Optik 126 (2015) 882–887

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Human action recognition via compressive-sensing-based dimensionality reduction

Jun Jiang [a,b], Xiaohai He [a,*], Mingliang Gao [a], Xiaofei Wang [a], Xiaoqiang Wu [a]

[a] *College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China*
[b] *School of Computer Science and Technology, Southwest Petroleum University, Chengdu 610064, China*

ARTICLE INFO

Article history:
Received 9 February 2014
Accepted 17 February 2015

Keywords:
Compressive sensing
Dimensionality reduction
Action recognition

ABSTRACT

We propose a new dimensionality reduction method called compressive sensing with Gaussian mixture random matrix (CS-GMRM), in which a novel measurement matrix using Gaussian mixture distribution is constructed and is proved to satisfy the restricted isometry property. The CS-GMRM method projects high-dimensional vector spaces into low-dimensional ones via a single matrix multiplication. In particular, the proposed method removes the need of a training process, preserves the metric information of the original vector space, and requires a low level of computational complexity. We apply our method to the problem of recognizing human action from video sequences. Experimental results show that the proposed method is simultaneously highly effective and highly efficient for action recognition, and outperforms the state-of-the-art dimensionality reduction methods.

1. Introduction

In recent years, human action recognition has attracted intensive attention from computer vision community and machine learning community due to the demands of various applications such as intelligent video surveillance, online video information retrieval, human-computer interaction, and smart robotics [1]. To obtain high accuracy, many action recognition approaches tend to extract rich information features from the actions which are usually represented by very high dimensional vectors. However, very high dimensional features would not only cause poor recognition performance due to the incapability of revealing the intrinsic properties of data, but also result in high computational complexity which could be unacceptable in many situations such as real-time applications. Therefore, in order to extract the inherent properties hidden in the high-dimensional data and reduce the computational complexity, the dimensionality reduction is an indispensable part of many action recognition methods [2–5].

A large number of dimensionality reduction techniques, such as principal component analysis (PCA), multidimensional scaling (MDS), isometric feature map (ISOMAP), and local linear embedding (LLE), have been proposed in the past decades [6]. PCA is a popular linear projection technique that uses low-dimensional latent data variables to represent the raw high-dimensional data variables with maximal preserved variance [7]. MDS is also a linear method that attempts to reduce the dimension of data in a distance-preserving manner [8,9]. ISOMAP and LLE both belong to nonlinear methods. ISOMAP models the input data as a high-dimensional manifold, utilizes the graph distance to approximate the geodesic distance, and then maps the manifold into low-dimensional embedding by keeping the distance relationship among the data points intact [10,11]. LLE works in a similar way to ISOMAP as it builds a graph representation of the input data. It constructs the low-dimensional manifold of the data by keeping local geometry properties, in which the point can be reconstructed by a linear combination of its nearest neighbors [12,13]. Nevertheless, these methods require learning from a training set by tuning parameters and are thus subject to limited generalization. In addition, a large amount of information might be lost owing to lack of data variables, which will turn to degrade the recognition accuracy. Finally, most conventional dimensionality reduction approaches involve extremely time-consuming operations, such as eigenvalue decomposition, which lead to high computational complexity.

Recently, compressive sensing (CS) arising as a novel field of information theory, has provided a fundamentally new approach to recovering certain signals that are sparse from a set of measurements far fewer than the number of measurements required by Shannon sampling theorem [14–16]. Implicitly, CS can be applied to data compression and has been used in image and signal processing [17,18].

* Corresponding author. Tel.: +86 2885462766.
E-mail addresses: scu_hxh@126.com, hxh@scu.edu.cn (X. He).

http://dx.doi.org/10.1016/j.ijleo.2015.02.053

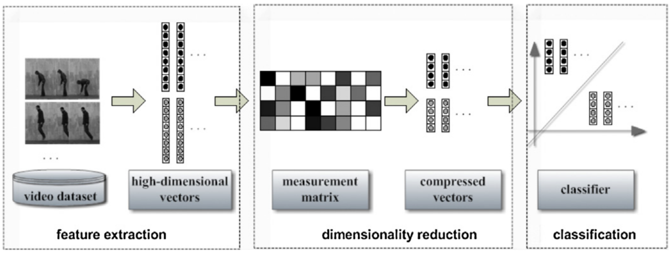

Fig. 1. Main components of the human action recognition framework.

In this paper, we propose a new dimensionality reduction method called compressive sensing with Gaussian mixture random matrix (CS-GMRM), in which a novel measurement matrix using Gaussian mixture distribution is constructed and is shown to satisfy the restricted isometry property (RIP). In addition, we use the firefly algorithm (FA) [19] to optimize the construction of the GMRM. The proposed CS-GMRM method maps high-dimensional vector spaces into low-dimensional ones via a single matrix multiplication without any training process. We then propose a human action recognition framework incorporating our CS-GMRM method. The main components of the framework are shown in Fig. 1. Experimental results on two benchmark datasets show that the proposed CS-GMRM method works quite well in human action recognition, and performs favorably against the state-of-the-art dimensionality reduction methods in terms of effectiveness and efficiency.

The rest of the paper is organized as follows: Section 2 introduces the background theory concerning the CS and the FA, respectively. In Section 3, we present the proposed method and the action recognition framework in detail. Experimental studies on benchmark action datasets are given in Section 4. We conclude the paper in Section 5.

2. Background theory

2.1. CS Preliminaries

Suppose that a vector $\boldsymbol{x}$ in R^N is K-sparse with respect to some bases. The measurements of $\boldsymbol{x}$, denoted as a vector $\boldsymbol{y}$ in R^M, $M \ll N$, is defined as

$$\boldsymbol{y} = \Phi \boldsymbol{x}, \tag{1}$$

where $\Phi \in R^{M\times N}$ is a measurement matrix. Furthermore, if Φ satisfies the RIP, then $\boldsymbol{x}$ can be exactly reconstructed from $\boldsymbol{y}$ with overwhelming probability. Φ is said to satisfy the RIP if there exists a constant $\delta \in (0, 1)$ such that

$$(1-\delta)||\boldsymbol{x}||_2^2 \le ||\Phi\boldsymbol{x}||_2^2 \le (1+\delta)||\boldsymbol{x}||_2^2 \tag{2}$$

holds for all K-sparse vectors in R^N [14].

Generally, verifying RIP of the measurement matrix Φ needs an exhaustive search over $\binom{N}{K}$ combinations, which results in great computational complexity. In practice, we use a more easily computable metric, i.e., the coherence of a matrix, to evaluate the performance of a measurement matrix. The coherence of the measurement matrix Φ, $\mu(\Phi)$, is defined as

$$\mu(\Phi) = \max_{1\le i\le j\le N} \frac{|\langle \varphi_i, \varphi_j\rangle|}{||\varphi_i||_2||\varphi_j||_2}, \tag{3}$$

where φ_i, φ_j are two columns of Φ, and $\langle \varphi_i, \varphi_j\rangle$ denotes the inner product of two columns [20]. Low coherence is expected and is intimately related to the RIP [21].

2.2. Firefly algorithm

Firefly algorithm is a metaheuristic technique for solving optimization problems [19], which mimics the behavior of social fireflies. Typically, three simplified rules are adopted [22]:

(1) All fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
(2) Attractiveness is proportional to their brightness, thus each fly tends to move toward the brighter one. The attractiveness is proportional to the brightness which decreases with increasing distance between flies. If there is no brighter one than a particular firefly, it will move randomly.
(3) The brightness of a firefly is somehow related to the analytical form of the objective function.

The two essential parts of the FA are: variation of light intensity and movement dominated by attractiveness. The variation of light intensity I is formulated as

$$I_i = f(\boldsymbol{x}_i), \quad 1 \le i \le n \tag{4}$$

where $f(\boldsymbol{x}_i)$ denotes the objective function (fitness function), $\boldsymbol{x}_i$ represents a solution (the position of a firefly i), and n is the total number of flies. The movement of a fly i, which is attracted to another brighter fly j, is formulated as

$$\boldsymbol{x}_i^{t+1} = \boldsymbol{x}_i^t + \beta_0 e^{-\gamma r^2}(\boldsymbol{x}_i^t - \boldsymbol{x}_j^t) + \alpha\varepsilon_i^t, \tag{5}$$

where r is the Euclidean distance between firefly i and firefly j, β_0 is the attractiveness at $r = 0$, α and γ are two constants, and ε_i is a random variable obeying a Gaussian distribution or a Lévy distribution [23].

3. Proposed method

3.1. Measurement matrix for dimensionality reduction

Designing a measurement matrix satisfying RIP is of central importance in dimensionality reduction. Motivated by Gaussian mixture models (GMM) presented in [24], we construct a novel measurement matrix $\Phi_{GM} \in R^{MN}$ called Gaussian mixture random matrix (GMRM), each entry, denoted as ϕ_{ij}, of which is an independent and identically distributed random variable following the Gaussian mixture distribution

$$p(\phi_{ij}) = \sum_{l=1}^{k} w_l N(\phi_{ij}|0, 1/(kMw_l)), \tag{6}$$

where k is the number of distributions, w_l is the weight of the lth Gaussian distribution satisfying $0 < w_l < 1$ and $\sum_{l=1}^{k} w_l = 1$, and $N(\phi_{ij}|0, 1/(kMw_l))$ represents a Gaussian distribution with mean 0 and variance $1/(kMw_l)$. The projection from $\boldsymbol{y} = \Phi_{GM}\boldsymbol{x}$ is indeed a dimensionality reduction and so likely loses information. However, if Φ_{GM} satisfies the RIP, then the metric structure of the original high-dimensional space is preserved in the low-dimensional compressed space, and little distance information is lost after the projection [14].

Proposition 1. Given $\varepsilon \in (0, 1)$, there exist positive constants a, b depending only on ε such that Φ_{GM} satisfies the RIP with probability $\ge 1 - 2e^{-bM}$ if $M \ge aK\log(N/K)$.

Proof. We first show $E(||\Phi_{GM}\boldsymbol{x}||_2^2) = ||\boldsymbol{x}||_2^2$. Let Φi denote the ith row of Φ_{GM}, $i = 1, \ldots, M$, then