map to the decoder, which tends to increase with the number of edges. Thus, two blocks may achieve similar levels of sparsity for a given threshold, but the block where more edges are identified may require a higher overall rate. As an alternative, the work in [6] seeks to find the optimal edge map for each block by considering this overhead.

A key observation in our work is that the optimal GBT (which [6] attempts to obtain) may not provide optimal performance if CS used. This is because performance depends both on the level of sparsity and on the incoherence between the sparsity basis and the measurement matrix, which is very important for reconstruction as studied in [7, 8]. Thus, for any GBT chosen as the sparsifying basis, $\boldsymbol{\Psi}$, we can compute the mutual coherence, $\mu(\boldsymbol{\Phi\Psi})$. Based on the mutual coherence, the minimum number of measurements for perfect reconstruction can be computed by $M = O(K\mu^2(\boldsymbol{U})N \log N)$ [7]. The bound on the number of measurements decreases as $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ becoming increasingly incoherent. Thus, if we can derive how μ changes for different GBT's, then we can also compute a bound on the number of measurements, which would then helps us design a cost function to select the best GBT given that CS will be used.

3. OPTIMAL GBT FOR COMPRESSED SENSING

3.1. Bound on the mutual coherence

In this work, we derive bound on the mutual coherence for the given GBT matrix, $\boldsymbol{\Psi}$, and Hadamard matrix, $\boldsymbol{\Phi}$. Since both matrices are deterministic, it is straightforward that the mutual coherence is also deterministic, so the mutual coherence could be computed for each candidate GBT. However, GBT construction is a complex operation as it requires finding all the eigenvectors of the Laplacian matrix. The complexity grows as the size of graph (equivalently, the size of block in depth map compression) increases. Even if there exist only a few GBTs that are truly useful and for those we could precompute the mutual coherence, the number of candidates of useful GBTs also increases with the size of graph, which leads to larger memory requirements. Thus it would be desirable to avoid having to construct GBTs at every stage of the search for the optimal GBT. In what follows, we derive upper and lower bounds on the mutual coherence then use their average to estimate the mutual coherence of the block.

We first derive the upper bound of the mutual coherence.

Theorem 3.1. *For a given graph $G(V,E)$, the mutual coherence, μ, between Hadamard sensing matrix, $\boldsymbol{\Phi}$ and a graph-based transform matrix, $\boldsymbol{\Psi}$, satisfies*

$$\mu \leq \sqrt{\frac{\max_{\forall i} N_{G_i}}{N}},$$

where N_{G_i} denotes the size of signal (equivalently, the number of pixels) in the group i. If a graph is connected, then all the pixels belong to one group (segment) thus the mutual coherence is bounded by 1 because the DC component of the Hadamard basis is identical to the eigenvector corresponding to the zero eigenvalue of the graph Laplacian. In contrast, a fully disconnected graph where all the pixels are separated by edges can achieve the minimum bound for the mutual coherence. However, this increases the overhead to encode the edge map so that the coding gain is limited. The proof is trivial because all the entries of Hadamard matrix are $\pm 1/\sqrt{N}$ and all the basis functions of GBT (columns of $\boldsymbol{\Psi}$) are normalized to 1, thus the maximum absolute value of the inner-products is bounded by the maximum size of group normalized by N. Next, a lower bound on mutual coherence is derived.

Theorem 3.2. *For a given graph $G(V,E)$, mutual coherence, μ, between an arbitrary sensing matrix, $\boldsymbol{\Phi}$ and a graph-based transform, $\boldsymbol{\Psi}$, satisfies*

$$\mu \geq \max_{\forall k} \sqrt{\frac{\sum_{(l,m)\in E} \left(\boldsymbol{\Phi}(k,l) - \boldsymbol{\Phi}(k,m)\right)^2}{2|E|}},$$

where $|E|$ indicates the number of links between pixels, which equals to 24 (total possible number of edges in 4×4 block) - the number of edges in the block. The numerator of the bound is a squared sum of difference of $\boldsymbol{\Phi}(i,j)$ corresponding to connected pixels where no edge exists. The bound indicates that the lower bound of the mutual coherence increases as more pixels corresponding to high variation of $\boldsymbol{\Phi}$ are connected.

The proof is based on the fact that $\boldsymbol{x^T L x} = \sum_{(i,j)\in E}(x(i) - x(j))^2$, for any $\boldsymbol{x} \in \Re^V$. Let $\boldsymbol{x}^T = \boldsymbol{\Phi}(k,:)$. Since $L = \boldsymbol{\Psi\Lambda\Psi}^T$,

$$\boldsymbol{x^T L x} = \boldsymbol{\Phi}(k,:)(\boldsymbol{\Psi\Lambda\Psi}^T)\boldsymbol{\Phi}(k,:)^T \tag{2}$$

$$= \sum_{(l,m)\in E} \left(\boldsymbol{\Phi}(k,l) - \boldsymbol{\Phi}(k,m)\right)^2, \ \ k \in \{1,2,\ldots,N\} \tag{3}$$

(2) can be expressed as follows:

$$\boldsymbol{\Phi}(k,:)(\boldsymbol{\Psi\Lambda\Psi}^T)\boldsymbol{\Phi}(k,:)^T = \boldsymbol{U}(k,:)\boldsymbol{\Lambda}\boldsymbol{U}(k,:)^T = \sum_i \lambda_i \boldsymbol{U}(k,i)^2, \tag{4}$$

where $\boldsymbol{U} = \boldsymbol{\Phi\Psi}$ and λ_i is i^{th} eigenvalue of Laplacian matrix of a given graph. From (3) and (4),

$$\sum_{(l,m)\in E} \left(\boldsymbol{\Phi}(k,l) - \boldsymbol{\Phi}(k,m)\right)^2 = \sum_i \lambda_i \boldsymbol{U}(k,i)^2 \tag{5}$$

$$\leq \left(\sum_i \lambda_i\right)\left(\max_{\forall i} \boldsymbol{U}(k,i)\right)^2, \tag{6}$$

where $\sum_i \lambda_i = Trace(\boldsymbol{L}) = 2|E|$ because the total sum of diagonal entries in $\boldsymbol{L}$ is the twice of the total number of links between pixels. From (6), we have

$$\max_{\forall i} |\boldsymbol{U}(k,i)| \geq \frac{\sum_{(l,m)\in E} \left(\boldsymbol{\Phi}(k,l) - \boldsymbol{\Phi}(k,m)\right)^2}{2|E|} \tag{7}$$

Thus, the lower bound of the mutual coherence is derived:

$$\mu = max_{i,j}|\boldsymbol{U}(i,j)| = \max_{\forall(k,i)} |\boldsymbol{U}(k,i)|$$
$$\geq \max_{\forall k} \sqrt{\frac{\sum_{(l,m)\in E} \left(\boldsymbol{\Phi}(k,l) - \boldsymbol{\Phi}(k,m)\right)^2}{2|E|}} \quad (8)$$

Note that both lower bound, μ_{lower}, and upper bound, μ_{upper}, can be computed without constructing GBT. The upper bound is determined by the maximum size of group in the graph and the lower bound by the edge map and the given Hadamard sensing matrix. To approximate the mutual coherence between the two bases, we take the average $\mu_{avg} = \frac{\mu_{lower}+\mu_{upper}}{2}$. Since the mutual coherence is the maximum correlation between two bases, the mutual coherence can be misleading especially when only a few correlations are large but the others are small. Thus, instead of looking at the maximum correlation, the average of a certain amount of the largest correlation provides better estimate for the CS performance as studied in [10]. Thus, μ_{avg} can be used as an alternative metric to the mutual coherence. The averaged mutual coherence will be used to approximate the rate for CS measurements to find optimal adjacency matrix, which will be covered in the following section.

3.2. Iterative GBT construction for CS

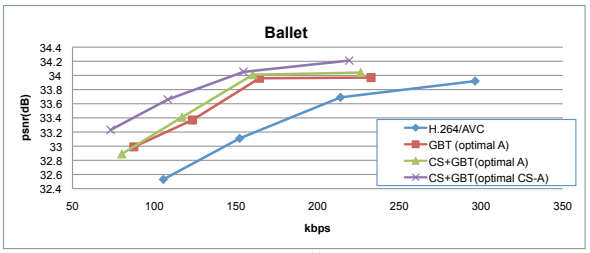
To find the best sparsifying basis, $\boldsymbol{\Psi}$, we iteratively evaluate a series of adjacency matrices using their average mutual coherence, μ_{avg}. We assume 4-neighbor connectivity in 4-by4 block, for simplicity, so that there exist 12 horizontal edges and 12 vertical edges. Instead of searching the whole space of 2^{24} possible adjacency matrices, we propose a greedy algorithm to find the optimal adjacency matrix. By defining a cost function in (9), the cost for removing each edge can be calculated. At the initial state, there exist edges when the pixel values between neighboring pixels are different. Thus, in the initial graph, all the links between pixels with the same value are connected. The algorithm iteratively finds a link with the minimum pixel difference at every iteration then add the link if the updated cost is smaller than the one in previous iteration. After searching all the links excluding the links in the initial state, the approach can find the optimal adjacency matrix. The cost function to be used in the algorithm is defined as:

$$\begin{aligned} \text{Cost} &= \text{Cost}_{\text{measurement\_rate}} + k\text{Cost}_{\text{edge\_rate}} \\ &= \log_2\left(\frac{\left(\sum_i a_{ij}(x_i - x_j)^2\right)\mu_{avg}^2}{2Q^2}\right) + km. \end{aligned} \quad (9)$$

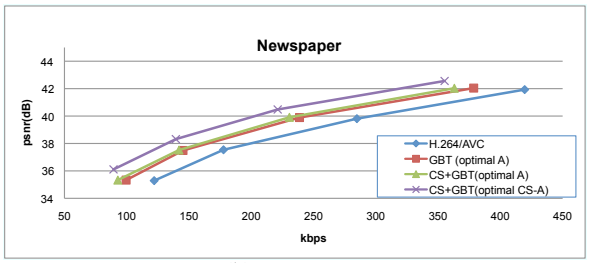
In the cost function Q is a quantization step size. The edge rate is that needed to code the adjacency matrix, which can be represented using 24 bits then compressed using entropy coding. The scaling factor k can be applied to control the trade-off between the coefficient rate and edge rate, which is empirically determined in our experiment. $\mathbf{x}$ is a vector representing the input depth map block thus x_i is the value of pixel i. a_{ij} is the corresponding element in the adjacency matrix thus $\sum_{i,j} a_{ij}(x_i - x_j)^2$ is a squared sum of difference between connected pixels which estimates the cost of GBT coefficient thus it approximates the sparseness of GBT coefficients. Note that the cost function is identical to the one proposed in [6] except for μ_{avg}^2. The average mutual coherence, μ_{avg}, is involved to estimate the rate of measurements because the number of measurements is proportional to $K\mu^2 \log N$ as studied in [7]. Note that we can ignore $\log N$ term because total number of pixels in each block does not change during the algorithm.

4. EXPERIMENTS

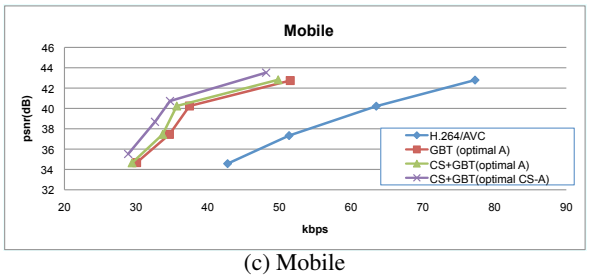
The experiment is based on H.264/AVC reference software JM17.1. For simplicity, only 4×4 transform block size is used in our experiments, but it can be easily extended to other block sizes. As test sequences in our experiments, we use only intra-frames of depth map sequences Ballet, Newspaper, and Mobile. With RD optimization with respect to H.264/AVC, GBT and CS-GBT, the encoder chooses the best mode and transmits extra bits to signal transform mode for each block. For CS-GBT, the encoder encodes 4 Hadamard measurements corresponding to 4 lowest frequency bases. To reconstruct depth map from Hadamard measurements, MOSEK C-library [11] is employed to solve L_1 minimization, which is then integrated into JM17.1.



(a) Ballet



(b) Newspaper

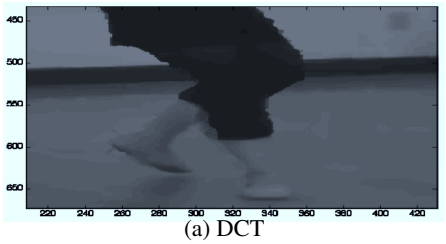


(c) Mobile

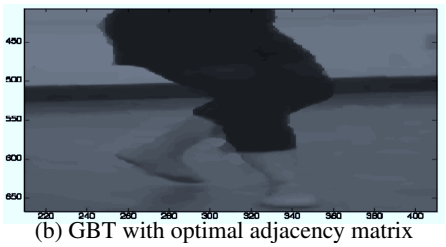
Fig. 1. RD curve comparison of i) H.264/AVC ii) GBT and CS-GBT with optimal adjacency matrix [6] iii) CS-GBT with optimal adjacency matrix with averaged mutual coherence discussed in Section 3.2 for different sequences: (a) Ballet (b) Newspaper (c) Mobile

For comparison, we construct GBT matrix using two different greedy algorithms with different cost metric; i) GBT construction without mutual coherence [6] ii) GBT construction with mutual coherence discussed in Section 3.2. The scaling factor in (9) is empirically chosen as 0.03 which equals to the one in the cost function of [6]. For both cases, the resulting adjacency matrices are entropy coded and sent to the decoder. The decoder can construct the equivalent GBT matrix from the losslessly-encoded adjacency matrix (equivalently, edge map). For CS-GBT approach, one can choose between DCT, GBT, and CS to achieve the best performance. For example, for each block, the RD cost can be calculated for DCT, GBT, and CS. Then the best one is selected. The overhead indi-

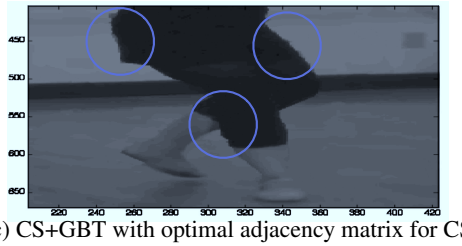
cating the chosen transform is encoded into the bitstream for each block, and the optimal adjacency matrix is provided only for blocks coded using GBT or CS. We consider QP values of 24, 28, 32, and 36 to encode depth maps. As a reference, we also compare those approaches to H.264/AVC for the depth map compression. The reconstruction quality is evaluated by PSNR calculated by comparing the ground truth video and the synthesized video using the decoded depth maps.



(a) DCT



(b) GBT with optimal adjacency matrix



(c) CS+GBT with optimal adjacency matrix for CS

Fig. 2. Perceptual improvement in Ballet sequence: comparison of i) H.264/AVC ii) GBT and CS-GBT with optimal adjacency matrix [6] iii) CS-GBT with optimal adjacency matrix with averaged mutual coherence discussed in Section 3.2

From the comparison of RD curves in Fig. 1, it is observed that significant bitrate savings can be achieved using GBT alone and that further gains can be achieved with CS-GBT.

From the RD curves, it is shown that CS-GBT approach using optimal adjacency matrix considering the average mutual coherence outperforms H.264/AVC. Also, our proposed approach shows better performance than GBT and CS-GBT using optimal adjacency matrix proposed in [6]. Noticeable PSNR improvement over other methods is observed because, with our optimal adjacency matrix for CS, more blocks are chosen to be coded using Hadamard measurements. The performance also depends on the amount of strong edges in a frame and the level of noise around the edges. Among three sequences in our experiment, Mobile sequence contains stronger edges along the object boundary with relatively less noise compared to other sequences, thus it shows the best performance. Also, the perceptual improvement in Ballet sequence is shown in Fig. 2. As marked by blue circles, we can notice clear edges reconstructed by our proposed approach. The results for three different sequences are shown in Table 1 and Table 2 in terms of BD-PSNR and BD-bitrate.

Table 1. BD-PSNR/bitrate results of CS-GBT compared to H.264/AVC.

Sequence	BD-PSNR	BD-bitrate
Ballet	0.9	-49.4
Newspaper	1.5	-26.8
Mobile	9.2	-42.8

Table 2. BD-PSNR/bitrate results of CS-GBT with optimal adjacency matrix for CS compared to CS-GBT with optimal adjacency matrix [6]

Sequence	BD-PSNR	BD-bitrate
Ballet	0.3	-7.8
Newspaper	0.9	-16.1
Mobile	2.4	-9.7

5. CONCLUSION

For depth map compression, we propose a novel CS approach where the adaptive GBT is used as a block-adaptive sparsifying basis. Based on the observation that maximum sparsity does not guarantee optimal performance when using CS, we propose a greedy algorithm that selects for each block a GBT that minimizes a metric that takes into consideration both the edge structure of the block and the characteristics of the CS measurement matrix, using an estimate of average mutual coherence. As compared to coding using H.264/AVC, the proposed approach applied to intra-frames shows a significant gain for interpolated views.

6. REFERENCES

[1] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth-map coding," in *Proc. of 28th Picture Coding Symposium, PCS 2010*, Nagoya, Japan, Dec. 2010.

[2] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *Proc. of IEEE Int. Conf. Image Proc., ICIP 2009*, Cairo, Egypt, Nov. 2009.

[3] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view," in *Proc. of IS&T/SPIE Electronic Imaging, VIPC 2010*, San Jose, CA, USA, Jan. 2010.

[4] M. Sarkis and K. Diepold, "Depth map compression via compressed sensing," in *Proc. of IEEE Int. Conf. Image Proc., ICIP 2009*, Cairo, Egypt, Nov. 2009.

[5] J. Duan, L. Zhang, R. Pan, and Y. Sun, "An improved video coding scheme for depth map sequences based on compressed sensing," in *International Conference on Multimedia Technology (ICMT)*, Hangzhou, Aug. 2011.

[6] W.-S. Kim, "3-d video coding system with enhanced rendered view quality," in *Ph.D. dissertation, University of Southern California*, 2011.

[7] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," in *Inverse Problems*, June 2007.

[8] S. Lee and A. Ortega, "Joint optimization of transport cost and reconstruction for spatially- localized compressed sensing in multi-hop sensor networks," *APSIPA*, Dec. 2010.

[9] D. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," in *Elsevier: Appplied and Computational Harmonic Analysis*, Apr. 2010, vol. 30, pp. 129–150.

[10] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[11] MOSEK C-library, "www.mosek.com," .