J. Vis. Commun. Image R. 33 (2015) 286–300

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Joint modeling and reconstruction of a compressively-sensed set of correlated images

Kan Chang [a,*], Baoxin Li [b]

[a] School of Computer and Electronic Information, Guangxi University, Nanning, Guangxi 530004, China
[b] Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:
Received 5 October 2014
Accepted 28 September 2015
Available online 9 October 2015

Keywords:
Compressive sensing
Correlated images
Intra-image correlation
Inter-image correlation
Inter-channel correlation
Total variation
Non-local means
Group sparsity

ABSTRACT

Employing correlation among images for improved reconstruction in compressive sensing is a conceptually attractive idea, although developing efficient modeling strategies and reconstruction algorithms are often the key to achieve any potential benefit. This paper presents a novel modeling strategy and an efficient reconstruction algorithm for processing a set of correlated images, jointly taking into consideration inter-image correlation, intra-image correlation and inter-channel correlation. The approach starts with joint modeling of the entire image set in the gradient domain, which supports simultaneous representation of local smoothness, nonlocal self-similarity of every single image, and inter-image correlation. Then an efficient algorithm is proposed to solve the joint formulation, using a Split-Bregman-based technique. Furthermore, to support color image reconstruction, the proposed algorithm is extended by using the concept of group sparsity to explore inter-channel correlation. The effectiveness of the proposed approach is demonstrated with extensive experiments on both grayscale and color image sets. Results are also compared with recently proposed compressive sensing recovery algorithms.

1. Introduction

The theory of compressive sensing (CS) [1,2], which enables sub-Nyquist sampling rates, has attracted considerable research interests from signal processing communities in recent years. By incorporating signal acquisition with compression, it can significantly improve the energy efficiency of sensors. Based on the CS theory, several real CS-based imaging systems have been built, including the single-pixel camera [3], the compressive spectral imaging system [4], and the high-speed video camera [5], etc. However, perfectly reconstructing the images from a small number of random measurements is still a big challenge due to the ill-posed nature of inverse problems.

In this paper, we study the problem of reconstructing a set of correlated images, each of which is independently acquired by the CS technique. Such a set of images could be multi-view images which represent a scene from different view points, or a series of video frames which are taken at different time points. The former situation occurs in scenarios where the cameras are densely deployed but with strong power consumption constraints, e.g., wireless multimedia sensor networks (WMSN), while the latter one could appear in scenarios like commercial video processing cameras, video surveillance or dynamic magnetic resonance imaging (MRI) for medical imaging. It is no doubt that one can treat each image independently and use the reconstruction algorithms designed for single CS-imaging systems, e.g., [6–9]. However, if we can take advantage of the high correlation between images, the quality of the reconstructed images could be potentially improved.

There have been many studies focussing on the problem of CS reconstruction of correlated images. We roughly review three kinds of methods as follows:

- Techniques of the first kind rely on joint sparsity of the whole image set. For examples, Ma et al. [10] applied a modification of the approximate message passing algorithm and proposed to incorporate three-dimensional (3-D) dual-tree complex wavelet transform during reconstruction. Li et al. [11] recovered video cubes by minimizing total variation (TV) of the pixelwise discrete cosine transform (DCT) coefficients along the temporal direction. Hosseini et al. [12] proposed an alternative TV regularization model that utilizes high-order-of-accuracy differential finite impulse response (FIR) filter. Then the proposed TV was extended to tensorial representation to encode video features by decomposing each space–time dimension.

* Corresponding author.
E-mail addresses: changkan0@gmail.com (K. Chang), baoxin.li@asu.edu (B. Li).

http://dx.doi.org/10.1016/j.jvcir.2015.09.020
1047-3203/

- The second kind of techniques relies on the estimation of geometric transformations between images. For instance, Park and Wakin [13] used a manifold lifting algorithm to obtain a first estimation of the transformation parameters and a standard CS algorithm for the estimation of the images. In order to reduce computational complexity, Thirumalai and Frossard [14] estimated correlation between neighboring images directly from the linear measurements by the graph cut algorithm. Puy et al. [15] modeled each observed image as the sum of a geometrically transformed background image and a foreground image. Then a reconstruction method was proposed to jointly estimate the background and foreground images and the transformation parameters by minimizing a nonconvex functional.
- Techniques of the third kind use the neighboring images (usually spatially or temporally close to the current one) to represent the current image. The most direct ways to do representation are disparity estimation (DE)/disparity compensation (DC) for multi-view imaging and motion estimation (ME)/ motion compensation (MC) for video sequences. Mun et al. [16] performed block-based ME between reconstructed neighboring frames. To guarantee a more precise prediction, iteratively processing estimation–reconstruction was needed. Trocan et al. [17] applied the optical flow method to perform DE instead of the simple block matching method, and proposed to reconstruct the residual value in TV domain. Similar method was also proposed in [18], where the prediction from a neighboring image was applied as an adaptive transform to form a joint optimization problem for all images. Other ways to do representation include Karhunen–Loève transform (KLT)-based method [19], dictionary learning-based method [20] and multi-hypothesis (MH) prediction-based method [21], etc.

Among the three kinds of techniques discussed above, the last one usually leads to a better quality of reconstruction. The reason lies in that the correlation within image set is explored in image–image pattern, providing guaranteed robustness. Although inter-image correlation is extensively studied for CS reconstruction of correlated images, intra-image correlation, i.e., structured prior knowledge of images, is not explored well or is neglected in most existing methods. When inter-image correlation is not reliable, the quality of reconstruction would drop (this situation happens frequently, e.g., large disparity/motion occurs in the image sets, or the images are sampled at low subrates, making accurate estimation of inter-image correlation impossible). On the other hand, if color images are considered, there also exists strong inter-channel correlation which can be taken advantage of. Unfortunately, as far as we know, none of the methods designed for CS reconstruction of correlated images considers inter-channel correlation, while works like [22–24] only focuss on single color image reconstruction.

In this paper, we propose a novel strategy for high-fidelity CS reconstruction of correlated image set by jointly characterizing inter-image, intra-image and inter-channel correlation. Our main contributions are listed as follows. Firstly, we establish a joint modeling (JM) for correlated images, where inter-image correlation and intra-image structured prior knowledge including local smoothness and nonlocal self-similarity are explored simultaneously to ensure a more reliable and robust estimation. Then a new Split-Bregman-based algorithm is developed to efficiently solve the inverse problem where the minimization functional is formulated by using JM under regularization-based framework. After that, for CS reconstruction of color images, by utilizing the concept of group-sparsity, inter-channel correlation is additionally taken into consideration, and the proposed algorithm is further extended as a joint-channel reconstruction manner. Here we need to point out that the goal of joint modeling in this paper is to explore different kinds of correlation within correlated image sets, so that the compressively-sensed image sets can be accurately recovered; while the modeling in [25], the sparse representation-based method in [26], and the learning methods in [27,28] aim to extract multiple features for applications such as image annotation and human motion recognition.

The remainder of the paper is organized as follows. Section 2 briefly reviews the related background knowledge. Section 3 elaborates the design of JM for correlated images. The new functional which contains regularization term formed by JM, and the details of solving inverse problem are provided in Section 4. Section 5 extends the proposed CS reconstruction algorithm to support reconstruction of color images. Extensive experimental results and discussion are given in Section 6. Finally, we summarize this paper in Section 7.

2. Background

Consider an imaging system which independently takes measurements for each image in a correlated image set. The sensing procedure for the ith image can be written as

$$\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{u}_i \tag{1}$$

where $\mathbf{u}_i$ is an N-length vector, which denotes the original ith image; $\mathbf{y}_i$ contains M measurements of $\mathbf{u}_i$ and $\mathbf{\Phi}_i$ is the measurement matrix, whose size is $M \times N$. We call M/N the subrate of CS, and decreasing the subrate causes the system to become highly ill-conditioned and eventually underdetermined.

To reconstruct the underlying images from such an underdetermined system, one common way is to employ image prior knowledge for regularizing the solution to the following minimization problem

$$\min_{\mathbf{u}_i} \mathbf{\Psi}(\mathbf{u}_i) \quad s.t. \quad \mathbf{y}_i = \mathbf{\Phi}_i \mathbf{u}_i \tag{2}$$

where $\mathbf{\Psi}$ is the regularization term denoting image prior. Usually, this problem is further transformed to the following unconstrained optimization problem

$$\min_{\mathbf{u}_i} \mathbf{\Psi}(\mathbf{u}_i) + \frac{\lambda}{2} \|\mathbf{y}_i - \mathbf{\Phi}_i \mathbf{u}_i\|_2^2 \tag{3}$$

where λ is the regularization parameter that controls the tradeoff between the data fidelity term and the regularization term.

The most current CS reconstruction algorithms explore the prior knowledge that a natural image is sparse in some domains, such as DCT, wavelets and the gradient domain utilized by the TV norm [29]. Particularly, TV-based CS reconstruction algorithms, which explore image local structural information and achieve state-of-the-arts results, turn (3) to

$$\min_{\mathbf{u}_i} \|\mathbf{D}\mathbf{u}_i\|_1 + \frac{\lambda}{2} \|\mathbf{y}_i - \mathbf{\Phi}_i \mathbf{u}_i\|_2^2 \tag{4}$$

where $\mathbf{D} = [\mathbf{D}_h^T, \mathbf{D}_v^T]^T$ and $\mathbf{D}_h,\ \mathbf{D}_v$ denote horizontal and vertical finite difference operators, respectively.

In natural images, there might exist many pixels which are very similar to each other in values but spatially far away from a current pixel. This phenomenon is named as the nonlocal structure [30], and it has been successfully utilized to enhance the performance of TV-based CS reconstruction by applying the nonlocal means (NLM) filter via regularization, which is represented as

$$\mathbf{\Psi}(\mathbf{u}_i) = \|\mathbf{u}_i - \mathbf{w}_i \mathbf{u}_i\|_p^p \tag{5}$$

where $\mathbf{w}_i$ is the weight matrix. Assume that $u_{i(j_1)},\ u_{i(j_2)}$ are two pixels at index j_1 and j_2 in image $\mathbf{u}_i$, respectively, the weight at (j_1, j_2) in $\mathbf{w}_i$ is calculated by