TWO-DIRECTIONAL TWO-DIMENSIONAL DISCRIMINANT LOCALITY PRESERVING PROJECTIONS FOR IMAGE RECOGNITION

Jiwen Lu and Yap-Peng Tan

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

We propose in this paper an improved manifold learning method called two-directional two-dimensional discriminant locality preserving projections, \((2D)^2\)-DLPP, for efficient image recognition. As the existing method of two-dimensional discriminant locality preserving projections (2D-DLPP) mainly relies upon the local structure information in the rows of images, we first derive an alternative 2D-DLPP algorithm that makes use of the information in the columns. Exploiting the local structure and discriminant information in both the rows and the columns, we develop the \((2D)^2\)-DLPP method for efficient image feature extraction and dimensionality reduction. Experimental results on two benchmark image datasets show the effectiveness of the proposed method.

Index Terms— Locality preserving projections, two-directional two-dimensional analysis, image recognition.

1. INTRODUCTION

Locality preserving projections (LPP) has been shown to be efficient for image feature extraction and dimensionality reduction [1]. The aim of LPP is to seek an embedding that can best describe the essential manifold and preserve the local structure of images. As LPP is relatively insensitive to outliers, it has gained popularity in many applications of pattern recognition and computer vision, such as face recognition [1] and scene analysis [2]. However, a LPP-based image representation method needs to convert 2D images into 1D vectors, a step that compromises the structural information of images and usually leads to the problem of “curse of dimensionality”.

To overcome this shortcoming, an improved LPP technique called two-dimensional locality preserving projections (2DLPP) [3, 4] has been recently proposed to directly project each image, rather than a lexicographically ordered vector, under a specific projection criterion. The effectiveness of 2DLPP is evidenced from experiments on several image databases [3, 4]. As 2DLPP is a unsupervised learning algorithm, it is suboptimal for image recognition and thus two-dimensional discriminant locality preserving projections (2D-DLPP) [5] has been more recently proposed to exploit discriminant information in 2DLPP, which has been successfully applied in facial expression recognition.

Like many other 2D dimensionality reduction methods, however, 2D-DLPP suffers from one major shortcoming: it needs many more coefficients for image representation when compared to LPP [3, 4] and DLPP [6]. Consider, for example, an image of size \(128 \times 128\), the number of coefficients required by 2D-DLPP is \(128 \times d\), where \(d\) is usually larger than 3 for satisfactory performance. Although this problem may be alleviated by applying PCA after 2D-DLPP, this additional dimensionality reduction may unduly compromise the image structure and the recognition performance.

Similar to some conventional two-dimensional subspace learning methods such as 2DPCA [13], 2DLDA [14] and 2DLPP [3, 4], 2D-DLPP also performs dimensionality reduction only in row direction. In other words, 2D-DLPP mainly relies on the local structure in the rows of the images. Inspired by the similar work done on principal component analysis (PCA) [7], Fisher’s linear discriminant analysis (LDA) [8] and locality preserving projections [9], in this paper we first derive an alternative 2D-DLPP that exploits the local image structure in the other (the column) direction, and then develop the proposed \((2D)^2\)-DLPP algorithm to perform DLPP in both the row and column directions, and evaluate its performance using two benchmark image datasets—ORL face database [10] and PolyU palmprint database [11, 12]—for face and palmprint recognition.

The remainder of this paper is organized as follows. Section 2 briefly reviews the existing 2D-DLPP algorithm. In Section 3, we derive the alternative 2D-DLPP method and propose the \((2D)^2\)-DLPP method. In Section 4, we present the experimental results to show the effectiveness of the proposed \((2D)^2\)-DLPP method. In Section 5, we conclude the paper by highlighting our contribution.

2. 2D-DLPP

Consider a training set consisting of \(N\) images \(X_i\) of \(m \times n\) pixels, where \(i = 1, 2, \cdots, N\). The 2D-DLPP minimizes an objective function defined as [5]:

\[
J = \frac{\sum_{s=1}^{C} \sum_{i,j=1}^{N_s} (Y^s_i - Y^s_j)^T(Y^s_i - Y^s_j)S^s_{ij}}{\sum_{i,j=1}^{C} (M_i - M_j)^T(M_i - M_j)P_{ij}}
\]  

(1)
where \( Y_s^i \) and \( Y_s^j \) denote the low-dimensional representation
of \( X_i \) and \( X_j \) in the \( s \)th class, \( M_i \) and \( M_j \) are the mean samples
of \( Y \) in the \( i \)th and \( j \)th classes, respectively, \( C \) is the
number of classes, \( N_s \) denotes the number of training samples
in the \( s \)th class, \( S_{ij}^s \) and \( P_{ij} \) are two affinity matrices,
defined as

\[
S_{ij}^s = \begin{cases} 
\exp \left( -\frac{\|X_i^s - X_j^s\|^2}{t_1} \right) & \text{if } L_{X_i} = L_{X_j} = s \\
0 & \text{otherwise} \end{cases}
\]

and

\[
P_{ij} = \exp \left( -\frac{\|F_i - F_j\|^2}{t_2} \right)
\]

where \( F_i = \frac{1}{N_i} \sum_{k=1}^{N_i} X_k^i \) and \( F_j = \frac{1}{N_j} \sum_{k=1}^{N_j} X_k^j \) are the
mean samples of \( X \) in the \( i \)th and \( j \)th classes, \( L_{X_i} \) and \( L_{X_j} \)
are the class labels of \( X_i \) and \( X_j \), \( t_1 \) and \( t_2 \) are two empirically
pre-specified parameters, respectively.

Let \( V \) be the transformation matrix and \( Y_i = X_i V \),
\( i = 1, 2, \ldots, N \). By simple algebraic manipulations as
shown in [5], one can reduce the numerator and denominator
mean samples of \( X \) to specifically pre-specified parameters, respectively.

Let \( F \) smaller than zero. Let \( F \) positive semi-definite, the eigenvalues obtained from (4) are no
smaller than zero. Let \( v_1, v_2, \ldots, v_d \) be the eigenvectors of
(4) corresponding to the \( d \) smallest eigenvalues ordered according to \( 0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d \). An \( n \times d \) transformation matrix \( V = [v_1, v_2, \ldots, v_d] \) can be obtained to project each
\( m \times n \) image \( X_i \) into an \( m \times d \) feature matrix \( Y_i \), as follows:

\[
Y_i = X_i V, \quad i = 1, 2, \ldots, N
\]

3. PROPOSED (2D)\(^2\)-DLPP

As 2D-DLPP mainly relies on the local structure in the rows
of the images, we can easily derive an alternative 2D-DLPP
that exploits the column directional information for feature
extraction. Moreover, we consider perform feature extraction
in both row and column directions, and thus introduce a new
two-directional two-dimensional discriminant locality
preserving projections \((2D)\(^2\)-DLPP) method, which performs
dimensionality reduction and feature extraction both in the
row and in column directions of image matrices.

Let \( Z = VX \), and \( Z \) be a \( q \times m \) transformation matrix
to be sought, where \( X = [X_1, X_2, \cdots, X_N] \) is an \( m \times N \) matrix obtained by arranging all the training images in a row form. Similarly, we can have an alternative of 2D-DLPP and
its objective can be formulated as follows:

\[
J(Z) = \sum_{s=1}^{C} \sum_{i,j=1}^{N_s} (Z_i^s - Z_j^s) (Z_i^s - Z_j^s)^T S_{ij}^s
\]

\[
\sum_{i,j=1}^{C} (M_i - M_j) (M_i - M_j)^T W_{ij}
\]

where \( Z_i^s \) and \( Z_j^s \) denote the projected features of \( X_i \) and \( X_j \) in the \( s \)th class, respectively. After simple algebraic manipulations, we obtain the projections of this alternative 2D-DLPP
through solving the following generalized eigenvalue problem:

\[
X \lambda = \eta F H F^T \eta
\]

To obtain an efficient DLPP-based image representation
that exploits the local information and reduces the dimensions
in both the row and column directions, we propose to
seek transformation matrices \( V \) and \( W \) of size \( n \times d \) and \( q \times m \), respectively, that project each \( m \times n \) image \( X_i \) into a \( q \times d \) feature matrix,
given as

\[
T_i = WX_i, \quad i = 1, 2, \ldots, N
\]

by optimizing the objective function

\[
J(T) = \sum_{s=1}^{C} \sum_{i,j=1}^{N_s} (T_i^s - T_j^s)^T (T_i^s - T_j^s) S_{ij}^s
\]

\[
\sum_{i,j=1}^{C} (Q_i - Q_j)^T (Q_i - Q_j) W_{ij}
\]

where \( Q_i \) and \( Q_j \) are the means of samples of \( T \) in the \( i \)th
class and \( j \)th class, respectively.

To the best of our knowledge, there is no closed-form solution
to (10). Hence, we apply a stepwise strategy to solve it.
We propose to (i) obtain the \( n \times d \) transformation matrix
\( V \) by solving the generalized eigenvalue problem (4) using
the \( m \times n \) training images \( X_i \), \( i = 1, 2, \ldots, N \); (ii) use \( V \) to
project the training image \( X_i \) into \( n \times d \) feature matrices \( Y_i \);
and (iii) obtain the \( q \times m \) transformation matrix \( W \) by solving
the generalized eigenvalue problem (7). Alternatively, we can
obtain matrix \( W \) by solving (7) first and then matrix \( V \) by
solving (4). Our empirical study has shown that similar performance in representation and recognition can be attained
regardless of which transformation matrix is obtained first.

For efficient representation, the transformation matrix \( V \)
and \( W \) can be used to project an \( m \times n \) image \( X_k \) into a \( q \times d \) feature matrix \( T_k = WX_k V \). For recognition, we apply a
nearest neighbor classifier on the distance between \( T_k \) and \( T_i \)
(\( i \) is the feature matrix of training image \( X_i \)) as

\[
d(T_k, T_i) = \|T_k - T_i\|_2 = (\sum_{x=1}^{q} \sum_{y=1}^{d} (T_k^{x,y} - T_i^{x,y})^2)^{1/2}
\]
where $T_{x,y}^k$ and $T_{x,y}^l$ denote the $(x, y)$ element of matrices $T_k$ and $T_l$, respectively.

4. EXPERIMENTAL RESULTS

We carried out several experiments on two benchmark image databases—ORL face database [10] and PolyU palmprint database [11, 12]—to evaluate the performance of the proposed $(2D)^2$-DLPP method in comparison with DLPP [6], 2D-DLPP [5], $(2D)^2$PCA [7], and $(2D)^2$LDA [8], and other conventional feature extraction methods such as PCA, LDA, 2DPCA [13] and 2DLDA [14]. The experiments were conducted on a PC with 3.4 GHz CPU and 1GB memory.

4.1. Results on ORL database

The ORL database contains 400 images from 40 subjects, and each subject has ten different images. The images of some of the subjects were acquired at different times. Furthermore, the images were taken with a tolerance for face tilting and rotation by up to $20^\circ$ and a variation in image scaling by up to 10%. All images are in gray levels and normalized to a resolution of $112 \times 92$ pixels. Fig. 1 shows ten samples of one subject in the database.

For each subject, we randomly selected $k$ images to construct the training set and the remaining images as the testing set. We performed two comparative experiments with $k = 2$ and 5, and empirically selected the parameter $t$ by the cross validation strategy. Table 1 shows the performance of the evaluated feature extraction methods.

Table 1. Top recognition rate (%) and training time (s) with corresponding reduced dimension obtained by each method on ORL face database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Dim</th>
<th>$k = 2$</th>
<th></th>
<th>Time</th>
<th>CRR</th>
<th>Dim</th>
<th>$k = 5$</th>
<th></th>
<th>Time</th>
<th>CRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>30</td>
<td>0.188</td>
<td>78.8</td>
<td>0.766</td>
<td>90.5</td>
<td>30</td>
<td>0.188</td>
<td>78.8</td>
<td>0.766</td>
<td>90.5</td>
</tr>
<tr>
<td>LDA</td>
<td>39</td>
<td>0.125</td>
<td>75.9</td>
<td>0.609</td>
<td>92.0</td>
<td>39</td>
<td>0.125</td>
<td>75.9</td>
<td>0.609</td>
<td>92.0</td>
</tr>
<tr>
<td>DLPP</td>
<td>30</td>
<td>0.188</td>
<td>80.9</td>
<td>0.641</td>
<td>93.5</td>
<td>30</td>
<td>0.188</td>
<td>80.9</td>
<td>0.641</td>
<td>93.5</td>
</tr>
<tr>
<td>2DPCA</td>
<td>112×3</td>
<td>0.016</td>
<td>86.6</td>
<td>0.016</td>
<td>96.0</td>
<td>112×3</td>
<td>0.016</td>
<td>86.6</td>
<td>0.016</td>
<td>96.0</td>
</tr>
<tr>
<td>2DLDA</td>
<td>112×3</td>
<td>0.016</td>
<td>85.9</td>
<td>0.016</td>
<td>96.5</td>
<td>112×3</td>
<td>0.016</td>
<td>85.9</td>
<td>0.016</td>
<td>96.5</td>
</tr>
<tr>
<td>2DLPP</td>
<td>112×3</td>
<td>0.016</td>
<td>88.1</td>
<td>0.016</td>
<td>97.0</td>
<td>112×3</td>
<td>0.016</td>
<td>88.1</td>
<td>0.016</td>
<td>97.0</td>
</tr>
<tr>
<td>Alter. 2DPCA</td>
<td>5×92</td>
<td>0.016</td>
<td>84.7</td>
<td>0.016</td>
<td>94.5</td>
<td>5×92</td>
<td>0.016</td>
<td>84.7</td>
<td>0.016</td>
<td>94.5</td>
</tr>
<tr>
<td>Alter. 2DLDA</td>
<td>5×92</td>
<td>0.016</td>
<td>84.4</td>
<td>0.016</td>
<td>95.0</td>
<td>5×92</td>
<td>0.016</td>
<td>84.4</td>
<td>0.016</td>
<td>95.0</td>
</tr>
<tr>
<td>Alter. 2D-DLPP</td>
<td>5×92</td>
<td>0.016</td>
<td>84.9</td>
<td>0.016</td>
<td>97.0</td>
<td>5×92</td>
<td>0.016</td>
<td>84.9</td>
<td>0.016</td>
<td>97.0</td>
</tr>
<tr>
<td>$(2D)^2$PCA</td>
<td>5×5</td>
<td>0.031</td>
<td>88.1</td>
<td>0.031</td>
<td>96.5</td>
<td>5×5</td>
<td>0.031</td>
<td>88.1</td>
<td>0.031</td>
<td>96.5</td>
</tr>
<tr>
<td>$(2D)^2$LDA</td>
<td>5×5</td>
<td>0.031</td>
<td>85.9</td>
<td>0.031</td>
<td>97.5</td>
<td>5×5</td>
<td>0.031</td>
<td>85.9</td>
<td>0.031</td>
<td>97.5</td>
</tr>
<tr>
<td>$(2D)^2$-DLPP</td>
<td>5×5</td>
<td>0.031</td>
<td>88.4</td>
<td>0.031</td>
<td>98.5</td>
<td>5×5</td>
<td>0.031</td>
<td>88.4</td>
<td>0.031</td>
<td>98.5</td>
</tr>
</tbody>
</table>

To evaluate the effects of face alignment on the proposed $(2D)^2$-DLPP method, we prepared the face images in two different ways: one is normalizing each image to align the two eyes at the same height, and the other is simply cropping from each image (without alignment) a subregion to include the main part of the face. Each processed image is of size $64 \times 64$ pixels and some samples are shown in Fig. 2. We then randomly selected 5 images of each subject to construct the training set and the remaining images as the testing set and applied DLPP, 2DLPP, alternative 2D-DLPP and $(2D)^2$-DLPP methods to perform face recognition.

Table 2. Top recognition rate (%) and training time (s) with corresponding reduced dimension obtained by each method on ORL face database with/without alignment.

<table>
<thead>
<tr>
<th>Method</th>
<th>Aligned ($64 \times 64$)</th>
<th>Cropped ($64 \times 64$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim</td>
<td>Time</td>
<td>CRR</td>
</tr>
<tr>
<td>DLPP</td>
<td>30</td>
<td>0.266</td>
</tr>
<tr>
<td>2D-DLPP</td>
<td>64×3</td>
<td>0.016</td>
</tr>
<tr>
<td>Alter. 2D-DLPP</td>
<td>5×64</td>
<td>0.031</td>
</tr>
<tr>
<td>$(2D)^2$-DLPP</td>
<td>5×5</td>
<td>0.063</td>
</tr>
</tbody>
</table>

We can see from Table 2 that the proposed $(2D)^2$-DLPP always attains the highest correct recognition rate with the same and fewer coefficients (dimensions) among all the methods under comparison. Furthermore, similar to other 2D-based feature representation methods, $(2D)^2$-DLPP is also faster than 1D-DLPP method. From Table 2, we can easily see that without proper alignment, the recognition performance of DLPP, 2D-DLPP and the alternative 2D-DLPP methods reduces significantly, while that of the proposed $(2D)^2$-DLPP method can still maintain over 90% correct recognition rate. In other words, $(2D)^2$-DLPP appears to more robust than the other comparison method when the face samples are not perfectly aligned.

4.2. Results on PolyU palmprint database

The PolyU palmprint database [11, 12] contains the palmprints of 100 subjects with six samples from each subject. These palmprint images were collected in two sessions, and three samples were acquired in each session. Fig. 3 shows six cropped palmprint images of size $128 \times 128$ from one subject.

We randomly selected 4 palmprint images of each subject to construct the training set and the remaining 2 as the testing set. Table 3 shows the recognition performance of $(2D)^2$-DLPP versus other DLPP-based feature extraction methods. The superiority of the proposed $(2D)^2$-DLPP method is evidenced for its highest recognition accuracy despite using the
same or fewer coefficients (dimensions) for feature representation.

To further reveal the relationship between the accuracy and dimension of the feature matrices, we conducted a series of experiments with different feature dimensions using the DLPP and the proposed \((2D)\)-DLPP methods. It is easy to see from the results shown in Fig. 4 that the proposed \((2D)\)-DLPP consistently achieves better recognition accuracy than the DLPP method under different feature dimensions.

5. CONCLUSIONS

We have proposed in this paper an efficient image representation and recognition method called \((2D)\)-DLPP. The main difference between the proposed method and the existing 2D-DLPP method is that the latter only relies on the local structure in the row of the images, while our proposed method exploits the local structure in both the image rows and columns. As a result, the proposed method requires fewer coefficients for image representation and attains better recognition accuracy than the existing 2D-DLPP method and other dimensionality reduction methods. Experimental results on benchmark face and palmprint databases clearly show the efficacy of the proposed method.

6. REFERENCES