Optik 125 (2014) 1149–1153

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Principal component analysis algorithm in video compressed sensing

Sheng Liu, Mingming Gu*, Qingchun Zhang, Bing Li

College of Automation, Harbin Engineering University, Harbin 150001, PR China

ARTICLE INFO

Article history:
Received 27 March 2013
Accepted 29 July 2013

Keywords:
Principal component analysis
Compressed sensing (CS)
Video compression
Eigenvector

ABSTRACT

An algorithm of principal component analysis in video compressed sensing is proposed in the paper. Aiming at the compressed sensing problems of video sequences, the inter-frame correlation among the images is analyzed and the transform coefficients with lower value are removed according to the energy concentration characteristics of principal component analysis. Therefore, the sparse realization of video signals in the form of principal component analysis is accomplished and the possibility of the transformation being used in compressed sensing algorithm is verified. Finally, simulation results show that, with the comparison of the traditional algorithm based on wavelet transform, the proposed algorithm can not only improve the reconstructed quality and the visual effects of the video sequence, but also save the sampling resources. Moreover, it is more suitable for stream transmission of multimedia.

1. Introduction

With the advent of information age, people has increasing demand on the amount of information, especially image, video and other multimedia information, which results in a high cost of sampling. On the other hand, after the sampling process, the signal will be compressed in order to reduce the cost of transmission and storage, and the process will cause the waste of sampling resources again. Thus, the codec technology of video signals in the field of sampling is an important issue to be solved in the encoding and transmission study of streaming media.

Recently, scholars in relevant area come up a new theory of compressed sensing (also referred to as CS for short) [1,2]. In traditional sampling theory – Shannon Theorem, the sampling frequency must be higher than twice the highest frequency in order not to make the signal distortion. Yet compressed sensing theory shows, when the signal is compressible or sparse, its compressed representation can be accessed to directly, which can omit the sampling of abundant useless information. For CS, assuming that signal can be exactly recovered from incomplete information through a random measurement process, the representation of signal in a form of small set of data can be acquired. Each measured value of original signal passed through measurement matrix is considered important or unimportant equivalently and loss of a few can still perfectly reconstructs the original signal, therefore, the method can effectively [3].

CS sampling is a statistical technique of data acquisition and estimation, mainly used in the sampling and compression field of sparse data. Because of the compressibility of video images in some transform domain and their fairly superior sparsity of residual, CS theory has important applications in video encoding [4]. At present, most of the papers for CS sampling studied in image compression field, while seldom pay much attention to operating against autologous characteristic of video sequence. Paper [5,6] proposed a video codec scheme based on compressed sensing, but they both used traditional wavelet basis to get the images sparse in the processing of the key frames. In this paper, principal component analysis algorithm for video compressed sensing is presented based on the inter-frame correlation among the video images. The algorithm can remove inter redundancy to a large extent to get better compression ratio and reconstruction quality.

2. Compressed sensing model

Compressed sensing theory was first proposed by Candès et al. in 2004. The main idea of the theory is that the signal can be observed in a lower frequency as long as it is sparse after some orthogonal transform. In this case, we can get the compression form of the original signal with the minimum number of measurements. Then the measurements help to reconstruct the original signal. At this point, the number of measurements is only determined by the characteristics of the signal, and not restricted by Nyquist frequency. Consequently, compressed sensing is suitable for sampling of the signal with high bandwidth.

Consider a discrete signal $x(n)$ $(n = 0,1,\ldots,N-1)$ of length N as an $N \times 1$ dimensional column vector in $\mathbb{R}^N$ space, denoted by X, which can be represented by a linear combination of a set of orthogonal basis $\left\{\psi_i\right\}_{i=1}^{N}$. Then $\Psi = \left[\psi_1, \psi_2 \cdots, \psi_N\right]$ is an $N \times N$ basis matrix consisted of column vector ψ_i. If X is sparse on this set of the basis,

* Corresponding author. Tel.: +86 13895721487; fax: +86 45182588758.
E-mail address: coffeewanzi@163.com (M. Gu).

0030-4026/$ – see front matter
http://dx.doi.org/10.1016/j.ijleo.2013.07.120

the transform coefficients $\Theta = \left[\theta_1, \theta_2, \cdots \theta_N\right]^T$ are equivalently or similarly K-sparse expressions under the basis Ψ through $\Theta = \Psi^T X$. At this point of view, there are K nonzero coefficients and $K \ll N$. We can write the signal X as a linear combination of basis Ψ:

$$X = \Psi\Theta = \sum_{i=1}^{N} \theta_i \psi_i \quad (1)$$

M measurements of $Y = [y_1, y_2, \ldots, y_M]^T$ can be observed. Since $M < N$, this is a dimension reduction process, expressed by linear projection as:

$$Y = \Phi X = \Phi\Psi\Theta \quad (2)$$

Φ is an $M \times N$ measurement matrix. The equation is underdetermined because the dimension M of the measurement Y is lower than that of the original signal with N. That is, we cannot get exact X when Y is known. However, Θ is K-sparse makes it possible to solve the problem with optimal ℓ_0 norm, so as to make sure we have the minimum number of nonzeros in Θ [7], which considers the problem

$$\min_{\Theta} \left\|\Theta\right\|_0 \quad \text{s.t.} \quad Y = \Phi\Psi\Theta \quad (3)$$

On purpose to reconstruct the signal accurately, Candès and Tao proposed and proved that the measurement matrix must satisfy the Restricted Isometry Property (RIP) [8]. That means, for any K-sparse signal x, we have

$$(1-\delta_K)\|x\|_2^2 \le \left\|\Phi_T x\right\|_2^2 \le (1+\delta_K)\|x\|_2^2 \quad (4)$$

where $0 < \delta_K < 1$, $\left|T\right| \le K, T \subset \left\{1, 2, \ldots N\right\}$. where $0 < \delta_K < 1|T| \le K, T \subset \{1, 2, \ldots, N\}$. Φ_T is an $M \times |T|$ submatrix consisting of the related columns of Φ indexed by T.

Paper [9] points out that the equivalent condition for Restricted Isometry Property (RIP) is the observation operator and the basis matrix are irrelated. Since Gaussian random matrix is not related to most orthogonal basis matrices, it is chosen to be the measurement operator. Then irrelation condition can be satisfied well [10]. Otherwise, Rademacher matrix whose elements are all plus or minus 1 and local Fourier matrix can also meet the condition.

The reconstruction of the signal is to find the solution to the optimization problem of ℓ_0 norm which is an NP-hard problem. In such circumstances, we usually change to the consideration of optimization problem of ℓ_1 norm [11]. Then we have

$$\min_{\Theta} \left\|\Theta\right\|_1 \quad \text{s.t.} \quad Y = \Phi\Psi\Theta \quad (5)$$

At present, the main reconstruction algorithm includes basis pursuit (BP), matching pursuit (MP), orthogonal matching pursuit (OMP) and so on.

3. Principal component analysis of video sequences

Compression degree of the data depends on how the redundancy can be removed. And the redundancy is measured by means of correlation. Inter-frame image of the video sequence has a large temporal redundancy, also has a large correlation. Principal component analysis is built on the basis of statistical properties, also known as eigenvectors transform, K-L transform and Hotelling transform. It has outstanding advantage of good decorrelation. The transform determines its transformation matrix according to the statistical characteristics of the image (covariance matrix of the image). Therefore the signal correlation in transform domain can be removed entirely to make the image has the best matching effect. Stated thus, principal component analysis transform, known as the best transform based on minimum mean square error (MSE), occupies an important position in digital image compression technology.

For a single image in the video sequence, there is correlation between adjacent pixels because of intra-frame redundancy. Then principal component analysis can be adopted to get the transformation and compression of the whole image. In the meantime, adjacent frames in the video sequence also have correlation by reason of inter redundancy between natural images of the sequence. With the principle above, we can utilize principal component analysis in the transformation process of the entire video sequence.

Consider an I-frame video sequence composed of images of size $W \times L$ and the sequence can be represented as

$$squ = f_i(x, y) \quad (6)$$

$1 \le x \le L, 1 \le y \le W, 1 \le i \le I$.

Divide the images into blocks of size $n \times n$, and assume W, L are both integral multiples of n. So each image is divided into LW/n^2 blocks. The vector $\mathbf{X}_i^t$ can be generated by row stacking or column stacking of block image $f_i^t(x, y)$, which is the image of the tth block in the ith frame. Here we have

$$\mathbf{X}_i^t = (f_i^t(1,1), f_i^t(1,2) \cdots f_i^t(1,n), f_i^t(2,1), \cdots f_i^t(2,n) \cdots f_i^t(n,1) \cdots f_i^t(n,n))^T \quad (7)$$

where $1 \le t \le LW/n^2$, $1 \le i \le I$.

If the mean vector is defined as

$$\mathbf{m}_f = \mathrm{E}\left\{\mathbf{X}\right\} \quad (8)$$

and the covariance matrix of $\mathbf{X}$ is

$$\mathbf{C}_f = E\{(\mathbf{X} - \mathbf{m_f})(\mathbf{X} - \mathbf{m_f})^T\} \quad (9)$$

For the tth block image in the image sequence of I frames, its mean vector and covariance matrix can be written

$$\mathbf{m}_f^t = E\left\{\mathbf{X}\right\} = \frac{1}{I}\sum_{i=1}^{I} \mathbf{X}_i^t \quad (10)$$

$$\mathbf{C}_f^t = E\left\{(\mathbf{X} - \mathbf{m_f})(\mathbf{X} - \mathbf{m_f})^T\right\} \approx \frac{1}{I}\sum_{i=1}^{I}(\mathbf{X}_i^t - \mathbf{m}_f^t)(\mathbf{X}_i^t - \mathbf{m}_f^t)^T$$
$$\approx \frac{1}{I}\sum_{i=1}^{I}\mathbf{X}_i^t\mathbf{X}_i^{tT} - \mathbf{m}_f^t\mathbf{m}_f^{tT} \quad (11)$$

where $\mathbf{m}_f^t$ is a vector of n^2 elements, $\mathbf{C}_f^t$ is a square matrix of n^2 order.

If $\lambda_i^t (i = 1, 2, \ldots n^2)$ are the eigenvalues of covariance matrix $\mathbf{C}_f^t$ in descending order, $\mathbf{e}_i^t = \left[\mathbf{e}_{i1}^t, \mathbf{e}_{i2}^t, \ldots, \mathbf{e}_{in^2}^t\right]^T (i = 1, 2, \ldots, n^2)$ are the corresponding eigenvectors of $\mathbf{C}_f^t$, then transformation matrix $\mathbf{A}^t$ of the image block is

$$\mathbf{A}^t = \begin{pmatrix} e_{11}^t & e_{12}^t & \cdots & e_{1n^2}^t \\ e_{21}^t & e_{22}^t & \cdots & e_{2n^2}^t \\ \vdots & \vdots & \ddots & \vdots \\ e_{n^21}^t & e_{n^22}^t & \cdots & e_{n^2n^2}^t \end{pmatrix} \quad (12)$$

Centralize the image vector, which means calculation of the difference between the original image vector $\mathbf{X}^t$ and the mean vector $\mathbf{m}_f^t$, such that

$$\hat{\mathbf{X}}^t = \mathbf{X}^t - \mathbf{m}_f^t \quad (13)$$

Accordingly, the principal component analysis is

$$\mathbf{S} = (\mathbf{A}^t)^T \hat{\mathbf{X}}^t \quad (14)$$