Signal Processing: Image Communication 54 (2017) 93–106

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Block compressive sensing of image and video with nonlocal Lagrangian multiplier and patch-based sparse representation

Trinh Van Chien[a,1], Khanh Quoc Dinh[a], Byeungwoo Jeon[a,*], Martin Burger[b]

[a] School of Electrical and Computer Engineering, Sungkyunkwan University, Republic of Korea
[b] Institute for Computational and Applied Mathematics, University of Münster, Germany

ARTICLE INFO

Keywords:
Block compressive sensing
Distributed compressive video sensing
Total variation
Nonlocal means filter
Sparsifying transform

ABSTRACT

Although block compressive sensing (BCS) makes it tractable to sense large-sized images and video, its recovery performance has yet to be significantly improved because its recovered images or video usually suffer from blurred edges, loss of details, and high-frequency oscillatory artifacts, especially at a low subrate. This paper addresses these problems by designing a modified total variation technique that employs multi-block gradient processing, a denoised Lagrangian multiplier, and patch-based sparse representation. In the case of video, the proposed recovery method is able to exploit both spatial and temporal similarities. Simulation results confirm the improved performance of the proposed method for compressive sensing of images and video in terms of both objective and subjective qualities.

1. Introduction

Current video coding techniques, such as HEVC [1], are designed to have low-complexity decoders for broadcasting applications; this is based on the assumption that large amounts of resources are available at the encoder. However, many emerging real-time encoding applications, including low-power sensor network applications or surveillance cameras, call for an opposite system design that can work with very limited computing and power resources at the encoder. Distributed video coding (DVC) [2] is an alternate solution for a low-complexity encoder, in which the encoding complexity is substantially reduced by shifting the most computationally-intensive module of motion estimation/motion compensation to the decoder. Nonetheless, other than the encoding process, the processes of image/video acquisition also need to be considered to further reduce the complexity of the encoder [2] because current image/video applications capture large amounts of raw image/video data, most of which are thrown away in the encoding process for achieving highly compressed bitstream. In this context, compressive sensing (CS) has drawn interest since it provides a general signal acquisition framework at a sub-Nyquist sampling rate while still enabling perfect or near-perfect signal reconstruction [3]. More clearly, a sparse signal that has most entries equal to zero (or nearly zero) can be sub-sampled via linear projection onto sensing bases; this can be reconstructed later by a sophisticated recovery algorithm, which basically seeks its K-sparse approximation (i.e., the K largest magnitude coefficients). Consequently, CS leads to simultaneous signal acquisition and compression to form an extremely simple encoder. Despite its simplicity, its recovery performance is heavily dependent on the recovery algorithm, in which some of the important factors are properly designing the sparsifying transforms and deploying appropriate denoising tools.

Although many CS recovery algorithms have been developed, including NESTA (Nesterov's algorithm) [4], gradient projection for sparse reconstruction (GPSR) [5], Bayesian compressive sensing [6–8], smooth projected Landweber (SPL) [9], and total variation (TV)-based algorithms [10–12], their reconstructed quality has yet to be improved much, especially at a low subrate. For better CS recovery, Candes [13] proposed a weighted scheme based on the magnitude of signals to get closer to ℓ_0 norm, while still using ℓ_1 norm in the optimization problems. In a similar manner, Asif et al. [14] adaptively assigned weight values according to the homotopy of signals. As another approach, the authors in [15–17] utilized local smoothing filters, such as Wiener or Gaussian filters, to reduce blocking artifacts and enhance the quality of the recovered images. Despite these improvements, the performances of the aforementioned approaches are still far from satisfactory because much of the useful prior information of the image/video signals (e.g., the non-local statistics) was not taken into full account.

More recent investigations have sought to design a sparsifying transform to sparsify the image/video signal to the greatest degree

* Corresponding author.
E-mail address: bjeon@skku.edu (B. Jeon).
[1] He is now with Communication Systems Division, Department of Electrical Engineering (ISY), Linköping University, Sweden.

http://dx.doi.org/10.1016/j.image.2017.02.012
Received 26 March 2016; Received in revised form 28 February 2017; Accepted 28 February 2017
Available online 02 March 2017
0923-5965/

because the CS recovery performance can be closer to that of sampling at the full Nyquist rate if the corresponding transform signal is sufficiently sparse [3]. The direct usage of predetermined transform bases, such as the discrete wavelet transform (DWT) [7,15], discrete cosine transform (DCT) [8,15], or gradient transform [10–12,17], is appealing due to their low complexity. However, predetermined transform bases cannot produce sufficient sparsity (i.e., the number of zero or close-to-zero coefficients is limited) for the signal of interest, thereby limiting their recovery performance. Because image and video signals are rich in nonlocal similarities (i.e., a pixel can be similar to other pixels that are not located close to it), usage of those nonlocal similarities [18] can generate a higher sparsity level to achieve better recovery performance; this is known as a patch-based sparse representation approach [19]. Note that this approach originally showed much success in image denoising [19–22] and researchers have incorporated this idea into CS frameworks. Xu and Yin [23] proposed a fast patch method for whole-image sensing and recovery under a learned dictionary, while Zhang et al. [24] took advantage of hybrid sparsifying bases by iteratively applying a gradient transform and a three-dimensional (3D) transform [20]. By using the concept of decomposition, the authors in [25] also used a 3D transform for cartoon images to enhance the recovery quality. The 3D transform can be considered a global sparsifying transform because it is used for all patches of the recovered images. Dong et al. [26], motivated by the success of data-dependent transforms for patches (referred to as local sparsifying transforms) such as principal component analysis (PCA) or singular-value decomposition (SVD), proposed a method to enhance the sparsity level with the *logdet* function to bring ℓ_1 norm closer to ℓ_0 norm, similar to the work of Candes [13]. Metzler et al. [27] acquired a local sparsifying transform via block matching [21] and demonstrated the effectiveness of applying denoising tools to the CS recovery of the approximate message passing (AMP) method. However, because of the frame sensing that accesses the entire image at once, the work described in [23,24,26,27] requires extensive computation and huge amounts of memory for storing the sensing matrix [28]; thus, these approaches are not suitable as sensing schemes for real-time encoding applications or large-scale images/video.

Alternatively, block compressive sensing (BCS) has been developed to deal more efficiently with large-sized natural images and video by sensing each block separately using a block sensing matrix with a much smaller size. The compressive sensor can instantly generate the measurement data of each block through its linear projection rather than waiting until the entire image is measured, as is done in frame sensing. The advantages of BCS are discussed in [9,28–30]. However, in BCS, the recovery performance has yet to be substantially improved in comparison to that of frame sensing. To address this problem on the sensing side, a Gaussian regression model between the coordinates of pixels and their gray levels can be used to achieve better performance compared to traditional Gaussian matrices [31]. Additionally, Fowler et al. [32] developed an adaptive subrate method (i.e., multi-scale BCS) to exploit the different roles of wavelet bands. On the recovery side, for example, Dinh et al. [33] designed overlapped recovery with a weighted scheme to reduce the blocking artifacts caused by block recovery. Chen et al. [34] used the Tikhonov regularization and residual image information to enhance the smooth projected Landweber [9]. Furthermore, to enrich the details of recovered images, the K-SVD algorithm [19] was used in [35]. By sharing the same idea in [23,24,26,27,35] where nonlocal similarities are exploited to design the local sparsifying transform, group-based sparse representation (GSR) [36] can achieve better recovery performance (in terms of the peak signal-to-noise-ratio (PSNR)) than other algorithms that were previously designed for BCS. However, its recovered images still contain many visual artifacts since the nonlocal searching and collecting patches based on the initial recovered images produced by [34] often have poor quality at low subrates. Consequently, this implies that more efforts are required for improving both the objective and subjective quality.

This paper attempts to improve the recovery performance of the BCS framework by using TV minimization, which is good at preserving edges [10], with multiple techniques consisting of reducing blocking artifacts in the gradient domain, denoising the Lagrangian multipliers, and enhancing the detailed information with patch-based sparse representation. Furthermore, the proposed recovery methods are easily extendible to compressive sensing and encoding problems of video [37–41]. Specifically, our main contributions are summarized as follows.

- For BCS of images, we propose a method, referred to as multi-block gradient processing, that addresses the blocking artifacts caused by block-by-block independent TV processing during recovery. Furthermore, based on our observation that both image information (e.g., edges and details) and high-frequency artifacts and staircase artifacts are still prevalent in the Lagrangian multiplier of the TV optimization, we propose a method to reduce such artifacts by denoising the Lagrangian multiplier directly with a nonlocal means (NLM) filter. Because the direct application of the NLM filter is not effective in preserving local details with low contrast [18], we further propose enriching these low-contrast details through an additional refinement process that uses patch-based sparse representation. We propose using both global and local sparsifying transforms because the single usage of either transform limits the effective sparse basis and achievement of a sufficient sparsity level for noisy data. The proposed recovery method demonstrates improvements for BCS of images compared to previous works [7,8,15,16,34–36].
- For BCS of videos, we extend the proposed method to a compressive video sensing problem known as block distributed compressive video sensing (DCVS). An input video sequence is divided into groups of pictures (GOP), each of which consists of one key frame and several non-key frames. These undergo block sensing by a Gaussian sensing matrix. The proposed method first recovers the key frame using the proposed recovery method. Then, for non-key frames, side information is generated by exploiting measurements of the non-key and previously recovered frames in the same GOP. Improved quality of the non-key frames is sought by joint minimization of the sparsifying transforms and side information regularization. Our experimental results demonstrate that the proposed method performs better than existing recovery methods designed for block DCVS, including BCS-SPL using motion compensation (MC-BCS-SPL) [38] or BCS-SPL using multi-hypothesis prediction (MH-BCS-SPL) [39].

The rest of this paper is organized as follows. Section 2 briefly presents works related to the BCS framework with some discussion. The proposed recovery method for BCS of images is described in Section 3, and its extension to the block DCVS model is addressed in Section 4. Section 5 evaluates the effectiveness of the proposed methods compared to other state-of-the-art recovery methods. Finally, our conclusions are drawn in Section 6.

2. Block compressive sensing

In the BCS framework, a large-sized image u is first divided into multiple non-overlapping (small) blocks. Let a vector $\bar{u}_k$ of length n denote the kth block, which is vectorized by raster scanning. Its $m \times 1$ measurement vector b_k is generated through the following linear projection by a sensing matrix A_B:

$$b_k = A_B \bar{u}_k \tag{1}$$

A ratio (m/n) denotes the subrate (or sub-sampling rate, i.e., the measurement rate). BCS is memory-efficient as it only needs to store a small sensing matrix instead of a full one corresponding to the whole image size. In this sense, block sampling is more suitable for low-complexity applications.