Pattern Recognition 63 (2017) 182–192

Contents lists available at ScienceDirect

Pattern Recognition

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

Joint representation classification for collective face recognition[☆]

Liping Wang[a,*], Songcan Chen[b]

[a] Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

ARTICLE INFO

Keywords:
SRC
JRC
IQM
Practical IQM

ABSTRACT

In recent years, many representation based classifications have been proposed and widely used in face recognition. However, these methods code and classify testing images separately even for image-set of the same subject. This scheme utilizes only an individual representation rather than the collective one to classify such a set of images, doing so obviously ignores the correlation among the given set of images. In this paper, a joint representation classification (JRC) for collective face recognition is presented. JRC takes the correlation of multiple images as well as a single representation into account. Even for an image-set mixed with different subjects, JRC codes all the testing images over the base images simultaneously to facilitate recognition. To this end, the testing images are aligned into a matrix and the joint representation coding is formulated as a generalized $l_{2,q} - l_{2,p}$ matrix minimization problem. A unified algorithm, named by iterative quadratic method (IQM), and its practical implementation are developed specially to solve the induced optimization problem for any $q \in [1, 2]$ and $p \in (0, 2]$. Experimental results on three public databases show that the JRC with practical IQM not only saves much computational cost but also achieves better performance in collective face recognition than state-of-the-art methods.

1. Introduction

Recently, representation coding based classification and its variants have been developed for facial image recognition (FR) [1–5]. These schemes have achieved a great success in FR and have boosted the applications of image classification [6,7]. The main idea can be carried out by two steps: 1) coding a testing sample as a linear combination of all the training samples, then 2) classifying the testing sample to the subject with the most compact representation evaluated by coding errors. The equations frequently used for representation coding can be uniformly accommodated to the following framework

$$\min_{x} \| y - Ax \|_q^q + \lambda \| x \|_p^p, \quad 1 \le q \le 2, 0 < p \le 2, \tag{1}$$

where $A \in R^{m\times d}$ is the dictionary of coding atoms and $y \in R^m$ is a testing sample. $x \in R^d$ is the representation vector. With the solution x^* to (1), y is identified as follows

$$\text{identity}(y) = \arg\min_{1\le i\le I} \{\| y - A\hat{x}_i^* \|_2\}, \tag{2}$$

where I denotes the number of classes. $\hat{x}_i^*$ is the recovered coefficient vector associated with class i which is extracted from x^* by keeping the ith class coefficients while the other entries are set to zero [1].

Different pairs of $q \in [1, 2]$ and $p \in (0, 2]$ result in different representation coding formulas. Sparse representation based classification (SRC) [1] is the most known one which uses the l_1-regularized least square problem ($q = 2, p = 1$ in (1)) to sparsely code the query image. The experimental results [1] exhibit the amazing recognition performance of SRC. But the authors of [2] argued that SRC over emphasized the importance of l_1-norm sparsity but ignored the effect of collaborative representation. Consequently, a collaborative representation based classification with l_2-regularized least square (CRC-RLS) was presented which is the special case of (1) with $q = p = 2$. Anyway, CRC-RLS's coding problem is easier to solve for its smoothness than that of SRC. Moreover, Wright et al. [3] ever used $l_1 -$ norm to measure the coding fidelity of y over A, which is another special case of (1) with $q = p = 1$. Compared with the works mentioned above, the representation and regularization measurements of (1) are extended to $\|\cdot\|_q\,(1 \le q \le 2)$ and $\|\cdot\|_p\,(0 < p \le 1)$ respectively. This modification provides possibility to adaptively choose the best formula for different applications. Moreover, the computational experiences [8–10] have showed that fractional norm $l_p (0 < p < 1)$ exhibits sparser pattern than l_1-norm. Then the unified generalization formula (1) is expected to achieve better performance.

On the other hand, Eq. (1) uses a coding vector to represent the

☆ The work is partially supported by the Chinese grants NSFC11471159, 61661136001, 11611130018 and Natural Science Foundation of Jiangsu Province (BK20141409).
* Corresponding author.
E-mail addresses: wlpmath@nuaa.edu.cn (L. Wang), s.chen@nuaa.edu.cn (S. Chen).

http://dx.doi.org/10.1016/j.patcog.2016.10.004
Received 19 May 2015; Received in revised form 19 June 2016; Accepted 5 October 2016
Available online 08 October 2016
0031-3203/

testing samples one by one. Given a collection of query images $y_1, y_2, \ldots, y_n \in R^m$, Eq. (1) codes each y_j independently by all the training samples A as

$$y_j \approx Ax_j, \quad 1 \le j \le n. \tag{3}$$

Then y_j is assigned by (2) based on its most compact coding coefficient x_j^*. Obviously the recognition of y_j depends on the single representation coding x_j^* individually but takes no account of the correlation with other testing samples (y_l, $l \neq j$). Even though different frontal faces take on different appearances, the pixel intensity values taken from facial images have direct correlation [11]. Similar images are located together while dissimilar images are spaced far apart which plays an important role in recognizing facial images [12]. In many applications, a great number of images for each known subject have been collected from video sequence or photo album. Face recognition has to be conducted with a set of probe images rather than a single one [13–15]. Collective face recognition or image-set based face recognition seems important and necessary.

Compared with regular face recognition, image-set based face recognition is much less studied. Few image-set based face recognition methods [13–15] ever explored the set-to-set classification after the testing images have been pre-separated according to different classes. To the best of our knowledge, collective face recognition for image-set mixed with different subjects has never been straightly concerned due to the challenging complexity. In this paper, we propose a joint representation coding based classification (JRC) for collective face recognition. To make sufficient use of correlation among the given set of images, we consider to jointly represent all the testing samples simultaneously over the training sample base. Here we employ matrix instead of vector as the coding variable to evaluate the distribution of feature space. The joint representation scheme is eventually formulated as a $l_{2,q} - l_{2,p}$ matrix minimization which covers the vector framework (1). To solve the matrix optimization problem with generalized measurements, a unified algorithm and its practical implementation are proposed and the convergence behavior is accordingly analyzed. Experiments on three public face datasets validate the improvement of JRC over the state-of-the-arts.

In short, the main contributions of this paper lie in:

(1) A joint representation coding based classification (JRC) is presented which implements collective images representation coding simultaneously. This approach is more economical and efficient in computational cost and CPU time. Moreover, JRC can handle collective face recognition but the testing images are not necessarily pre-separated according to classes which is different from the set-to-set approaches employed in [13–15].
(2) Joint coding technique takes account of the correlation hidden in the multiple testing face images. The generalized measurements $q \in [1, 2]$ and $p \in (0, 2]$ in representation coding Eq. (12) provide possible adaption to different applications. For example when $0 < p \le 1$, all the testing images are jointly represented by the training samples in sparse pattern. The recovered largest row coefficients are distinguished according to different subjects but jointly clustered with respect to images of the same subject.
(3) To solve the joint representation $l_{2,q} - l_{2,p}$ matrix optimization problem (12), a uniform algorithm is developed for any $q \in [1, 2]$ and $p \in (0, 2)$. The algorithm makes objective function strictly decrease until it converges to the optimal solution. To the best of our knowledge, it is an innovative approach to solve such a generalized $l_{2,q} - l_{2,p}$ matrix minimization problem.

This paper is organized as follows. In the second section, a joint representation based classification (JRC) will be established. The third section is dedicated to an iterative quadratic (IQM) algorithm for solving the joint matrix optimization problem induced by JRC. Some computational details are considered in the fourth section and a practical implementation is developed. The experimental results are reported in the fifth section while the convergence analysis on IQM is presented in Appendix B.

2. Joint representation classification for collective face recognition

2.1. Joint representation formulation

Suppose that we have I classes of subjects in the facial image dataset. $A_i \in R^{m\times d_i} (1 \le i \le I)$ denotes the i-th class, and each column of A_i is a sample of class i. Hence all the training samples are aligned by $A = [A_1, A_2, \ldots, A_I] \in R^{m\times d}$, where $d = \sum_{i=1}^{I} d_i$. Denote $Y = [y_1, y_2, \ldots, y_n] \in R^{m\times n}$ all the query images, we propose to jointly represent the image set simultaneously by

$$Y \approx AX, \tag{4}$$

where $X = [x_1, x_2, \ldots, x_n] \in R^{d\times n}$ stands for the joint coding matrix. As far as the columns are concerned, system (4) is an easy consequence of (3). To measure the fidelity of the joint coding system (4), we consider X in another sense. Let $A^i \in R^d$ and $Y^i \in R^n$ be the ith ($i = 1, 2, \ldots, m$) row vectors of matrix A and Y respectively, formula (4) is equivalent to

$$X^T (A^i)^T \approx (Y^i)^T \quad \text{for } i = 1, 2, \ldots, m. \tag{5}$$

It is noticed that A and Y array the sampled images column by column, hence their rows span the training and testing feature spaces respectively. In feature extraction view, the collective coding matrix X also plays approximation projecting role from the training feature space to the testing feature space. Traditional least square regression aims to minimize the error

$$\min_X \sum_{i=1}^{m} \| X^T (A^i)^T - (Y^i)^T \|_2^2 \quad \text{or } \min_X \sum_{i=1}^{m} \| A^i X - Y^i \|_2^2. \tag{6}$$

(6) can be easily reformulated as

$$\min_X \sum_{i=1}^{m} \| (AX - Y)^i \|_2^2 \tag{7}$$

where $(AX - Y)^i$ is the ith row vector of $AX - Y$. Actually we prefer a uniform generalization of (7) in the sense

$$\sum_{i=1}^{m} \| (AX - Y)^i \|_2^q \quad (1 \le q \le 2). \tag{8}$$

Under the assumption that joint representation and feature distribution share the similar pattern for all testing facial images, we use the following regularization

$$\sum_{i=1}^{d} \| X^i \|_2^p \quad (0 < p \le 2), \tag{9}$$

where X^i is the ith row vector of X for $i = 1, 2, \ldots, d$. Combining (8) and (9), we present the joint representation formulation as follows

$$\min_X \sum_{i=1}^{m} \| (AX - Y)^i \|_2^q + \lambda \sum_{i=1}^{d} \| X^i \|_2^p, \quad (1 \le q \le 2, 0 < p \le 2). \tag{10}$$

When the number of testing samples in Y is 1, collective representation Eq. (10) is reduced to the single coding Eq. (1). Compared with coding vector x, joint coding matrix X expands each coefficient entry to a row vector which naturally reflects the integral structure of dataset. The row vector norm $\| X^i \|_2$ gives the joint coefficient of all the testing images over the ith training samples. Then the joint row coefficient vector $\{\| X^i \|_2\}_{i=1}^d$ somewhat measures the correlation of different classes. To illustrate the joint pattern of JRC, we randomly choose 3 images of two classes in Georgia-Tech database for testing (see Fig. 1). In the left images, 600 (12 images each of 50 classes)