Multi-Criteria in Discriminant Analysis to Find the Dominant Features

Arif Muntasa*, Indah Agustien Siradjuddin, Rima Tri Wahyuningsrum
Informatics Engineering Department, Engineering Faculty, University of Trunojoyo Madura, Ry Telang Po. Box 2 Kamal, Bangkalan, Indonesia
*Corresponding author, e-mail: arifmuntasa@if.trunojoyo.ac.id

Abstract

A crucial problem in biometrics is enormous dimensionality. It will have an impact on the costs involved. Therefore, the feature extraction plays a significant role in biometrics computational. In this research, a novel approach to extract the features is proposed for facial image recognition. Four criteria of the Discriminant Analysis have been modeled to find the dominant features. For each criterion is an objective function, it was derived to obtain the optimum values. The optimum values can be solved by using generalized the Eigenvalue problem associated to the largest Eigenvalue. The modeling results were employed to recognize the facial image by the multi-criteria projection to the original data. The training sets were also processed by using the Eigenface projection to avoid the singularity problem cases. The similarity measurements were performed by using four different methods, i.e. Euclidian Distance, Manhattan, Chebyshev, and Canberra. Feature extraction and analysis results using multi-criteria have shown better results than the other appearance method, i.e. Eigenface (PCA), Fisherface (Linear Discriminant Analysis or LDA), Laplacianfaces (Locality Preserving Projection or LPP), and Orthogonal Laplacianfaces (Orthogonal Locality Preserving Projection or O-LPP).

Keywords: multi-criteria, discriminant analysis, features extraction, singularity problem, facial recognition

1. Introduction

Humans can record human faces with storing important features. Humans have also been able to recognize a person's face with a very fast time. The process was difficult to implement on a computer. Many approaches have been modeled to imitate the performance of the human brain, for both obtaining the dominant features and the decision making. Researchers have developed the biometrics to mimic the human intelligence. The biometrics computational problem is the image dimensionality. If image dimensionality used is higher, then cost taken is also more expensive. Many algorithms have been improved to reduce the curse dimensionality and also obtain the highest acceptance rate, for both holistic methods [1-11] and featured-based approach [12], and even combination of them (hybrid method) [13]. The Principal Component Analysis is the oldest and the simplest of the appearance approach. Due the most straightforward approach, the Principal Component Analysis (PCA) has been improved by many researchers, i.e. the Linear Discriminant Analysis [14-16], the Locality Preserving Projection (LPP or Laplacianfaces), the Discriminative Common Vector [17], Regularized Discriminant Analysis [18], LDA-Based Algorithms [19], and Kernel Principal Component Analysis [20].

The PCA is subspace method to project the original sample sets to the Eigenvector of the covariance matrix. It can demonstrate any particular facial image in the coordinate space of the Eigenface. The projection results can significantly reduce the image dimensionality and also produce the dominant features to recognize the object. Nevertheless, it has the weakness. The Principal Component Analysis can only efficiently work, when the number of training sets is not larger than the image dimensional [20]. If it does not occur, then it will fail to reduce the dimensionality.

The LDA is one of appearance method as the development result of the Principal Component Analysis. It can map and reduce the dimensionality become a number of classes. The LDA is not also depended on the number of training sets but depends on the number of classes. The LDA can optimize the projection results by optimization of the between and within
class scatter. However, the LDA has also the weakness. It cannot capture the manifold non-
linear structure. It cannot optimally perform dimensionality reduction when the training classes
are more than the image dimensionality.

The LPP performs the data projection of the original sample sets to the Eigenvector of
the Affinity Matrix. It was generated by the Heat Kernel matrix. The features extraction results of
the LPP are the local structure, where they can represent and retain the local manifold.
However, the LPP cannot perfectly restore the integration of the features. The LPP approach
also has the similar weakness with the PCA, which is the performance results depend on the
training sets used as the training sets.

In this research, multi-criteria in the Discriminant Analysis were proposed. Multi-criteria
is a way to capture the manifold structure from four directions. For each direction is described
by using the objective function. The generating results of the objective function were calculated
by using generalized the Eigenvalue problem corresponding to the largest Eigenvalue. Four
directions capturing has proved that the dominant features produced can preserve the manifold
structure so that they can present the object.

The arrangement of the paper is composed as follows. The second section explains the
proposed approach in detail. The similarity measurements are written in the third part. The
fourth section represents the Experimental results. The fifth part discusses and compares the
proposed approach results in other methods. The last section resumes the results of the
research.

2. Research Method

The Linear Discriminant Analysis is the method to maximize the values between-class
scatter and to minimize within-class scatter. It is the enhancement result of the Principal
Component Analysis. However, the Linear Discriminant Analysis or well known as LDA also has
a limitation, which is manifold non-linear structure was difficult to capture. Therefore, it is
necessary to be improved. In this research, four different directions are proposed to obtain the
projection space by multi-criteria in Discriminant Analysis. Suppose $C$ represented classes, and
$i$ stated class index. The members of $C$ class are $X_1, X_2, X_3, \ldots, X_C$. The proposed method has
optimized the Between-Class Scatter ($S_b$) and Total-Class Scatter ($S_b+S_w$) to obtain the
dominant features. It can be performed by derivation of the multi-criteria in Discriminant
Analysis, which are $A_1^*, A_2^*, A_3^*$ and $A_4^*$. They can be stated as maximum argument of the
Between-Class Scatter and Total-Class Scatter as follows:

\[
A_1^* = \operatorname{arg\ max}_a \left( \frac{A^T S_b A}{A^T S_w A} \right)
\]

\[
A_2^* = \operatorname{arg\ max}_a \left( \frac{A^T S_b A}{A^T S_w A} \right)^T
\]

\[
A_3^* = \operatorname{arg\ max}_a \left( \frac{A^T S_b A}{A^T (S_b + S_w) A} \right)
\]

\[
A_4^* = \operatorname{arg\ max}_a \left( \frac{A^T S_b A}{A^T (S_b + S_w) A} \right)^T
\]

The value of $A_1^*, A_2^*, A_3^*$ and $A_4^*$ can capture the object features from the different
direction. Total criteria can be obtained by adding Equation (1), (2), (3) and (4) as seen in the
following equation:
If the value of $A$ is set of criteria $\{A_1, A_2, A_3, \ldots, A_l\}$, then the objective function of the proposed method can be written as follows:

$$A^* = \arg \max_A \left( \frac{\text{trace}(A^T (S_b + S_w) A)}{\text{trace}(A^T S_w A)} + \frac{\text{trace}(A^T (S_b + S_w))}{\text{trace}(A^T S_w A)} \right)$$

The Equation (6) can be solved by splitting for each objective function and followed by the matrix trace. The matrix trace can be calculated by using generalized the Eigenvalue problem as shown in the following equation:

$$A_i^* = \arg \max_A \left( \frac{\text{trace}(A^T (S_b + S_w) A)}{\text{trace}(A^T S_w A)} \right)$$

$$A_2^* = \arg \max_A \left( \frac{\text{trace}(A^T (S_b + S_w))}{\text{trace}(A^T S_w A)} \right)$$

$$A_3^* = \arg \max_A \left( \frac{\text{trace}(A^T S_b A)}{\text{trace}(A^T (S_b + S_w) A)} \right)$$

$$A_4^* = \arg \max_A \left( \frac{\text{trace}(A^T S_b A)}{\text{trace}(A^T (S_b + S_w) A)} \right)$$

The objective function of the Equation (7), (8), (9), and (10) can be stated respectively as follows:

$$J(A) = \frac{A^T (S_b + S_w) A}{A^T (S_w) A}$$

$$J(A) = \left( \frac{A^T (S_b + S_w) A}{A^T S_w A} \right)$$

$$J(A) = \frac{A^T S_b A}{A^T (S_b + S_w) A}$$

$$J(A) = \frac{\left( A^T S_b A \right)^T}{\left( A^T (S_b + S_w) A \right)^T}$$
Minimizing of the value of \( J(A) \) as shown in Equation (11) can be obtained by derivation of the function of \( J \) to \( A \) and it is set to 0 value as follows:

\[
\frac{d}{dA} J(A) = \left( \frac{d}{dv} A^T (S_w + S_o) A \right)^* \left( A^T (S_o) A \right) - \left( \frac{d}{dv} A^T (S_o) A \right)^* \left( A^T (S_w + S_o) A \right)
\]

\[
0 = \frac{2 (S_w + S_o) A^* \left( A^T (S_o) A \right) - 2 (S_o) A^* \left( A^T (S_w + S_o) A \right)}{\left( A^T (S_o) A \right)^2}
\]

\[
0 = 2 (S_w + S_o) A^* \left( A^T (S_o) A \right) - 2 (S_o) A^* \left( A^T (S_w + S_o) A \right)
\]

\[
0 = (S_w + S_o) A^* \left( A^T (S_o) A - (S_o) A^* \left( A^T (S_w + S_o) A \right) \right)
\]

\[
0 = A^T (S_w) A^* (S_w + S_o) A - \left( A^T (S_w + S_o) A \right)^* (S_o) A
\]

\[
0 = (S_w + S_o) A - \frac{A^T (S_w + S_o) A}{A^T S_w A} * (S_o) A
\]

\[
\lambda A = (S_w + S_o)^* (S_w)^{-1} A
\]

The maximum value of Equation (15) can be obtained by minimization the value of \( S_w \) and maximization the value of \( S_w + S_o \). The derivation result of the Equation (15) can be solved by using the Generalized Eigenvalue problem corresponding to the largest Eigenvalue as follows:

\[
\lambda A = (S_w + S_o)^* (S_w)^{-1} A
\]

\[
\lambda A = S_w^* \left( S_w + S_o \right)^{-1} A
\]

\[
\lambda A^T = \left( S_w^* \right)^* \left( S_w + S_o \right)^{-1} A^T
\]

The calculation results of Equations (16), (17), (18) and (19) are summed to obtain the dominant features as follows:

\[
A^* = \sum_{i=1}^{4} A_i^*
\]

The value of \( S_w \) and \( S_o \) can be represented as follows:

\[
S_w = \sum_{i=1}^{C} \frac{1}{(N_i - 1)} \sum_{x \in A_i} (X - \mu_i)^2 (X - \mu_i)^T
\]

Four similarity measurements have been used to obtain the matching results, i.e. the Euclidian Distance (d1), Manhattan (d2), Chebyshev (d3) and Canberra (d4). The similarity measurement methods can be represented in the following equation:
3.1. Evaluation the Proposed Method on the University of Bern Facial Image Database

The University of Bern Facial Image (UoB) Database involved thirty persons, for each person has ten images with different poses. The original size of the UoB facial image database is 512 pixels for height and 342 pixels for width [22]. In this research, all of images were resized into 140 pixels for height and 120 pixels for width. The image sample of the UoB can be seen in Figure 1. It has the different poses for each class, but it has the same expressions, which is normal expression.

Four scenarios have been performed to evaluate the proposed method, these are using two, three, four, and five facial images, they have been chosen randomly. The experimental results are depending on the training sets used, for both the number of training sets and poses. In this case, the number of features used is twenty-nine. As known, the proposed method has been produced C - 1 classes, C represents number of classes of the training sets, which are thirty classes. For each scenario, twenty experiments have been conducted, which are using ten until twenty-nine features as the similarity measurements. The maximum recognition rates
produced by the proposed method using two, three, four and five the training sets are 75.42%, 87.62%, 96.11%, and 98.67%, respectively. The highest recognition rate has occurred when using five training sets. The experimental results of the proposed method demonstrated that the highest recognition rate has occurred when using five training sets, whereas the similarity used is the Euclidian Distance. The lowest recognition rate occurred when using two training sets. Figure 2a, 2b, 2c, and 2d depict the experimental results two training sets using two, three, four, and five training sets (see Figure 2). The usage of the features influenced the recognition rate results. The maximum features used are number of classes minus one (30 – 1 = 29) features.

Figure 1. The University of Bern Facial Image [22]

Figure 2. Experimental Results Using the Proposed Method on the University of Bern Facial Image Database

Figure 3. Image Tests were error recognized [22]
The smallest acceptance rate was produced by using two training sets. The similar features of the different classes induced the recognizing errors. The similar features are caused by the similar image of the testing and the training set found as seen in Figure 3, the first column is the testing set, whereas the second column is the image matching found. In Figure 3 shows that the image matching found are false. However, the similarity between images to each other will also produce the similar features.

3.2. Evaluation the Proposed Method on the YALE Facial Image Database

The second evaluation is experimental using the YALE facial image database. To evaluate the proposed method, scenario used is the same as the first assessment. The YALE facial image database has the smaller classes than the University of Bern facial image database [23]. The YALE facial image database has fifteen classes, for each class has eleven image poses. The YALE is facial image database that has the different poses, lightings and expressions as seen in Figure 4. In this research, the same scenario was used to evaluate the proposed method, which is using four scenarios. For each scenario, the proposed method was tested using two until fourteen features (the number of classes minus one). The experimental results using the proposed method has produced recognition rate 81.48% (Figure 5a), 90% (Figure 5b), 90.47% (Figure 5c), and 91.11% (Figure 5d) for two, three, four, and five training sets respectively.
The recognition rate has increased proportional to the training sets used, though the recognition rate increasing is not significant. Recognition errors were caused by the image variants such as expressions, but the lighting effect can be still recognized. The experimental results in detail can be seen in Figure 5. In Figure 5 can also be shown that the smaller features used have produced, the lower recognition rate, the more features used has also contributed the larger acceptance rate as seen in Figure 5.

3.3. Evaluation the Proposed Method on the ORL Facial Image Database

The last evaluation is executed by using the ORL facial image database. The ORL facial image database has four hundred images. They are gained from forty persons, four each person has ten different poses. The variant of poses consists of expressions, accessories, and position of the facial pose. The expressions of the ORL face image database are smiling, neutral, open and close eyes. The accessories of the ORL is using glasses or not, whereas the position of the facial pose is described by pose owned for each person such as right, left, up and down [24]. Figure 6 demonstrated ten persons of the ORL face image database with the different poses, expressions and accessories.

In this research, the dominant features of the proposed method results have been reduced to number class minus one (40 - 1 = 39). The proposed method was evaluated by using two, three, four and five training sets as demonstrated in Figure 7, i.e. 7a, 7b, 7c, and 7d respectively.

![Figure 6. The ORL Facial Image [24]](image)

![Figure 7. Experimental Results Using the Proposed Method on the ORL Facial Image Database](image)
Twenty-one until thirty-nine features has been utilized for similarity measurements. The evaluation results have produced the recognition 85.62%, 91.07%, 97.08%, and 97.5% for two, three, four and five training sets respectively. The usage of features has affected the recognition rate obtained. The higher recognition rate can be only achieved by using the more dominant features. The more dominant features reduced, the lowest recognition rate obtained. The number of training sets has also influenced the recognition rate obtained. The highest recognition rate was obtained when experimental results using five training sets and thirty-five until thirty-nine dominant features, which is 97.5%. It means, only five of two hundred images were false recognition.

3.4. Discussion and Comparison to Other Methods

The proposed method has proved that multi-criteria of Discriminant Analysis can be implemented to recognize the facial image. The objective function of the proposed method as written in Equation (11), (12), (13) and (14) can be derived and solved by using generalized the Eigenvalue problem. Four criteria of the proposed method modeled have produced the dominant features that can be implemented to recognize the facial image. The proposed method was assessed by using three facial image databases, which are the UoB, the YALE, and the ORL face image databases. The recognition rate has displayed that the proposed method has produced the larger acceptance rate on the ORL face image database than other facial image databases, which are the University of Bern and the YALE face image databases. The proposed method has produced the lowest recognition rate than the University of Bern and the ORL face image databases. The recognition errors on the YALE face image database were caused by lighting effect. It means the proposed method cannot normalize lighting effect. The acceptance rate results of the proposed method were also compared to the other methods, which are the Principal Component Analysis or well-known Eigenfaces, Fisherface, Laplacianfaces, and Orthogonal Laplacianfaces (O-Laplacianfaces) as shown in Table 2, 3, and 4.

<table>
<thead>
<tr>
<th>Training Set Used</th>
<th>Recognition Rate (%)</th>
<th>The Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eigenface</td>
<td>Fisherface</td>
</tr>
<tr>
<td>2</td>
<td>66.25</td>
<td>66.67</td>
</tr>
<tr>
<td>3</td>
<td>71.43</td>
<td>71.43</td>
</tr>
<tr>
<td>4</td>
<td>83.33</td>
<td>81.67</td>
</tr>
<tr>
<td>5</td>
<td>92.00</td>
<td>91.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training Set Used</th>
<th>Recognition Rate (%)</th>
<th>The Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eigenface</td>
<td>Fisherface</td>
</tr>
<tr>
<td>2</td>
<td>43.5</td>
<td>45.7</td>
</tr>
<tr>
<td>3</td>
<td>48.9</td>
<td>64.5</td>
</tr>
<tr>
<td>4</td>
<td>52.2</td>
<td>72.7</td>
</tr>
<tr>
<td>5</td>
<td>57.8</td>
<td>77.5</td>
</tr>
</tbody>
</table>

The proposed method has proved that the experimental results outperformed to other methods for all facial image databases, the University of Bern, the YALE, and the ORL facial image databases.
4. Conclusion

Three databases were employed to evaluate the proposed method. The evaluation results showed that the proposed method can recognize the facial image. The recognition rate results are larger than other methods such as Eigenface, Fisherface, Laplacianfaces, and Orthogonal Laplacianfaces. The highest acceptance rates were achieved when using five training sets for the University of Bern, the YALE, and the ORL facial image databases. The recognition result has produced the lowest acceptance rate when using the YALE face. Recognition errors on the YALE facial image database were caused by lighting effect. Furthermore, the limitation of the proposed method can be recovered by using the Retinex methods to remove the lighting effect.

References