DISPARITY-COMPENSATED TOTAL-VARIATION MINIMIZATION FOR COMPRESSED-SENSED MULTIVIEW IMAGE RECONSTRUCTION

Ying Liu, Chen Zhang, and Joohee Kim

Department of Electrical and Computer Engineering, Illinois Institute of Technology
Chicago, IL, 60616

ABSTRACT

Compressed sensing (CS) is the theory and practice of sub-Nyquist sampling of sparse signals of interest. Perfect reconstruction may then be possible with much fewer than the Nyquist required number of data. In this paper, we consider a distributed multi-view imaging system where each camera at a different location performs independent compressed sensing acquisition of the target scene. At the decoder, we propose a disparity-compensated total-variation (TV) minimization algorithm to jointly reconstruct the multiple views. Experimental results show that the proposed joint decoding algorithm outperforms significantly independent-view decoding as well as disparity-compensated residue-view reconstruction algorithm.

Index Terms— Compressed sensing, multi-view imaging, sparse representation, total-variation minimization, disparity compensation

1. INTRODUCTION

Multi-view images are captured by a network of cameras distributed in a 3D scene. Compared to conventional 2D images, multi-view images offer a richer description of the captured scene because they convey both the texture and the 3D scene information. Multi-view imaging has found usage in applications such as surveillance, 3D television and robotics [1].

In multi-view imaging systems, it is feasible to measure the raw data at each sensor, but joint, collaborative compression of that data among the sensors would require costly communication. Recently, inspired by distributed source coding (DSC) [2] and compressed sensing (CS) [3]-[5], CS based distributed multi-view imaging systems have been proposed. In such systems, each camera independently captures and compresses one view of the scene by taking a small number of (random [5] or deterministic [6]) linear measurements, and high quality reconstruction of the multi-view images is achievable by exploiting multi-view image sparsity at the decoder. Typically, joint signal sparsity model is established and the reconstruction problem can be solved via convex optimization [7].

In [8], distributed compressed sensing and joint decoding algorithm is used for satellite image reconstruction. For two-view reconstruction scenario, [9] proposed to utilize the block discrete cosine transform (DCT) and the disparity-compensated view difference as the sparse penalty. Although signal sparsity is increased, the algorithm did not achieve competitive reconstruction quality due to the poor DCT basis. In [10], the decoder predicts a single view from initially reconstructed neighbor view(s) via disparity compensation (DC), then the sparse residue is recovered by two-dimensional total-variation (2D-TV) minimization, and is added back to the prediction. Although the algorithm achieves superior reconstruction quality after a few iterations, it does not consider spatial sparsity in the residue-view recovery stage.

In this paper, we propose a disparity-compensated multi-view TV minimization algorithm to jointly recover all the views from independent compressive samples. Initially, each view is reconstructed individually via 2D-TV minimization. A group of disparity maps are then estimated from the initial reconstructions, and a prediction model is established in which each view is predicted by its neighbor view(s). In the joint decoding stage, all views are recovered simultaneously using the sum-up of each view's 2D-TV and each residue view's 2D-TV as the sparse constraint.

The remainder of this paper is organized as follows. In Section 2, our simple CS encoder is introduced. In Section 3, the proposed disparity-compensated joint-view TV minimization decoder is developed. Experimental results and performance analysis are presented in Section 4. Finally, a few conclusions are drawn in Section 5.

2. A SIMPLE COMPRESSED SENSING ENCODER

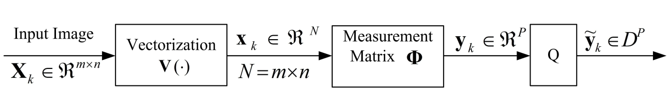

Fig. 1. Compressive sensing at the k^{th} camera with quantization alphabet $\mathcal{D}$.

978-1-4673-6997-8/15/$31.00

In this paper, we propose a practical CS multi-view image acquisition system that performs pure, direct CS encoding. In the simple encoder block diagram shown in Fig. 1, the k^{th} view $\mathbf{X}_k$ of size $m \times n$ is viewed as a vectorized column $\mathbf{x}_k = \mathbf{V}(\mathrm{X}_k) \in \mathbb{R}^N$, $N = mn$. Compressive sampling is performed by projecting $\mathbf{x}_k$ onto a $P \times N$ random measurement matrix $\mathbf{\Phi}$,

$$\mathbf{y}_k = \mathbf{\Phi}\mathbf{x}_k, \tag{1}$$

where $\mathbf{\Phi}$ is generated by randomly permuting the columns of an order-q, $q \geq N$ and multiple-of-four, Walsh-Hadamard (WH) matrix followed by arbitrary selection of P rows from the q available WH rows (if $q > N$, only N arbitrary columns are utilized). This class of WH measurement matrices has the advantage of easy implementation (antipodal ± 1 entries), fast transformation, and satisfactory reconstruction performance. To quantize the elements of the resulting measurement vector $\mathbf{y}_k \in \mathbb{R}^P$ (block $\mathbf{Q}$ in Fig. 1), in this work we follow a simple adaptive quantization approach of two codeword lengths. A positive threshold $\eta_k > 0$ is chosen such that 1% of the elements in $\mathbf{y}_k$ have magnitude above η_k. For every measurement vector $\mathbf{y}_k$, $k = 1, 2, ...$, 16-bit uniform scalar quantization is used for elements with magnitudes larger than η_k and 8-bit uniform scalar quantization is used for the remaining elements. The resulting quantized values $\widetilde{\mathbf{y}}_k$ are then indexed and transmitted to the decoder.

3. DISPARITY-COMPENSATED JOINT DECODER

3.1. Initial Decoding

Initially, multi-view images are reconstructed independently from their individual de-quantized CS measurements $\widehat{\mathbf{y}}_k$ via 2D-TV minimization [11], [12] namely,

$$\widehat{\mathbf{x}}_k = \arg\min_{\mathbf{x}} \left(\|\widehat{\mathbf{y}}_k - \mathbf{\Phi}\mathbf{x}\|_2^2 + \alpha \|\mathbf{G}\mathbf{x}\|_{\ell_1} \right), \tag{2}$$

where the linear transform matrix $\mathbf{G}$ operates on the vectorized image $\mathbf{x} = \mathbf{V}(\mathbf{X})$ and generates a vector consisted of horizontal gradient values $x_{i,j} - x_{i,j-1}$ and vertical gradient values $x_{i,j} - x_{i-1,j}$ over all pixels in image $\mathbf{X} \in \mathbb{R}^{m \times n}$.

3.2. Disparity Compensation

After the initial reconstruction of all views $\widehat{\mathbf{x}}_k$, $k = 1, ..., K$ are obtained, they are used to estimate a group of disparity maps. Let the k^{th} view $\widehat{\mathbf{x}}_k$ be the base view, $1 < k \leq K$, then the geometry relation between $\widehat{\mathbf{x}}_k$ and its left neighbor $\widehat{\mathbf{x}}_{k-1}$ can be described with a disparity map $\mathbf{d}_k^{k-1}$, where the subscript k represents the base view index, and the superscript $k-1$ represents the reference view index. With the aid of $\mathbf{d}_k^{k-1}$, every pixel $\widehat{x}_k(i,j)$ in $\widehat{\mathbf{x}}_k$ can be predicted by a matching point in $\widehat{\mathbf{x}}_{k-1}$ in the following form

$$\widehat{x}_k^{\mathrm{p}}(i,j) = \widehat{x}_{k-1}(i, j + d_k^{k-1}(i,j)). \tag{3}$$

Similarly, for $1 \leq k < K$, the geometry relation between the base view $\widehat{\mathbf{x}}_k$ and its right reference view $\widehat{\mathbf{x}}_{k+1}$ can be described with disparity map $\mathbf{d}_k^{k+1}$, such that every pixel $\widehat{x}_k(i,j)$ in the base view can be predicted by a matching point in its right neighbor $\widehat{\mathbf{x}}_{k+1}$ in the form of

$$\widehat{x}_k^{\mathrm{p}}(i,j) = \widehat{x}_{k+1}(i, j - d_k^{k+1}(i,j)). \tag{4}$$

In this work, we only consider the case that multi-view images are rectified. For efficient disparity estimation, we adopt a two-stage algorithm. In the first stage, a group of coarse disparity maps are generated via local block-matching. In the second stage, the coarse disparity maps are used to compute the mutual information, which is then considered as the matching cost in the semi-global stereo matching algorithm [13] that further refines the disparity maps.

After the disparity maps are obtained, a prediction model for each view in the group of multi-view images can be established via DC as shown in Fig. 2. For $2 \leq k \leq K-1$, $\mathbf{x}_k$ can be predicted from its left and right adjacent views $\mathbf{x}_{k-1}$ and $\mathbf{x}_{k+1}$. The prediction for pixel $x_k(i,j)$ in $\mathbf{x}_k$ is given by

$$x_k^{\mathrm{p}}(i,j) = \frac{1}{2} x_{k-1}(i, j + d_k^{k-1}(i,j)) + \frac{1}{2} x_{k+1}(i, j - d_k^{k+1}(i,j)) \tag{5}$$

For $k = 1$ and $k = K$, $x_k^{\mathrm{p}}(i,j)$ can be predicted by

$$x_k^{\mathrm{p}}(i,j) = \begin{cases} x_{k+1}(i, j - d_k^{k+1}(i,j)), & k = 1, \\ x_{k-1}(i, j + d_k^{k-1}(i,j)), & k = K. \end{cases} \tag{6}$$

In matrix form, the predicted k^{th} view $\mathbf{x}_k^{\mathrm{p}}$ with $x_k^{\mathrm{p}}(i,j)$ as its $(i,j)^{\text{th}}$ element is given by

$$\mathbf{x}_k^{\mathrm{p}} = \begin{cases} \mathbf{D}_k^{k+1}\mathbf{x}_{k+1}, & k = 1, \\ \frac{1}{2}\mathbf{D}_k^{k+1}\mathbf{x}_{k+1} + \frac{1}{2}\mathbf{D}_k^{k-1}\mathbf{x}_{k-1}, & 2 \leq k \leq K-1, \\ \mathbf{D}_k^{k-1}\mathbf{x}_{k-1}, & k = K, \end{cases} \tag{7}$$

where $\mathbf{D}_k^{k+1}$ and $\mathbf{D}_k^{k-1}$ are the DC operators that lead to the expressions in (5) and (6).

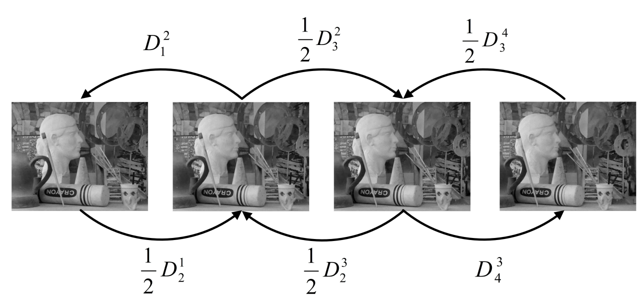

Fig. 2. Illustration of disparity compensation (DC) when $K = 4$.

3.3. Joint Decoding

When adjacent views are highly correlated and disparity maps are accurate enough, the residue between original view $\mathbf{x}_k$ and its prediction $\mathbf{x}_k^{\mathrm{p}}$, i.e., $\mathbf{x}_k - \mathbf{x}_k^{\mathrm{p}}$ also has small 2D-TV. Hence, we propose to combine the 2D-TV of individual views and the 2D-TV of DC residues as the sparse-inducing penalty in the reconstruction problem.

Consider a group of K views $\mathbf{x} = [\mathbf{x}_1^T, ..., \mathbf{x}_K^T]^T$. The observed CS measurements at the decoder can be modeled as

$$\widehat{\mathbf{y}} = \widetilde{\mathbf{\Phi}}\mathbf{x} + \mathbf{n}, \tag{8}$$

where $\widehat{\mathbf{y}}$ is the concatenation of K de-quantized measurement vectors $\widehat{\mathbf{y}}_k$, $k = 1, ..., K$, $\widetilde{\mathbf{\Phi}}$ is the block diagonal matrix with K diagonal elements

$$\widetilde{\mathbf{\Phi}} = \mathrm{diag}\{\mathbf{\Phi} \quad \ldots \quad \mathbf{\Phi}\}, \tag{9}$$

and $\mathbf{n}$ is the quantization noise. The view residue after DC operation is

$$\underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_k \\ \vdots \\ \mathbf{f}_K \end{bmatrix}}_{\triangleq \mathbf{f}} = \underbrace{\begin{bmatrix} \mathbf{I} & -\mathbf{D}_1^2 & & & \\ -\frac{1}{2}\mathbf{D}_2^1 & \mathbf{I} & -\frac{1}{2}\mathbf{D}_2^3 & & \\ & & \cdots & & \\ & -\frac{1}{2}\mathbf{D}_k^{k-1} & \mathbf{I} & -\frac{1}{2}\mathbf{D}_k^{k+1} & \\ & & \cdots & & \\ & & & -\mathbf{D}_K^{K-1} & \mathbf{I} \end{bmatrix}}_{\triangleq \widetilde{\mathbf{D}}} \underbrace{\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k \\ \vdots \\ \mathbf{x}_K \end{bmatrix}}_{\triangleq \mathbf{x}},$$

where $\widetilde{\mathbf{D}}$ is the DC operator for the group of K views. If the disparity estimation is accurate enough, $\mathbf{f}_k$ shall have sparse 2D-TV, which can be utilized to enhance the sparse representation for multi-view image reconstruction. The resultant reconstruction problem can be formulated as

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 + \alpha_1\|\widetilde{\mathbf{G}}\mathbf{x}\|_1 + \alpha_2\|\widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x}\|_1, \tag{10}$$

where $\widetilde{\mathbf{G}}$ is the block diagonal matrix with the gradient operator $\mathbf{G}$ as the K diagonal elements

$$\widetilde{\mathbf{G}} = \mathrm{diag}\{\mathbf{G} \quad \ldots \quad \mathbf{G}\}. \tag{11}$$

At first, problem (10) is converted into an equivalent variant through variable splitting technique by introducing auxiliary variables $\mathbf{w}$ and $\mathbf{u}$:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|\widehat{\mathbf{y}} - \widetilde{\mathbf{\Phi}}\mathbf{x}\|_2^2 + \alpha_1\|\mathbf{w}\|_1 + \alpha_2\|\mathbf{u}\|_1 \tag{12}$$

$$\text{subject to} \quad \widetilde{\mathbf{G}}\mathbf{x} = \mathbf{w}, \widetilde{\mathbf{G}}\widetilde{\mathbf{D}}\mathbf{x} = \mathbf{u}.$$

Then, the problem in (12) can be formulated as minimizing an augmented Lagrangian function, and solved by the alternating direction method (ADM) [14],[15].

4. EXPERIMENTAL RESULTS

In this section, we experimentally study the performance of the proposed disparity-compensated joint multi-view image decoder by evaluating the perceptual quality as well as the peak signal-to-noise ratio (PSNR) of reconstructed images. Two data sets, *Art* and *Doll*, with a resolution of 370×463 pixels are used. Each data set contains 7 rectified views. Processing is carried out only on the luminance component.

At our trivial, pure CS encoder side, each view is handled as a vectorized column of length $N = 171310$ multiplied by a $P \times N$ randomized partial WH matrix $\mathbf{\Phi}$. The sensing matrix $\mathbf{\Phi}$ is generated only once to encode all views in each data set. The elements of the captured P-dimensional measurement vector are quantized and then transmitted to the decoder. In our experiments, the CS ratio $\frac{P}{N} = 0.125, 0.25, 0.375, 0.5, 0.625$ are used to produce the corresponding rates $1.01, 2.02, 3.03, 4.04$, and 5.05 bits per pixel (bpp) [1]. At the decoder side, all 7 views in each data set are jointly reconstructed by the proposed joint decoding algorithm. In our experimental studies, three reconstruction algorithms are examined: (*i*) the proposed DC joint decoder; (*ii*) the DC residue-view decoder [10]; and *iii*) the independent decoder that recovers each individual view via 2D-TV minimization.

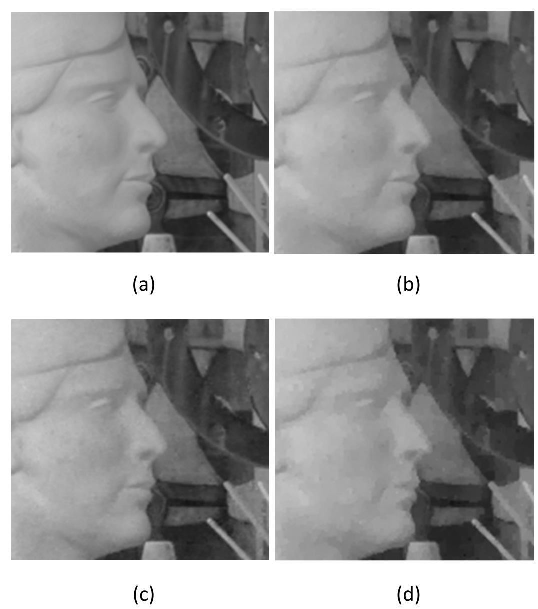

Fig. 3. Reconstruction of the 3^{rd} view of *Art* data set: (a) original view; (b) proposed DC joint decoder; (c) DC residue decoder [10]; and (d) independent decoder ($\frac{P}{N} = 0.25$).

[1]Considering the quantization scheme described in Section 2, the bit rate can be calculated as $(16 \times 0.01P + 8 \times 0.99P)/N$ bpp.

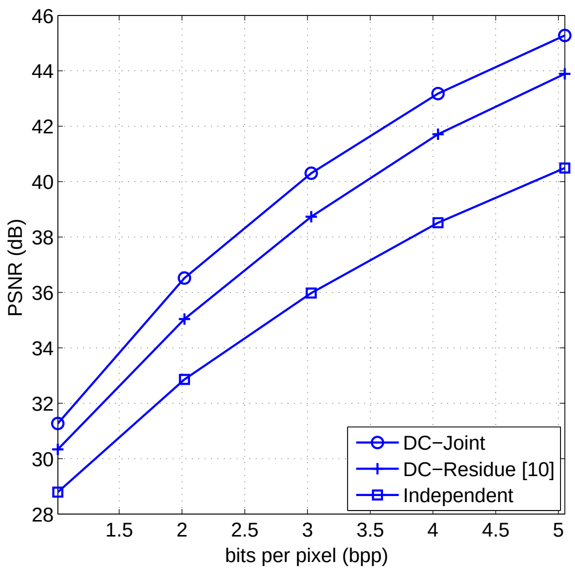

Fig. 4. Rate-distortion studies of *Art* data set.

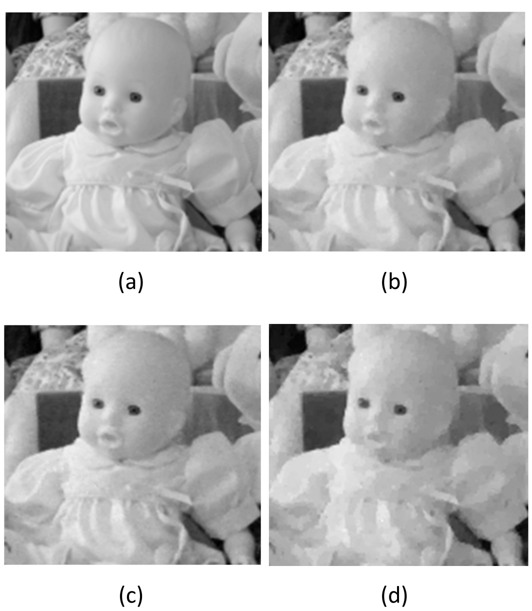

Fig. 5. Reconstruction of the 3^{rd} view of *Doll* data set: (a) original view; (b) proposed DC joint decoder; (c) DC residue decoder [10]; and (d) independent decoder ($\frac{P}{N} = 0.25$).

Fig. 3 shows the decodings of the 3^{rd} view of the *Art* data set. The decoders in comparison are the proposed DC joint-view decoder (Fig. 3(b)), the DC residue-view decoder [10] (Fig. 3(c)), and the independent 2D-TV decoder (Fig. 3(d)). It can be observed that the texture details of the scene are well preserved with our proposed decoder, and are blurred with the independent view decoder. Although the perceptual

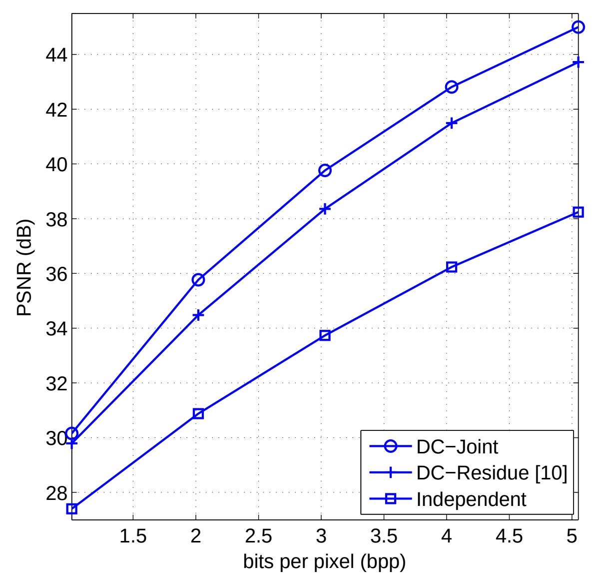

Fig. 6. Rate-distortion studies of *Doll* data set.

quality differences between Fig. 3 (b) and (c) are not pronounced in the pdf formatting, Fig. 4 shows quantitatively around 1.5 dB PSNR improvement of the proposed DC joint decoder compared with the DC residue decoder, and around 4 dB PSNR improvement compared to the independent view decoder at median-to-high bit rates. Similar conclusions can be drawn from Figs. 5 and 6 for the *Doll* data set.

5. CONCLUSIONS

We proposed a joint reconstruction algorithm for distributed compressed-sensed multi-view images. Initially, each view is independently recovered via 2D-TV minimization. Afterwards, a set of disparity maps are estimated from the initial reconstructions, and utilized in the disparity-compensated TV minimization stage for joint-view reconstruction. Experimental studies demonstrate that our proposed decoding algorithm outperforms significantly the independent view decoder, as well as the disparity-compensated residue-view decoder.

6. REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen. and C. Zhang, “Multiview imaging and 3DTV,” *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 10-21, Nov. 2007.

[2] D. Slepian, J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Info. Theory*, vol. 19, no. 3, pp. 471-480, Jul. 1973.

[3] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[5] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[6] K. Gao, S. N. Batalama, D. A. Pados, and B. W. Suter, "Compressive sampling with generalized polygons," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4759-4766, Oct. 2011.

[7] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207-1223, Aug. 2006.

[8] M. B. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," *Proc. Picture Coding Symp. (PCS 2009)*, Chicago, IL, May 2009.

[9] Q. Dai, C. Fu, X. Ji,"Robust joint reconstruction in compressed multi-view imaging," *Proc. Picture Coding Symp. (PCS 2012)*, Kraków, Poland, May 2012.

[10] M. Trocan, E. W. Tramel, J. E. Fowler, and B. Pesquet, "Compressed-sensing recovery of multiview image and video sequences using signal prediction," *Multimed Tools Appl.* vol. 72, no. 1, pp. 95-121, Sept. 2014.

[11] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, pp. 259-268, 1992.

[12] J. Yang, Y. Zhang, and W. Yin, "An efficient TVL1 algorithm for deblurring of multichannel images corrupted by impulsive noise," *SIAM J. Sci. Comput.*, vol. 31, pp. 2842-2865, 2009.

[13] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual *Information," IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 328-341, Feb. 2008.

[14] C. Li, W. Yin, and Y. Zhang, "TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms," URL: http://www.caam.rice.edu/ optimization/L1/TVAL3/.

[15] J. Zhang, S. Liu, R. Xiong, S. Ma, D. Zhao, "Improved total variation based image compressive sensing recovery by nonlocal regularization," *IEEE Intern. Symp. of Circuits and Systems (ISCAS 2013)*, pp. 2836-2839, Beijing, China, May 2013.