Abstract—Local features, such as local binary patterns (LBP), have shown better performance than global feature in the problem of face recognition. However, the methods to extract the local features are usually given as fixed, and also neglect the class labels of the training samples. In this paper, we propose a novel algorithm to learn a discriminate local feature from the small patches of the face image to boost the face recognition. The pixels of each image patch and its neighboring patches are both used to construct the local feature. The pixel vector of each patch is mapped to new subspaces by a transformation matrix, and mapped pixel vectors the neighboring patches are also combined to obtain the local feature vector. The subspace mapping parameter and the neighboring patch combination parameter are learned to minimize the distances of local features between the same person, and at the same time to maximize that between different persons. We perform experiments on some benchmark face image database to show the advantage of the proposed method.

Index Terms—Face Recognition, Image Patches, Contextual Feature

I. INTRODUCTION

Face recognition is a traditional pattern recognition problem, and it has attracted much attention from both the research and industrial areas [1]–[5]. This problem is composed of two different procedures, which are feature extraction [6]–[9] and the classifier learning [10]–[13], as is shown in Figure 1.

![Figure 1. Face recognition procedures.](image)

Other pattern recognition problems usually assume that the number of classes is fixed. However, in face recognition problem, the registrant person number is always increasing, and it makes the transitional methods not suitable to this problem anymore. Up to now, on one hand, the most popular classification method used in this field is nearest neighbors classifier [14]–[16]. Moreover, it can also be transferred to a binary classification problem, and the traditional binary classification methods, such as SVM [17]–[19], and Adaboost [20]–[22], can also be used in this problem. On the other hand, feature extraction is also very important. In real world applications, many factors, including expressions, poses, occlusions and illuminations, etc. can affect the face images [23], [24], which are shown in Figure 2. Thus, it is very important to extras the robust visual features from the face images to represent the useful information for recognition. Many feature extraction methods have been developed, which can be classified to two different types — the global features and the local features.

- The global features include the principal component analysis [25]–[27], linear discriminate analysis [25]–[27], etc. They aim to extract the features from the global view of the image.
- The local features are different from the global features, they focus on extract robust features robust to local changes. This type of features include the Gabor wavelets [28]–[30], the local binary patterns (LBP) [31]–[33], and scale-invariant feature transform (SIFT) [34]–[36]. Moreover, the combination of them can improve the recognition performance significantly.

In this paper, we propose to use the small image patch as local feature of face image for face recognition. Moreover, contextual information has been shown to be very useful in both representation and ranking problems. Inspired by the past works on contextual learning [37], [38], we developed a novel learning algorithm to explore the information of contextual image patches. Following Wang et al. [39], who used the neighborhood of the data sample to represent the contextual information, we also include the neighboring patches as the contextual patches. The framework of the proposed local feature learning method is shown in Figure 3. It first find the neighboring patches of one centerline patch, and then map the pixel vectors of these patches to a subspace using a transformation matrix, and finally combines the mapped vectors to a final local feature vector. Using these discriminate local features, we adopt the bag-of-features [40]–[42] method to represent a face image. The main contributions of this work are of three folds:

1. We proposed a novel patch pixel vector mapping method. A transformation matrix is learned for this purpose, and it can extract more useful information for the face recognition problem.
2. We also proposed a novel contextual patches combi-
between class scatter, which is defined as criterion [43]–[45]. Under this criterion, we maximize the mapped pixel matrix as

\[ w \]

Applying it to all the vectors in a pixel matrix, we have a transformation vector \( w \). Besides the image patch itself, we also explore the contextual image patches surrounding it. The pixel vectors of the neighboring \( i \)-th are denoted as \( \mathbf{x}_i, \mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_c} \), where \( c \) is the number of contextual image patches. The pixel vector of an image patch and that of its contextual patches are further organized as a pixel matrix to represent it better, denoted as

\[ X_i = [\mathbf{x}_i, \mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_c}] \in \mathbb{R}^{d \times (c+1)} \tag{1} \]

To map the pixel vectors into a subspace, we can use a transformation vector \( w \in \mathbb{R}^d \) to map it, \( y = w^\top x \). Applying it to all the vectors in a pixel matrix, we have the mapped pixel matrix as

\[ y_i = w^\top X_i = [w^\top \mathbf{x}_i, w^\top \mathbf{x}_{i_1}, \ldots, w^\top \mathbf{x}_{i_c}] \in \mathbb{R}^{1 \times (c+1)} \tag{2} \]

To learn the transformation vector, we use the Fisher criterion [43]–[45]. Under this criterion, we maximize the between class scatter, which is defined as

\[ S_b = \sum_{j=1}^{L} ||y^j - y||^2 \]

\[ = \sum_{j=1}^{L} ||w^\top X^j - w^\top X||^2 \]

\[ = \sum_{j=1}^{L} (w^\top (X^j - X)(X^j - X)^\top w) \]

\[ = w^\top \left( \sum_{j=1}^{L} (X^j - X)(X^j - X)^\top \right) w \tag{3} \]

where \( L \) is the number of classes, \( y^j \) is the mean mapping vector of \( j \)-th class and \( y \) is the total mean vector over the entire training set. \( X^j \) is the mean patch pixel matrix of the \( j \)-th class, and \( X \) is the total mean patch pixel matrix of the entire training set,

\[ X^j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i^j \]

\[ X = \frac{1}{\sum_{j=1}^{L} n_j} \sum_{j=1}^{L} \sum_{i=1}^{n_j} X_i^j \tag{4} \]

where \( X_i^j \) is the \( i \)-th patch matrix of the \( j \)-th class, and \( n_j \) is the number of patches in the \( j \)-th class.

At the same time, we also minimize the within class scatter, which is defined as

\[ S_w = \sum_{j=1}^{L} \sum_{i=1}^{n_j} ||y_i^j - y^j||^2 \]

\[ = \sum_{j=1}^{L} \sum_{i=1}^{n_j} ||w^\top X_i^j - w^\top X^j||^2 \tag{5} \]

\[ = w^\top \left( \sum_{j=1}^{L} \sum_{i=1}^{n_j} (X_i^j - X^j)(X_i^j - X^j)^\top \right) w \]

The problem is formulated as

\[ \max_w \frac{w^\top \left( \sum_{j=1}^{L} \sum_{i=1}^{n_j} (X_i^j - X^j)(X_i^j - X^j)^\top \right) w}{w^\top \left( \sum_{j=1}^{L} \sum_{i=1}^{n_j} (X_i^j - X^j)(X_i^j - X^j)^\top \right) w} \tag{6} \]

and it could be solved as a generalized eigenvalue problem [46]–[48].
Let $w_1, w_2, \ldots, w_l$ be the $l$ eigenvectors of (7) corresponding to the largest eigenvalues ordered. We obtain the transformation matrix $W = [w_1, w_2, \ldots, w_l]$ map the $d \times (c + 1)$ pixel matrix into $l \times (c + 1)$ low dimensional feature matrix, as

$$Y_i = W^TX_i = [W^T x_i, W^T x_{i_1}, \ldots, W^T x_{i_l}] \in \mathbb{R}^{l \times (c+1)} \quad (8)$$

### B. Contextual Patch Combination

We also propose to combine the mapped contextual patches with the centralized patch to obtain the final local feature, as follows,

$$z_i = \alpha_0 W^Tx_i + \sum_{k=1}^c \alpha_k W^T x_{i_k} = W^T X_i \alpha_i \quad (9)$$

where $\alpha_0$ and $\alpha_k$ are the combination weights. We put them in a weight vector as $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_c]^\top \in \mathbb{R}^{c+1}$. To learn the weight vector, we also use the Fisher criterion. The between class scatter is rewritten as

$$S_b = \sum_{j=1}^l \|z^j - z\|^2$$

$$= \sum_{j=1}^l \|W^T X^j \alpha - W^T X \alpha\|^2$$

$$= \sum_{j=1}^l \left( W^T (X^j - X) \alpha \alpha^\top (X^j - X)^\top W \right) \quad (10)$$

Similarly, the within class scatter, which is rewritten as

$$S_w = \sum_{j=1}^L \sum_{i=1}^{n_j} ||z_i^j - z^j||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L W^T (X^j - X) \alpha \alpha^\top (X^j - X)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} ||W^T X_i \alpha - W^T X^j \alpha||^2$$

$$= \sum_{j=1}^L \sum_{i=1}^{n_j} W^T (X_i^j - X^j) \alpha \alpha^\top (X_i^j - X^j)^\top W$$
For $t = 1 : T$
- Solve $w(t)$ as in (16) by using $\alpha(t)$;
- Solve $\alpha(t+1)$ as in (14) by using $w(t)$;
- Endfor
- Output: Image Patches $w(T)$ and $\alpha(T+1)$;

III. EXPERIMENTS

In the experiment, we evaluate our method on a large scale face image database.

A. GTAV Face Database

The GTAV Face Database [49] is used in our experiment. It is a face database created mainly to test the robustness of face recognition algorithms when strong pose and illumination variations are imposed. There are images of 44 persons, and for each person 27 images are available. Different pose views and different illuminations are applied to the images of each person. The different pose views include $0^\circ, \pm 30^\circ, \pm 45^\circ, \pm 60^\circ$ and $\pm 90^\circ$, while the three different illuminations include environment or natural light, strong light source from an angle of $45^\circ$, and an almost frontal mid-strong light source. Moreover, different expressions are also included. Each image is of size of $240 \times 320$. Some examples are given in Figure 4.

![Figure 4. Examples of GTAV Face Database.](image)

B. Recognition Results

We compare the proposed CPF method against some other face feature extraction methods, including Gabor features, SIFT, and LBP, etc. The 10-fold cross validation is used as the experiment protocol [50]–[52]. The recognition results are given in Figure 5. From this figure, we can see that the proposed learning base local feature CPF achieves much higher recognition rates than the other local features, such as Gabor, LBP, and SIFT. The results indicate that proposed method can extract more discriminative information than the fixed methods for face recognition problem.

![Figure 5. Face recognition results of different local features.](image)

We also compare different distance measures to compare the local feature based image representation. The bag-of-features based method is applied to each image to represent it to a normalized histogram. Thus we compared a few popular distance metrics to compare the histograms, including $\chi^2$ distance [53], [54], histogram intersection kernel (HIK) [55], [56], $L_2$ distance [57], [58], and the earth mover’s distance (EMD) [59], [60]. The results are given in Figure 6. It can be seen that with EMD distance, CPF achieves the best recognition rates, and outperforms HIK significantly. The $L_2$ norm distance and $\chi^2$ distance also achieve significantly higher recognition rates than HIK.

![Figure 6. Face recognition results of different distance measures.](image)
IV. CONCLUSION

In this paper, we proposed a novel local feature CFP for the problem of face recognition. It uses both the pixels the pixels of small image patch, and its contextual information to boost the recognition performance. The pixel transformation parameter, and the contextual patch combination parameter are learned by using the Fisher criterion, so that the differences between different persons could be maximized, while that of the images of the same person could be minimized. The encouraging experimental results show that the proposed CFP outperforms other local features and works well with different distance functions.

REFERENCES


Wenjing Liao received her B.E. degree in Computer Science fromGuizhou University of Finance and Economics, China in June 2005 and her M.E. degree in Software Engineering from SiChuan University, China inDecember2010. Her current research interest includes data mining technology and algorithm research and applied cryptography.