Pattern Recognition Letters 45 (2014) 46–54

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

An adaptive rank-sparsity K-SVD algorithm for image sequence denoising ☆,☆☆

Yin Kuang [a,b], Lei Zhang [a,*], Zhang Yi [a]

[a] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China
[b] College of Computer Science, Chengdu Normal University, Chengdu 611130, PR China

ARTICLE INFO

Article history:
Received 12 August 2013
Available online 17 March 2014

Keywords:
Image sequence denoising
Rank-sparsity representation
Matrix decomposition
K-SVD
Adaptivity

ABSTRACT

In this paper, we propose an algorithm for the removal of additive white Gaussian noise (AWGN) from a given image sequence. By extending a frame in the spatial and temporal dimensions, the sequence is transformed into the volumetric data in which each frame includes both the spatial and temporal correlation. Image sequence denoising is then formulated as an optimization problem that can be iteratively solved by constructing a rank-sparsity representation on a propagated dictionary. The proposed algorithm effectively trains this dictionary by adaptively determining the required number of iterations. Restoration of the volumetric data is adaptively determined in terms of the noise level. The results on some standard data sets show that the proposed algorithm outperforms the K-singular value decomposition (K-SVD) algorithm and the sparse K-SVD algorithm. If a sequence is characterized by global motion (the moving objects in a scene with similar trajectories, i.e., they moves as a unit) or high motion activity, the performance of the proposed algorithm is comparable to that of block-matching and 4-D filtering (BM4D) and video block-matching and 4-D filtering (V-BM4D).

1. Introduction

Denoising is a fundamental problem of image processing. In recent decades, many approaches have been investigated from diverse points of view [16,13,10,7,18,1,23,27,28]. Image sequence (video) denoising is the extended version of this problem, because an image sequence always encloses the inherent temporal correlation between frames (images). However, in practice, some approaches ignore the temporal correlation enclosed in an image sequence and process each frame separately [2,4]. Other image sequence denoising methods explore the high temporal correlation in an image sequence to achieve better performance [6,12,29].

In general, many approaches of image sequence denoising can be divided into two categories depending on the utilization of temporal correlation. The first kind is the motion compensated filters that treat the motion compensation and filtering as two independent problems [5]. Motion compensation is either explicitly applied during preprocessing or implicitly incorporated into the filtering [5]. After preprocessing, the temporal nonstationarity in an image sequence is removed. Then, the estimated trajectories can be applied for denoising either in the signal domain [14] or transformed domain [34,17]. For these motion compensated filters, a motion compensation is assumed to be helpful when dealing with the dynamic nature of the image sequence. Therefore, they are expected to outperform their non-motion compensated counterparts. However, this is not always true. A motion compensation may be unnecessary or even counterproductive for denoising, because it may propose some inaccurate trajectories that will lead to blur and information loss, especially in an image sequence that contains high levels of noise. Moreover, a motion compensation is itself a difficult problem that adds an additional computational cost.

The other category of image sequence denoising methods is the spatio-temporal approaches that attempt to use the temporal correlation without motion compensation. Most of the spatio-temporal approaches are extended from classic 2-D filters [5], such as the techniques proposed by Buades et al. [6,12,29,30]. These spatio-temporal filters tend to be less sensitive to nonstationarity in both space and time, because they take advantage of the correlation in both directions. This fact implies that it is crucial to make full use of both the spatial and temporal correlation to maximize performance. Spatio-temporal filters can also adapt their parameters for denoising. Because there is no one set of parameters that can

☆ This paper has been recommended for acceptance by C. Luengo.

☆☆ This work was supported by National Basic Research Program of China (973 Program) under Grant No. 2011CB302201, and Specialized Research Fund for the Doctoral Program of Higher Education of China under Grants Nos. 20100181120030 and 20120181130007.

* Corresponding author. Tel./fax: +86 2885400618.
E-mail address: leizhang@scu.edu.cn (L. Zhang).

http://dx.doi.org/10.1016/j.patrec.2014.03.003
0167-8655/© 2014 Elsevier B.V. All rights reserved.

fit all sequences, even at a fixed noise level [29], many approaches use adaptive statistical estimation [2,4], adaptive selection of neighborhood size [3], or adaptive smoothing [19] to achieve better results.

One of the most successful non-motion compensated spatio-temporal filters is the approach reported by Protter and Elad [29]. This technique extends the work by [16,15] with several modifications. The results reported by Protter and Elad [29] demonstrated that a propagated dictionary can help speed-up the algorithm and lead to an improved denoising performance. This is because the similarity between two adjacent frames can reduce the number of iterations (denoted by K) required to train the dictionary. A further conclusion is that K should not be constant, but rather depends on the noise level [29]. However, no quantitative result is given.

The recent interest in denoising is related to low rank representation (LRR), which shows an excellent performance on many benchmark data sets [21,20,31]. LRR is able to automatically correct corrupted data [21], so it is mainly applied to reveal the actual segmentation of data (in the presence of noise or noise free) that are drawn from a union of multiple subspaces [21,20,31,22]. Compared with sparse representation (SR), LRR is more robust to noise and outliers, because LRR is better at capturing the global structure of data [21,20]. At the same time, the applications based on low-rank and sparse matrix decompositions have been reported for object detection [32], image classification [33], image inpainting [11], and dynamic magnetic resonance imaging restoration [26].

In this paper, we propose an algorithm similar to the K-singular value decomposition (K-SVD) algorithm, based on the foundational work by [16,15,29], to remove additive white Gaussian noise (AWGN) from image sequences. We propose three extensions to the original algorithm. The first is the rank-sparsity representation produced by solving an optimization problem, in which the representation is combined from LRR and SR. This is motivated by the conclusion that the exact solution to the problem of decomposing a matrix into the sum of a low-rank matrix and a sparse matrix can be found by minimizing the sum of the nuclear norm and the l_1 norm [8,9]. Unlike some other methods [32,8,9,29], the low-rank matrix and the sparse matrix are not separately used for different purposes. In fact, we propose that the sum of the low-rank matrix and the sparse matrix can be regarded as the hybrid representation matrix of signals on a specific dictionary, i.e., the identity matrix. Thus, signals are assumed to be linearly restored by the hybrid representation matrix on a redundant dictionary that is adaptively learned from the noisy signals. This is similar to the assumption by the authors of [16,15,29]. But we are interested in the rank-sparsity representation matrix of signals for training dictionaries.

The second extension relates to the adaptivity of K. We describe a method that experimental determines K in terms of the similarity between two adjacent frames in the transformed volumetric data. On the contrary, K is a constant in [16,15,29].

The last extension relates to the adaptivity of λ, the parameter that balances the method of signal restoration. According to the noise level, λ is adaptively determined, unlike [16,15,29] where it is constant.

The rest of this paper is organized as follows: some related work is presented in Section 2, and the proposed method is discussed in Section 3. Section 4 contains the experimental results obtained by the proposed algorithm. Finally, our conclusions are given in Section 5.

2. Related work

In this section, we will present some fundamental preliminaries. For clarity, denote a matrix $A = [a_1 \cdots a_m]$, where a_i $(1 \leqslant i \leqslant m)$ is the ith column vector of A. Denote a column vector $v = [v_1 \cdots v_n]^T$. For a given index $I = \{i_1, \ldots, i_p\}$, denote a sub-matrix of A and a sub-vector of v by $A_I = [a_{i_1} \cdots a_{i_p}]$ and $v_I = [v_{i_1} \cdots v_{i_p}]^T$, respectively, where $p \leqslant \min(m,n)$. Define $A_{I,J}$ as a sub-matrix of A, including the rows and columns indexed by I and J, respectively, where $J = \{j_1, \ldots, j_q\}$, and $q \leqslant m$.

2.1. K-SVD algorithm for image denoising

A successful approach for image denoising is the K-SVD algorithm, also named sparse and redundant representation modeling [16,15]. This approach assumes that a signal can be represented as the linear combination of a few atoms chosen from a redundant dictionary. Denoising is then implemented by iteratively learning the redundant dictionary from a noisy image.

More specifically, the image denoising task is described as the following optimization problem:

$$\langle x^*, D^*, c_i^* \rangle = \arg\min_{x,D,c_i} \left\{ \lambda\|x - y\|_2^2 + \sum_{i=1}^{M} \mu_i \|c_i\|_0 + \sum_{i=1}^{M} \|Dc_i - R_i x\|_2^2 \right\}, \tag{1}$$

where $x \in \mathbb{R}^N$ is a noise-free image and $y \in \mathbb{R}^N$ is the noisy counterpart, contaminated by AWGN $v \in \mathbb{R}^N$ with standard deviation σ, M is the number of patches extracted from a given image by using $R_i \in \mathbb{R}^{n \times N}, c_i \in \mathbb{R}^k$ is the representation of the ith patch on the dictionary $D = [d_1 \cdots d_k] \in \mathbb{R}^{n \times k}$ $(d_i \in \mathbb{R}^n, 1 \leqslant i \leqslant k,$ and $k > n)$, μ_i is the patch-specific weight, and λ is the parameter to control the restoration error.

Eq. (1) is iteratively solved using two steps. The first step is the sparse coding by

$$\begin{cases} c_i^* = \arg\min_{c_i} \|c_i\|_0 \\ \text{s.t. } \|Dc_i - R_i y\|_2^2 \leqslant (\rho\sigma)^2, \end{cases} \tag{2}$$

where $1 \leqslant i \leqslant M$, and ρ is the noise gain parameter.

The second step is to update D. Define $C = [c_1 \cdots c_M] \in \mathbb{R}^{k \times M}$ as the sparse representation matrix constructed by stacking all $c_i (1 \leqslant i \leqslant M)$. Define $B = [b_1 \cdots b_M] \in \mathbb{R}^{n \times M} (b_i = R_i y, 1 \leqslant i \leqslant M)$ as the input of the K-SVD denoising algorithm. For d_j $(1 \leqslant j \leqslant k)$, define $I_j = \{i | C_{j,i} \neq 0,\ 1 \leqslant i \leqslant M\}$. Then, the update is implemented by approximately solving the following optimization problem:

$$\begin{cases} \langle d_j^*, g^* \rangle = \arg\min_{d_j, g} \|E - d_j g^T\|_2^2 \\ \text{s.t. } \|d_j\|_2 = 1, \end{cases} \tag{3}$$

where $g^T = C_{j,I_j}$, $E = B_{I_j} - \sum_{i=1, i \neq j}^{k} d_i C_{i,I_j}$.

After K iterations, the output x^* is obtained by solving the following optimization problem:

$$x^* = \arg\min_x \left\{ \lambda\|x - y\|_2^2 + \sum_{i=1}^{M} \|Dc_i - R_i x\|_2^2 \right\}. \tag{4}$$

Clearly, the closed-form solution of (4) is

$$x^* = \left(\lambda\Pi + \sum_{i=1}^{M} R_i^T R_i \right)^{-1} \left(\lambda y + \sum_{i=1}^{M} R_i^T D c_i \right), \tag{5}$$

where $\Pi \in \mathbb{R}^{N \times N}$ is the identity matrix.

2.2. From image denoising to image sequence denoising

The K-SVD algorithm for image denoising is extended for image sequence denoising by Protter and Elad [29]. Denote the tth frame of the noise-free and noisy image sequences (where there are T frames) by $x(t) \in \mathbb{R}^N$ and $y(t) \in \mathbb{R}^N$, respectively, where $1 \leqslant t \leqslant T$. Define $b_i(j) \in \mathbb{R}^n$ as the temporal extension of b_i, extracted from