Multiple samples dictionary learning and locality constrained coding for face recognition

Ling Gan, and Ruifang Wang[a]

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[a]1512460579@qq.com

Keywords: Face recognition, dictionary learning, virtual sample, locality constrain.

Abstract: This paper proposes a new dictionary learning algorithm, Multiple Samples Dictionary Learning and Locality Constrained Coding algorithm (MSDL-LCC), to solve the problems that the insufficient number of training samples when learning a dictionary and the lack of discriminative power of the test coefficient. The proposed algorithm first generates virtual training samples for the origin training data, and then uses all the training samples to learn a dictionary. Finally the learned dictionary is used to encode the test samples under local constraint to obtain a coefficient matrix with discriminative power. Experimental results show that the proposed MSDL-LCC algorithm framework outperforms some previous state-of-the-art dictionary learning algorithms on the LFW and AR face databases.

1. Introduction

In the past few years, Sparse Representation (SR) has attracted the attention of many scholars in the field of face recognition [1,2]. The main idea of SR is that samples can be represented by a linear combination of a large number of "atoms" where these "atoms" are the column elements of the learned dictionary. Dictionary learning (DL), as one of the most important steps in SR, is widely studied by various researchers in recent years.

Various DL methods had been devised for recognition tasks. The K-SVD algorithm is one of the most famous DL algorithms [3]. It does not update the entire dictionary matrix at once when updating the dictionary, but uses the step of updating the atoms one by one. It improves the convergence speed of the algorithm and is widely used. However, this algorithm only emphasizes the reconstruction of training samples, so it is not suitable for classification tasks. Zhang et al. [4] considered the discriminability of the dictionary and proposed D-KSVD algorithm for classification tasks. Label information has a great effect on improving the recognition rate. Jiang et al. [5] added the label consistency constraint by connecting label information and each atom based on the K-SVD algorithm. On this base, they also added classification error term and proposed two kinds of Label Consistency K-SVD (LC-KSVD) algorithm. Xu et al. [6] proposed a sample diversity DL algorithm. By generating virtual versions of the original training samples, and using all training samples to the process of DL, the problem of insufficient training samples is solved.

How to obtain sparse and discriminative coding matrix is another research hot point of SR in recent years. Locality is more essential than sparsity, since locality leads to sparsity but not necessary vice versa [7]. Therefore, more and more researchers focused on locality information preservation. Min et al. [8] proposed Laplacian Regularized Locality-constrained Coding algorithm to address the problem of vector quantization errors. This algorithm obtained locality information by computing a Laplacian matrix. Wang et al. [9] proposed a new coding algorithm which added the local relationship between the neighbor feature points. In the above methods, the number of neighbors was usually a fixed constant and it would change with different conditions. Therefore, Yuan et al. [10] proposed a novel feature coding algorithm, which encoded features based on adaptive coding bases.

In this paper, dictionary learning and coefficients code of test samples are considered for better recognition rates. We adopt a method of generating virtual training samples to increase the number of training samples. During dictionary training, we not only consider the reconstruction error terms of the original training samples, but also consider the reconstruction error terms of the virtual training samples. Besides, locality constraint is added in test procedure by preserving the distance information between test samples and atoms. Therefore, we propose a new algorithm framework, multiple samples dictionary learning and locality constrained coding (MSDL-LCC) for recognition tasks.

The organization of the rest paper is as follows. Section 2 introduces the framework of SR. Section 3 introduces the proposed MSDL-LCC algorithm. Section 4 depicts the experimental results and Section 5 concludes the paper.

2. Sparse Representation

Given a signal set $\mathrm{Y} = [y_1, \dots, y_n] \in R^{d\times n}$, learning a dictionary for Y can be accomplished by solving the following objective function:

$$\min_{D,X} ||Y - DX||_2^2, ||X||_0 < T_0 \tag{1}$$

Where D is the learned overcomplete dictionary, X is the representation matrix of Y. $||\mathrm{X}||_0$ is the l_0 form of X. And T_0 is a constant, which controls the degree of sparsity. The K-SVD algorithm [3] is always used to solve the problem in formulation of (1).

After the dictionary matrix D is generated, for a test signal vector y_i,its coefficient vector x_i can be obtained by solving the following problem:

$$x_i = \min_{x_i} ||y_i - Dx_i||_2^2, ||x_i||_0 < T_0 \tag{2}$$

Where x_i represents the number of nonzero elements. And the famous Orthogonal Matching Pursuit algorithm (OMP) [11] is usually used to solve the problem in formulation of (2).

3. The Proposed Algorithm

3.1 Notation

Given a training set $\mathrm{Y} = [y_1, \dots, y_N] \in R^{d\times N}$, and a test set $\mathrm{H} = [h_1, \dots, h_M] \in R^{d\times M}$, we assume that the training set and the test set contains samples from C classes and all of them consists of all samples. Let $\mathrm{D} = [d_1, \dots, d_K] \in R^{d\times K}$ be the learned dictionary matrix, where d_i represents an atom in dictionary. $\mathrm{X} = [x_1, \dots, x_N] \in R^{K\times N}$ and $\mathrm{W} = [w_1, \dots, w_M] \in R^{K\times M}$ are the coding matrix of training and test data. $\mathrm{B} = [b_1, \dots, b_N] \in R^{N\times C}$ is the label matrix of train samples. If the label vector of x_i is $b_i = [0, \dots, 1, 0, \dots 0] \in R^{1\times C}$, where the place of nonzero element in b_i is j, x_i belongs to the $\mathrm{j}-\mathrm{th}$ class.

3.2 The MSDL-LCC Algorithm

This section will introduce the MSDL-LCC algorithm proposed in this paper in detail. Since the proposed algorithm involves dictionary train step and test step, we describe them separately below. And the overall process of the MSDL-LCC algorithm are shown in Algorithm1.

3.2.1 MSDL

According to [6], virtual samples can be calculated using the principle of mirroring. Specifically, for $\mathrm{i}-\mathrm{th}$ data y_i, its virtual version is defined as

$$y_i'(p,q) = y_i(p, Q - q + 1), (p = 1, \dots, P; q = 1, \dots, Q) \tag{3}$$

Where y_i' is virtual data. P and Q are the number of rows and columns of the training sample. $y_i(p,q)$and $y_i'(p,q)$ represent the pixel values of y_i and y_i' in the $\mathrm{p}-\mathrm{th}$ row and $\mathrm{q}-\mathrm{th}$ column, respectively.

By constructing the error terms of both origin data and virtual data, the objective function of the DL model is:

$$\min_{D,X} ||Y - DX||_2^2 + \alpha||Y' - DX||_2^2 + \beta||X||_2^2, s.t. ||d_i||^2 = 1, i = 1, \dots, K \tag{4}$$

Where the first term is the reconstruction error of the original training sample, the second term is the reconstruction error of the virtual training sample Y', and the third term is the constraint term for the coding matrix X. α and β are regularization parameters, used to balance the relationship between the three terms.

The optimization of formulation of (3) is divided into two steps, fixed D to solves X and fixed X to solves D.When we fix D, we can calculate X by taking the first-order derivation of (4) and setting it to zero. This led to

$$-D^T(Y - DX) + \alpha(-D^T)(Y' - DX) + \beta X = 0 \tag{5}$$

Thus, the optimal X is obtained by

$$\mathrm{X} = (D^T\mathrm{D} + \alpha D^T\mathrm{D} + \beta\mathrm{I})^{-1}(D^TY + \alpha D^TY') \tag{6}$$

Similarly, when we fix X,we can calculate D by

$$\mathrm{D} = (\mathrm{Y}X^T + \alpha Y'X^T)(XX^T + \alpha XX^T)^{-1} \tag{7}$$

3.2.2 MSDL

In the testing phase, this paper considers the local information between the test sample and the learned dictionary D, and adds a local constraint term to the objective function of the coding coefficients of the test samples. The objective function of the LCC model is

$$\min_w ||h_i - Dw||_2^2 + \gamma||p_i \odot w||_2^2 \tag{8}$$

Where $h_i \in R^{d\times 1}$ represents the $\mathrm{i-th}$ test sample of test matrix $\mathrm{H} \in R^{d\times M}$, w_i is the coding coefficient of h_i and γ is the regularization parameter. The first item is the reconstruction error term. The second term is a local constraint, where p represents the locality adaptor for each atom proportional to its similarity to the descriptor h_i and $\odot$ denotes the element-wise multiplication. Specifically,

$$p_{ij} = \sqrt{\exp(\frac{L(w,d_{ij})}{\sigma})} \tag{9}$$

In (9), $\mathrm{L}(\mathrm{w}, d_{ij}) = ||w - d_{ij}||_2$ and $\sigma = 1$.The larger the value of p_{ij}, the further the distance between the test sample and the dictionary atoms.

For (8), We solved its closed solution. Its first-order differential form of w is

$$2(\gamma\mathrm{diag}(p_i)^2\mathrm{w} - D^Th_i + D^T\mathrm{Dw}) \tag{10}$$

Let formulation of (10) be 0, we can get

$$\mathrm{w} = (D^TD + \gamma diag(p_i)^2)^{-1}D^Th_i \tag{11}$$

Where $diag(p_i)$ represents the diagonal matrix of the $\mathrm{i-th}$ distance vector p_i.

Algorithm1 Procedure of MSDL-LCC

Input: Training samples set $\mathrm{Y} = [y_1, \dots, y_N]$,test samples $\mathrm{H} = [h_1, \dots, h_M]$, label matrix of training samples $\mathrm{B} = [b_1, \dots, b_N]$,parameters: α, β ,and γ, iterations T_{max}.

Compute virtual samples set $Y' = [y_1', \dots, y_N']$ by using (3).

Use K-SVD algorithm to initialize dictionary D_0 and coding coefficient matrix X_0.

for $\mathrm{i} = 1{:}\,T_{max}$

  Obtain coefficients matrix X by using (6).

  Compute dictionary D by using (7).

end for

Compute a linear classifier coefficient G by using X and label matrix B by using (12).

for $\mathrm{i} = 1{:}\,\mathrm{M}$

  Compute coefficients matrix w_i of testing data h_i by using (11).

  Obtain the label information of h_i by computing the largest number of Gw_i.

end for

Output: The Label information of test sample.

3.2.3 Classification method

Following the work of [12], a linear classifier was used for classification tasks in our algorithm and the label of test sample w_i can be obtained by using the following steps.

A classifier parameter G is obtained by using the coding coefficient matrix X and label matrix B of the training samples through the formulation of (12):

$$\mathrm{G} = \mathrm{B}X^T(XX^T + I)^{-1} \tag{12}$$

For a test sample h_i,its coefficient vector w_i was obtained using the LCC model proposed in the previous section. Then a label vector l was calculated using Gw_i.The label of test data w_i is the index corresponding to the largest number of the label vector l.

4. Experiments results

In order to verify the performance of our proposed algorithm framework, we perform experiments on two public face databases: the Labeled Faces in the Wild(LFW) [6] and the AR face database [5],and compare them with some state-of-the-art methods, including LC-KSVD1 [5],LC-KSD2 [5]and MFI-DL [6].All experiments are performed on the same platform, specifically, software Matlab 2016a and Windows 7 system, hardware Intel Core i5-3470 CPU and 8GB ram. All results are average values after 10 runs and the symbol $\pm$ denotes the standard deviation of the average recognition rates.

4.1 Experimental results on the LFW database

The LFW face database contains more than 13,000 face images. Following [6], we use a subset of the LFW database to experiment. It consists of 86 people, each of whom had about 11-20 images, which together made up to 1251 images. The resolution of these images is 32×32 and some examples of images are shown in Fig.1.

Figure 1. Examples of images from the LFW face database.

We select six images of each people as training samples and take the rest as test data. Three parameters are set to:$\alpha = 10^{-3}, \beta = 10^{-3}, \gamma = 0.05$. And the average recognition rates are shown in Table 1. It can be found that the recognition of our algorithm is higher other algorithms.

Different numbers of atoms also directly affect the recognition accuracy, so we also make experiments with changing the number of atoms while keeping the other parameters constant. Fig.2 depicts the average recognition rate of our algorithm and three previous algorithms on the LFW database under different numbers of dictionary atoms($\mathrm{K} = 86,172, \ldots ,430$).

Table1. Average recognition rates on the LFW database.

Algorithms	Average recognition rates (%)
LC-KSVD1	26.15±0.009
LC-KSVD2	26.49±0.013
MFI-DL	31.35±0.012
Ours	*32.59±0.008*

4.2 Experimental results on the AR face database

The AR database contains 126 persons, and each person contains 26 face images taken under different lighting conditions and other different external conditions (such as whether wearing sunglasses or a scarf). Our experiments use its subset, which consists of 50 males and 50 females, with a total of 2600 images. Some examples of the AR faces are shown in Fig.3.

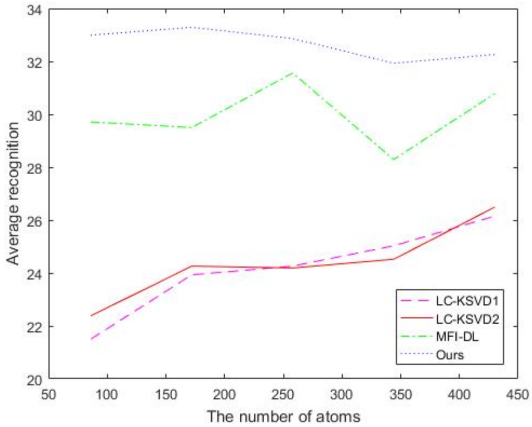

Figure 2. Average recognition rates with different numbers of atoms on the LFW database.

It can be seen from figure 2 that the recognition rate of the proposed algorithm is significantly higher than that of other algorithms.

Figure 3. Examples of images from the AR face database.

We randomly choose 8 face images from every category as training samples and the rest of them are used for testing. The parameters are respectively set to: $\alpha = 10^{-3}, \beta = 10^{-3}, \gamma = 10^{-4}$.The average recognition rates are denoted in Table 2.

Table2. Average recognition rates on the AR database.

Algorithms	Average recognition rates (%)
LC-KSVD1	83.81±0.010
LC-KSVD2	84.31±0.009
MFI-DL	88.67±0.010
Ours	*91.18±0.006*

Since the number of atoms will affect the recognition rate, experiments are performed on the recognition rate of the algorithm under different numbers of atoms. Figure 4 depicts the average recognition rate of our algorithm and three previous algorithms on the AR database under different numbers of dictionary atoms ($K = 100,200, \ldots ,800$). It can be seen from Figure 4 that the recognition rate of proposed algorithm is higher than other 3 algorithms under different numbers of atoms.

5. Conclusion

In this paper, we propose the MSDL-LCC algorithm to solve the following two problems: insufficient number of training samples when training the dictionary and information loss of the coefficient matrix caused by sparse coding during the test step. Experimental results on two public face database show that the MSDL-LCC model obtain higher recognition rates than many state-of-the-art methods. In addition, learning a discriminative and robust dictionary by using multiple layers of structure will be our next research content.

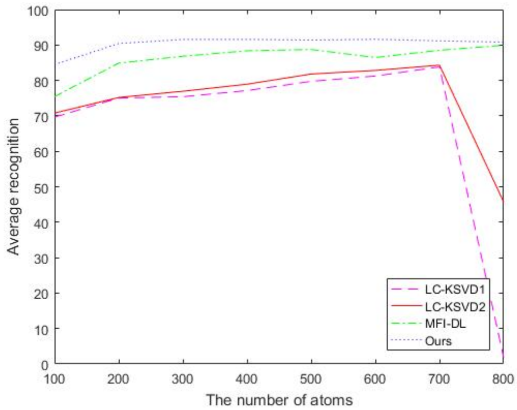

Figure 4. Average recognition rates with different numbers of atoms on the AR face database.

References

[1] Heinsohn, D., Villalobos, E., Prieto, L., & Mery, D. 2019. Face recognition in low-quality images using adaptive sparse representations. *Image and Vision Computing*, 46-58.

[2] Fan, Z., Zhang, D., Wang, X., Zhu, Q., & Wang, Y. 2018. Virtual dictionary based kernel sparse representation for face recognition. *Pattern Recognition*, 1-13.

[3] Aharon, M., Elad, M., & Bruckstein, A. M. 2006. $rm K$ -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, 54(11), 4311-4322.

[4] Zhang, Q., & Li, B. 2010. Discriminative K-SVD for dictionary learning in face recognition. *computer vision and pattern recognition*.

[5] Jiang, Z., Lin, Z., & Davis, L. S. 2013. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11), 2651-2664.

[6] Xu, Y., Li, Z., Zhang, B., Yang, J., & You, J. 2017. Sample diversity, representation effectiveness and robust dictionary learning for face recognition. *Information Sciences*, 171-182.

[7] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. S., & Gong, Y. 2010. Locality-constrained Linear Coding for image classification. *computer vision and pattern recognition*.

[8] Min, H., Liang, M., Luo, R., & Zhu, J. 2016. Laplacian regularized locality-constrained coding for image classification. *Neurocomputing*, 1486-1495.

[9] Wang, X., Li, Y., & Xie, W. 2018. A Dual Locality-Constrained Linear Coding Algorithm for Image Classification. *IEEE Access*, 19247-19254.

[10] Yuan, Y., Li, B., & Meng, M. Q. 2017. WCE Abnormality Detection Based on Saliency and Adaptive Locality-Constrained Linear Coding. *IEEE Transactions on Automation Science and Engineering*, 14(1), 149-159.

[11] Tropp, J. A., & Gilbert, A. C. 2007. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53(12), 4655-4666.

[12] Pham, D., & Venkatesh, S. 2008. Joint learning and dictionary construction for pattern recognition. *computer vision and pattern recognition*.