Compressive Sensing Reconstruction Using Collaborative Sparsity among Color Channels

Satoshi Sato[1], Motonori Ishii[2], Yoshihisa Kato[2], Kunio Nobori[1], Takeo Azuma[1]
1. Advanced Research Division, Panasonic Corporation
3-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
2. Automotive & Industrial Systems Company, Panasonic Corporation
1 Kotari-yakemachi, Nagaokakyo City, Kyoto 617-8520, Japan
{ sato.satoshi, ishii.m, kato.yoshihisa, nobori.kunio, azuma.takeo }@jp.panasonic.com

Abstract

This study describes a reconstruction method for compressive sensing using collaborative sparsity among multi-frame images and color channels. The proposed method reduces the artifact for compressive sensing and obtains better image quality. Experimental results reveal that the proposed method reconstructs 6.1 dB higher quality images than the conventional one for complex texture at occlusion boundaries.

1. Introduction

Compressive sensing [1], which consists of much more efficient sampling than Nyquist-Shannon theorem and perfect reconstruction in theory, is promising to achieve very low power consumption digital cameras [2] and high-resolution light field camera [3].

Compressive sensing utilizes many kinds of prior-information, such as total variation [4, 5], linear transformation [6], and non-local similarity, to reconstruct high quality images. As one of the most effective prior among them, Zhang et al. [7] have proposed a reconstruction method using non-local similarity, which is called collaborative sparsity. It is based on Block Matching 3D Filtering (BM3D) process [8] known as image denoising. It enforces both local two-dimensional sparsity and non-local three-dimensional sparsity simultaneously, so that this method enables a natural image to be highly sparse in an adaptive hybrid space-transform domain. A drawback of this method, however, is the fact that the image quality for complex texture at occlusion boundaries tends to be degraded in the case of moving camera or multi-view cameras.

To overcome this issue, we apply new sparsity: collaborative sparsity among multi-frame images and color channels. The proposed method achieves higher degree of sparsity and obtains better image quality at occlusion boundaries. Experimental results reveal that the proposed method reconstructs 6.1 dB higher quality images than the conventional one in such the regions.

2. Compressive Sensing Reconstruction Using Collaborative Sparsity

In this section, the conventional reconstruction using collaborative sparsity is described.

2.1. Formulation of compressive sensing reconstruction using collaborative sparsity

We define the ideal solution of reconstruction images of size N as $\mathbf{X}$. Given M ($< N$) linear measurements vector $\mathbf{Y}$, the reconstruction of $\mathbf{X}$ from $\mathbf{Y}$ is formulated as the following equation:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \tag{1}$$

where $\mathbf{A}$ represents the random sampling matrix.

Evidently, there are infinitely many possible $\mathbf{X}$ because rank of $\mathbf{A} < N$. To obtain optimal images, compressive sensing introduces as prior-information that natural images have high degree of sparsity in general. Zhang et al. [7] have applied collaborative sparsity and formulated as the following constrained optimization problem:

$$\min_{\mathbf{X}} \left\{ \left\| \Psi_{\mathrm{L2D}} \mathbf{X} \right\|_1 + \alpha \left\| \mathbf{\Theta}_{\mathbf{x}} \right\|_1 \right\} \quad \text{subject to} \quad \mathbf{Y} = \mathbf{A}\mathbf{X}, \tag{2}$$

where Ψ_{L2D} denotes vertical and horizontal finite difference operator, $\left\| \mathbf{\Theta}_{\mathbf{x}} \right\|_1$ is nonlocal three-dimensional sparsity in transform domain Ψ_{N3D}, $\| \cdot \|_1$ is represent l_1 norm and α denotes a regularization parameter.

The former type of sparsity describes the smoothness, which is local two-dimensional sparsity in space domain.

The latter one denotes non-local three-dimensional sparsity in transform domain, which is the self-similarity of natural images, retaining the sharpness and edges effectively. This is defined as follows:

1. Divide the image $\mathbf{X}$ with size N into N_b overlapped blocks of size $N_1 \times N_1$ and each block is denoted by x_k, i.e., k =1, 2, ... , N_b.

2. Define $\mathbf{S}_{\mathbf{x}_k}$ the set including the N_2 best matching blocks to x_k in the $N_\mathrm{S} \times N_\mathrm{S}$ searching window, that is, $\mathbf{S}_{\mathbf{x}_k} = \left\{ S_{x_k \otimes 1}, S_{x_k \otimes 2}, \ldots, S_{x_k \otimes N2} \right\}$.

3. For every $\mathbf{S}_{\mathbf{x}_k}$, a group is formed by stacking the blocks belonging to $\mathbf{S}_{\mathbf{x}_k}$ into a three-dimensional array, which is denoted by $\mathbf{Z}_{\mathbf{x}_k}$.

4. Denote $\mathbf{T}^{3D}$ the operator of a three-dimensional transform, and $\mathbf{T}^{3D}\left(\mathbf{Z}_{\mathbf{x}_k}\right)$ the transform coefficients for $\mathbf{Z}_{\mathbf{x}_k}$ in domain Ψ_{N3D}. Let $\mathbf{\Theta}_{\mathbf{x}}$ be the column vector with size $K = N_1 \cdot N_1 \cdot N_\mathrm{b} \cdot N_2$ built from all the $\mathbf{T}^{3D}\left(\mathbf{Z}_{\mathbf{x}_k}\right)$ arranged in lexicographic order.

From this definition, it is obvious that if the matching blocks are more similar, the reconstruction quality be-

comes higher. Therefore, this method is particularly effective for video sequence (multi-frame) and light field (multi-view) images.

2.2. Drawback of the conventional method

This conventional reconstruction is assumed that every image piece has many matching blocks enough. However, in the case of moving camera or multi-view cameras, this assumption is sometimes broken and results are degraded at occlusion boundaries.

3. Reconstruction Using Collaborative Sparsity among Color Channels

We introduce collaborative sparsity among multi-frame images and color channels to overcome above mentioned issue. The proposed method finds more similar blocks than the conventional one and obtains better image quality for complex texture at occlusion boundaries. Additionally, to improve the reconstruction quality, our method captures multi-frame images with different random sampling by frame by frame.

In this section, we present our proposed method and its solution.

3.1. Prior-information based on color correlation

As mentioned above, collaborative sparsity is effective for multi-frame and multi-view images. Additionally, natural image has strong color correlation [9]. Then we introduce collaborative sparsity among multi-frame, multi-view camera and color channel images. We summarize the similarity for these images in Table 1. For complex texture, collaborative sparsity among multi-frame images is efficient. For complex texture at occlusion boundaries, however, image blocks are not similar among multi-view camera images. In contrast, they are very similar among color channel images as long as the color correlation is high enough. Our method searches matching blocks not only in image plane but also among multi-frame, multi-view and color channel images. Then the proposed method finds more similar matching blocks than the conventional one. As a result, better image quality is obtained at occlusion boundaries.

3.2. Formulation of reconstruction using collaborative sparsity among color channels

The proposed method is formulated as minimization of the following cost function:

$$\min_{\mathbf{X}}\left\{\frac{1}{2}\left\|\mathbf{AX}-\mathbf{Y}\right\|_2^2+\tau_{\mathrm{TV}}\mathrm{TV}(\mathbf{X})+\alpha\left\|\mathbf{\Theta}_{\mathbf{X}}\right\|_1\right\}, \tag{3}$$

where τ_{TV} and α denote regularization parameters.

This cost function consists of three terms: data-fidelity term, total variation term and non-local sparsity term.

The first one is data-fidelity term, which takes the difference random sampling of reconstruction images **X** and measurements vector **Y**.

The second one is total variation term, which is based on local two-dimensional sparsity of **X** [4].

The third one is non-local sparsity term, which is applied similar term in (2) to inter-color channels.

Table 1. Similarity among images.

	Simple Texture	Complex Texture	Occlusion Boundary
Intra-Color Channel	✓		
Multi-Frame (Multi-View)	✓	✓	
Among Color Channels	✓ (As long as the color correlation is high enough.)	✓	✓

3.3. Solution of the proposed method

Note that the cost function (3) is essentially non-convex and quite difficult to solve directly due to non-differentiability and non-linearity of non-local three-dimensional sparsity term. To solve it efficiently, we introduce Split Augmented Lagrangian Shrinkage Algorithm (SALSA) [10].

Using variable splitting, SALSA reformulates (3) as a constrained problem (4) for variables **U**, **V**:

$$\min_{\mathbf{U},\mathbf{V}} g(\mathbf{V}) \quad \text{subject to} \quad \mathbf{GU}+\mathbf{BV}=\mathbf{0}, \tag{4}$$

$$\mathbf{V}=\left(\mathbf{V}_1^{\,T},\mathbf{V}_2^{\,T},\mathbf{V}_3^{\,T}\right)^T, \tag{5}$$

$$g(\mathbf{V})=\frac{1}{2}\left\|\mathbf{AV}_1-\mathbf{Y}\right\|_2^2+\tau_{\mathrm{TV}}\mathrm{TV}(\mathbf{V}_2)+\alpha\left\|\mathbf{\Theta}_{\mathbf{V}3}\right\|_1, \tag{6}$$

$$\mathbf{G}=\begin{bmatrix}\mathbf{I}\\ \mathbf{I}\\ \mathbf{I}\end{bmatrix},\quad \mathbf{B}=\begin{bmatrix}-\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & -\mathbf{I}\end{bmatrix}. \tag{7}$$

By introducing new variable **D**, SALSA works as Table 2, where μ is a penalty parameter, τ_{c} is the parameter that depends on its convergence, and ST(x,b) stands for soft thresholding operator as follows:

$$\mathrm{ST}(x,b)=sign(x)\min\left(\left|x-b\right|,0\right). \tag{8}$$

4. Experimental Results

In this section, the experimental results for both video sequences and multi-view images are presented to evaluate the performance of the proposed method.

For all experiments, we use spatiotemporal random sampling whose elements are 0 or 1, which is reasonable for the fabrication of the image sensor. As is known, Gaussian random sampling is efficient to reconstruct high quality images. However, the circuit is complicated.

The proposed method compared with two reconstruction methods, i.e., total variation method [4], which doesn't utilize any non-local similarity, and reconstruction with collaborative sparsity among multi-frame images and intra-color channel, which is called multi-frame method. For all simulation, the stopping criterion at Table 2 is as following:

$$\left\|\mathbf{GU}^{(k)}+\mathbf{BV}^{(k)}\right\|_2^2\le Th\_m,\ \text{ or }\ k>250, \tag{9}$$

where *Th_m* denotes the convergence parameter.

At first, we evaluated the results for video sequences. We used four video sequences with 16 frames provided by the Institute of Image Information and Television Engi-

Table 2. Algorithm for the proposed method.

Formulation:

$$L(\mathbf{U},\mathbf{V},\mathbf{D}) = g(\mathbf{V}) + \frac{\mu}{2}\left\|\mathbf{GU}+\mathbf{BV}-\mathbf{D}\right\|_2^2,$$

$$\mathbf{D} = \left(\mathbf{D}_1{}^T, \mathbf{D}_2{}^T, \mathbf{D}_3{}^T\right)^T.$$

Algorithm:

Initialization: $k = 0,\ \mu > 0,\ \mathbf{V}^{(0)} = \mathbf{0},\ \mathbf{D}^{(0)} = \mathbf{0},\ \mathbf{p}^{(0)} = \mathbf{0}.$

Repeat:

$$\mathbf{U}^{(k+1)} = \arg\min_{\mathbf{U}} L\left(\mathbf{U}, \mathbf{V}^{(k)}, \mathbf{D}^{(k)}\right)$$

$$= \left\{\left(\mathbf{V}_1{}^{(k)} + \mathbf{D}_1{}^{(k)}\right) + \left(\mathbf{V}_2{}^{(k)} + \mathbf{D}_2{}^{(k)}\right) + \left(\mathbf{V}_3{}^{(k)} + \mathbf{D}_3{}^{(k)}\right)\right\}/3.$$

$$\mathbf{V}_1{}^{(k+1)} = \left(\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}\right)^{-1}\left\{\mathbf{A}^T\mathbf{Y} + \mu\left(\mathbf{U}^{(k+1)} - \mathbf{D}_1{}^{(k)}\right)\right\}$$

$$\mathbf{V}_2{}^{(k+1)} = \arg\min_{\mathbf{V}_2}\left\|\mathbf{V}_2 - \mathbf{g}^{(k+1)}\right\|_2^2 \Big/ 2\lambda + \mathrm{TV}(\mathbf{V}_2)$$

$$= \mathbf{g}^{(k+1)} - \lambda\cdot\mathrm{div}\,\mathbf{p}^{(k+1)}.$$

$$\mathbf{g}^{(k+1)} = \mathbf{U}^{(k+1)} - \mathbf{D}_2{}^{(k)},\ \lambda = \tau_{\mathrm{TV}}/\mu.$$

$$\mathbf{p}_{j,i}^{(k+1)} = \frac{\mathbf{p}_{j,i}^{(k)} + \tau_c\left(\nabla\left(\mathrm{div}\,\mathbf{p}^{(k)} - \mathbf{g}^{(k+1)}/\lambda\right)\right)_{j,i}}{\max\left(1, \left|\mathbf{p}_{j,i}^{(k)} + \tau_c\left(\nabla\left(\mathrm{div}\,\mathbf{p}^{(k)} - \mathbf{g}^{(k+1)}/\lambda\right)\right)_{j,i}\right|\right)}.$$

$$\mathbf{V}_3{}^{(k+1)} = \arg\min_{\mathbf{V}_3}\frac{1}{2}\left\|\mathbf{V}_3 - \mathbf{r}^{(k+1)}\right\|_2^2 + \frac{\alpha}{\mu}\left\|\mathbf{\Theta}_{\mathbf{V}_3}\right\|_1$$

$$= \Omega_{\mathrm{N3D}}\left(\mathrm{ST}\left(\mathbf{\Theta}_{\mathbf{r}^{(k+1)}}, \frac{K\alpha}{N\mu}\right)\right).$$

$$\mathbf{r}^{(k+1)} = \mathbf{U}^{(k+1)} - \mathbf{D}_3{}^{(k)}.$$

$\mathbf{\Theta}_{\mathbf{r}^{(k+1)}}$ is the column vector built from all the $\mathbf{T}^{3D}\left(\mathbf{Z}_{\mathbf{r}^{(k+1)}}\right)$ arranged in lexicographic order.

Ω_{N3D} is the inverse operator of Ψ_{N3D}.

$$\mathbf{D}^{(k+1)} = \mathbf{D}^{(k)} - \mathbf{GU}^{(k+1)} - \mathbf{BV}^{(k+1)}.$$

Update iteration: $k \leftarrow k+1$.

Until some stopping criterion is satisfied.

$\mathbf{X} = \mathbf{U}$, return ($\mathbf{X}$).

neers as shown in Fig. 1. Figure 2 plots PSNR of all reconstruction images versus compression ratio (presented as the ratio of number of measurements to total pixels) from 0.125 to 0.500 for the three different approaches. Figure 3 depicts the reconstruction images in the rectangular region in Fig. 1. The compression ratio is 0.125. These figures show that the proposed method reconstructs complex texture more precisely than conventional methods although its improvement for whole images is small.

Next, we evaluated the results for multi-view images which have complex texture at occlusion boundaries. We used nine view-point images provided by Tsukuba University [11]. The compression ratio is 0.25. PSNR of all reconstruction sequences versus camera position is plotted in Fig. 4. This figure presents that the proposed method is 1.8 dB better in average than the conventional one. The reconstruction images for these three methods are presented in Fig. 5 and the right column of Fig. 5 depicts the enlarged results for complex texture at occlusion boundaries. Figure 5 (c) shows the multi-frame method cannot reconstruct the texture in the background near occlusion boundaries. In contrast, the proposed method preserves the texture as shown in Fig. 5 (d). The computational cost of the proposed method becomes three times because it searches matching blocks for three times wider than the multi-frame method.

Fig. 1. Experimental test images for video sequences.

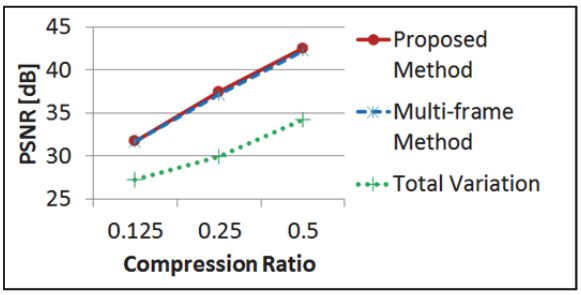
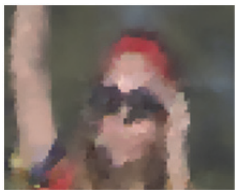


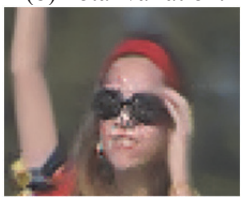
Fig. 2. PSNR vs. compression ratio.

(a) Experimental test image.



PSNR = 22.6 [dB].
(b) Total variation.

PSNR = 25.4 [dB].
(c) Multi-frame method.



PSNR = 27.2 [dB].
(d) The proposed method.

Fig. 3. Reconstruction images for video sequence (magnifying the rectangular region in Fig. 1.)

These experimental results present that the proposed method reconstructs 6.1 dB higher quality images than the conventional one.

These results reveal that the proposed method reconstructs higher quality images than the conventional one for complex texture at occlusion boundaries.

5. Future work

To realize our method, we are fabricating a prototype image sensor for spatiotemporal random sampling. The prototype chip microphotograph is shown in Fig. 6.

In our sensor, all pixels have cross bar switches and the pixel array is partitioned into blocks. Spatial sampling is performed on each block using the same 4x4 encoding matrix which controls cross bar switches. Signal adder sums up the pixel values corresponding to the non-zero elements of the encoding matrix. By varying the encoding matrix for each frame, our image sensor performs spatiotemporal random sampling whose elements are 0 or 1.

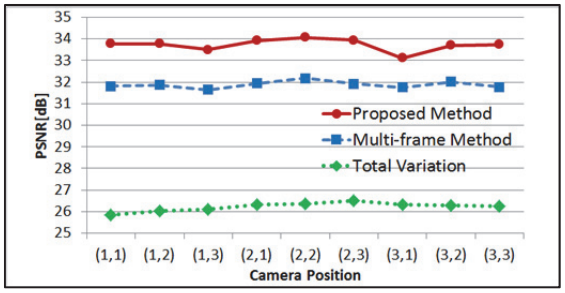
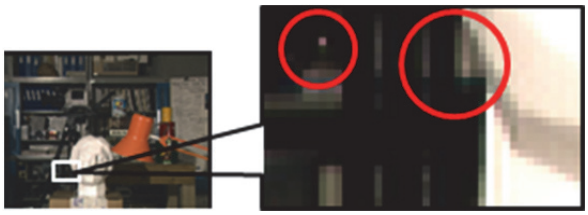
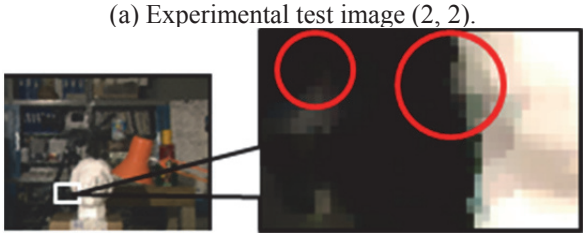


Fig. 4. Results of PSNR vs. camera position.

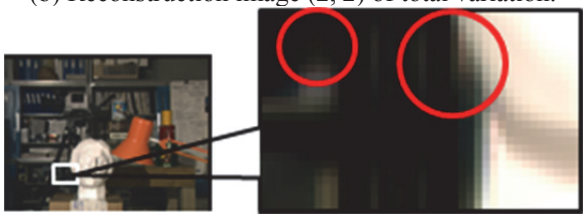


(a) Experimental test image (2, 2).



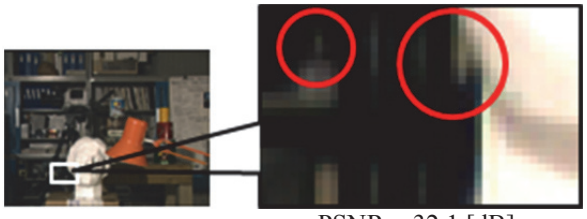
PSNR = 23.7 [dB].

(b) Reconstruction image (2, 2) of total variation.



PSNR = 26.0 [dB].

(c) Reconstruction image (2, 2) of multi-frame method.



PSNR = 32.1 [dB].

(d) Reconstruction image (2, 2) of the proposed method.

Fig. 5. Reconstruction images for multi-view images.

6. Conclusion

In this study, we proposed the new reconstruction method for compressive sensing. We introduced collaborative sparsity among multi-frame images and color channels. Since natural images have color correlation in general, the proposed method leads higher degree of sparsity. As a result, better image quality is achieved for complex texture at occlusion boundaries. Experimental results reveal that the proposed method reconstructs 6.1 dB higher quality images than the conventional method in such the region.

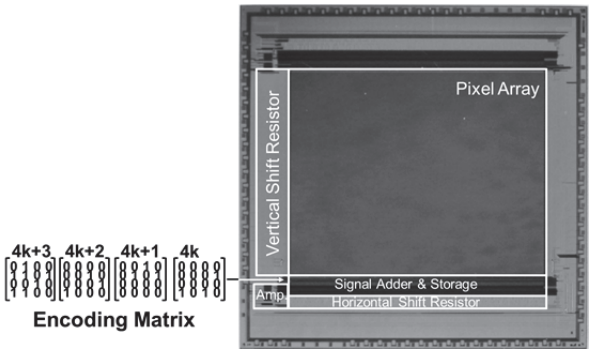


Fig. 6. Chip micrograph of image sensor.

For future work, we will evaluate on real data set which is captured by our prototype image sensor and speed up our algorithm to using parallel processing.

References

[1] D. L. Donoho, "Compressive Sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.

[2] Y. Oike, and AE Gamal, "CMOS Image Sensor With Per-Column ΣΔ ADC and Programmable Compressed Sensing", IEEE Journal of Solid-State Circuits, vol. 48, no. 1, pp. 318-328, 2013.

[3] K. Marwah, G. Wetzstein, Y. Bando, R. Raskar, "Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections", ACM Transactions on Graphics (SIGGRAPH), 2013.

[4] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications", Journal of Mathematical Imaging and Vision, vol. 20, iss. 1-2, pp. 89-97, 2004.

[5] L. I. Rudin, S. J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms", Physica D, vol. 60, pp. 259–268, 1992.

[6] J. Ma, "Improved Iterative Curvelet Thresholding for Compressed Sensing and Measurement", IEEE Transactions on Instrumentation & Measurement, vol. 60, no. 1, pp. 126-136, Jan.2011.

[7] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Compressed Sensing Recovery via Collaborative Sparsity", Proc. of IEEE Data Compression Conference, pp. 287-296, 2012.

[8] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering", IEEE Transactions on Image Processing, vol. 16, iss. 8, pp. 2080-2095, 2007.

[9] S. Ugawa, T. Azuma, T. Imagawa, Y. Okada, "Performance Evaluation of High Sensitive DRE Camera for Cultural Heritage in Subdued Light Conditions", Proc. of International Conference on Virtual Systems and Multimedia, pp. 141-147, 2010.

[10] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast Image Recovery Using Variable Splitting and Constrained Optimization", IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2345-2356, 2010.

[11] http://vision.middlebury.edu/stereo/