J. Vis. Commun. Image R. 31 (2015) 197–207

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Optimal-correlation-based reconstruction for distributed compressed video sensing ☆

Haixiao Liu [a], Bin Song [a,*], Fang Tian [a], Hao Qin [a], Xiao Liu [b]

[a] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China
[b] College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

ARTICLE INFO

Article history:
Received 6 March 2014
Accepted 30 June 2015
Available online 7 July 2015

Keywords:
Compressed sensing
Distributed video coding
Bregman iteration
Correlation model
Spatial/temporal redundancy
Joint reconstruction
Optimal-correlation-based
Inter-signal correlation

ABSTRACT

Distributed compressed video sensing (DCVS) is a framework that integrates both compressed sensing and distributed video coding characteristics to achieve a low-complexity video coding. However, how to design an efficient joint reconstruction by leveraging more realistic signal models is still an open challenge. In this paper, we present a novel optimal-correlation-based reconstruction method for compressively sampled videos from multiple measurement vectors. In our method, the sparsity is mainly exploited through inter-signal correlations rather than the traditional frequency transform, wherein the optimization is not only over the signal space to satisfy data consistency but also over all possible linear correlation models to achieve minimum-l_1-norm correlation noise. Additionally, a two-phase Bregman iterative based algorithm is outlined for solving the optimization problem. Simulation results show that our proposal can achieve an improved reconstruction performance in comparison to the conventional approaches, and especially, offer a 0.7–9.9 dB gain in the average PSNR for DCVS.

1. Introduction

Distributed video coding (DVC) [1] refers to a special video coding paradigm that encodes frames of a video sequence independently and decodes them jointly. As the temporal redundancies are exploited by the decoder exclusively, the computational burden is shifted from the encoder to the decoder, which makes DVC potentially applicable to many fields, e.g., wireless multimedia sensor networks (WMSN), video conferencing with mobile devices and surveillance systems. However, it still requires enormous data collection followed by data compression and thus, wastes valuable resources. Compressed sensing (CS) [2–4] is an innovative concept that has attracted considerable research interest in the signal processing community. It provides a new way to collect data incorporating both acquisition and compression, and consequently helps reduce the required number of measurements and transcend hardware limitations. Hence, the advantage of CS makes it a natural fit for DVC, due to the great reduction of sampling rate, power consumption and computational complexity.

Benefit from CS and DVC, distributed compressed video sensing (DCVS) [5–9] has recently emerged as a new way to directly capture video data via random projections at a low-complexity encoder, while performing joint reconstruction at a more complex decoder. The main challenge of DCVS is how to utilize the spatial/temporal redundancy in video at the decoder to achieve efficient reconstruction. One of the earlier works addressing DCVS was presented by Prades-Nebot et al. [5], in which a video sequence is divided into key frames and non-key (NK) frames. Key frames are intra encoded and decoded using traditional video compression standards; while NK frames are projected and recovered using CS techniques, with an adaptive redundant dictionary built by picking blocks from previously reconstructed frames. A similar method was proposed in [6], introduced as an inter-frame sparsity model. However, in these schemes, it is still required to capture huge amounts of raw video data for key frames, which are encoded using conventional compression algorithms.

In this work, we study the sparse reconstruction problem for compressively sampled videos from multiple measurement vectors, while the encoder still retains low computation complexity. Considering a WMSN scenario, a number of sensors measure video signals that are individually sparse in a certain basis and also correlated from sensor to sensor. Each signal is independently acquired via random projections and jointly reconstructed at a collection point. Then, it is expected that if the inter-signal correlation

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.
* Corresponding author. Fax: +86 29 88204409.
E-mail addresses: hxliu@stu.xdian.edu.cn (H. Liu), bsong@mail.xidian.edu.cn (B. Song), ftian@stu.xidian.edu.cn (F. Tian), hqin@mail.xidian.edu.cn (H. Qin), liux@njupt.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.jvcir.2015.06.020
1047-3203/

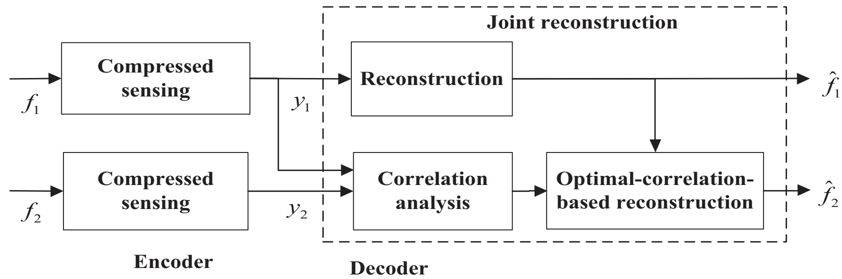

Fig. 1. The architecture of CS with joint reconstruction.

structure can be exploited at the decoder, the reconstruction performance could be significantly improved. In our work, the sparsity is mainly exploited through inter-signal correlations rather than the traditional frequency transform, wherein the optimization is not only over the signal space to satisfy data consistency but also over all possible linear correlation models to achieve minimum-l_1-norm correlation noise. In other words, since the conventional CS reconstruction is non-adaptive in the sense that no information but the sparsity of the signal is used, we try to utilize the temporal/spatial redundancy that exists in video signals to achieve an effective joint reconstruction, so that the required number of measurements can be substantially reduced (which then achieves a very low-complexity encoder) and the video quality could also be considerably improved.

1.1. Related work

Applications of CS to video sampling are still in an infant stage. One of the early works was to consider the whole sequence as a signal and apply 3-D wavelet transform to jointly explore the sparsity in the spatial and temporal domains [10]. It reconstructed the entire sequence at once, which makes this straightforward framework face several challenges, including a computationally expensive reconstruction process and huge memory required to store the random sampling operator, not to mention video sequences with many frames. One possible solution to these issues is to partition the frame into smaller blocks, i.e., block-based compressed video sensing. In [6], a novel two-phase measurement acquisition scheme with the inter-frame sparsity model was proposed for DCVS. However, in these methods, huge amounts of raw video data are still required for key frames encoding using conventional compression algorithms, which then wastes valuable resources. Chen and Kang proposed another DCVS framework in [7,8], wherein sparse representation was achieved through the K-SVD dictionary learning algorithm. In [28] the author presented a novel algorithm for learning a dictionary on a set of training signals using only compressive sensing measurements, which is a generalization of the well-known K-SVD algorithm and preserves its convergence properties. Specially, the dictionary is learned using samples extracted from previous recovered frames together with the side information. As soon as the trained dictionary is obtained, NK frames are reconstructed using the traditional CS recovery algorithm. However, signal reconstruction and sparse representation are designed as independent tasks in [7,8,28], and then, it has a negative impact in terms of consuming resources, since the sparse coefficient calculation has already been included in the process of dictionary learning. Zheng and Jacobs [11] presented a differencing method in order to take advantage of the spatial redundancy between neighboring frames, but it is only suitable for sequences that have small spatial changes. In [12], Ma et al. proposed a modified approximate message passing algorithm to recover the undersampled videos, by using the 3D dual-tree complex wavelet transform with some correlation noise. Liu et al. presented a motion-aware decoding method [13] for compressively sampled videos and a maximum frame rate video acquisition framework was proposed in [14]. Pudlewski et al. briefly discussed challenges involved in the transmission of video over a WMSN [15] and presented a cross-layer system that jointly controls the video encoding rate, the transmission rate, and the channel coding rate to maximize the received video quality [16]. Besides, a dictionary generation scheme for CS-based video sampling and a dictionary learning based DCVS reconstruction method were proposed in our previous work [17,18] respectively, and more recently, an adaptive alternating direction method of multipliers with its application to compressed video sensing was presented in [19,20].

1.2. Contributions

In this paper, we focus on the design of an efficient joint reconstruction algorithm in the DCVS framework, wherein more signal structures that go beyond simple sparsity are exploited to improve the reconstruction performance. Particularly, as illustrated in Fig. 1, assuming that the correlation between the two frames f_1 and f_2 can be described by a linear operator A with some correlation noise [21], then we could easily extend such model to the compressed domain, namely B. We show that when B is obtained through multiple measurements, there will be infinitely many models A that lead to B. To end this, we propose a novel optimal-correlation-based reconstruction method, in the sense that the optimization is not only over the signal space to satisfy data consistency but also over all possible models A to achieve the minimum-l_1-norm correlation noise. In this way, the sparsity is mainly exploited by means of inter-signal correlations, rather than through the frequency transform. Then, it is especially useful when the traditional basis could not provide sufficient sparse representation for CS reconstruction.

Another contribution of this paper is a two-phase Bregman [22–25] based iterative algorithm for solving the optimization problem. On one hand, the method obtains its solution by solving a sequence of unconstrained Lagrangian relaxation subproblems, and particularly it could converge very quickly when applied to l_1--regularization terms. On the other hand, it is more tolerant to numerical errors due to inaccurately solved subproblems, which enjoys an interesting error-forgetting property [25] in the sense that the errors do not accumulate and can even cancel each other under certain situations. Lastly, it is worth noting that in this paper we mainly focus on the joint reconstruction of signals from multiple measurement vectors and try to develop an efficient DCVS framework, which provides a novel fully low-complexity video sampling (and compression) paradigm without feedback channels and an alternative scheme adaptive to the environment where raw video data is not available, instead of competing compression performance against the current video compression standards or DVC architectures, which need raw data available for encoding.