FACE RECOGNITION BASED ON GENDER USING A MODIFIED METHOD OF 2D-LINEAR DISCRIMINANT ANALYSIS

[1]Fitri Damayanti, [2]Wahyudi Setiawan, [3]Sri Herawati, [4]Aeri Rachmad

[1,2,3,4]Faculty of Engineering , University of Trunojoyo Madura
Jalan Raya Telang PO BOX 2 Kamal, Bangkalan, East Java, Indonesia
*E-mail: fitri2708@yahoo.com

Abstract
Currently, the development of technology related to facial recognition is already widely used in applications introduction of biological data (biometrics) such as the introduction of gender. The introduction of gender based human face image is part of a biometric method for identifying an individual by taking facial features. The implementation of applications that require the introduction of gender is the process of market segmentation to determine the demographic trend of products marketed by sex. It also can be used to limit access to a room and others. The purpose of this study is to make implementation of the system is the introduction of gender on facial image or Gender Recognition to enter a form of facial images that can recognize a person's sex quickly, accurately and running well.
This study uses test data 200x200 pixels 400 facial images of 200 images of 200 male and female images. Face database training is taken from http://www.advancedsourcecode.com. This study aims to establish a system of recognition of sex based on facial image using feature extraction Two Dimensional Linear Discriminant Analysis Modification (Modification 2DLDA), which is the result of the development of methods of Two Dimensional Linear Discriminant Analysis (2DLDA) while 2DLDA is the development of methods of Linear Discriminant Analysis (LDA). The use of classification method is the Two Dimensional Correlation Coefficient which is the method of calculating the value of a strong existence or inexistence of linear relationship between two variables. The trial result with the highest degree of accuracy is at 91.45% while the lowest accuracy rate is 86.33%.

Keywords : Gender, Modification 2DLDA, 2D Correlation Coefficient.

INTRODUCTION

Face recognition today has become one of the areas that was observed and developed by experts in pattern recognition, and this is because of the vast use of facial recognition techniques in the application used by the public. The introduction of gender based human face image is part of a biometric method for identifying an individual by taking facial features. The researchers have conducted a study of existing techniques and propose a new technique that is better than the old one, until now many new techniques that have been proposed but these techniques are still not able to provide optimal accuracy. Two things that become the major problem in the face identification is the process of feature extraction from the sample of faces. Classification techniques are also used to classify faces based on features.

Many studies on the introduction of the gender are based on facial image. One of those studies was conducted by Sunita Kumari, Sambit Bakshi, Banshidhar Majhi in 2012 entitled "Gender periocular Classification Using Global ICA Features For Poor Quality Images". In this study, the gender recognition software for using Independent Component Analysis and data test uses Jaffe database (Kumari, Bakshi, and Majhi, 2012). Siddharta K, K Sandhya Rani in 2014 conducted a study entitled "Gender Recognition From Face Pictures with Native WLD Descriptor And Neural Network Approach". The study discussed the use of methods of Weber's Local Descriptor and Neural Networks for the introduction of gender (Siddharta & Rani, 2014). In 2013 Hadeel Fahad Mohamed Abdou Berbar Alrashed and conducted research entitled "Gender Recognnition Using Facial Images Eyes". This study used a 2D-Wavelet Transform, GLCM, DCT and SVM (Alrashed & Berbar, 2013).

Feature extraction is the process to get the distinguishing features, which distinguish a sample face with another face, for most applications of pattern recognition, feature extraction techniques are reliable for a key element in solving pattern recognition problems. Principal Component Analysis (PCA) Methods is used to perform face recognition (M Turk & Pentland, 1991). The method is aimed at projecting the data on the direction which has the largest variation, which is indicated by the eigenvector corresponding to the largest eigenvalue of the covariance matrix. The weakness of the PCA method is not optimal in the separation between the classes (Turk & Pentland, 1991). In 1991, Cheng et al introduce the method of Linear Discriminant Analysis (LDA) for face recognition. This method tries to find a linear subspace that maximizes separation of two classes according to Fisher Criterion J_F patterns. It can be obtained by minimizing the distance matrix distribution in the same class (within-class) and maximizes the distance matrix S_w distribution between classes (between-class) S_b simultaneously resulting in a maximum J_F Fisher Criterion. Linear discriminant Fisher will find a subspace where the classes are separated from each other by maximizing linear J_F Fisher Criterion. If the dimensions of the data is much higher than the number of training samples will cause S_w be singular. This is a drawback of the method of LDA (Belhumeur, Hespanha, & Kriegman, 1997).

There are many methods offered to overcome same class covariance (within the class) which is always singular (small sample problem) . In 1997, P.N. Belheumeur introduced fisherface method for face recognition. This method is a merger between PCA and LDA. Dimension reduction process was performed by the PCA before the LDA. This can overcome singular problem. But the drawback of this method is that during the process of PCA dimension reduction will cause the loss of some information that is useful in the process of discriminant LDA (Belhumeur, Hespanha, & Kriegman, 1997). Any other method that can overcome the problem are singular Direct-LDA, Null-space based LDA, Pseudo-inverse LDA, Two-stage LDA, LDA Regularized (Kong, Wang, Teoh, Wang, & Venkateswarlu, 2005). However, all of the LDA techniques use models based on vector data representation. Generate vectors typically have a higher dimension. The method of Two Dimensional Linear Discriminant Analysis (2DLDA) directly assess within-class scatter matrix of images without image transformation matrix into a vector, and it solves singular problems in within-class scatter matrix (Gao, Zhang, & Zhang, 2008). TDLDA uses fisher criterion to find the optimal projection discriminatory.

Methods 2DLDA finds the value of R and L values which are used to calculate the within-class and between-class scatter scatter. So, there are two values ie within-class scatter S^{R}_{W} and S^{L}_{W}, as well as to the value of between-class scatter there are also two S^{R}_{b} dan S^{L}_{b}. Calculation of two values within-class scatter and the two values between-class scatter causing computation is required even greater. Modification Method of Two Dimensional Linear Discriminant Analysis (Modified 2DLDA) directly assess without transformation matrix image into vector image. 2DLDA modification method calculates the value of within-class scatter and the between-class scatter values. This can reduce the computation time required. (Muntasa, 2014).

In the introduction of gender based on facial image, the classification process is as important as the process of feature extraction. Once the essential features of the face image is generated in the extraction process features, these features will be used for the classification process. Classification method used is the Two Dimensional Correlation Coefficient the method of calculating the value of a strong existence or inexistence of the linear relationship between two variables.

METHODOLOGY

This research integrates methods Modified Two Dimensional Linear Discriminant Analysis as feature extraction and methods Two Dimensional Correlation Coefficient as classification. The process in the introduction of gender based on facial image consists of two parts, namely the image of the training process and the testing process. Each face data base used is divided into two, partly used for training process and the rest was used for the testing process. In the training process there are processes Modification of Two Dimensional Linear Discriminant Analysis which are used to extract features, features that are selected during the training process used in the classification process and is also used to get features that are selected in the trial data.

Modified algorithm design Two Dimensional Linear Discriminant Analysis (Modified 2DLDA) is divided into two subsystems, namely training subsystems and subsystem testing. Here is a description of each subsystem.

Training Process Modification Two Dimensional Linear Discriminant Analysis

The process of *Two Dimensional Linear Discriminant Analysis* Modification (2DLDA Modification) ias divided into three steps, they are:counting the class average score and global average, then counting *within class scatter* matrix and *between class scatter* matrix, ang the last is counting matrix of extraction feature of training data.

Imagery Average

Here are the steps in the process of modification of Two Dimensional Linear Discriminant Analysis (Modified 2DLDA) against a database of training images to calculate the average value:

1. If a database of facial images are a set of n training image $A_i = [A_1, A_2, \ldots, A_n]$ $(i = 1,2,\ldots,n)$ with the dimensions of the image (r x c), then the total set of an image matrix is :

$$A_n = \begin{bmatrix} A_{(n)11} & A_{(n)12} & \cdots & A_{(n)1c} \\ A_{(n)21} & A_{(n)22} & \cdots & A_{(n)2c} \\ \cdots & \cdots & \cdots & \cdots \\ A_{(n)r1} & A_{(n)r2} & \cdots & A_{(n)rc} \end{bmatrix}. \quad (1)$$

This matrix is used as the input data. Other input data is the number of classes (k), the amount of data Smaller (n_i), and the amount of training data (n).

2. The next step is to calculate the average training image of class to i:

$$M_i = \frac{1}{n_i} \sum_{X \in \Pi_i} A \quad . \quad (2)$$

3. Calculate the average of all training image:

$$M = \frac{1}{n}\sum_{i=1}^{k}\sum_{X\in\Pi_i} A \quad . \qquad (3)$$

Within Class Scatter Matrix and Between Class Scatter Matrix

Here are the steps in the process of modification of Two Dimensional Linear Discriminant Analysis (Modified 2DLDA) against a database to calculate the training image within class scatter matrix and between class scatter

1. Calculate the between class scatter matrix.

$$S_b = \sum_{i=1}^{k} n_i (M_i - M)(M_i - M)^T \text{, the size of the matrix } (r \text{ x } r). \qquad (4)$$

S_b matrix size is smaller than the size of the matrix S_b on the classical LDA (Dimension x Dimension).

2. Calculate the within class scatter matrix.

$$S_w = \sum_{i=1}^{k}\sum_{x\in\Pi_i} (A - M_i)(A - M_i)^T, \text{ the size of the matrix } (r \text{ x } r). \qquad (5)$$

S_w matrix size is smaller than the size of the matrix S_w on classical LDA (Dimension x Dimension).

3. Calculate the generalized eigenvalue (λ_i) and *eigenvector* (V) of S_b and S_w in accordance with equation (6) using SVD. $Z = \left[\frac{S_w*(S_w)^T}{S_b*(S_b)^T}\right]\left[\frac{S_w*(S_w)^T}{S_b*(S_b)^T}\right]$ (6)

Feature Extraction

Training feature extraction is the last step of the imagery training process which is used to search extraction features in each image (feature image) in the image data base training. The steps are as follows:

1. Input the form of a matrix of training data:

$$A_n = \begin{bmatrix} A_{(n)11} & A_{(n)12} & \cdots & A_{(n)1c} \\ A_{(n)21} & A_{(n)22} & \cdots & A_{(n)2c} \\ \cdots & \cdots & \cdots & \cdots \\ A_{(n)r1} & A_{(n)r2} & \cdots & A_{(n)rc} \end{bmatrix}.$$

Other Input: projection matrix (V).

2. Calculate the matrix of feature extraction is $B=A*V$. (7)
3. Output: weight training matrix B.

Testing Process of Two Dimensional Linear Discriminant Analysis Modification

The testing process Modification of Two Dimensional Linear Discriminant Analysis consists of the process is the process of feature extraction of test data that aims to search ekstraski features on the test image. The steps are as follows:

1. Input the form of a matrix of test data matrix PC dimensions equal to that of training data matrix (r x c) :

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1c} \\ C_{21} & C_{22} & \cdots & C_{2c} \\ \cdots & \cdots & \cdots & \cdots \\ C_{r1} & C_{r2} & \cdots & C_{rc} \end{bmatrix}. \qquad (8)$$

Other Input: projection matrix (V).

2. Calculate the matrix of feature extraction $D=C*V$. (9)
3. Output: weight testing matrix D.

Process Classification of Two Dimensional Correlation Coefficient

The use of 2D Correlation Coefficient on this study is aimed to measure the distance between the image similarity testing with imagery training. Weight matrix testing the data will be processed by the 2D Correlation Coefficient which works by comparing the entire result of reduction in the value matrix to i weight training data and the value of the data matrix to i weight testing with its square root. Data are considered most similar is the maximum value of each data comparison results of testing and training (Soni & Joshi, 2012).

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (10)$$

Where :
A = weight training image.
B = weight testing image .
$\bar{A}\bar{A}$ = weighted average of the training image.
$\bar{B}\bar{B}$
= weighted average of the testing image.

RESULT AND DISCUSSION

The trial of the recognition system based on the sex of the face image is developed in this research is done on the basis of data taken from http://www.advancedsourcecode.com. This database contains 400 facial images of 200 facial images male and 200 female face images. The face image is 200x200 pixels. Tests on gender recognition system are based on facial image developed by separating the facial image data in a database into two sets of mutually exclusive (disjoint) i.e., the set of training images and test images. The calculation of the percentage of successful introduction of testing is done on a set image. Figure 1 shows some examples of facial images that are used as data testing and training data.

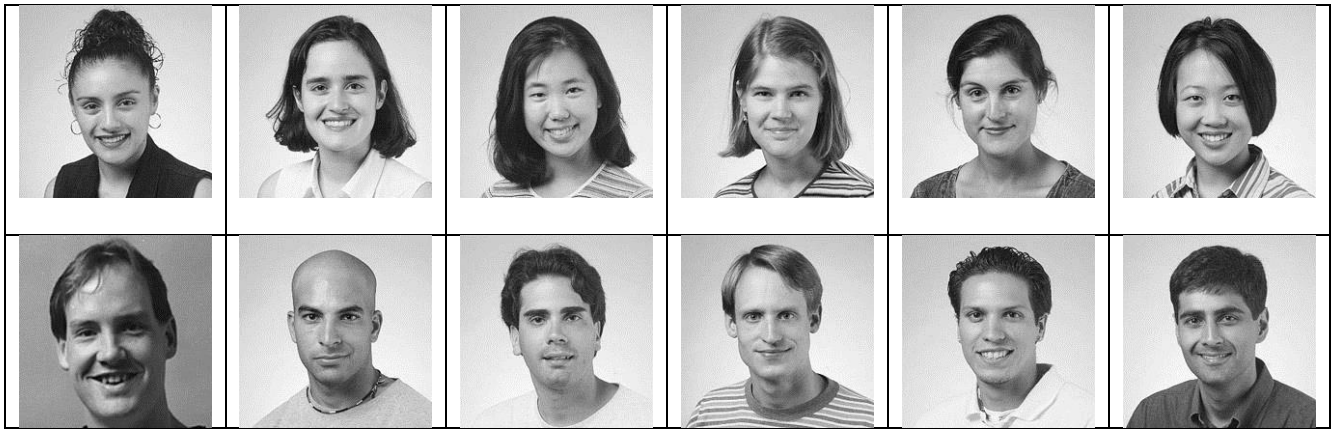

Figure 1. Example Imagery Used For Data Face Trial

Scenario testing is done by varying the amount of training data, and varying the taken-features. training data taken from the face image sequence, while for the trial using eight variations of the amount of training data, there are 8 scenarios:

- Scenario 1, the test took 150 facial images of each class as a member of the set of training data so that the other face image 50 in each class which are included in the test data.
- Scenario 2, the test took 140 facial images of each class as a member of the set of training data so that the other face image 60 in each class which are included in the test data.
- Scenario 3, the test took 130 facial images of each class as a member of the set of training data so that the other face image 70 in each class which are included in the test data.
- Scenario 4, the test took 120 facial images of each class as a member of the set of training data so that the other face image 80 in each class which are included in the test data.
- Scenario 5, the test took 80 facial images of each class as a member of the set of training data so that 120 more face images in each class which are included in the test data.
- Scenario 6, the test took 70 facial images of each class as a member of the set of training data so that 130 more face images in each class which are included in the test data.
- Scenario 7, the test took 60 facial images of each class as a member of the set of training data so that 140 more face images in each class which are included in the test data.
- Scenario 8, the test took 50 facial images of each class as a member of the set of training data so that 150 more face images in each class which are included in the test data.

Table 1 shows the recognition accuracy by using Modified 2DLDA and 2D Correlation method for retrieval features Coefficient 3, 5 and 10.

Table 1. Results of Testing

Scenario	Imagery Data		The number of features	Accuracy level
	Data Training	Data Testing		
1	300	100	3	90%
			5	85%
			10	83%
2	280	120	3	90%
			5	85%
			10	83%
3	260	140	3	90%
			5	87,14%
			10	85%
4	240	160	3	90,63%
			5	86,25%
			10	85%
5	160	240	3	90%
			5	90,42%
			10	87,5%
6	140	260	3	91,45%
			5	89,23%
			10	86,92%
7	120	280	3	89,64%
			5	87,5%
			10	83,21%
8	100	300	3	86,33%
			5	84,67%
			10	80,67%

To easily view the results of gender recognition accuracy by using facial image modification method 2DLDA and 2D Correlation Coefficient methods, it is depicted in the graph shown in Figure 2.

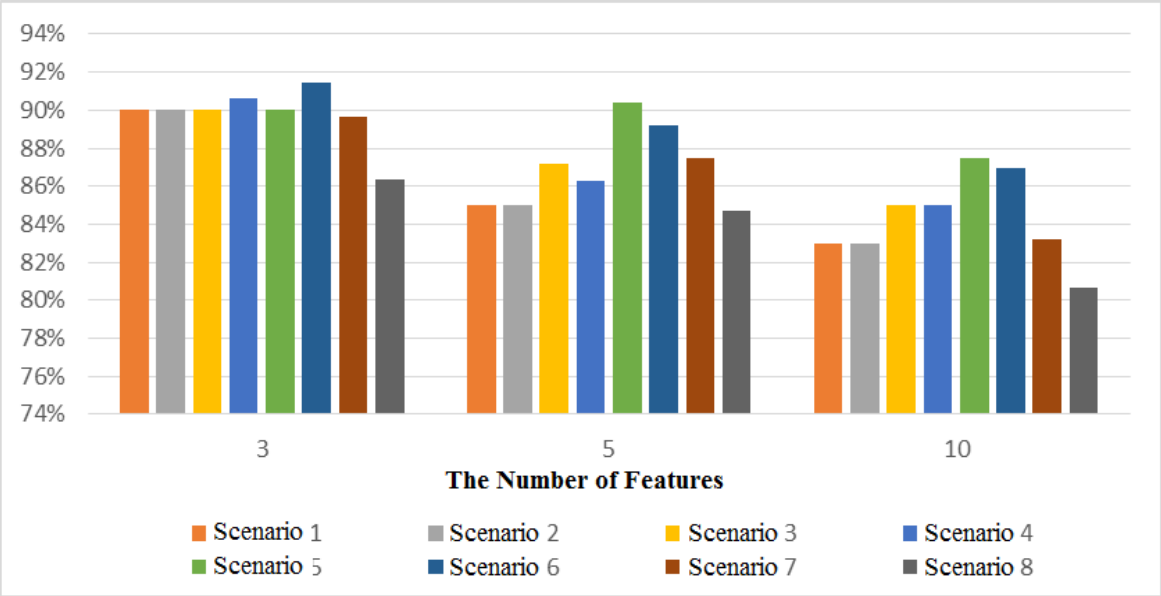

Figure 2. Graph Results Accuracy Recognition

There are three groups of tests to obtain comparative results of the gender recognition accuracy based on facial image. The first group uses the Local Binary Pattern (LBP) for preprocessing, Fisherface method for feature extraction and classification methods as the Euclidean Distance classification (Fenanda, Wahyuningrum, and Damayanti, 2015). The second group using Two Dimensional Linear Discriminant Analysis as a method of feature extraction and classification Support Vector Machine as (Damayanti & Rachmad, 2016). The third group uses methods Modification of Two Dimensional Linear Discriminant Analysis as feature extraction and 2D Correlation method Coefficient as classification.

Table 2 shows the comparison of the results of trials with modification method Coefficient 2DLDA and 2D Correlation with other methods

Table 2. Comparison of Results of Testing

Scenario	Accuracy		
	LBP – Fisherface - ED	*2DLDA - ED*	Modification method 2DLDA – *2D Correlation Coefficient*
1	75%	89%	90%
2	68,33%	88,75%	90%
3	67,86%	88%	90%
4	73,75%	86%	90,63%
5	60,83%	84%	90%
6	49,62%	83%	91,45%
7	45%	70%	89,64%
8	52%	68%	86,33%

CONCLUSION

After analysis test gender recognition system based on facial image, it can be concluded as follows:

1. The highest percentage of recognition accuracy rate of gender based on facial image uses Modified 2DLDA-2D Correlation Coefficient is by 91.45%, while using 2DLDA-ED is by 89%, using the LBP-Fisherface-ED is by 75%.
2. 2DLDA method of computing time is used to perform feature extraction of $O(n^3)$, whereas the method of modification 2DLDA only takes amounted $O(n^2)$
3. There are two important variables that affect the success rate of introduction, i.e. the number of training samples per class is used, and the number of features that are taken.
4. Incorrect classification which is closely the trials is caused by head shape, hair forms and expressions between both the image of women and men.

REFERENCES

Alrashed, H. F., & Berbar, M. A. (2013). Facial Gender Recognition Using Eyes Images. *International Journal Of Advanced Research In Computer And Communication Engineering*, 2, 2441-2445.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs Fisherfaces Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19 (7) : 711 – 720.

Damayanti, F., & Rachmad, A. (2016). Recognizing Gender Through Facial Image Using Support Vector Machine. *Journal of Theoretical and Applied Information Technology (JATIT)*, 88 (3), 607-612.

Fenanda, N., Wahyuningrum, R. T., & Damayanti, F. (2015). Introduction of Gender Based Facial Image Method Using Local Binary Pattern (LBP) and Fisherface. Bangkalan-Indonesia. Informatic Engineering University of Trunojoyo Madura.

Gao, Q. X., Zhang, L., & Zhang, D. (2008). Face Recognition Using FLDA With Single Training Image Per Person. *Applied Mathematics and Computation*. 205 : 726-734.

Kong, H., Wang, L., Teoh, E. K., Wang, J. G., & Venkateswarlu, R. (2005). A framework of 2D Fisher discriminant analysis : application to face recognition with small number of training samples. *IEEE Computer Society Conference On Computer Vision And Pattern Recognition*, (pp. 1083-1088).

Kumari, S., Bakshi, S., & Majhi, B. (2012). Periocular Gender Classification using Global ICA Features for Poor Quality Images. *Procedia Engineering*, 38, 945-951.

Muntasa, A. (2014). New Modelling of Modified Two Dimensional Fisherface Based Feature Extraction. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 12, 115-122.

Siddharta, & Rani, S. (2014). Gender Recognition From Face Pictures With Native WLD Descriptor And Neural Network Approach. *International Journal Of Advanced Research In Computer Science And Software Engineering*, 4(9), 898-901.

Soni, V., & Joshi, R. (2012). A Novel Dimension Reduction Technique Based On Correlation Coefficient. *International Journal Of Scientific And Technology Research* , 1, 122-124.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1): 71-86.