DETAIL-PRESERVING COMPRESSIVE SENSING RECOVERY BASED ON CARTOON TEXTURE IMAGE DECOMPOSITION

Thuong Nguyen Canh[1], Khanh Quoc Dinh[2], and Byeungwoo Jeon[3]
School of Electronic and Electrical Engineering, Sungkyunkwan University, Korea
{[1]ngcthuong, [2]dqkhanh, [3]bjeon}@skku.edu

ABSTRACT

In this paper, we propose a detail-preserving reconstruction method for total variation-based recovery in low subrate compressive sensing using cartoon texture image decomposition and residual reconstruction. It iteratively decomposes and reconstructs cartoon and texture image components separately. A nonlocal structure-preserving filter is utilized to reduce staircase artifacts while preserving nonlocal structures of image in the spatial domain. Experimental results show that the proposed method outperforms the conventional ones in terms of preserving small scale details of image.

Index Terms**—** Compressive sensing, total variation, image decomposition, nonlocal mean, split Bregman

1. INTRODUCTION

Recently compressive sensing (CS) [1] has drawn a lot of interest due to its capability of simultaneous sampling and compression. For natural images $f \in \mathbb{R}^{n^2\times 1}$ which are either sparse in pixel domain or in some transform domain, CS can be applied to dramatically reduce sampling cost by taking much smaller number of measurements $y \in \mathbb{R}^{m^2\times 1}$:

$$y = Af \qquad (1)$$

where $A \in \mathbb{R}^{m^2\times n^2}$ denotes a sensing matrix which satisfies restricted isometry property condition [1] and $m^2 \ll n^2$.

However, in case of multidimensional signals (*i.e.*, 2D-image, video, or hyper-spectral image), CS still requires high computational complexity due to the large size of the sensing matrix. In this regard, a block-based CS is preferred [13, 14]. While the block-based approach can preserve local characteristics of image well, however, it misses the global ones. Therefore, Duarte *et al*. introduced the Kronecker compressive sensing (KCS) which jointly models sensing matrices for each signal dimension [2]. That is, for 2D signal, $F \in \mathbb{R}^{n\times n}$, a sensing matrix is given as $A = R \otimes G^T$, where $\otimes$ denotes the Kronecker product, R and $G \in \mathbb{R}^{m\times n}$ represent sensing matrices for each dimension. The CS measurement in eq.(1) becomes $Y = RFG$.

Texture (or detail) is an important feature for many image processing tasks such as image reconstruction, classification, segmentation, *etc*. Normal compressed sensing encodes edge, texture, and smooth components all together and all of them are to be simultaneously reconstructed later. However, since texture and non-texture components (*i.e.*, edges and smooth) have different characteristic in terms of randomness or pattern repetitiveness [15], *etc,* it would be desirable to distinguish them in CS recovery. For instance, the total variation (TV)-based CS recovery [4-6] can produce high objective quality but may suffer from the staircase artifact and loss of image details. In this regards, we are motivated that their separate reconstruction might be able to improve detail-preserving performance of CS recovery.

In image processing literatures, noteworthy is a technique known as cartoon-texture image decomposition [7] which decomposes input image into different feature image components. It can split an image into cartoon (*i.e.,* edges and smooth components) and texture parts (*i.e.,* details components) as illustrated in Fig. 2. Note a natural image has rich nonlocal structures which have been extensively exploited in denoising, de-blurring, *etc*. Following this direction, the work [6] proposed new regularization terms to reduce the staircase artifact by using nonlocal means filter (NLM) [3] in spatial and gradient domains.

Note that at a low subrate, TV-based CS recovery easily loses texture since TV prefers piece-wise constant output, thus, the recovered image is expected to look like a cartoon image. In this paper, by employing the cartoon image decomposition and residual CS recovery, we propose a detail-preserving CS recovery method for low subrate applications. It iteratively decomposes and reconstructs cartoon and texture images rather than reconstructs the image as it is. A structure-preserving filter (e.g., NLM) is used to preserve nonlocal structures of images in the spatial domain.

The paper is as follow. Section 2 studies the cartoon-texture decomposition and residual recovery in CS. The proposed method is addressed in Section 3. Section 4 shows simulation results and this paper is concluded in Section 5. All notations are listed in Table 1.

Table 1. Notations in the proposed method

Y^k	Received measurement at *k*-th iteration, $Y^k = RF^kG$
F^k	Recovered image at *k*-th iteration
$(.)_C$ & $(.)_T$	Notations related to cartoon and texture images
R, G	Kronecker sensing matrices.
$\mathcal{G}(.)$	Structure-preserving filter like NLM [3] or BM3D[12]
$CSrec(.)$	CS reconstruction method

Acknowledgement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2011-001-7578).

2. CARTOON-TEXTURE IMAGE DECOMPOSITION AND RESIDUAL RECONSTRUCTION

2.1. Cartoon-Texture Decomposition and CS Recovery

According to the cartoon-texture image decomposition [7], an image F is decomposed into carton and texture images as:

$$F = F_C + F_T, \tag{2}$$

where F_C, F_T denotes a cartoon (showing edge and smooth components) and a texture (showing texture and/or noise components) images, respectively. The cartoon image is obtained by using a spatial adaptive filter, $F_C = \mathscr{g}(F)$, which eliminates not only noise but also texture while keeping strong edges. By subtracting the cartoon component from the input image, the texture component is found, $F_T = F - F_C$. This simple decomposition can be done by the split Bregman-based TV [8].

Under the KCS framework, suppose a recovered image at k-th iteration is decomposed into cartoon and texture as $F^k = F_C^k + F_T^k$. The signal measurement at k-th iteration is,

$$Y^k = RF^kG = R(F_C^k + F_T^k)G = RF_C^kG + RF_T^kG, \tag{3}$$
$$\Rightarrow Y^k = Y_C^k + Y_T^k,$$

At a low subrate, the TV-based recovered image is expected to look like a cartoon image.

$$F_C^k = F^{k+1} = CSrec(Y^k), \tag{4}$$

$$Y_C^k = RF_C^kG = RF^{k+1}G = Y^{k+1}, \tag{5}$$

From eq. (4), the corresponding texture measurement is

$$Y_T^k = Y^k - Y_C^k = Y^k - Y^{k+1}, \tag{6}$$

$$F_T^k = CSrec(Y_T^k) = CSrec(Y^k - Y_C^k), \tag{7}$$

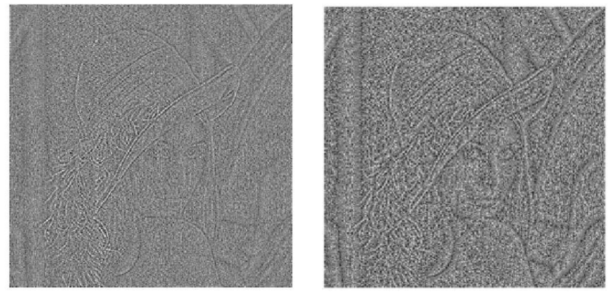
From eqs. (5), (7), we can formulate F_T^k as

$$F_T^k = CSrec(Y^k - Y^{k+1}), \tag{8}$$

Eq. (8) shows that the texture image F_T^k is nothing but recovery of residual measurement between two consecutive iterations. Surprisingly, texture information still remains in residual measurement $Y^k - Y^{k+1}$ as shown in Fig. 1.

2.2. Related Work

Actually residual reconstruction is not new in CS recovery. For example, Mun *et al.* [9] introduced a residual image reconstruction method for image and video for CS recovery. They formed residual measurement between the current block and the predicted block (obtained by motion estimation). Note that the blocks here are not texture image. Given an initial recovered image, Kim *et al.* [16] iteratively recovered the residual between the received and the recovered measurement $(Y^0 - Y^k)$ utilizing state-of-the-art BM3D filter [12]. While Kim's approach recovered texture only, in this paper, we iteratively decompose and recover not only texture but also cartoon image, and utilize the texture information from the very beginning iteration. In another hand, Guo *et al.* proposed a details-preserving method in [10], but it is different from the proposed method since it uses total generalized variation [17] and shearlet transform [18].



5-th Iteration *25*-th Iteration

Fig. 1. Recovery of residual measurement between two consecutive iterations of TV[5] at subrate 0.2.

3. PROPOSED RECONSTRUCTION METHOD

3.1. Total Variation Reconstruction

Under the KCS framework, the optimization problem of TV with spatial nonlocal regularization (TVNLR1) [6] is:

$$min_F \; TV(F) + \frac{\mu}{2}\|RFG - Y\|_2^2 + \frac{\gamma}{2}\|F - \mathscr{g}(F)\|_2^2, \tag{9}$$
$$TV(F) = \|\nabla_x F\|_1 + \left\|\nabla_y F\right\|_1$$

where μ, γ are constant parameters, ∇_x and ∇_y denote gradient operators, respectively in horizontal and vertical direction. In addition, eq.(9) is equivalent to TV[5] if γ is set to zero. By applying the split Bregman approach [4], we can solve this problem by replacing $V = F, D_x = \nabla_x F, D_y = \nabla_y F$, and adding parameters B_x, B_y, and W as:

$$\min_{F,V,D_x,D_y} TV(F) + \frac{\mu}{2}\|RFG - Y\|_2^2 + \frac{\nu}{2}\|F - V - W\|_2^2$$
$$+ \frac{\lambda}{2}\|D_x - \nabla_x V - B_x\|_2^2 + \frac{\lambda}{2}\left\|D_y - \nabla_y V - B_y\right\|_2^2$$
$$+ \frac{\gamma}{2}\|F - \mathscr{g}(F)\|_2^2, \tag{10}$$

where λ and ν are also constant. Eq. (10) is split into sub-problems F, V, D_x, D_y and solved via eigen-decomposition and shrinkage function [6]. For more details, refer to [5, 6].

3.2. Proposed Decomposition-based CS Recovery

In this section we address a proposed framework of image **D**ecomposition-based CS recovery using **T**otal **V**ariation (**DTV**) which separately reconstructs cartoon and texture images. Firstly, we recover an initial image by TV[5] then remove staircase artifacts by using a structure-preserving filter like NLM. Since the NLM filter is good at keeping strong edges while removing some textures, it will turn the TV-based CS recovery output into a cartoon image. Texture image is obtained by solving eq.(8) using a CS recovery method. Fig. 2 shows that most strong edge components are kept in the cartoon image, and textures and noise mostly exist in the texture part.

(a) Cartoon image, F_C^k (b) Texture image, F_T^k

Fig. 2. Cartoon and texture images of the proposed method (8-th iteration at subrate 0.2)

In image processing tasks, image texture is often modeled as repeated patterns [15], therefore texture image contains a lot of nonlocal structures. On the other hand, the texture image, which has structure but with high level of randomness [15], is also contaminated by noises. We try to eliminate noise, after summating cartoon and texture image ($F_C^{k+1} + F_T^{k+1}$), by using a structure-preserving filter like NLM in the spatial domain ($F^{k+1} = \mathcal{g}(F_C^{k+1} + F_T^{k+1})$). In general, a proper recovery method should be selected to reconstruct cartoon ($CSrec_C$) and texture ($CSrec_T$) images. In this regards, we select a TV-based method to reconstruct cartoon image [7, 8]. However, there is no existing recovery method designed particularly to recover texture image. Therefore, in this paper, we experimentally select TV[5] to reconstruct the texture image. The detail of the proposed framework is given in Table 2.

As shown in Fig. 1, structures in residual measurement $Y^k - Y^{k+1}$ exist from the very beginning to ending iteration. For this, current iteration starts by taking over the final results of the previous iteration with all intermediate values. Furthermore, we update the measurement $Y^{k+1} = Y^0 + Y^k - RF^{k+1}G$ as in [4]. Note that, it is natural to apply DTV scheme to conventional CS.

Table 2. Proposed reconstruction method

Input: Y, R, G and parameters for cartoon and texture
 Initialize: $Y_C^0 = Y^0 = Y$
while $\|F^{k+1} - F^k\|_2^2 / \|F^k\|_2^2 > tol$ and $k < K$
 Find cartoon part: $F_C^{k+1} = \mathcal{g}\left(CSrec_C(Y^k)\right)$
 Form texture measurement: $Y_T^k = Y^k - RF_C^{k+1}G$
 Find texture part: $F_T^{k+1} = CSrec_T(Y_T^k)$
 Combine: $F^{k+1} = \mathcal{g}(F_C^{k+1} + F_T^{k+1})$
 Update measurement: $Y^{k+1} = Y^0 + Y^k - RF^{k+1}G$
 $k = k + 1$
end while
Output F^{k+1}

Table 3. Description of various CS reconstruction methods

Algorithm	Descriptions
TV[5]	TV reconstruction [5]
TVNLR1[6]	TV[5] with spatial nonlocal regularization[6], h = 0.03
TV+BM3D[16]	TV[5] with post processing BM3D with σ = 10
DTV-NL*	DTV with TV[5] for both cartoon and texture; NLM for both cartoon and combine with h = 0.04
DTV-BM3D*	DTV with TV[5] for both cartoon and texture; BM3D with σ = 20 (cartoon) and 10 (combine)
DTV-NLR1*	DTV with TVNLR1[6] for cartoon, TV[5] for texture; BM3D with σ = 15 (cartoon) and 5 (combine)

Proposed method (*)

4. EXPERIMENTAL RESULTS

In this section, we validate the effectiveness of the proposed method (DTV-NL) by comparing it with conventional CS recovery methods such as TV[5], TVNLR1[6], TV+BM3D[16] as listed in Table 3. Since the proposed decomposition-based framework is not limited just to one combination of TV[5] and NLM as in DTV-NL, we also show experimental results of DTV with BM3D [12] (see DTV-BM3D) and TVNLR1[6] (see DTV-NLR1). KCS measurements of $n \times n$ image are obtained by Gaussian matrices R and G with size $\lceil n\sqrt{subrate} \rceil \times n$ where $\lceil . \rceil$ stands for the ceiling operator. We setup $\lambda = 1, \nu = 0.05, \mu = 1$ for TV[5], and $\gamma = 10$ for TVNLR1[6]. The NLM filter is used with a search window size 13×13 and patch size 7×7. The maximum iteration number is $K = 15$ and stopping criterion is selected as $tol = 5.10^{-3}$. For TV+BM3D, we use TV[5] to generate an initial image and reconstruct residual image. Values of PSNR and structural similarity index (SSIM) [11] are evaluated with test images of size 512×512 at subrates from 0.1 to 0.3. All results are obtained by averaging experiment results of five times.

As shown in Table 4, the proposed DTV-NL outperforms TV[5] by an average of 1.72 dB for all test cases. Compared to TVNLR1, DTV-NL produces the same PSNR in smooth images (Peppers) but achieves up to 0.28-0.48dB gain in weak edge image (Cameraman) and complex texture image (Lena). But it has 0.15-0.6dB lower PSNR in rich edge image (Barbara). However, DTV-NL still produces better visual quality than TVNLR1 in small scale details region as illustrated in Fig. 3 (see Barbara's scarf (nearby her hand)). Especially, in texture regions like Lena hair and her hat, DTV-NL preserves image details very well.

Even though TV+BM3D improves up to by 1.81dB over TV[5] but it smoothes out the texture of image. As shown in Fig. 3, TV+BM3D removes details in Lena's hat and eyelash. However, thanks to separate recovery of texture and cartoon images, even with a less effective filter like NLM, the proposed DTV-NL still offers better quality at low subrate like 0.2-0.3. Since the recovered image is output of texture preserving filter, a better filter will give higher performance. As expected, the proposed with a better filter like BM3D (DTV-BM3D) produces better visual quality and gains over TV+BM3D and DTV-NL by 0.57dB and 0.46dB, respectively.

Table 4. PSNR [dB] and SSIM comparison of various CS reconstruction methods

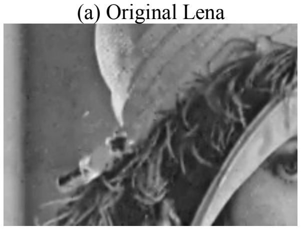
Image	Sub rate	TV[5]		TVNLR1[6]		TV+BM3D[16]		DTV-NL		DTV-BM3D		DTV-NLR1	
		PSNR	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*
Lena	**0.1**	28.25	0.773	30.04	0.818	30.04	0.813	30.04	0.824	30.21	0.827	**31.10**	0.847
	0.2	31.15	0.839	33.06	0.871	32.91	0.874	33.49	0.883	33.71	0.886	**34.44**	0.901
	0.3	33.23	0.878	35.02	0.902	35.25	0.905	35.50	0.911	35.82	0.915	**36.46**	0.926
Barbara	**0.1**	22.43	0.557	23.05	0.595	22.54	0.568	22.91	0.597	23.01	0.600	**25.15**	0.715
	0.2	23.95	0.639	26.45	0.757	25.28	0.709	25.85	0.743	26.35	0.765	**30.55**	0.884
	0.3	25.47	0.714	29.13	0.840	28.20	0.818	28.54	0.839	29.47	0.863	**33.88**	0.933
Peppers	**0.1**	29.10	0.778	30.38	0.804	30.08	0.799	30.15	0.806	30.19	0.807	**31.61**	0.833
	0.2	32.01	0.831	33.05	0.846	32.70	0.841	33.11	0.850	33.18	0.852	**34.25**	0.876
	0.3	33.69	0.861	34.39	0.871	34.23	0.867	34.45	0.870	34.70	0.876	**35.73**	0.900
Camera-man	**0.1**	28.66	0.834	30.87	0.870	30.81	0.871	30.21	0.865	30.98	0.880	**32.37**	0.901
	0.2	32.53	0.901	34.57	0.924	35.20	0.932	34.64	0.957	35.67	0.941	**36.19**	0.945
	0.3	35.30	0.935	37.24	0.951	37.84	0.955	37.52	0.955	**38.68**	0.963	38.57	0.963
Average		29.65	0.795	31.44	0.837	31.26	0.829	31.37	0.842	31.83	0.848	**33.36**	0.885


(a) Original Lena


(b) TVNLR1[6], 33.02dB


(a) Original Barbara


(b) TVNLR1[6] – 26.55dB


(c) TV+BM3D[16], 33.19dB


(d) DTV-NL, 33.42dB


(c) TV+BM3D[16], 25.31dB


(d) DTV-NL, 25.95dB


(e) DTV-BM3D – 33.75dB


(f) DTV-NLR1 – 33.95dB


(e) DTV-BM3D, 26.30dB


(f) DTV-NLR1 –30.61dB

Fig. 3. Visual quality comparison of various CS reconstruction methods for image Lena and Barbara at subrate 0.2.

Our experiments show that, with a better recovery method (e.g., TVNLR1), we can further improve performance of the proposed DTV scheme. That is, DTV-NLR1 gains over DTV-BM3D by 0.52-1.40dB in Lena, Peppers, and Cameraman images, and up to 4.41dB in Barbara at subrate 0.3. As illustrated in Fig. 3, the textures region (Lena's hair, small scale detail of Barbara's scarf) and smooth regions are preserved well. The proposed DTV-NLR1 even produces much better performance than TVNLR1. Therefore, the proposed DTV framework with a better filter and/or reconstruction algorithm is expected to achieve the further higher reconstruction performance (both subjectively and objectively).

5. CONCLUSION

This paper proposed a detail-preserving TV-based CS recovery method, especially for low subrate, based on cartoon-texture image decomposition. A structure-preserving filter was used to preserve nonlocal structure in cartoon and final recovered images in the spatial domains. The proposed method is also shown to preserve small-scale details of image well. In addition, the proposed decomposition-based recovery scheme is also tested to work successfully with other TV-based recovery method and structure-preserving filter.

6. REFERENCES

[1] D. Donoho, “Compressed sensing,” *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] M. Duarte and R. Baraniuk, “Kronecker compressive sensing,” *IEEE Trans. Image Process.*, vol.21, no.2, pp. 494–504, 2012.

[3] A. Buades, B. Coll, and J. Morel, “Image denoising methods. A new nonlocal principle,” *SIAM Review*, vol. 52, no. 1, pp. 113-147, 2010.

[4] T. Goldstein and S. Osher, “The split Bregman method for L1 regularized problems,” *SIAM J. on Imaging Sci*., vol. 2, no. 2, pp. 323-343, 2009.

[5] S. Shishkin, H.Wang and G. Hagen, “Total variation minimization with separable sensing operator,” in Proc. *Conf. on Image and Signal Process. (ICISP)*, pp. 86–93, 2010.

[6] T. N. Canh, K. Q. Dinh and B. Jeon, “Total variation reconstruction for Kronecker compressive sensing with a new regularization,” in Proc. *Pic. Coding Symp. (PCS),* pp. 261-264, 2013.

[7] A. Buades, T. M. Le, J. M. Morel, and L. A. Vese, “Fast cartoon + texture image filter,” *IEEE Trans. Image Process*., vol. 19, no. 8, pp. 1978-1986, 2010.

[8] J. F Cai, S. Osher, and Z. Shen, “Split Bregman methods and frame based image restoration,” *SIAM Multiscale Model. Simul.* vol. 8, no. 2, pp. 337-369, 2009.

[9] S. Mun and J. E. Fowler, “Residual reconstruction for block-based compressed sensing of video,” in Proc. *Data Comp. Conf. (DCC)*, pp. 183-192, 2011.

[10] W. Guo, J. Qin and W. Yin, A new detail-preserving regularity scheme, *Rice CAAM technical report 13-01*, 2013.

[11] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error measurement to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080-2095, 2007.

[13] S. Mun and E. Fowler, “Block compressed sensing of images using directional transforms,” in Proc. *IEEE Intern. Conf. on Image Process. (ICIP)*, pp. 3021-3024, USA, 2009.

[14] K. Q. Dinh, H. J. Shim, and B. Jeon, “Measurement coding for compressive Imaging based on structured measurement matrix,” in Proc. *IEEE Inter. Conf. on Image Process. (ICIP)*, pp. 10-13, 2013.

[15] W. Zuo, L. Zhang, C. Song, and D. Zhang, “Gradient histogram estimation and preservation for texture enhanced image denoising,” *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2459-2472, 2014.

[16] Y. Kim, H. Oh, and A. Bilgin, “Video compressed sensing using iterative self-similarity modeling and residual reconstruction,” *J. of Electron. Imaging,* vol. 22, no.2, pp. 021005, 2013.

[17] K. Bredies, K. Kunisch and T. Pock, “Total generalized variation,” *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp.492-526, 2010.

[18] G. Kutyniok and D. Labate, “Resolution of the wavefront set using continuous shearlets,” *Trans. Amer. Math. Soc.*, vol. 361, pp. 2719-2754, 2009.