

RESEARCH

Open Access

Adaptive dictionary learning based on local configuration pattern for face recognition



Dongmei Wei¹, Tao Chen¹, Shuwei Li¹, Dongmei Jiang², Yuefeng Zhao^{1*} and Tianping Li^{1*}

Abstract

Sparse representation based on classification and collaborative representation based classification with regularized least square has been successfully used in face recognition. The over-completed dictionary is crucial for the approaches based on sparse representation or collaborative representation because it directly determines recognition accuracy and recognition time. In this paper, we proposed an algorithm of adaptive dictionary learning according to the inputting testing image. First, nearest neighbors of the testing image are labeled in local configuration pattern (LCP) subspace employing statistical similarity and configuration similarity defined in this paper. Then the face images labeled as nearest neighbors are used as atoms to build the adaptive representation dictionary, which means all atoms of this dictionary are nearest neighbors and they are more similar to the testing image in structure. Finally, the testing image is collaboratively represented and classified class by class with this proposed adaptive over-completed compact dictionary. Nearest neighbors are labeled by local binary pattern and microscopic feature in the very low dimension LCP subspace, so the labeling is very fast. The number of nearest neighbors is changeable for the different testing samples and is much less than that of all training samples generally, which significantly reduces the computational cost. In addition, atoms of this proposed dictionary are these high dimension face image vectors but not lower dimension LCP feature vectors, which ensures not only that the information included in face image is not lost but also that the atoms are more similar to the testing image in structure, which greatly increases the recognition accuracy. We also use the Fisher ratio to assess the robustness of this proposed dictionary. The extensive experiments on representative face databases with variations of lighting, expression, pose, and occlusion demonstrate that the proposed approach is superior both in recognition time and in accuracy.

Keywords: Collaborative representation classification, Nearest neighbors, Local configuration pattern (LCP), Statistical similarity, Configuration similarity

1 Introduction

As a biological feature, the human face has been paid more attention because it can be easily captured with a common camera even without cooperation of the subject. Yet the performance of face recognition is affected by the expression, illumination, occlusion, pose, age change, and so on. There are still some challenges in the field of unrestricted face recognition.

Facial images are very high dimensional, which is bad for classification. So dimension reduction is carried out before classification. Principle component analysis (PCA) [1–4] has become the classic reducing dimension approach and has been used widely in image processing and pattern recognition fields. All training images make up covariance matrix, and these eigenvectors corresponding to the bigger eigenvalues of covariance matrix span the linear feature subspace. Through projecting high-dimensional original face images onto the low dimensional linear feature subspace, PCA performs the dimension reduction and preserves the global structure of an image.

* Correspondence: yuefengzhao@126.com; Sdsdltp@163.com

¹Shandong Provincial Engineering and Technical Center of Light Manipulations & Shandong Provincial Key Laboratory of Optics and Photonic Device, School of Physics and Electronics, Shandong Normal University, Jinan 250014, China

Full list of author information is available at the end of the article

Designed as a texture descriptor originally, local binary patterns (LBP) [5] is a simple and efficient algorithm, it can capture the local structural features that are very important for human texture perception. Researches show that LBP is discriminative and robust to illumination, expression, pose, misalignment, and so on. LBP has been a popular feature extraction approach and has been used successfully in face recognition [6–8]. Up to now, some improved approaches based on LBP have been presented, such as complete LBP (CLBP) [9], local Gabor LBP [10], multi-scale LBP [11], and so on. Because the intensity of the central pixel is independent to its LBP value, the information embodying the relationship between pixels in the same neighborhood is lost. Guo et al. [12] proposed the local configuration pattern (LCP) including both local structural and microscopic features. The local structural features are represented by the pattern occurrence histograms, just as LBP, and the microscopic features are described by optimal model parameters. The microscopic configuration features reveal the pixel-wise interaction relationships. It has been proved LCP is an excellent feature descriptor. In this paper, we use LCP to measure the similarities between face images and finally to find neighbors of testing image.

Many researches confirmed that sparse representation and collaborative representation are good at image processing, such as image reconstruction, image representation, and image classification. Li et al proposed an effective approach called patch matching-based multi-temporal group sparse representation to restore the missing information that should be contained in remote sensing images [13] (Patch Matching-Based Multitemporal Group Sparse Representation for the Missing Information Reconstruction of Remote-Sensing Images). Yang et al. [14] regarded face recognition as a globally sparse representation problem and proposed the face recognition (FR) approach known sparse representation based on classification (SRC), in which the over-complete dictionary is formed by training face images. Subsequently, To improve the recognition accuracy with occlusion, Wright et al. [15] subsequently introduced an identity matrix to code the outlier pixels that were occluded. Yang et al. [16] viewed the sparse coding as a sparsity constrained robust regression problem and proposed the robust sparse coding method. It is more robust to detect outliers than SRC. In 2010, Yang et al. [17] introduced Gabor features into SRC; they projected firstly the high-dimension facial images into the lower dimension Gabor-feature space and then used SRC to classify. This approach greatly decreased the size of the occlusion dictionary compare with that of SRC. As a result, recognition-time is reduced. Ou et al. proposed an approach of face recognition with occlusion named

structured sparse representation based classification [18]. In which, occlusion dictionary is not an identity matrix, but is learned from data and is smaller than that of SRC. In addition, the occlusion dictionary is independent of the training sample as possible, so the occlusion is sparsely represented by the learned occlusion dictionary and can be separated from the occluded image. The testing image is classified by the recovered non-occluded image. Ou also provided another occlusion dictionary learning algorithm named discriminative nonnegative dictionary learning, where the occlusions were estimated adaptively by the reconstruction errors and weights for different pixels were learned during iterative processing [19].

The resolution method of sparse representation vector (or sparse solution) is another crucial issue for sparse representation based on classification (SRC), which influences the recognition accuracy and recognition time. The first researchers, such as the above mentioned, hold the view that “sparsity” of the sparse representation vector is the most crucial for SRC and paid more attention to the “sparsity” of sparse solution. Because l_0 -norm minimization is NP-hard, they used l_1 -norm minimization to replace l_0 -norm minimization as the optimal solution (the sparsest solution). But l_1 -norm minimization is time-consuming, which is bad for recognition especially in the case of real-time identification. Ma et al [20] proposed a discriminative low-rank dictionary learning, in which the over-completed discriminative dictionary is composed of series of sub-dictionaries. Atoms of each sub-dictionary are training images from the same category, so they are linearly correlated and all sub-dictionaries are low-rank with a compact form. Zhang et al. [21] also regarded the feature coding problem as the low-rank matrix learning problem and researched local structure information among features on the image-level. Using the similarities among local features embraced in the same spatial neighborhood, an exacting joint representation of these local features *w.r.t.* the codebook was founded. Zhang et al [22] analyzed the mechanism of SRC and found that “collaboration representation” among categories is the crucial factor for SRC. He relaxed the demand on “sparsity” and used l_2 -norm to replace l_1 -norm as the sparse constraint condition and proposed the algorithm known collaborative representation based on classification with regularized least square (CRC_RLS). Gou et al investigated deeply approaches based on collaborative representation-classification and proposed several novel approaches of classification based on collaborative representation [23–26]. He proposed a two-phase probabilistic collaborative representation-based classification [25], in which the nearest representative samples are chosen first and then each testing sample is represented and classified by those chosen nearest samples. In Ref.

[26], the locality of data is employed to constrain the collaborative representation in order to represent faithfully testing images with the nearest samples. In this paper, we proposed a new and simple approach based on sparse representation and LCP features. We pay our attention to the over-completed compact dictionary. Atoms of this proposed dictionary are all nearest neighbors in LCP feature subspace, so structures of atoms are more similar to that of testing images and the atoms' number is greatly decreased compared with the training images, which will improve the recognition accuracy and reduce the recognition time.

This paper is organized as follows: related theories about SRC and LCP were described in Section 2. The proposed face recognition approach using a sparse representation-based adaptive nearest dictionary was described in Section 3. Experimental results were provided in Section 4 and conclusions were illustrated in Section 5.

2 Related

2.1 Sparse representation based on classification

Assume there are M training face images from Q subjects, each subject has m_q images, $M = \sum_{q=1}^Q m_q$. Let $\mathbf{v}_q \in \mathcal{R}^{d \times 1}$ denote the d -dimension vector stretched by the i th image from the q th class (subject) and \mathbf{y} represent the testing image vector. Let matrix $\mathbf{X}_q = [\mathbf{v}_{q,1}, \dots, \mathbf{v}_{q,m_q}]$ whose column vectors are the training images from the q th class, $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_q, \dots, \mathbf{X}_Q] \in \mathcal{R}^{d \times M}$ is the set of all training images.

Yang et al. [14] proved that face recognition can be regarded as sparse representation based on classification (SRC) if there were enough training face images (the number of training face images M should be greater than the dimension of image, i.e., $d < M$). In this case, the testing image \mathbf{y} can be as the sparse linear combination of all training images, i.e.,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \quad (1)$$

where $\boldsymbol{\beta} = [\beta_{1,1}, \beta_{1,2}, \dots, \beta_{1,m_1}^T, \dots, \beta_{Q,1}, \dots, \beta_{Q,m_Q}]^T \in \mathcal{R}^{M \times 1}$ is the representation coefficient vector.

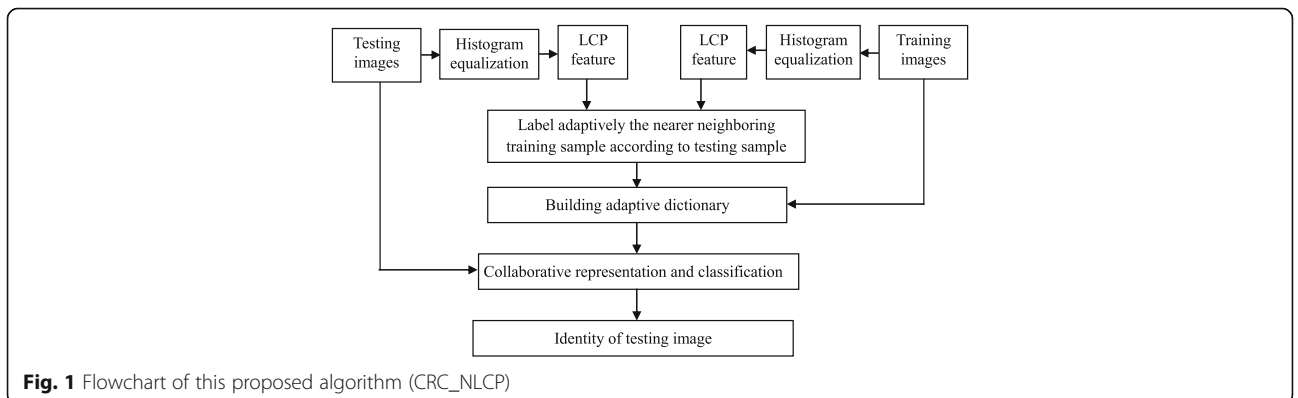
Sparse representation classification firstly encodes the testing image by all training images according to Eq. (1) and then classifies class-by-class. Specifically, if the testing image \mathbf{y} comes from the q th subject (class), entries of its $\boldsymbol{\beta}$ should be zero except for those associated with the q th class ideally, i.e. $\boldsymbol{\beta} = [0, \dots, 0, \beta_{q,1}, \beta_{q,2}, \dots, \beta_{q,m_q}, 0, \dots, 0]^T$ (T is transpose operation). In fact, any entry of $\boldsymbol{\beta}$ can be very small non-zero, so the identity of \mathbf{y} is determined by the residual between \mathbf{y} and its reconstruction by each class. Let $\boldsymbol{\beta}_q = [\beta_{q,1}, \beta_{q,2}, \dots, \beta_{q,m_q}]^T$ be the representation coefficient associated with the q th class. The reconstruction of the test image by the q th class training samples denotes by \mathbf{y}_q , $\mathbf{y}_q = \mathbf{X}_q \boldsymbol{\beta}_q$, the residual $r_q = \|\mathbf{y} - \mathbf{y}_q\|_2 = \|\mathbf{y} - \mathbf{X}_q \boldsymbol{\beta}_q\|_2$. The testing image is classified to the q th class if r_q is the minimum residual, i.e.,

$$\text{identity}(\mathbf{y}) = \arg \min_q r_q \quad (2)$$

How to get the representation coefficient vector from Eq. (1) is crucial for sparse representation. Sparse representation requires the dictionary to satisfy $d < M$, so Eq. (1) has more than one solution. The sparsest solution is generally considered as the optimal solution for classification. The optimal solution expressed $\hat{\boldsymbol{\beta}}$ is the solution of Eq. (3), while it is a NP-hard problem.

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0 \text{ s.t. } \mathbf{X}\boldsymbol{\beta} = \mathbf{y} \quad (3)$$

Theories of sparse representation and compressed sensing reveal that l_0 -norm minimization solution and l_1 -norm minimization solution were nearly equivalent when the solution is sparse enough. So Yang [14]



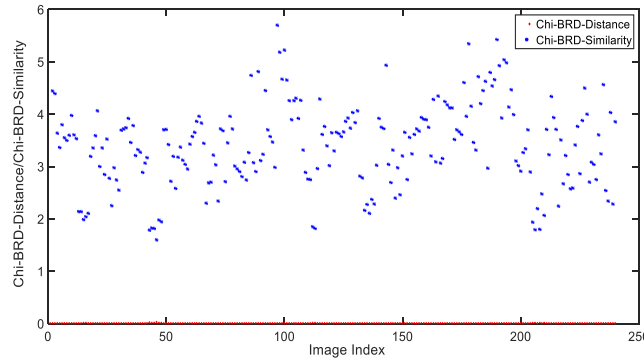


Fig. 2 Chi-BRD-distances and chi-BRD-similarities

employed l_1 -norm to replace l_0 -norm and estimated the optimal representation vector by Eq. (4)

$$\hat{\beta} = \arg \min_{\beta} \|\beta\|_1 \quad \text{s.t.} \quad \|\mathbf{X}\beta - \mathbf{y}\|_2 \leq \varepsilon \quad (4)$$

where ε denotes the error term including noisy and model error, and this is the well-known sparse representation based on classification (SRC).

Zhang employed l_2 -norm as the sparse constraint condition to solve representation vector using regularized least square and proposed CRC_RLS algorithm [22], in which the representation vector can be obtained by Eq. (5)

$$\hat{\beta} = \arg \min_{\beta} \{\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2\} \quad (5)$$

With regularized least squares method, the unique solution of Eq. (5) can be easily deduced as $\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$, where λ is a regularization parameter given experientially.

Assume $\mathbf{P} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T$, it is clear that \mathbf{P} is independent of \mathbf{y} , so \mathbf{P} can be seen as a projection matrix decided only by training images. Just projecting \mathbf{y} onto \mathbf{P} , $\hat{\beta}$ can be solved easily, $\hat{\beta} = \mathbf{P}\mathbf{y}$.

According to sparse representation theory, the over-completed dictionary should be satisfied that the number of features is greater than that of an atom. However, face recognition is a typical small-size-sample task and matrix \mathbf{X} made directly by face images could not be an over-completed dictionary. Generally, dimension reduction was performed before a sparse representation.

2.2 Local configuration pattern

In essence, LBP feature is the histogram build by LBP value of all pixels in an image. LBP of the given pixel is defined as follows [5]:

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} u(g_p - g_c) 2^p, \quad u(x) = \begin{cases} 1, & a \geq 0 \\ 0, & a < 0 \end{cases} \quad (6)$$

where R is the radius of the circle neighborhood centered at the given pixel, P is the number of neighboring samples spaced at regular intervals on this circle, g_c is gray of the center pixel, and g_p ($p = 0, 1, \dots, P-1$) denotes gray of its neighboring.

Regard the LBP as a circular binary string with P bits, the number of two bitwise transitions 0 and 1 is defined as U . If $U \leq 2$, then the LBP are defined as a uniform pattern, notated $\text{LBP}_{P,R}^{u2}$. Ojala et al. [5] defined the rotation-invariant uniform patterns $\text{LBP}_{P,R}^{riu2}$:

$$\text{LBP}_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} u(g_p - g_c), & U(\text{LBP}_{P,R}) \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \quad (7)$$

Let N_k denotes the occurrence of the k -pattern included in LBP, $N_k = \sum_{i=1}^w \sum_{j=1}^h \delta(\text{LBP}_{P,R}^{riu2}(i, j), k)$, ($k = 0, 1, \dots, 2^P$), where $\delta(x, y)$ is the Dirac function. Histogram \mathbf{H} expresses the LBP feature vector of the image.

$$\mathbf{H} = [N_1, \dots, N_k, \dots, N_S] \quad (8)$$

where S is the maximum value of LBP pattern.

Although the LBP feature vector can capture the statistical feature of the image and is robust for illumination, it neglects the relationship between neighboring pixels, which leads to wrong classification results when images have the same LBP feature but the different gray variations between the center pixel and its neighboring pixels. Guo et al. [12]. introduced the microscopic feature (MiC) information into LBP and proposed a local configuration pattern (LCP).

Table 1 The description on both face databases

Database name	AR	ORL
Number of class	100	40
Number of image per subject	14	10
Total number of image	1400	400
Number of training image per class	7	6
Total number of training-image	700	240
Number of testing image per class	7	4
Total number of testing image	700	160
Size of image	60 × 43	112 × 92

Let $\mathbf{A} = (a_0, \dots, a_{P-1})$ presents a parameter vector and $\mathbf{g} = (g_0, \dots, g_{P-1})$ presents a neighboring vector of center pixel. The central pixel can be reconstructed by $\sum_{p=0}^{P-1} a_p g_p$,

the reconstruction error $|g_c - \sum_{p=0}^{P-1} a_p g_p|$ will be minimum when parameter vector \mathbf{A} reaches optimum.

Assume there are N_L pixels whose patterns are all equal to L , let $c_i^L (i = 1, \dots, N_L)$ denote the gray value of the i th center pixel belonging to L pattern, these N_L gray values made up the matrix $\mathbf{C}_L = (c_1^L, c_2^L, \dots, c_{N_L}^L)$. Vector $\mathbf{g}_i^L = (g_{i,0}^L, g_{i,1}^L, \dots, g_{i,P-1}^L)$ consists of the gray values of neighboring pixels of c_i^L , matrix $\mathbf{G}_L = \begin{pmatrix} \mathbf{g}_1^L \\ \vdots \\ \mathbf{g}_{N_L}^L \end{pmatrix} =$

$\begin{pmatrix} g_{1,0}^L, \dots, g_{1,P-1}^L \\ \vdots \\ g_{N_L,0}^L, \dots, g_{N_L,P-1}^L \end{pmatrix}$, $\mathbf{A}_L = (a_0^L, \dots, a_{P-1}^L)$ represents the optimal parameters corresponding to the L pattern. \mathbf{C}_L , \mathbf{A}_L , and \mathbf{G}_L satisfy Eq. (9)

$$\mathbf{C}_L = \mathbf{A}_L \mathbf{G}_L \quad (9)$$

With overwhelming probability $N_L \gg P$, hence the matrix \mathbf{G}_L is over-determined and the optimal parameter vector \mathbf{A}_L can be computed by the least squares.

$$\mathbf{A}_L = \mathbf{C}_L \mathbf{G}_L^T (\mathbf{G}_L^T \mathbf{G}_L)^{-1} \quad (10)$$

A small probability event ($N_L \leq P$) is regarded as unreliable and its model parameters are set to zero. The

Fourier transform of \mathbf{A}_L is $\phi_L(k)$, $\phi_L(k) = \sum_{p=0}^{P-1} a_p^L e^{-j2\pi kp/P}$ ($k = 0, \dots, P-1$). MiC features consist of the amplitude of the Fourier transform ($|\phi_L(k)|$), denoted by Φ_L .

$$\Phi_L = (|\phi_L(0)|, \dots, |\phi_L(P-1)|) \quad (11)$$

Appending N_L to Φ_L , LCP feature corresponding to L pattern can be obtained by the Eq. (12). The LCP feature vector of an image can be obtained by concatenating all S LCP features, it is denoted by \mathbf{F}_{LCP} .

$$\mathbf{F}_{LCP} = [\Phi_1, N_1, \dots, \Phi_S, N_S] \quad (12)$$

3 Methods

According to sparse representation theories, the more similar the atom is to the testing sample, the sparser the representation vector is, and the greater recognition accuracy of the dictionary is.

Let $\eta = m_q/M$ describe the sparsity of the representation vector. If all classes have the same number of training samples, then $\eta = 1/Q$, else $\min_q(m_q)/M \leq \eta \leq \max_q(m_q)/M$, where $\min_q(m_q)$ and $\max_q(m_q)$ are the minimum and the maximum number of training samples for each class. The representation vector is sparser with the smaller η . Clearly, we want to make η small enough. There are two ways to make η smaller. One is to augment categories and the other is to reduce the sample size of each class. For the former, it is difficult to ask more enough volunteers to take face picture. So we use the second way to increase the sparsity. A testing image may be not represented sufficiently if we discard randomly some training images, which are bad for the recognition accuracy. In this paper, we recommend to utilize the nearest neighboring images of the testing image as atoms to build a dictionary so that the testing image can be represented sparsely and faithfully. SRC then

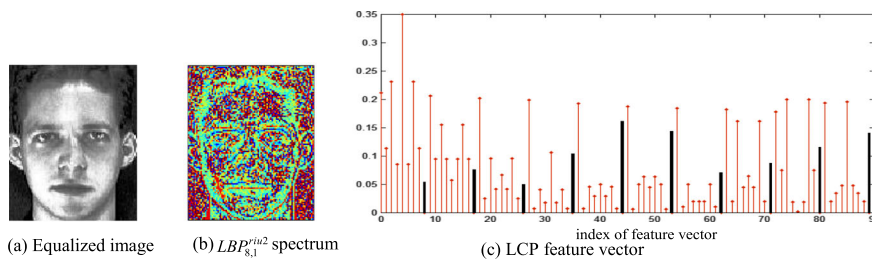


Fig. 3 One face image and its LCP feature vector. **a** Equalized image. **b** $LBP^{riu2}_{8,1}$ spectrum. **c** LCP feature vector

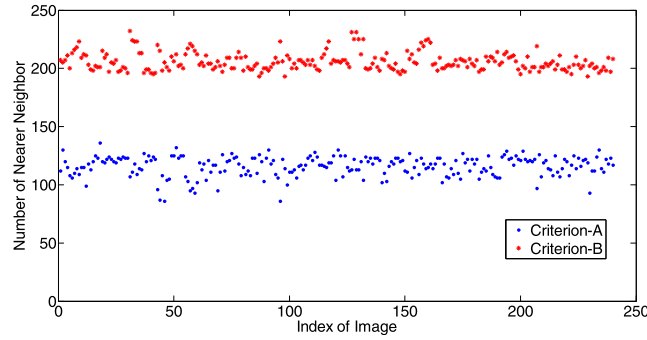


Fig. 4 Scatter diagram of nearest neighbors of training images from ORL

are performed on this proposed dictionary. We call this proposed approach for FR as “Collaborative representation classification based adaptive nearer-neighbor-dictionary” (CRC_NLCP). Figure 1 shows the flowchart of this proposed algorithm.

The proposed algorithm includes two stages, the first is to learn adaptively the over-completed dictionary, and the second stage is the SRC process. Dictionary learning includes rough similarity measurement, nearest neighbor criterion, and dictionary atom selection.

3.1 Similarity measurement

As the aforementioned, the LCP feature includes the statistical feature (LBP feature) described by the histogram and the microscopic configuration feature described by the parameter vector (MiC feature). So we measure the similarity among images in these two feature subspaces respectively.

3.1.1 Statistical similarity

In essence, the LBP feature vector is a frequency histogram, so the chi-square distance is a good measurement

to estimate the similarity between histograms. Hu et.al [27] proved that the distance-based chi-square is not robust for partial occlusion and they may lose the co-occurrence relations that benefit to improve the recognition accuracy. Hence, he proposed a bin ratio-based distance (BRD).

Assuming $\tilde{\mathbf{H}}^A = [u_1^A, \dots, u_k^A, \dots, u_Q^A]$ and $\tilde{\mathbf{H}}^B = [u_1^B, \dots, u_k^B, \dots, u_Q^B]$ are respectively two unit histogram vectors corresponding the image **A** and image **B**, BRD of these two images are defined as Eq. (13):

$$d_{BRD}(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B) = \sum_{i=1}^Q \sum_{j=1}^Q \left(\frac{u_i^B u_j^A - u_j^B u_i^A}{u_i^A + u_i^B} \right)^2 \quad (13)$$

where the numerator term $(u_i^B u_j^A - u_j^B u_i^A)$ embodies the ratio-relations and the differences of bins included in the same histogram, and the denominator is the standard term. He also proved the combination of BRD and chi-square distance, notated as d_{χ^2-BRD} and defined by Eq. (14), is superior to both them for classification.

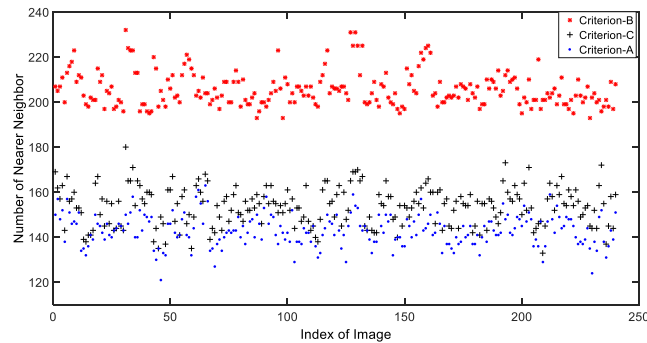


Fig. 5 Scatter diagram of nearest neighbors of training images from AR

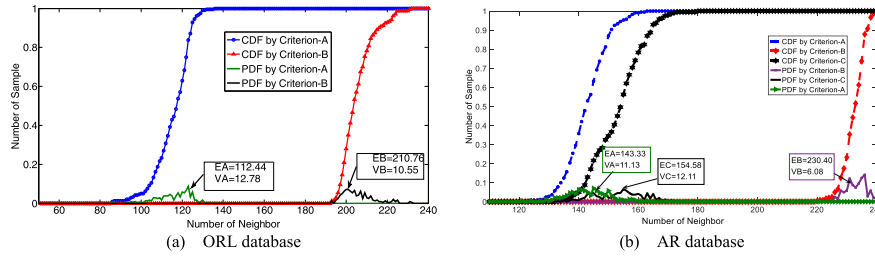


Fig. 6 Cumulative distribution function (CDF) and probability density function (PDF) of the nearer-neighbor-number. **a** ORL database. **b** AR database

$$dmsubsup(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B)^{\chi^2_{\text{BRD}}} = d_{\chi^2}(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B)^{\chi^2} - 2 \left\| \tilde{\mathbf{H}}^A + \tilde{\mathbf{H}}^B \right\|^2 \sum_{i=1}^Q \frac{(u_i^A - u_i^B)^2 u_i^A u_i^B}{(u_i^A + u_i^B)^3} \quad (14)$$

where $d_{\chi^2}(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B)^{\chi^2} = 2 \sum_{i=1}^Q \frac{(u_i^A - u_i^B)^2}{u_i^A + u_i^B}$ is a chi-square distance of $(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B)$. For the unit length histograms, values of $dmsubsup^{\chi^2_{\text{BRD}}}$ are all very small, for example, the maximum value is shown in Fig. 2 (red rhomb) is 0.025, and differences $dmsubsup^{\chi^2_{\text{BRD}}}$ are even smaller, which means $dmsubsup^{\chi^2_{\text{BRD}}}$ has less discrimination feature and is not fit to measure the statistical similarity. We defined a new statistical similarity measurement named “chi-square-BRD-similarity” and notated as ξ to estimate the similarity of images in the LBP feature subspace.

$$\xi(A, B) = -\log \left[dmsubsup(\tilde{\mathbf{H}}^A, \tilde{\mathbf{H}}^B)^{\chi^2_{\text{BRD}}} \right] \quad (15)$$

Clearly, Chi-square-BRD-similarity (ξ) is the minus logarithm of $dmsubsup^{\chi^2_{\text{BRD}}}$, it expands nonlinearly the value range of $dmsubsup^{\chi^2_{\text{BRD}}}$. Figure 2 shows the chi-BRD-distance ($dmsubsup^{\chi^2_{\text{BRD}}}$) indicated by red rhombs

and Chi-square-BRD-similarity (ξ) denoted by blue stars between one image and the others from the ORL face database. Clearly, ξ is more discriminative than $dmsubsup^{\chi^2_{\text{BRD}}}$. The larger ξ , the more similar two images are. Chi-BRD-distance between the image and itself is zeros while the chi-square-BRD-similarity is infinity.

3.1.2 Configuration similarity

In the MiC feature subspace, we use Euclidean distance to measure the configuration similarity between two images \mathbf{x}_1 and \mathbf{x}_2 , notating $\rho(\mathbf{x}_1, \mathbf{x}_2)$.

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = \|\Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_2)\|_2 \quad (16)$$

Clearly, two images are more similar if their Euclidean distance is smaller.

3.2 Neighboring images selection criteria

According to Eq.(7)~(11), we can obtain unit histogram $\tilde{\mathbf{H}}_{q,i}$ and Mic feature $\tilde{\Phi}_{q,i}$ from training image $\mathbf{v}_{q,i}$ and $\tilde{\mathbf{H}}_y$ (unit histogram of testing image) and $\tilde{\Phi}_y$ (MiC feature of testing image). Let $\xi_{q,i}$ and $\rho_{q,i}$ describe respectively the statistical similarity and configuration similarity between testing image y and $\mathbf{v}_{q,i}$, $\xi_{q,i} = \xi(\tilde{\mathbf{H}}_y, \tilde{\mathbf{H}}_{q,i})$ can be obtained by Eq. (13)~(15); $\rho_{q,i} = \|\Phi_y - \Phi_{q,i}\|_2$ can be computed according to Eq. (16).

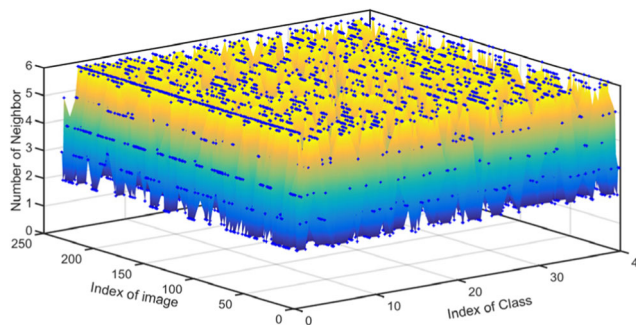
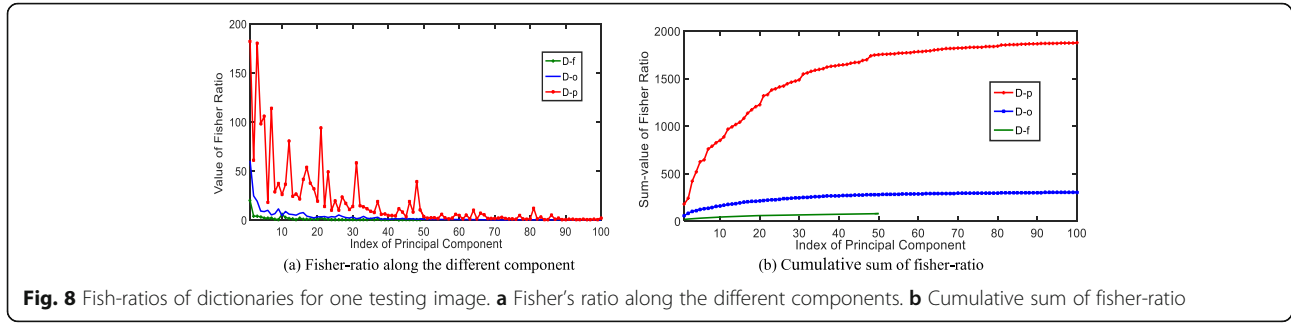


Fig. 7 The number of nearest neighbors in each category



Define $\bar{\xi}_q = \frac{1}{m_q} \sum_{i=1}^{m_q} \xi_{q,i}$ as “class-average-statistical similarity.” If $\xi_{q,i} \geq \bar{\xi}_q$, then the training image $\mathbf{v}_{q,i}$ is labeled as statistical neighboring. Similarly, define ‘class-average-configuration similarity’ as $\bar{\rho}_q = \frac{1}{m_q} \sum_{i=1}^{m_q} \rho_{q,i}$. $\mathbf{v}_{q,i}$ is labeled as configuration neighboring if $\rho_{q,i} \leq \bar{\rho}_q$.

There are two criteria to label the nearest neighboring samples. One, named criterion-A, is that the nearest neighboring image should be labeled as both statistical neighboring and configuration neighboring. The other, named criterion-B, is that the nearest neighboring image was labeled as either statistical neighboring or configuration neighboring. Whatever criterion is used, there must be enough training samples in each class to be labeled as the nearest neighbor. If not, the complementary criterion, named criterion-C, must be performed. In this paper, the complementary criterion (criterion-C) is to sort respectively the training images in LBP and MiC subspace, and then keep labeling the closer ones in both two subspaces until the quantity of the nearest neighbor in each class reaches the requirement. Generally, the number of the nearest neighbor in each class should be between 2 and half of the number of training images in each class. The algorithm of labeling the nearest neighbor is summarized as the algorithm 1.

Algorithm 1: Nearest Neighbor Labeling

- Step1 Project all face image into LCP subspace and extract the LBP and Mic features, $\bar{\mathbf{H}}_{q,i}, \bar{\mathbf{H}}_p, \bar{\Phi}_{q,i}$ and $\bar{\Phi}_p$;
- Step2 Compute the statistical similarity $\xi_{q,i}$ and the configuration similarity $\rho_{q,i}$ between \mathbf{y} and all training image $\mathbf{v}_{q,i}, \mathbf{v}_{p,i}$. Then compute the class-average-statistical similarity $\bar{\xi}_q$ and class-average-configuration similarity $\bar{\rho}_q$;
- Step3 Label nearest neighboring images using different criterion class by class:
- A) The nearest neighboring image satisfies both $\xi_{q,i} \geq \bar{\xi}_q$ and $\rho_{q,i} \leq \bar{\rho}_q$;
 - B) The nearest neighboring image satisfies either $\xi_{q,i} \geq \bar{\xi}_q$ or $\rho_{q,i} \leq \bar{\rho}_q$;
 - C) Keep labeling the nearest neighbor image in LBP and MiC until the number of nearest neighbors is between 2 and $m_q/2$.
-

3.3 Building an adaptive dictionary

For this proposed approach, it is crucial to fast find and label these training samples that are more similar to the inputting testing image. So we compute the similarity in the lower-dimension LCP feature subspace. However,

compared to the face image, LCP feature vector loses much information, which are bad for SRC. We employ the original face image as an atom to make up the dictionary, which means all atoms of the dictionary are these face-image-vectors labeled in the LCP subspace, but not LCP-feature-vectors. That is to say, sparse representation and classification are performed in face image space but not in feature subspace. The experimental results shown in the later section also verify that SRC based on image vector is better than based on the LCP feature vector.

Use these training images labeled the nearest neighbor as the atoms to build the over-completed dictionary. In order not to be confused with the ordinary dictionary notated by \mathbf{X} (defined in Section 2.1), we use \mathbf{R} to notate this proposed adaptive nearest neighbor dictionary. $\mathbf{R} = [\mathbf{R}_1, \dots, \mathbf{R}_q, \dots, \mathbf{R}_Q]$, where $\mathbf{R}_q = [\mathbf{v}_{q,j_1}, \mathbf{v}_{q,j_2}, \dots, \mathbf{v}_{q,j_{\kappa_q}}] \in \mathbb{R}^{d \times \kappa_q}$ ($1 \leq j_1, j_2, \dots, j_{\kappa_q} \leq m_q$), whose column vectors are the nearest neighbors selected from the q th subject's training images. Integer κ_q indicates that there are κ_q nearest neighbor images are chosen in the q th class. Usually, $\kappa_1 \neq \kappa_2 \neq \dots \neq \kappa_Q$.

3.4 Assessment of the adaptive dictionary

The over-completed dictionary greatly affects the performance of SRC. We employ Fisher's ratio as the criterion to quantitatively evaluate the validity of the dictionary. Using the within-class scatter and the between-class scatter, Fisher's ratio directly assesses the class separation performance. For the q -class and p -class, their Fisher's ratio along the l -principal component vector direction $F_{q,p,l}$ is defined

$$F_{p,q,l} = \frac{\mathbf{w}_l^T \mathbf{S}_b \mathbf{w}_l}{\mathbf{w}_l^T \mathbf{S}_w \mathbf{w}_l} \quad (17)$$

where $\mathbf{S}_b = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T$ is the between-class scatter matrix of the class p and class q , $\boldsymbol{\mu}_p$ is the p -class mean, $\boldsymbol{\mu}_q$ is the q -class mean; $\mathbf{S}_w = \frac{1}{2} \sum_{j=p,q} (\frac{1}{m_j} \sum_{i=1}^{m_j} (\mathbf{x}_{j,i} - \boldsymbol{\mu}_j)(\mathbf{x}_{j,i} - \boldsymbol{\mu}_j)^T)$ is the within-class scatter matrix for the

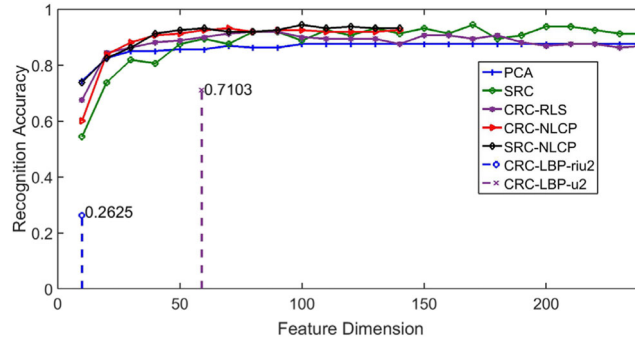


Fig. 9 The curve of recognition accuracy with feature-dimension

class p and class q , $x_{j,i}$ is the i th image from the j th class, m_j is the number of the images in the class j ; w_l is the eigenvector corresponding to the eigenvalue λ_l , w_l^T express the transposition of w_l . The coefficients $1/2$ and m_j remove the interference brought by the different sample size of each class.

From Eq. (17), we can see that $F_{p,q,l} = F_{q,p,l}$. For the whole dictionary, the fisher ratio along l -principal component is defined as the mean of $F_{p,q,l}$, notated F_l :

$$F_l = \frac{2}{Q(Q-1)} \left(\sum_{i=1}^Q \sum_{j=i+1}^Q F_{p,q,l} \right) \quad (18)$$

Integer Q , defined in Section 2.1, represents the number of class included in the dictionary.

4 Results and discussion

In this section, the performance of the proposed algorithm is verified based on two standard face databases, AR [28] and ORL [29]. Both face databases involve some changes in gender, illumination, pose, expression, glass, and time. The detailed data used in this paper are list in Table 1.

This proposed adaptive dictionary is different for different testing images, so different sets of training

samples have a little influence on experimental results. We randomly selected training samples to repeat the proposed algorithm for multiple times and presented the average result in the later of this section. All algorithms are coded by MATLAB 2015 and performed on the same computer with 2.6GHz CPU and 8G RAM. In this paper, we never use any parallel computing or GPU.

4.1 Nearest neighbor image selection based on LCP features

In order to reduce computation while retaining neighbor attribution, we employ the rotation-invariant uniform pattern to compute the similarities between images. Set $P = 8$, $R = 1$, and the length of the LCP feature vector is 90. Figure 3 presents an example about one face image and its LCP feature vector, where (a) is the equalized image, (b) is the rotation-invariant uniform LBP pattern spectrum for $P = 8$ and $R = 1$, and (c) is the LCP feature vector corresponding to (c). The ten groups of red vertical line '†' denote the ten normalized MiC feature vectors respectively, the ten black solid vertical lines describe the frequencies of all patterns in LBP.

In this paper, all images were equalized in advance to reduce the effects of illumination. The number of

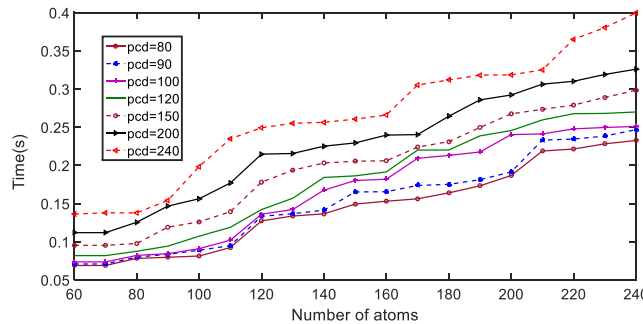


Fig. 10 Solving times for a different number of atoms

Table 2 Recognition accuracy on ORL

	Best recognition rate	Optimal number of features
PCA	0.900	100
SRC	0.920	100
CRC_RLS	0.931	95
CRC_LBP	0.706	10
CRC_NLCP	0.975	95

nearest neighbors is different for different testing images. Figure 4 represents the scatter diagrams of the nearer-neighbor-number for the face database ORL. The numbers of nearest neighbor both on criterion-A and on criterion-B (given in algorithm 1) are reasonable. It is unnecessary to perform criterion-C.

Figure 5 represents the scatter diagrams of the nearer-neighbor-number for the face database AR. For some images, its numbers of nearest neighbor labeled by criterion-A are relatively small to the number of classes. So the number labeled criterion-C is shown in Fig. 5.

For clarity, we presented the Cumulative Distribution Function (CDF) and Probability Density Function (PDF) of the nearer-neighbor-number for AR and ORL in Fig. 6, where EA and VA are mean and variance respectively of the numbers. Obviously, the number of nearest neighbor labeled by criterion-A is mainly between 130 and 150 in AR database, it is much less than the total number of training images.

For each image in the database ORL, the number of its nearest neighbors in each class is shown in Fig. 7. It is clear that the number of nearest neighbors labeled by our method is almost half of that of training samples and the nearest neighbor images are distributed in all categories. So the number of categories is not reduced, but the number of atoms in each category is reduced, which reduces the dictionary thickness and improves the sparsity.

4.2 Adaptive dictionary

Face image is a high dimension with lots of discriminative information. But the LCP feature dimension is 90 and LBP feature dimension is only 10 in this paper. So much information is missed in the LCP or LBP feature subspace, which leads to classify wrong.

Table 3 Recognition accuracy without occlusion on AR

Feature dimension	10	54	120	300
PCA	–	0.680	0.701	0.713
SRC	–	0.833	0.895	0.933
CRC_RLS	–	0.805	0.900	0.938
CRC_LBP	0.57.0	–	–	–
CRC_NLCP	–	0.838	0.943	0.950

Table 4 Recognition accuracy with occlusion on AR

Algorithm	Wears glasses	Wears scarf
PCA	0.400	0.147
SRC	0.480	0.543
CRC_RLS	0.530	0.777
CRC_NLCP	0.735	0.826

We use fish-ratio defined by Eq. (17) to measure the discrimination of the dictionary. In order to compare, three dictionaries are created. As baseline, the first dictionary, denoted by \mathbf{D}_O , is composed by all training original facial images and is stationary for all testing images; the second dictionary proposed in this paper and denoted by \mathbf{D}_p is composed by the nearest neighbor facial images, and the third dictionary denoted by \mathbf{D}_f is composed by the nearest neighbor LCP feature vectors. Atom feature dimensions of both \mathbf{D}_O and \mathbf{D}_p are equal to the size of the facial image and are much larger than the atoms' number, but for the dictionary \mathbf{D}_f , the atom feature dimension is dependent on the type of LBP.

According to Eq. (18), fish-ratio is defined independently along each principal component vector direction. We compute the fish-ratios of the three dictionaries along the bigger principal component vector directions. As an example, Fig. 8 presents the fish-ratios of different dictionaries for the same testing images from ORL. The length of the LCP feature vector is 90, so we compute the fish-ratio of the first 90 principal component directions for the \mathbf{D}_f and for the other dictionaries, the principal component is less than the number of the atom. Figure 7b shows the cumulative sum of fish-ratio from the first-principal component to the given-principal component. We can conclude that the classification performance of the proposed dictionary in the paper is superior to the other two dictionaries.

Since the sparse dictionary requires the number of atoms to be greater than the feature number of atoms, the feature number of atoms is the image resolution, and the image resolution is far greater than the number of images; PCA dimensionality reduction is required to calculate the sparse coefficient.

4.3 Recognition accuracy

The testing image is first collaboratively represented with its neighbors labeled in the LCP feature subspace,

Table 5 Time to solve one SRV on database ORL(s)

Method	Number of atoms		
	240	150	120
$\ell_1 \sim \ell_2$	0.2674	0.1775	0.1463
RLS	0.0077	0.0027	0.0015

Table 6 Time to solve one SRV on database AR (s)

Method	Feature dimension	Number of atoms			
		700	500	300	200
$\ell_1 \sim \ell_s$	100	0.4974	0.3345	0.2362	0.1982
	200	0.5482	0.3844	0.2542	--
	300	0.6890	0.4834	--	--
	400	0.8241	0.5921	--	--
RLS	290	0.0804	0.0408	0.0173	--
	450	0.0870	0.0420	--	--

and then it is classified using its over-completed compact dictionary. In this paper, we use the recognition accuracy to measure the recognition performance.

$$\text{Recognition accuracy} = \frac{\text{Number of the testing samples classified correctly}}{\text{Total number of testing samples}} \quad (19)$$

Figure 9 shows the curve between the right recognition rate and feature-dimension for different dictionaries on ORL. In which, PCA, SRC, and CRC_RLS have the same fixed dictionary \mathbf{D}_O ; CRC_NLCP and SRC_NLCP have the same proposed adaptive dictionary \mathbf{D}_P , and the difference between them is the former used collaborative representation while the later used sparse representation. CRC_LBP-riu2 and CRC_LBP-u2 employ the dictionary \mathbf{D}_F and they employed different kinds of LBP.

To be clearer, we listed the best recognition rate and the corresponding optimal feature number in Table 2. The proposed algorithm has the greatest recognition accuracy reached 97.5% and a relatively small number of features 95. CRC_LBP has the least recognition accuracy rate because the number of features is too small.

Table 3 shows the recognition accuracy of the OR database for different numbers of feature. Obviously, the proposed algorithm is the best one for any dimension, and the recognition rate was improved by 0.5–4.3% compared with the second best. The robustness of this proposed method to occlusion (wears glasses or scarf) is authenticated on AR database and the results are presented in Table 4. Our algorithm recognition accuracy is increased by 20.5% for glass and by 4.9% for scarf. Each algorithm has different requirements for feature dimension, and the classification will not be performed if feature dimension is not satisfactory.

4.4 Recognition time

For those approaches based on the sparse representation, recognition time consists of a solution to representation vector (SRV) and pattern matching. In this paper, we employ the nearest neighbor classifier, which is the simplest and fast, to classify. The recognition time was

mainly decided by SRV. Neither $\ell_1 \sim \ell_s$ used in SRC nor regularized least squares (RLS) used in CRC_RLS, the time to solve SRV is mainly determined both by the number of dictionary atoms and by feature dimension. Any increase in either of the two numbers will make the solution time longer. Figure 10 provided the curve of the time spent on one SRV with the number of atoms and feature dimensions on the database ORL.

For ORL, when feature dimension equals to 95, the time to solve one SRV with different numbers of atoms and different methods is listed in Table 5. For database AR, the time is given in Table 6. Experiment results show that the recognition time is greatly reduced as the number of atoms decreases and that the RLS method is much faster than $\ell_1 \sim \ell_s$.

In addition, the number of atoms must be greater than the feature dimension, if not, there will be no solution.

5 Conclusion

Over-complete dictionaries are very important for sparse representation based on classification. In this paper, we presented an adaptive dictionary learning approach. We choose separately the training images that are closer to the testing image from each class as the atoms to adaptively make up the over-complete dictionary. The proposed dictionary changes with the different testing images. The closer training images are called nearest neighbors labeled class by class in the LCP feature subspace, so atoms are more similar to the testing in structure, which increases the recognition accuracy. In addition, the number of nearest neighboring images is much smaller than that of total training images, which greatly reduced the recognition time. Fisher's ratio also shows the proposed adaptive dictionary is more discriminatory. Experiment results also show that collaborative representation classification based on this proposed adaptive nearest neighbor dictionary is excellent in recognition accuracy and in recognition time.

To accurately and quickly identify the input testing images, we pay more attention to build an adaptive dictionary. The main idea is to find the nearest neighbors of testing images in each class. If the number of training samples is small for certain classes, then the advantages of this proposed algorithm will disappear. In addition, only Euclidean distance is selected to measure MiC similarity, and we did not investigate other distance formulas such as "cosine coefficient distance." We also did not consider the image noise. In the next work, we will employ different distances such as "Jeffreys and Matusita distance," "cosine coefficient distance," "Canberra distance," and "generalized Dice coefficients" to find similarity between images and investigate influence brought by noise.

Abbreviations

PCA: Principle component analysis; FR: Face recognition; LBP: Local binary patterns; LCP: Local configuration pattern; SRC: Sparse representation based on classification; CRC_RLS: Collaborative representation based classification with regularized least square; RLS: Regularized least square; MiC: Microscopic feature; BRD: Bin ratio-based distance; SRV: Representation vector; CDF: Cumulative distribution function; PDF: Probability density function

Acknowledgements

The authors acknowledge the support of Shandong Province Science Key Research and Development Project (2016GGX101016) and the Innovation Group of Jinan (2018GXRC010).

Author's contributions

Dongmei Wei conceived the algorithm and designed experiments. Dongmei Wei, Taochen, and Shuwei Li perform the experiments. Dongmei Wei, Yuefeng Zhao, and Dongmei Jiang analyzed the results. Dongmei Wei and Tianping Li drafted the manuscript. The authors read and approved the final manuscript.

Funding

The work was supported by Shandong Province Science Key Research and Development Project (2016GGX101016) and the Innovation Group of Jinan (2018GXRC010).

Availability of data and materials

Please contact the corresponding author for data requests.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Shandong Provincial Engineering and Technical Center of Light Manipulations & Shandong Provincial Key Laboratory of Optics and Photonic Device, School of Physics and Electronics, Shandong Normal University, Jinan 250014, China. ²School of Electronic Information, Qingdao University, Qingdao, China.

Received: 8 November 2019 Accepted: 16 March 2020

Published online: 07 May 2020

References

1. M. Turk, A. Pentland, Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991). <https://doi.org/10.1162/jocn.1991.3.1.71>
2. M. Slavković, J. Dubravka, Face recognition using eigenface approach. *Serbian Journal of Electrical Engineering* **9**(1), 121–130 (2012)
3. M.A.-A. Bhuiyan, Towards Face Recognition Using Eigenface. *Int. J. Adv. Comput. Sci. Appl.* **7**(5), 25–31 (2016)
4. A. A. Alorfi, *Performance evaluation of the PCA versus improved PCA (IPCA) in image compression, and in face detection and recognition*[C]. Future Technologies Conference. 2017.
5. T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **24**(7), 971–987 (2002). <https://doi.org/10.1109/tpami.2002.1017623>
6. M.A. Rahim et al., Face Recognition Using Local Binary Patterns (LBP). *Global Journal of Computer Science & Technology* (2013)
7. Xie S, Shan S, Chen X, et al. V-LGBP: Volume based local Gabor binary patterns for face representation and recognition[C]. *International Conference On Pattern Recognition*, 2008: 1–4 2013. doi: <https://doi.org/10.1109/ICPR.2008.4761374>
8. B. Yang, S. Chen, A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. *Neurocomputing* **120**(10), 365–379 (2013). <https://doi.org/10.1016/j.neucom.2012.10.032>
9. Zhenhua, G., Z. Lei, and Z. David, A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *Image Processing IEEE Transactions*, 2010. 19(6): p. 1657–1663. doi: <https://doi.org/10.1109/TIP.2010.2044957>
10. W. Zhang et al., Local Gabor Binary Patterns Based on Kullback–Leibler Divergence for Partially Occluded Face Recognition. *IEEE Signal Processing Letters* **14**(11), 875–878 (2007). <https://doi.org/10.1109/lsp.2007.903260>
11. W. Wang, F.F. Huang, J.W. Li, Face description and recognition using multi-scale LBP feature. *Opt. Precis. Eng.* **16**(4), 696–705 (2008). <https://doi.org/10.1080/02533839.2008.9671389>
12. Y. Guo, G. Zhao, M. Pietikäinen, *Local Configuration Features and Discriminative Learned Features for Texture Description* (2014). https://doi.org/10.1007/978-3-642-39289-4_5
13. X. Li, H. Shen, H. Li, et al., Patch Matching-Based Multitemporal Group Sparse Representation for the Missing Information Reconstruction of Remote-Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(8), 3629–3641 (2017). <https://doi.org/10.1109/JSTARS.2016.2533547>
14. A.Y. Yang et al., Feature selection in face recognition: A sparse representation perspective. *IEEE Trans. Pattern Anal. Mach. Intell.* (2007)
15. J. Wright et al., in *IEEE International Conference on Automatic Face & Gesture Recognition*. Demo: Robust face recognition via sparse representation (2009). <https://doi.org/10.1109/TPAMI.2008.79>
16. M. Yang et al., in *Computer Vision and Pattern Recognition*. Robust sparse coding for face recognition (2011). <https://doi.org/10.1109/CVPR.2011.5995393>
17. M. Yang, L. Zhang, in *European Conference on Computer Vision*. Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary (2010). https://doi.org/10.1007/978-3-642-15567-3_33
18. W. Ou, X. You, D. Tao, P. Zhang, et al., Robust face recognition via occlusion dictionary learning. *Pattern Recogn.* **47**, 1559–1572 (2014). <https://doi.org/10.1016/j.patcog.2013.10.017>
19. W. Ou, X. Luan, J. Gou, et al., Robust discriminative nonnegative dictionary learning for occluded face recognition. *Pattern Recogn. Lett.* **107**, 41–49 (2018). <https://doi.org/10.1016/j.patrec.2017.07.006>
20. Ma L, Wang C, Xiao B, et al. Sparse representation for face recognition based on discriminative low-rank dictionary learning[C]. *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. IEEE, 2012. doi: <https://doi.org/10.1109/CVPR.2012.6247977>
21. Zhang T, Ghanem B, Liu S, et al. Low-Rank Sparse Coding for Image Classification[C]. *2013 IEEE International Conference on Computer Vision (ICCV)*. IEEE Computer Society, 2013. doi: <https://doi.org/10.1109/ICCV.2013.42>
22. L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition? **2011**(5), 471–478 (2011). <https://doi.org/10.1109/ICCV.2011.6126277>
23. J. Gou, B. Hou, W. Ou, Q. Mao, et al., Several robust extensions of collaborative representation for image classification. *Neurocomputing* **348**, 120–133.27 (2019). <https://doi.org/10.1016/j.neucom.2018.06.089>
24. J. Gou, L. Wang, Z. Yi, Y. Yuan, W. Ou, et al., Discriminative Group Collaborative Competitive Representation for Visual Classification. *IEEE International Conference on Multimedia and Expo (ICME)* (2019). <https://doi.org/10.1109/ICME.2019.00255>
25. J. Gou, L. Wang, B. Hou, Y. Yuan, et al., Two-phase probabilistic collaborative representation-based classification. *Expert Syst. Appl.* **133**, 9–20 (2019). <https://doi.org/10.1016/j.eswa.2019.05.009>
26. J. Gou, L. Wang, Z. Yi, Y. Yuan, et al., A New Discriminative Collaborative Neighbor Representation Method for Robust Face Recognition. *IEEE Access* **6**(74713), 74727 (2018). <https://doi.org/10.1109/ACCESS.2018.2883527>
27. W. Hu et al., Bin Ratio-Based Histogram Distances and Their Application to Image Classification. *Pattern Analysis & Machine Intelligence IEEE Transactions* **36**(12), 2338–2352 (2014). <https://doi.org/10.1109/tpami.2014.2327975>
28. AR. Available from: http://rvll.ech.purdue.edu/~aleix/aleix_face_DB.html.
29. ORL. Available from: <http://www.cam-orl.co.uk>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.