Face Recognition and Gender Classification Using Orthogonal Nearest Neighbour Feature Line Embedding

Regular Paper

Gang-Feng Ho¹, Ying-Nong Chen¹, Chin-Chuan Han²,* and Kuo-Chin Fan¹,3

1 Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan
2 Department of Computer Science and Information Engineering, National United University, Miaoli, Taiwan
3 Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan
* Corresponding author E-mail: cchan@nuu.edu.tw

Received 18 Jun 2012; Accepted 20 Jul 2012

© 2012 Ho et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract In this paper, a novel manifold learning algorithm for face recognition and gender classification - orthogonal nearest neighbour feature line embedding (ONNFLE) - is proposed. Three of the drawbacks of the nearest feature space embedding (NFSE) method are solved: the extrapolation/interpolation error, high computational load and non-orthogonal eigenvector problems. The extrapolation error occurs if the distance from a specified point to one line is small when that line passes through two farther points. The scatter matrix generated by the invalid discriminant vectors does not efficiently preserve the locally topological structure - incorrect selection reduces recognition. To remedy this, the nearest neighbour (NN) selection strategy was used in the proposed method. In addition, the high computational load was reduced using a selection strategy. The last problem involved solving the non-orthogonal eigenvectors found with the NFSE algorithm. The proposed algorithm generated orthogonal bases possessing more discriminating power. Experiments were conducted to demonstrate the effectiveness of the proposed algorithm.

Keywords Nearest Feature Line; Face Recognition; Gender Classification; Extrapolation; Orthogonal Basis

1. Introduction

Face recognition (FR) and gender classification (GC) are widely used in many commercial systems for monitoring pedestrians who watch advertisements on TV walls or else for monitoring the behaviour of buyers at vending machines. The ID and gender of pedestrians is automatically identified for further application. Though FR and GC are two different applications, gender classification could be considered to be a two-class classification problem - a special case of FR. Three approaches - classifier design, image enhancement and feature discriminative analysis - are proposed to achieve better performance.

The first step is to design a good classifier. Neural network-based classifiers are frequently used in FR. A principle component neural network with a self-
configurable systolic architecture to automatically update eigenfaces whenever training faces change was proposed in [1]. Fuzzy measures and fuzzy logic improved the aggregation operator in neural networks [2]. The modular neural network outputs are integrated and performance is improved. Similarly, a radial basis function enhances the neural network’s performance with an incremental learning mechanism [3]. The re-training process is unnecessary because new data is locally selected and updated by a regressor. From the results of [3], a higher average accuracy and a much lower computational load were achieved. A neuro-fuzzy quantification method [4] is proposed to describe the lyashi of human faces, increasing accuracy by 2-8%. An AdaBoost-based classifier is another popular classifier through the boosting method. Weak classifiers are integrated to construct a strong classifier and several strong classifiers are cascaded to enhance performance for gender classification [5-7]. Besides this, certain tricks were also utilized for the improvement of efficiency. For example, a look-up-table is used in [6] and a 93% identification rate was achieved by pixel comparisons in [7].

Second, images are enhanced for the improvement of recognition by using image processing methods. A set of weighted colour-component features is found from various colour spaces and integrated by a boosting selection algorithm for face recognition [8]. Local colour vector binary patterns (LCVBPs) [9] consisting of colour-norm patterns are extracted from the norm values of a pixel and its four neighbours. In addition, colour angular patterns are also extracted from the angles between the vectors of colour features. 3-D face features are extracted from the shape-based method [10] against the variations of face expressions, illumination and poses.

Third, eigenspace projection methods are successfully used for feature extraction. Recently, manifold learning algorithms have attracted much attention because samples are always presented in a manifold structure in high dimensional feature spaces [11]. By preserving non-linear local structures, the discriminative power of manifold learning-based classifiers is greatly increased when compared to conventional global Euclidean structure preserving methods, such as principle component analysis (PCA) [12] and linear discriminative analysis (LDA) [13]. A third-order tensor colour representation and a tensor discriminant colour space (TDCS) model [14] preserve the spatial structure of colour images. A colour space transformation matrix is obtained to maximize Fisher’s criterion. Images are convoluted by a complete set of bases - e.g., blur kernels - to construct a new subspace against blur variance [15]. A view-manifold-based TensorFace (V-TensorFace) and a kerneled TensorFace (K-TensorFace) [16] generate a continuous view manifold structure of unseen views in the tensor-based analysis. Other, remaining, problems with FR include occlusion and poor image quality. Outliers caused by occlusion and illumination variations are solved by kernel discriminant analysis [17] and the spectral-graph-based method [18]. In addition, low-resolution images are enhanced and identified in video surveillance. The non-linear mapping from low-resolution images to high-resolution images is constructed with the canonical correlation analysis and radial basis functions [19]. For gender classification, another PCA variant, called principle geodesic analysis (PGA), was proposed. Sample data with a manifold structure in Euclidean spaces was transformed into data in Riemannian spaces by using exponential log maps [20]. Furthermore, the linear discriminant techniques for gender classification have been reviewed in [21]. When the training samples are sufficient, the SVM-based classifier outperforms the others. On the other hand, if the training samples were only a few, the linear approaches achieved better results. In addition, the accuracy rates of gender classification are significantly improved by automatic face alignment [22].

In this paper, we focus on the study of feature discriminative analysis. An algorithm - orthogonal nearest neighbour feature line embedding (ONNFLE) - was modified based on previous work (e.g., nearest feature line embedding (NFLE) [23]). With the NFLE method, two problems - extrapolation/interpolation error and high computational load - were remedied during the computation of the class scatterers. Those discriminant vectors which were only a small distance from a specified point on the feature lines were chosen for scatter computation. An extrapolation error occurs when the distance is small while two prototypes are far away from the specified point. At the same time, the near neighbours are not chosen due to the large distances involved. This selection strategy violates the rule of local structure preservation. To remedy this problem, the ONNFLE algorithm modified the selection rule. A point-to-line (P2L) adjacency matrix was constructed to preserve the local topology structure. Three characteristics - NFL-based measurement, neighbourhood structure preservation and class separability - were presented in ONNFLE. In addition, orthogonal projection bases were obtained in ONNFLE.

The rest of this paper is organized as follows: In section 2, eigenspace projection methods, LPP, modified NFL and NFLE, are briefly reviewed. The algorithm ONNFLE is presented to obtain the orthogonal projection matrix in section 3. In section 4, two experiments - face recognition and gender classification - are presented to show the effectiveness of the proposed method. Finally, conclusions are given in section 5.
2. Eigenspace projection methods

Given $N$ training samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{R}^d$ consisting of $C$ classes, new samples in a low-dimensional space were obtained by the linear projection $y_i = \mathbf{W}^T \mathbf{x}_i$, where $\mathbf{W}$ is a linear projection matrix that needed to be found.

2.1. Locally preserving projection algorithm

Locally preserving projection (LPP) [11] is a popular and effective manifold learning algorithm for preserving the manifold structure of samples. The transformation matrix $\mathbf{W}^*$ is obtained by solving the following minimization problem:

$$
\mathbf{W}^* = \arg\min_{\mathbf{W}} \sum_{i,j} \| y_i - y_j \|_2^2 S_{i,j}
= \arg\min_{\mathbf{W}} \sum_{i,j} \| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j \|_2^2 S_{i,j}
= \arg\min_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W})
$$

where $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix in which, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, $\mathbf{D} = \sum_i \mathbf{S}_{i,i}$ and $\mathbf{S}$ is the similarity matrix. The transform matrix $\mathbf{W}^*$ is given for the eigenvectors with the smallest corresponding eigenvalues by solving the general eigenvalue problem: $\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W}$. This minimization ensures that two ‘close’ samples in a high resolution space are close enough in a low resolution space. According to the consequences presented in [24], the minimization form in the LPP algorithm has the same form as in the Locally Linear Embedding (LLE) algorithm below [25]:

$$
\mathbf{W}^* = \arg\min_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W})
= \arg\min_{\mathbf{W}} \text{tr}(\mathbf{Y}(\mathbf{D} - \mathbf{S}) \mathbf{Y}^T)
= \arg\min_{\mathbf{W}} \text{tr}(\mathbf{Y}(\mathbf{I} - \mathbf{M})^T (\mathbf{I} - \mathbf{M}) \mathbf{Y}^T)
= \arg\min_{\mathbf{W}} \sum_i \| \mathbf{y}_i - \sum_j \mathbf{M}_{i,j} \mathbf{y}_j \|_2^2,
$$

in which $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, $\mathbf{S}_{i,j} = (\mathbf{M} + \mathbf{M}^T - \mathbf{M}^T \mathbf{M})_{i,j}$ and $\mathbf{I}$ is an identity matrix. In [24], the optimal weights in $\mathbf{M}$ were obtained by solving an least-squares problem with two constraints: (1) $\sum_j \mathbf{M}_{i,j} = 1$, and (2) $\mathbf{M}_{i,j} = 0$, if $\mathbf{y}_j$ does not belong to the neighbours of $\mathbf{y}_i$.

2.2. Modified nearest feature line method

Li showed the discriminative power of nearest linear combination (NLC) in face classification [26]. Generally, the NLC algorithm generates virtual prototypes for matching using straight lines passing through prototype pairs. The capacity of prototype representation is increased because of infinite virtual prototypes. However, two problems - extrapolation and interpolation errors - exist for the NLC algorithm, as shown in Fig. 1. Consider two feature line points $L_{2,3}$ and $L_{4,5}$ generated from two prototype pairs $(\mathbf{x}_2, \mathbf{x}_3)$ and $(\mathbf{x}_4, \mathbf{x}_5)$, respectively. Points $f_{2,3}(\mathbf{x}_1)$ and $f_{4,5}(\mathbf{x}_1)$ are two projection points of lines $L_{2,3}$ and $L_{4,5}$ for a query point $\mathbf{x}_1$. From Fig. 1, it is clear that point $\mathbf{x}_1$ is close to points $\mathbf{x}_2$ and $\mathbf{x}_3$, but far away from points $\mathbf{x}_4$ and $\mathbf{x}_5$. However, the distance $\| \mathbf{x}_1 - f_{4,5}(\mathbf{x}_1) \|$ for line $L_{4,5}$ is smaller than that for line $L_{2,3}$ - i.e., $\| \mathbf{x}_1 - f_{2,3}(\mathbf{x}_1) \|$. The discriminant vector for line $L_{4,5}$ to point $\mathbf{x}_1$ was selected rather than the other one. In addition, a great deal of computational time was needed due to the vast number of feature lines in the classification phase - i.e., $C_2^{N-1}$ possible lines.

A local nearest neighbour classifier (LNCC) [27] was designed to solve the extrapolation inaccuracy problem of the NLC algorithm. The feature lines passing through the nearest two prototypes of a probe were chosen for matching. Both the extrapolation inaccuracy and high computation problems were mitigated. However, the proposed method decreased the classification ability of the NLC-based classifier. A rectified nearest feature line segment (RNFLS) [28] algorithm - another modified version - was proposed to remedy both extrapolation and interpolation inaccuracy. Line segments trespassing on the territory of different classes were removed. The specified subface for a class was constructed from the satisfied line segments. However, RNFLS requires high computation because a vast number of segments must be checked. The NLC, LNCC and RNFLS algorithms check the possible feature lines in the classification phase, which is an impractical strategy for many systems. In the NFLE algorithm [23], the NLC strategy is performed in the training phase rather than in the classification phase. The point-to-line distance metric is embedded into the transformation matrix. As shown in [23], not only is the
matching time reduced but so too is classification performance improved.

2.3. Nearest feature line embedding

The NFLE algorithm is a new linear transformation method for face recognition. The point-to-line (p-2-l) strategy originated from the NLC approach [26]. The class scatter using the p-2-l strategy is represented as a Laplacian matrix, defined as follows:

\[
W^* = \arg \min_{W} \sum_{i \in \alpha} \| y_i - f_{m,a}(y_i) \|^2 w_{m,a}(y_i) = \arg \min_{W} \text{tr}(W^T X L X^T W). 
\]

Here, \( f(y_i) \) is a projection point on line \( L_{m,a} \) for point \( y_i \) and weight \( w_{m,a}(y_i) \) represents the connectivity strength for point \( y_i \) and line \( L_{m,a} \). The projection point \( f_{m,a}(y_i) \) is represented as a linear combination with points \( y_m \) and \( y_a \) :

\[
f_{m,a}(y_i) = y_m + t_{m,a}(y_i - y_m).
\]

The discriminant vector is represented by a new matrix notation using simple algebra operations:

\[
y_i - f_{m,a}(y_i) = y_i - \sum_{j} M_{ij} y_j.
\]

According to the consequences of [24], the class scatter in EQ. (3) was represented as a Laplacian matrix. Furthermore, the matrices of the within-class scatter \( S_w \) and the between-class scatter \( S_b \) were calculated to maximize Fisher’s criterion. For more details, see [23]. Since the NLC measure increases the representational capacity of the prototypes, the algorithm NFLE using the NLC measure also increases the discriminative power in the feature spaces. NFLE preserves, therefore, much more information than LPP. In addition, the class scatter in NFLE is represented as a Laplacian matrix. Consequently, the topological locality in original spaces is preserved in low-dimension feature spaces.

Though the p-2-l strategy was successfully adopted in the training phase instead of the classification phase for the NFL-based classifier, some drawbacks remained and limited its performance. Three problems are: (1) extrapolation/interpolation inaccuracy: NFSE may not preserve the locality precisely when prototypes are far away from the probes. Thus, the discriminant vector from point \( x_i \) to feature line \( L_{4,5} \) was selected instead of that from point \( x_i \) to line \( L_{2,3} \), as shown in Fig. 1; (2) high computation complexity: a large number of feature lines are generated when the number of training samples \( N \) is large; and (3) non-orthogonal bases: non-orthogonal eigenvectors are generated from the NFLE algorithm, which does not reconstruct the original images and the intrinsic structure properly.

3. Orthogonal nearest neighbour feature line embedding

In this section, an algorithm - Orthogonal Nearest Neighbour Feature Line Embedding (ONNFL) - is proposed. Three problems - extrapolation/interpolation inaccuracy, high computational complexity and non-orthogonal bases - are all considered as having more discriminant power when obtaining the transformation.

3.1 Nearest neighbour feature line embedding (NNFL)

In order to overcome the extrapolation and interpolation inaccuracy, the feature lines for a query point were generated from the \( K \) nearest neighbourhood prototypes. More specifically, when two points \( x_m \) and \( x_a \) belonged to the nearest neighbours of a query point \( x_i \), a straight line passed through points \( x_m \) and \( x_a \) was called the nearest neighbour feature line (NNFL).

The discriminant vector \( x_i - f_{m,a}(x_i) \) was chosen for the scatter computation. The selection strategy for discriminant vectors in NNFL was designed as follows:

(1) The within-class scatter \( S_w \) : feature lines were generated from the \( k_i \) nearest neighbours within the same class for the computation of the within-class scatter, i.e., a set \( F_{k_i}^w(x_i) \).

(2) The between-class scatter \( S_b \) : similarly, \( k_j \) nearest neighbours in different classes for a specified point \( x_j \) were selected to generate the feature lines and calculate the scatter \( S_b \), i.e., a set \( F_{k_j}^w(x_j) \).

The proposed algorithm is a simple and effective method for alleviating extrapolation and interpolation errors. In addition, the scatter matrices were also represented as a Laplacian matrix. The complexity of NNFL was more efficient than that of NFL. Consider \( N \) training samples, \( C_2^{N-1} \) possible feature lines generated and a distance of
$C_2^{N-1}$ calculated for a specified point. The $K_1$ nearest feature lines were chosen from all possible lines in order to calculate the class scatter. The time complexity was $O(N^2)$ for line generation and $O(2N^2 \log N)$ for distance sorting. At the same time, when the $k$ nearest prototypes were chosen for line generation, where $k \ll N$, the time complexity for selecting the $K_1$ nearest feature lines was $O(k^2) + O(2k^2 \log k)$. Extra overhead $O(N \log N)$ was needed for finding the $k$ nearest prototypes. When $N$ was large, the traditional method needed more time to calculate the class scatter.

### 3.2 Orthogonal nearest neighbour feature line embedding (ONNFLE)

According to the consequences of [29], non-orthogonal basis functions make it difficult to reconstruct the original data. In addition, orthogonal basis functions have more locality preserving power than non-orthogonal ones. The proposed ONNFLE method was modified from the NFLE approach in which the within-class scatter $S_w$ and the between-class scatter $S_b$ were calculated to maximize Fisher’s criterion $S_b/S_w$. ONNFLE is expected to have more discriminating power than NFLE; in other words, ONNFLE tries to find a transformation matrix $W^*$ which is composed of $m$ orthogonal basis vectors $W^* = [w_1, \ldots, w_m]$ and satisfies the following constraints:

$$W^* = \arg \max_w \frac{W^T S_b W}{W^T S_w W} \quad \text{subject to} \quad w_i^T w_j = 0, i \neq j, i, j = 1, 2, \ldots, m.$$  

The derivation of orthogonal bases is very similar to that in [29]. The optimization in Eq. (6) is equivalent to maximizing $W^T S_b W$ with an additional constraint $W^T S_w W = 1$. Generally, this constraint is a positive definite matrix. The ratio of $W^T S_b W$ and $W^T S_w W$ in Eq. (4) remained unchanged when $W^T S_w W$ was normalized to one. The following generalized eigenvalue problem was solved:

$$S_b W = \lambda S_w W \quad \text{(7)}$$

Vector $w_1$ was the first eigenvector with respect to the largest eigenvalue of the matrix $(S_w)^{-1} S_b$. When the first $m-1$ orthogonal eigenvectors were generated, the $m^{th}$ eigenvector was obtained by maximizing the following objective function:

$$f(w_m) = \frac{w_m^T S_b w_m}{w_m^T S_w w_m}$$

with the constraints

$$w_m^T w_1 = w_m^T w_2 = \cdots = w_m^T w_{m-1} = 0, \text{ and } w_m^T w_m = 1.$$ 

EQ. (8) was formulated using the Lagrange multipliers to include all constraints:

$$L_m = w_m^T S_b w_m - \lambda \left( w_m^T S_w w_m - 1 \right) - \alpha_1 w_m^T w_1 - \cdots - \alpha_{m-1} w_m^T w_{m-1}$$

The parameters $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}, \lambda$ and $w_m$ were derived as follows: first, the partial derivative of $L_m$ with respect to $w_m$ was set as zero, i.e., $\partial L_m / \partial w_m = 0$.

$$2S_b w_m - 2\lambda S_w w_m - \alpha_1 w_1 - \cdots - \alpha_{m-1} w_{m-1} = 0 \quad \text{(10)}$$

Next, EQ. (11) was obtained by multiplying the left side of Eq. (10) with a term $w_j^T$ and the orthogonal basis constraints $w_m^T w_1 = w_m^T w_2 = \cdots = w_m^T w_{m-1} = 0$ were required. Parameter $\lambda$ was calculated from the following equation:

$$2w_m^T S_b w_m - 2\lambda w_m^T S_w w_m = 0 \Rightarrow \lambda = \frac{w_m^T S_b w_m}{w_m^T S_w w_m} \quad \text{(11)}$$

These $m-1$ equations were obtained next by successively multiplying the left side of Eq. (10) with $m-1$ terms $w_i^T S_w w_i^T S_w \cdots w_i^T S_w$, as shown in Eq. (12). These $m-1$ equations were represented in a single matrix form to obtain the parameters $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}$ from Eq. (13).

$$U^{(m-1)} \alpha^{(m-1)} = 2 \left[ W^{(m-1)} \right]^T \left( S_w \right)^{-1} S_b w_m \Rightarrow \alpha^{(m-1)} = \frac{2}{U^{(m-1)}} \left( W^{(m-1)} \right)^T \left( S_w \right)^{-1} S_b w_m. \quad \text{(13)}$$

Here,

$$\alpha^{(m-1)} = [\alpha_1, \ldots, \alpha_{m-1}]^T, \quad U^{(m-1)}(i, j) = w_j^T \left( S_w \right)^{-1} w_j, \quad W^{(m-1)} = [w_1, \ldots, w_{m-1}]^T, \quad \text{and} \quad U^{(m-1)} = [u^{(m-1)}(i, j)].$$

www.intechopen.com

Gang-Feng Ho, Ying-Nong Chen, Chin-Chuan Han and Kuo-Chin Fan: Face Recognition and Gender Classification Using Orthogonal Nearest Neighbour Feature Line Embedding
Third, the left side of EQ. (10) is multiplied by a term $(S_w^{-1})^\top$ to derive the eigenvector $w_m$. Similarly, it is represented in matrix notation as:

$$2(S_w^{-1})^\top S_B w_m - 2\lambda w_m - (S_w^{-1})^\top W^{(m-1)} \alpha^{(m-1)} = 0$$  \hfill (14)

Parameter $\alpha^{(m-1)}$ was substituted into EQ. (14) as follows:

$$\left\{I - (S_w)^{-1} W^{(m-1)} (U^{(m-1)})^\top (W^{(m-1)})^\top \right\} (S_w)^{-1} S_B w_m = \lambda w_m$$  \hfill (15)

From EQ. (15), when the parameter $\lambda$ in EQ. (11) was maximized, the eigenvector $w_m$ was obtained from the eigenvector of matrix $R_w$ with the largest eigenvalue. Finally, the orthogonal projection bases $W = [w_1, \ldots, w_m]$ were repeatedly generated using the above process.

4. Experimental results

In this section, the experimental results of face recognition and gender classification are presented to show the effectiveness of the proposed NNFLE and ONNFLE methods. Besides this, they are compared with four state-of-the-art methods. The algorithms were evaluated using three public face databases: the dataset CMU [30] for face recognition and the datasets XM2VTS [31] and Max-Planck [32][33] for gender classification. The performance was also evaluated according to the ROC curves for face recognition and the accuracy rates for gender classification.

4.1. Face recognition

The CMU [30] database was composed of 68 people and 170 images per individual with PIE variations for evaluation. Four state-of-the-art methods - LPP [11], orthogonal LPP [29], NFLE [23] and orthogonal NFLE - were implemented for comparison. In the experiments, several images were randomly selected from the data sets for training while the others were used for testing. The face-only images (sized 32-by-32 pixels) were cropped from the original ones to eliminate the influence of hair and background. In addition, the projection matrix $W^\top$ was constructed from the eigenvectors of $S_B^{-1} S_B$ with the largest corresponding eigenvalues when Fisher’s criterion was maximized. Since the dimension of a feature space is sometime larger than the number of training samples, the matrix $S_B$ is singular and non-invertible. This is the well-known “small-sample-size” (S3) problem. The dimensionality of feature vectors was reduced by the PCA transformation to avoid the small-sample-size problem. More than 99% of the feature information was kept in the PCA process. After the PCA transformation, the optimal projection transformations were obtained for the four state-of-the-art algorithms and the proposed ONNFLE and NNFFLE methods. All of the testing samples were matched with the trained prototypes using the NN matching rule. The algorithms were run ten times so as to obtain the average rates. The highest rates of the implemented algorithms for the various training samples are tabulated in Table 1. Moreover, the ROC curves of the recognition rates versus the reduced dimensions for the database CMU are shown in Figs. 2 to 5. From Table 1, the proposed NNFFLE outperformed the other two algorithms for 6, 7 and 8 training samples. This implies that much more local information is preserved in the proposed method.
Figure 2. The recognition rates of various algorithms for various training samples.
4.2 Gender classification

For gender classification, two face datasets were used for evaluation. The dataset XM2VTS [31] was composed of 295 people (viz., 153 males and 142 females) and 12 images per individual with various expressions. As with the experiments for face recognition, the four state-of-the-art methods were implemented for comparison. In the experimental configurations, the data set was separated into two parts: \( T_n \) for training and \( T_m \) for testing. Sixty images (30 males and 30 females) and eighty images (40 males and 40 females) were randomly selected from the data set for training. The other 3480 images (1806 males and 1674 females) and 3460 images (1796 males and 1664 females) were used for testing. The NN matching strategy was adopted for matching the testing samples. The average rates were obtained by running the algorithms ten times. The highest recognition rates and their corresponding reduced dimensions for various algorithms are tabulated in Table 2. The proposed ONNFLE method outperformed the other algorithms both in the 60 and the 80 training samples. The NN strategy preserved more of the local topological structure than the other algorithms. In addition, the orthogonal bases reconstructed the images better than the non-orthogonal bases.

<table>
<thead>
<tr>
<th>Method</th>
<th>5 Trains</th>
<th>6 Trains</th>
<th>7 Trains</th>
<th>8 Trains</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPP</td>
<td>66.94</td>
<td>68.72</td>
<td>72.94</td>
<td>78.13</td>
</tr>
<tr>
<td>(65)</td>
<td>(55)</td>
<td>(60)</td>
<td>(55)</td>
<td></td>
</tr>
<tr>
<td>OLPP</td>
<td>71.55</td>
<td>73.79</td>
<td>77.87</td>
<td>79.83</td>
</tr>
<tr>
<td>(70)</td>
<td>(60)</td>
<td>(70)</td>
<td>(70)</td>
<td></td>
</tr>
<tr>
<td>NFLE</td>
<td>70.80</td>
<td>72.36</td>
<td>76.77</td>
<td>77.17</td>
</tr>
<tr>
<td>(70)</td>
<td>(60)</td>
<td>(45)</td>
<td>(50)</td>
<td></td>
</tr>
<tr>
<td>ONFLE</td>
<td>78.75</td>
<td>79.41</td>
<td>82.51</td>
<td>81.39</td>
</tr>
<tr>
<td>(70)</td>
<td>(60)</td>
<td>(70)</td>
<td>(70)</td>
<td></td>
</tr>
<tr>
<td>NNFLE</td>
<td>69.69</td>
<td>73.17</td>
<td>77.75</td>
<td>82.12</td>
</tr>
<tr>
<td>(65)</td>
<td>(50)</td>
<td>(50)</td>
<td>(55)</td>
<td></td>
</tr>
<tr>
<td>ONNFLE</td>
<td>78.02</td>
<td>80.41</td>
<td>84.08</td>
<td>86.20</td>
</tr>
<tr>
<td>(70)</td>
<td>(70)</td>
<td>(70)</td>
<td>(70)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The recognition performance on the CMU database. (%)

<table>
<thead>
<tr>
<th>Methods</th>
<th>( T_n = 60, P_m = 3480 )</th>
<th>( T_n = 80, P_m = 3460 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPP</td>
<td>72.53 (16)</td>
<td>79.57 (20)</td>
</tr>
<tr>
<td>OLPP</td>
<td>79.46 (15)</td>
<td>81.56 (19)</td>
</tr>
<tr>
<td>NFLE</td>
<td>72.68 (11)</td>
<td>79.88 (19)</td>
</tr>
<tr>
<td>ONFLE</td>
<td>80.01 (12)</td>
<td>81.86 (18)</td>
</tr>
<tr>
<td>NNFLE</td>
<td>73.31 (16)</td>
<td>82.73 (15)</td>
</tr>
<tr>
<td>ONNFLE</td>
<td>82.53 (10)</td>
<td>83.34 (15)</td>
</tr>
</tbody>
</table>

Table 2. The performance of gender classification on the database XM2VTS. (%)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Highest accuracy rates (%)</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted PGA</td>
<td>91.75 (30)</td>
<td>142-by-124 pixels 160 training samples 40 testing samples [20]</td>
</tr>
<tr>
<td>Supervised Weighted PGA</td>
<td>97.00 (10)</td>
<td></td>
</tr>
<tr>
<td>Supervised PGA</td>
<td>91.25 (10)</td>
<td></td>
</tr>
<tr>
<td>Standard PGA</td>
<td>91.25 (20)</td>
<td></td>
</tr>
<tr>
<td>Real AdaBoost</td>
<td>96.25</td>
<td></td>
</tr>
<tr>
<td>Gentle AdaBoost</td>
<td>95.75</td>
<td></td>
</tr>
<tr>
<td>Modest AdaBoost</td>
<td>94.50</td>
<td></td>
</tr>
<tr>
<td>PCA+NNFLE</td>
<td>95.75 (10)</td>
<td>43-by-37 pixels 100 training samples 100 testing samples</td>
</tr>
<tr>
<td>PCA+ONNFLE</td>
<td>97.00 (10)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. The gender classification performance on the MaxPlanck database.

5. Conclusions

In this study, a nearest neighbour selection strategy was adopted through the ONNFLE method to alleviate extrapolation and interpolation inaccuracies. In addition, high computational complexity was reduced by using the NN selection strategy. The projection transformation was obtained from orthogonal bases with more discriminating powers in the feature spaces. According to the experimental results of both face recognition and gender classification, the ONNFLE algorithm outperformed the other state-of-the-art methods.
6. References


