SMILE STAGES CLASSIFICATION BY USING FEATURE EXTRACTION BASED ON 2DPCA AND 2DLDA IN ORTHODONTIC REHABILITATION

Rima Tri Wahyuningrum (1), I Ketut Eddy Purnama (2), Mauridhi Hery Purnomo (3)

(1) Department of Informatic Engineering, Trunojoyo University
Jl. Raya Telang Po Box 2 Kamal, Bangkalan 69162, Indonesia
Email: rima_tw@yahoo.co.id
(2, 3) Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember
Jl. Arif Rahman Hakim, Keputih Sukolilo, Surabaya 60111, Indonesia
Email: (2) hery@ee.its.ac.id, (3) ketut@ee.its.ac.id

Abstract - This paper discusses classification of smile stages based on 2DPCA for feature extraction and then compares the results with the 2DLDA to distinguish the smiling faces at Stage I, Stage III and Stage IV. For the classification process, SVM multiclass was applied with method of one against one and one against all. Thus, 90 smiling faces data which has been validated by dentist specializing in tooth conservation (30 Stage I, 30 Stage III and 30 Stage IV) is being used for learning and testing evaluation in the experimental process. The experimental results show that not only computational efficiency, but also the recognition rates of 2DLDA + SVM one against all are higher than others.

Keywords: smile stages classification, 2DPCA, 2DLDA, SVM one against one and one against all.

1. INTRODUCTION

Many serious studies have been conducting about the smile expression and/or recognition or classification. Philips presented the classification of smile patterns to identify various smile patterns in dentistry [1]. The identification is important in order to get the common understanding among dentists in dental cure and oral surgery, especially in improving patient’s smile or orthodontic rehabilitation. Hence, some complicated oral facial surgeries, periodontal fractions, traumatic and tooth fractures can be easily grouped and indexed. This technology will be very useful for patients before and after surgery. Hence, dentists have been using it as an important foundation in order to have common perception in dental cure and in conducting oral surgery likewise orthodontic. Furthermore, it can be used to easily group and index some complicated problems such as periodontal furcations, traumatic tooth fractures and complicated oral facial surgeries. Whitehill et al. pointed out that the smile expression of a face, in the form of an image and a video, can be detected [2]. They used the DFAT Dataset as training sample and GENKI as test sample. The classification was using Gentleboost which was able to yield detection correctness of 87.5%. Meanwhile, Wojdel and Rothkrantz proposed the combination of fuzzy system and artificial neural network to recognize the expression of oral features [3]. They used 3 different expressive conditions, i.e., smile, sad, and normal. The edge detection was applied for oral part. Fuzzy system was being used to get the inner lip gradient direction and intensity in the form of vector, while the neural network back propagation worked for the classification process.

This paper compares the two feature extraction techniques, 2DPCA (Two Dimensional Principal Component Analysis) and 2DLDA (Two Dimensional Linear Discriminant Analysis) based on smile stages classification. Before them, LDA (Linear Discriminant Analysis) [4, 5] is a well known scheme for feature extraction and dimension reduction. It has been used widely in many applications such as face recognition [6], image retrieval [7], micro array data classification [8], etc. The objective of LDA is to find the optimal projection (transformation) so that the ratio of the determinants of the between class and the within class scatter matrices of the projected samples reaches its maximum. A difficulty in using LDA is the very high dimensional nature of the image vector. The traditional solution to this problem is to utilize PCA (Principal Component Analysis) as a preprocessing step aiming to reduce the dimensionality of the vector space [6]. After all the image vectors are projected into subspace consisting of the principal component, the LDA algorithm can perform well in the subspace. However, since the projection criterion of the PCA and that of the LDA are essentially different, the preprocessing procedure to reduce the dimensionality using the PCA could result in the loss of some important discriminatory information for the LDA algorithm that follows the PCA. Actually, Chen et al. [9] have shown that the null space of the within class scatter matrix contains...
valuable discriminatory information. In view of this, called direct LDA (DLDA) algorithm [9,10] have been proposed to avoid the possible loss of useful information.

An alternative way to handle the above problem is to directly project the image matrix under a specific projection criterion, rather than using the stretched image vector. Yang et al. [11] have shown that a “two dimensional” PCA (2DPCA) can be constructed in a straightforward manner based on the image matrix projection. The size of the scatter matrices for 2DPCA scheme is either only m x m or n x n for an image of size m x n, instead of the size mn x mn in the classic PCA scheme. Therefore, the 2DPCA scheme is much faster in computation than the conventional PCA algorithm [12]. Meanwhile, Cahyono et al. had presented the classification of smile stages using PCA, LDA, and SVM. Unfortunately, the results were not maximal and they needed more computational time [13].

In this research, 2DPCA, a straightforward image projection technique, is developed for the image feature extraction. However, like PCA and 2DPCA are only good at image representation rather than discrimination. To overcome the shortcoming in 2DPCA and meanwhile to solve the (Singular Small Sample) SSS problem in LDA based algorithm, 2DLDA is proposed. 2DLDA overcomes the singularity problem implicitly, while achieving efficiency. The key difference between LDA and 2DLDA lies in the model for data representation [14]. LDA works with data in vectorized representations of data, while 2DLDA algorithm works with data in matrix representation. Extensive experimental work shows the 2DLDA framework outperforms higher than 2DPCA, PCA + LDA, and PCA.

2. SYSTEM OVERVIEW

The proposed system is shown in Fig. 1, that are consists of three modules: image preprocessing, feature extraction, and classification. The entire system flows are briefly described as follows. The first module employs smiling faces data. The data is manually cropped against a face area and produces spatial coordinate [5.90816 34.0714 39.3877 15.1020]. The coordinate is being employed as a reference for the automatically cropping process against all other face data. Next, first experiment the feature extraction performs 2DPCA to compute the projection matrices to reduce the dimension of smile images and generate the smile feature vectors. Second experiment the feature extraction performs 2DLDA to find the optimal projection direction so that the projected vectors in the m dimensional space reaches its maximum class separability. Finally, the classification employs SVM classifier with method one against one and all against all to two feature extractions before it.

3. THE ALGORITHM

3.1. 2D-PCA (Two Dimensional Principal Component Analysis)

Suppose that there are N training mouth images denoted by m x n matrices $A_i$ (i = 1, 2, ..., N). Yang et al. [11] presented the image covariance matrix $G_i$ as

$$G_i = \frac{1}{N} \sum_{i=1}^{N} (A_i - \bar{A})^T (A_i - \bar{A})$$  \hspace{1cm} (1)

where $\bar{A}$ is the mean image of all training images:

$$\bar{A} = \frac{1}{N} \sum_{i=1}^{N} A_i$$  \hspace{1cm} (2)

The optimal projection axis $X_{opt}$ composes of the eigenvectors of $G_i$ corresponding to the first d largest eigenvalues. Then the optimal projection axis, $X_1, X_2, ..., X_d$ are used for feature extraction. For a given image sample $A$,

$$Y_k = A X_k \hspace{1cm} k = 1, 2, ..., d$$  \hspace{1cm} (3)

$Y_1, ..., Y_d$ are called the principal component (vectors) of the sample image $A$. The principal component vectors obtained are used to form an m x d matrix $B = \{Y_1, ..., Y_d\}$, which is called the feature matrix or feature image of the image sample $A$.

3.2.2. DLDA (Two Dimension Linear Discriminant Analysis)
Discriminant Analysis)

The matrix representation in 2DLDA leads to an eigen-decomposition on matrices with much smaller sizes, more specifically with sizes $r \times r$ and $c \times c$, which are much smaller than the matrices in LDA. This dramatically reduces the time and space complexities of 2DLDA over LDA. As follows the algorithm 2DLDA:

Input: $A_1, \ldots, A_n, L, l_1, l_2$

Output: $L, R, B_1, \ldots, B_n$

1. Compute the mean $M_i$ of $i$th class for each $i$
   
   $$M_i = \frac{1}{n_i} \sum_{x \in c_i} X$$

2. Compute the global mean
   
   $$M = \frac{1}{n} \sum_{j=1}^{n} X$$

3. $R_0 = (I_L, 0)^T$

4. For $j$ from 1 to $I$

5. $S_w^{R} \leftarrow \sum_{i=1}^{k} (X - M_i)R_{j-1}^{T}R_{j-1}(X - M_i)^T$

6. Compute the first $l_1$ eigenvectors
   
   $\{\phi^L_{i1}, \ldots, \phi^L_{il_1}\}$

7. $L_i \leftarrow [\phi^L_{i1}, \ldots, \phi^L_{il_1}]$

8. $S_w^{L} \leftarrow \sum_{i=1}^{k} (X - M_i)^T L_i L_i^T (X - M_i)$

9. Compute the first $l_2$ eigenvectors $\{\phi^R_{i1}, \ldots, \phi^R_{il_2}\}$

10. $R_0 = (I_L, 0)^T$

11. End For

12. $L \leftarrow L_1, R \leftarrow R_1$

13. $B_i \leftarrow L_i R_i$, for $i = 1, \ldots, n$

14. Return $(L, R, B_1, \ldots, B_n)$

3.3. Multiclass Classification Using Support Vector Machines (SVM)

The technique of SVMs, first proposed in the late seventies [15] [16] is receiving increasing attention. The main idea of this technique is to construct a decision surface that lead to the larger separation margin between positive and negative regions. Traditional techniques such as multilayer perceptron neural networks try to minimize the empirical risk, (frequency of errors made by the learning machine on the training samples set). On the other side, the SVM technique searches for structural risk minimization, that implies the realization of the best generalization performance by matching the machine capacity to the available training data for the problem at hand. Therefore, the goal of this technique is to find, among the networks with the minimum training error, the simpler one (the one with the least complexity).

Originally this technique was proposed to perform binary classification, so extension methods are necessary to make it possible to deal with multiple classes. Therefore for this goal: combining several binary classifiers (one against one, one against all, DAG, among others). In this research, for comparison one against one method and one against all method.

3.3.1 One Against One

This method was introduced by Knerr [17], and later, a modification named Max Wins strategy was proposed by Friedman [18]. If $n$ is the number of classes, it’s necessary to train one binary classifier for each of the possible two class combinations. This procedure will generate $n(n-1)/2$ binary classifiers.

In the testing stage, the unknown sample $x$ is submitted to all binary classifiers. Each classifier decides in favor of one or other class; then we predict $x$ in the class with the largest vote (max wins strategy). The advantages of this method are its easy understanding and implementation, besides a good performance. The disadvantage is the huge number of binary classifiers, which means a great memory load. Furthermore, there is a great computational load in the test step because each unknown sample must be submitted to all of the $n(n-1)/2$ binary classifiers [20]. Method of one against one is using 3 SVM biner, as shown in table 1 and Fig. 2.

<table>
<thead>
<tr>
<th>Table 1. The 3 SVM binaries using one against one method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1 = 1$</td>
</tr>
<tr>
<td>Class 1</td>
</tr>
<tr>
<td>Class 1</td>
</tr>
<tr>
<td>Class 2</td>
</tr>
</tbody>
</table>
3.3.2 One Against All

This was probably the earliest used implementation for SVM multi-class classification [19]. If \( n \) is the number of classes, this method constructs \( n \) SVM models. The \( i \)th SVM is trained with all of the samples of the \( i \)th class with positive labels, and all other examples with negative labels. Again, in the testing step, the unknown sample \( x \) is submitted to all of the \( n \) classifiers. In general, only one of the classifiers will give a positive value for the separate class, and this is the classification criterion. In some cases, it possible that more than one classifier give a positive output for the separated class; in this case, the one with the highest output is selected. The advantage of this method is the small number of classifiers, leading to memory savings and faster classification. However, the training stage is very time consuming, because each SVM must be trained with all training samples [bics]. Method of one against all is using 3 SVM biner, as shown in table 2 and Fig. 3.

Table 2. The 3 SVM binaries using one against all method

<table>
<thead>
<tr>
<th>( y_I = 1 )</th>
<th>( y_I = -1 )</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>No Class 1</td>
<td>( f^1(x) = (w^1)x + b^1 )</td>
</tr>
<tr>
<td>Class 2</td>
<td>No Class 2</td>
<td>( f^2(x) = (w^2)x + b^2 )</td>
</tr>
<tr>
<td>Class 3</td>
<td>No Class 3</td>
<td>( f^3(x) = (w^3)x + b^3 )</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSIONS

For the experiment, the proposed 2DPCA and 2DLDMA method uses the 90 smiling faces data, i.e., 30 data for Stage I, 30 data for Stage III, and 30 data for Stage IV (see Fig. 4 and Fig. 5). The data itself has been validated by dentist specializing in tooth conservation. Initially, every face data has the size of 50x50 pixels at each stage. It is manually cropped against a face data at oral area. This process causes the face data size reduction into 40x16 pixels.

Figure 4. Visualization of smile at stage I

As the method of three-fold cross validation is being applied, data at each stage is divided into 3 groups. The first 2/3 data (20 data) becomes the training data, while the last 1/3 data (10 data) work as the testing data. Those groups are being rotated with no overlap, thus all of them have the experience of becoming testing data.

Figure 5. Classification of smile for grouping and indexing when a patient is in preoperative and postoperative respectively [3]

Figure 6. Comparison 2DPCA+SVM 1vs1 and 2DPCA+SVM 1vsAll.
Comparison 2DLDA + SVM 1 vs 1 and 2DLDA + SVM 1 vs all

Figure 7. Comparison 2DLDA+SVM 1vs1 and 2DLDA+SVM 1vsAll.

The experimental evaluation results show that the 2DLDA + SVM one against all outperform the others significantly. In the experiments with those three groups of data then calculate the average of three groups, shown in Table 3 and Fig. 7 the average of classification accuracy 2DPCA+SVM one against all 94.44%, while 2DLDA+SVM one against one only gives 91.11% from 20 largest eigenvalues. Meanwhile, shown in Table 3 and Fig. 7 the 2DLDA+SVM one against all can achieve the average of classification accuracy 96.67%, while 2DLDA+SVM one against one only gives 95.56% at dimension of row projection (p) = 14 and dimension of column projection (q) = 3. As listed in Table 3, 2DPCA+SVM 1vs1 needs less time than the others.

Table 3. Comparison between 2DPCA, 2DLDA and SVM Multiclass

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Average of Classification Accuracy (%)</th>
<th>Time of feature extraction + classification (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+SVM 1vs1*</td>
<td>83.33</td>
<td>57.91</td>
</tr>
<tr>
<td>PCA+LDA+SVM 1vs1*</td>
<td>94.44</td>
<td>76.87</td>
</tr>
<tr>
<td>2DPCA+SVM 1vs1**</td>
<td>91.11</td>
<td>2.71</td>
</tr>
<tr>
<td>2DPCA+SVM 1vsAll*</td>
<td>94.44</td>
<td>2.52</td>
</tr>
<tr>
<td>2DLDA+SVM 1vs1**</td>
<td>95.56</td>
<td>1.39</td>
</tr>
<tr>
<td>2DLDA+SVM 1vsAll**</td>
<td>96.67</td>
<td>1.07</td>
</tr>
</tbody>
</table>

* Cahyono et al. [5]
** Proposed system

5. CONCLUSIONS

In this paper, we discussed classification of smile stages based on 2DPCA for feature extraction and then compared the results with the 2DLDA to distinguish the smiling faces at Stage I, Stage III and Stage IV. After implementing the feature extraction and SVM multiclass techniques for smile face data, the important advantages of 2DPCA + SVM 1 vs all over the others are proven. Firstly, it is simpler and more straightforward to use for image feature extraction since it is based on the image matrix. 2DPCA + SVM one against all has better classification accuracy in all experiments. Table 3 shows that 2DPCA + SVM one against all could give 96.67% while 2DLDA + SVM one against one can only give 95.56% of accuracy. Meanwhile, 2DPCA + SVM one against all could give 94.44% while 2DPCA + SVM one against one can only give 91.11% of accuracy. Finally, the 2DPCA + SVM one against all is computationally more efficient than the others and it can improve the speed of image feature extraction and classification significantly. Table 3 also shows that 2DPCA + SVM one against all only needs 1.07s, while 2DPCA + SVM one against all needs 1.39s of computation time. Meanwhile, shows that 2DPCA + SVM one against all only needs 2.52s, while 2DLDA + SVM one against all needs 2.71s of computation time.

REFERENCES

16. V. Vapnic, Estimation of dependencies based on empirical data, Nauka, Moscow, Russia, 1979.