J. Vis. Commun. Image R. 43 (2017) 21–29

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A two-stage learning approach to face recognition ☆

Xiao Dong [a], Huaxiang Zhang [a,b,*], Jiande Sun [a,b], Wenbo Wan [a,b]

[a] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong Province, China
[b] Institute of Data Science and Technology, Shandong Normal University, Jinan 250014, Shandong Province, China

ARTICLE INFO

Article history:
Received 6 July 2016
Revised 14 October 2016
Accepted 17 December 2016
Available online 23 December 2016

Keywords:
Collaborative representation
Collaborative coefficient
Face recognition
Small sample size

ABSTRACT

This paper introduces the Collaborative Representation (CR) techniques to small sample size conditions, and propose a Two-Stage learning approach to face recognition based on Collaborative Representation (TSCR). Based on the assumption that the same class samples should lie in the same subspace, we first use the unlabeled samples as dictionary atoms to construct each labeled sample, and obtain the collaborative coefficients by CR. The unlabeled sample with the largest collaborative coefficient is assigned the same class label as the reconstructed labeled sample, and is added to the labeled data set. This process is repeated until about half of the unlabeled samples are labeled and added to the labeled dataset. After that, we employ the original CR approach to classify the left unlabeled samples based on the newly labeled dataset. Experimental results demonstrate that the proposed TSCR is effective on face recognition.

1. Introduction

Face recognition(FR) is such a challenging yet interesting problem that it has attracted researchers from different fields, and various techniques have been proposed to handle different problems in FR, such as illumination, pose, occlusion and small sample size [1]. The dimensionality reduction methods for face images mainly include Eigenface, Laplacianfaces and Fisherface [2,3], and are widely used in pattern recognition and computer vision. A typical face recognition approach includes two steps: feature extraction and classification. Since Wright et al. [4] proposed a different face recognition approach SRC based on sparse representation techniques, SRC and its variants [5–8] have been widely applied to face recognition problems. In SRC, a test sample is linearly reconstructed by all training samples and is classified based on the reconstruction errors associated with each class. Different from classical classification methods, SRC does not need feature extraction.

Recent studies show that, Collaboration Representation (CR) [9] using l_2 sparse constraint rather than l_1 sparse constraint in the optimization problem has been successfully applied to face recognition. In fact, face recognition is a small sample size (SSS) problem, and the dimension of the training samples is larger than the number of training samples. In order to overcome this issue, a Collaborative Representation based Classification (CRC) scheme [10] was proposed by replacing the l_1 regularization [11] with the l_2 regularization. CR codes a testing sample by linear combination of all the training samples with regularized least square and performs classification based on the reconstruction errors. In recent years, deep learning [12–14] has become a hot research topic for its outstanding performance on face recognition. Deep learning techniques learn a multilayer neural network using a large scale of training data, and many parameters need to be adjusted in the training process. We focus on small sample size face recognition problems in this work.

All the above mentioned methods are based on the assumption that sufficient training samples are available. In reality, sample annotation is time consuming and laborious, and it is not easy to get enough labeled samples. Under the condition with small sample size, most of the traditional methods can not obtain satisfying performance.

As a matter of fact, many unlabeled samples are easily available and can be used for enhancing discrimination. Semi-supervised learning approaches [15] make the utmost of the correlations among the labeled and unlabeled samples. Considering the discriminative information of the collaborative coefficients in CR, we can make the utmost of these information for classification. In order to solve the SSS problem, unlabeled samples can be labeled by some approaches to expand the limited labeled sample set, thus increasing the number of the labeled samples, and reliable classifiers can be trained based on the expanded training dataset. Face images can be represented by different features, such as Eigenface, LBP (Local Binary Pattern) [16] and Gabor [17], and each feature

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: School of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong Province, China.
E-mail addresses: huaxzhang@hotmail.com, huaxzhang@163.com (H. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2016.12.006
1047-3203/

representation shows different recognition performance. We propose a two-stage learning approach to face recognition under small sample size conditions based on collaborative representation techniques, and do experiments under different feature representations. The proposed approach makes full use of the discriminative information of unlabeled samples to increase the classification performance. The contributions of this paper is summarized as follows:

(1) A two-stage classification approach is proposed. Different from the original CRC, it uses most of unlabeled samples to reconstruct the few labeled samples and expands the labeled sample set for the classification of the left unlabeled samples.
(2) The approach makes use of the collaborative coefficients to classify almost half of the unlabeled samples.

The rest of this paper is organized as follows. In Section 2, related works are introduced. The proposed TSCR is described in Section 3. In Section 4, the experimental performance of TSCR is evaluated and compared with other approaches. Section 5 concludes the paper.

2. Related work

In this section, we introduce the classical SRC and CRC approach for face recognition.

2.1. Sparse representation classifier

We concisely review SRC in the following. For the training samples from c classes, let $X = [X_1, X_2, X_3, \ldots, X_c]$, where $X_i \in R^{m\times n_i}$, c is the number of classes, and n_i is the number of data in the ith class. The number of training samples is $n = \sum_{j=1}^{c} n_j$. Each column of X_i denotes a training sample of the ith class. Given a test sample $y \in R^m$, we represent it as

$$y \approx X\alpha \tag{1}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_c)$ is the sparse representation coefficient vector. The SRC algorithm is summarized in Algorithm 1.

Algorithm 1. SRC

1: **Input:** the training data X.
2: Represent y over X via l_1 minimization.

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 s.t. \|y - X\alpha\|_2^2 < \varepsilon \tag{2}$$

3: Compute the residual

$$e_i(y) = \|y - X_i\hat{\alpha}_i\|_2 \tag{3}$$

where $\hat{\alpha}_i$ is the coefficient vector associated with the ith class
4: **Output:** $identity(y) = \arg\min_i e_i(y)$.

2.2. Collaborative representation classifier

SRC needs to solve a l_1-regularized minimum problem [11], and it is too expensive and time-consuming. Researches [10] show that the classifier can get competitive result with less computation cost if we replace the l_1-regularization term with l_2-regularization term in Eq. (2).

Given a test sample $y \in R^m$ and a training dataset $X \in R^{m\times n}$, the CRC coding model is represented as

$$\hat{\alpha} = \arg\min_{\alpha}\{\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_2^2\} \tag{4}$$

where λ is a regularized parameter, and $\alpha \in R^{n\times 1}$ is the collaborative representation of y in terms of X.

The role of the regularization term is twofold: it makes the least square solution stable, and introduces a certain amount of sparsity to the solution $\hat{\alpha}$. The solution of regularized least square based collaborative representation in Eq. (4) can be analytically derived

$$\hat{\alpha} = Py \tag{5}$$

where

$$P = (X^TX + \lambda I)^{-1}X^T. \tag{6}$$

The CRC algorithm is summarized in Algorithm 2.

Algorithm 2. CRC

1: Normalize the columns of X to have unit l_2-norm.
2: Code y over X by

$$\hat{\alpha} = (X^TX + \lambda I)^{-1}X^Ty \tag{7}$$

3: Compute the residual

$$e_i = \|y - X_i\hat{\alpha}_i\|/\|\alpha_i\| \tag{8}$$

where $\hat{\alpha}_i$ is the coefficient vector associated with the ith class
4: **Output:** $identity(y) = \arg\min_i e_i$.

3. Two-stage learning approach based on CR

Traditional supervised learning methods employ a lot of labeled samples as atoms of a dictionary to construct a test sample, but it is not easy to get enough labeled samples in SSS problems. Since most traditional methods including SRC and CRC can not obtain satisfactory performance under this scenario, we can naturally think of using the unlabeled samples to reconstruct the labeled ones so as to increase the number of training samples. In our algorithm, a large number of unlabeled samples are used to linearly reconstruct each labeled one. According to the subspace assumption [18], samples of the same class exist in the same subspace, and should have bigger representation coefficients than samples in different classes. The representation coefficients have discriminative information and can be used to classify unlabeled samples. At the first stage, each labeled sample is reconstructed by all the unlabeled ones, and the one with the largest representation coefficient is assigned the same label as the labeled reconstructed sample. After the label assignment, we have a new labeled and unlabeled sample set. The above process is repeated until a termination condition is satisfied. This stage enlarges the size of the labeled sample set. The label process in the first stage is shown in Fig. 1. In the second stage, the expanded labeled samples are served as dictionary atoms and are used to reconstruct the left unlabeled samples. For the sake of simplicity, we use a collaborative representation classifier to get the labels of the left unlabeled samples.

We do several experiments on YaleB dataset, and each time, we randomly choose one labeled sample and linearly reconstruct it by all the unlabeled ones. The results show that almost all the unla-