Combining Local And Global Features For Face Recognition

J Huang*#, P C Yuen*, J H Lai# and C H Li*

*Department of Computer Science
Hong Kong Baptist University

#Department of Mathematics
Zhongshan University China

Abstract

This paper addresses the problem of face recognition using combined classifier. The multi-expert approach has been used in the last two decades and proved to be a proper way to improve recognition performance. One of the crucial problems in this approach is how to combine the output of each classifier to draw a final conclusion, as the scale of each classifier is different. In turn, normalization is required. Three methods are proposed in this paper. They are linear method, exponential method and linear-exponential method, in normalizing and combining the output of each classifier. We have performed experiments in combining three global features face recognition algorithms, namely principal component analysis, independent component analysis, spectroface, and one local feature algorithm, namely Gabor wavelet. Four combination rules proposed by Kittler et al. [3] are employed to evaluate the performance. The face database from Yale University is used. Experimental results show that exponential normalization method and linear-exponential normalization method with sum rule give the best performance.

Keywords: Face recognition, combining classifier

1. Introduction

Face recognition research has been started in the late 70's and is one of the active and exciting researches in computer science and information technology areas since 1990. Basically, there are two major approaches in automatic recognition of faces by computer [1], namely, constituent-based recognition and face-based recognition.

A number of face recognition algorithms have been developed in the last decade. The common approach is to develop a single, sophisticated and complex algorithm to handle one or some of the face variations. However, developing a single algorithm to handle all variations is not easy. As such, this paper proposes to employ the theory of combining classifier [6] that make use of different types of features and classifiers to draw a conclusion.

2. Brief Review on Existing Methods

2.1 Review on Classifier Combination

In this paper, four combination rules proposed by Kittler et al. [3] are employed to evaluate the performance. The four classifier combination schemes are Product Rule, Sum Rule, Min Rule, Max Rule.

2.2 Review on Existing Face Recognition Methods

We have used three global features face recognition algorithms, namely principal component analysis, independent component analysis, Spectroface, and one local feature algorithm, namely Local Gabor wavelet.

PCA is used to find a low dimensional representation of data. For some important details of PCA you can refer to [5]. Independent

Component Analysis is a statistical signal processing technique. The concept of ICA can be seen as a generalization of the PCA, which only impose independence up to the second order. The basic idea of ICA is to represent a set of random variables using basis functions where the components are statistically independent or as independent as possible [7]. Spectroface method combines the wavelet transform and the Fourier transform for feature extraction [4].

In Local Gabor Wavelet method, we use 15 points on important landmark on the face for recognition. Figure 1 shows the 15 points used on the face image. These points lie in the corner or none smooth places of important landmarks on the face image, so these locations contain more information than other points.

Figure 1: Points on the face image

3. Proposed method

We have reviewed four popular face classifiers, and output of each classifier has different scales. Spectroface, PCA and ICA use distance measurement for classification, and local Gabor wavelet use similarity measurement for classification. To combine the four classifiers, the distance measurement and the similarity measurement from the outputs of different classifiers should be normalized to the same scale. Transformation method is proposed to solve the problem. In addition, these transforms must not affect the order of the transformed classifier ranking. So these transform should be monotone functions.

3.1 Two basic transformations for scale normalization

Ackermann and Bunke [6] presented some basis transformation for scores, the following two transformations have been proposed in [6]. Consider the following problem, the original data are in the range of *DataIn*=[α_1, α_2], and we want to convert them to the range of *DataOut*=[β_1 , β_2]. We can use Linear transformation with the following equation

$$DataOut = \beta_1 + \frac{(DataIn - \alpha_1)}{(\alpha_2 - \alpha_1)} * (\beta_2 - \beta_1) \cdot \quad (1)$$

A logistic transformation can be performed with the following steps. First, use the linear transformation (1) to convert the input data into scope S=[0.0, 100.0]. Then a logistic transformation is given by,

$$S^1 = \frac{\exp(\alpha + \beta S)}{1 + \exp(\alpha + \beta S)} \cdot \quad (2)$$

Generally, the parameters α >0 and β >0 can be determined empirically. These two parameters control intersection with the X-axis and slope respectively. These two basic transformations will be used in our developed methods in the following.

3.2 Proposed Normalization Methods

To solve the combining problem, three methods are proposed for converting the distance measurement to similarity measurement (or probability estimate) with scale normalization.

We denote the distance between pattern Z_i and the training sample Z_j be d_{ij} . S_{ij} is the similarity between them and p_{ij} is the estimated probability that pattern Z_i belongs to the class of training sample Z_j. We denote the mean value of the squares of all distances as σ, the mean of all the distances as σ^I:

$$\sigma = \frac{\sum_{i,j} d_{ij}^2}{N} \quad (3)$$

$$\sigma^I = \frac{\sum_{i,j} d_{ij}}{N}, \quad (4)$$

where N is the total number of the distances.

3.2.1 Method 1: Linear Normalization Method

The linear normalization method consists of two steps. First, we use the linear transformation to convert the input data $d_{ij} \in [\alpha_1, \alpha_2]$ into output data scope [β_1=0.0, β_2=100.0]. From (1), we can get:

$$d_{ij}^I = \frac{d_{ij} - \alpha_1}{\alpha_2 - \alpha_1} * 100, \quad (5)$$

By substituting (5) into (2), we have:

$$d_{ij}^{II} = \frac{\exp(\alpha + \beta d_{ij}^I)}{1 + \exp(\alpha + \beta d_{ij}^I)}. \quad (6)$$

As we know that the similarity between two patterns is reverse ratio to the distance between them. So an inverse relationship can be denoted as the following:

$$Similarity = \frac{1}{distanace}. \quad (7)$$

Substitutes (6) into (7), we get:

$$S_{ij} = \frac{1 + \exp(\alpha + \beta d_{ij}^I)}{\exp(\alpha + \beta d_{ij}^I)}. \quad (8)$$

We can see that S_{ij} is inversely proportional to d_{ij}. But if $\exp(\alpha + \beta d_{ij})$ is big enough, then S_{ij} would be almost the same. So we should select the parameters α, β carefully. In our experiments, we find that it is difficult to find appropriate parameters α, β, because we do not know the scale of each output. So we updated it as follows, First, we convert d_{ij} in equation (5) into scope [0.0,10.0], then substitute (5) into (4), we get:

$$\sigma^U = \frac{\sum_{i,j} d_{ij}^I}{N}. \quad (9)$$

Second, we compute the similarity as follows,

$$S_{ij}^U = \frac{\exp(\sigma^U)}{\exp(\sigma^U) + \exp(\alpha + \beta d_{ij}^I)}. \quad (10)$$

Here, we convert d_{ij} into scope [0.0, 10.0] because we do not want $\exp(\sigma^U)$ to be too large. Now, we can select the parameter α, β easily. For example α=0, β=1.

We can also normalize the similarity measurement to estimate probability measurement. This is done in the following manner. Uses the linear transformation (1) to convert $S_{ij}^U \in [S1, S2]$ into scope [0.0, 1.0], we have:

$$p_{ij} = \frac{S_{ij}^U - S1}{S2 - S1}. \quad (11)$$

3.2.2 Method 2: Exponential Normalization Method

In statistical view, we can compute the similarity like the following:

$$S_{ij}^I = \frac{1}{\exp(\frac{d_{ij}^2}{\sigma})} \quad (12)$$

σ is mean square distances. As $\frac{d_{ij}^2}{\sigma}>0$, so

$0<S'_{ij}<1$, and S'_{ij} is inversely related to d_{ij}.

We can also convert the similarity measurement to estimated probability measurement.

If $S'_{ij}\in[S1', S2']$, using (1), we have:

$$p'_{ij}=\frac{S'_{ij}-S1'}{S2'-S1'}\ . \quad (13)$$

3.2.3 Method 3: Linear Exponential Normalization Method

In the same way, we can compute the similarity like the following:

$$S''_{ij}=\frac{1}{\exp(\frac{d_{ij}}{\sigma'})}\ . \quad (14)$$

σ' is the mean of the distances, as $\frac{d_{ij}}{\sigma'}>0$,

so $0<S''_{ij}<1$, S''_{ij} is inversely related to $\frac{d_{ij}}{\sigma'}$.

We can also convert the similarity measurement to estimated probability measurement.

If $S''_{ij}\in[S1'', S2'']$, use (1), we have:

$$p''_{ij}=\frac{S''_{ij}-S1''}{S2''-S1''}\ . \quad (15)$$

4. Experimental Results

Face images from Yale University are used for evaluating the proposed method. In Yale database, there are 15 persons and each person consists of 11different facial expressions, illumination and small occlusion (by glasses). And the resolution of all images is 128×128. One of the individual's face images is shown in Figures 2.

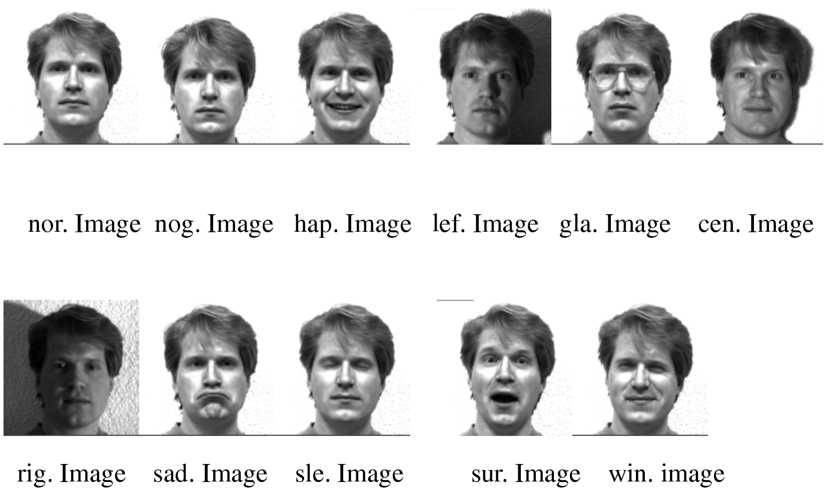

Figure 2: images of one person from Yale database

4.1.1 Results on Yale database

In this experiment, only the normal (nor.) images are used for training and all other images are used (not including nog. image and hap. image) for testing. Light variation images (left (lef.), right (rig.) and glass (gla.) are included in testing samples. Table 1 shows the results on Yale database. It can be seen that the highest accuracy is 87.5% using the spectroface method.

	Rank1 (%)	Rank2 (%)	Rank3 (%)
Spectroface	87.5000	93.3333	94.1667
PCA	73.3333	78.3333	85.8333
ICA	70.0000	79.1667	85.0000
Local Gabor wavelet	71.6667	77.5000	80.0000

Table 1: Results on original Yale database

Now let us see the results on combining classifiers. Figures 1-3 show the results of linear normalization method (method 1), exponential normalization method (method 2) and linear exponential normalization method (method 3). For every table, we will list three ranks accuracy for four rules of similarity measurement and estimated probability measurement. It can be seen that, for all three normalization methods, sum rule gives the best results. The rank 1 accuracy for linear normalization, exponential normalization and linear exponential normalization are 93.33%, 94.16% and 94.16% using sum rule. Comparing the best result obtained by Spectroface as shown in Table 1, there is 6% - 7% increment. This also demonstrates that combining different classifiers will provide better results.

We have also performed similar experiment and analysis on Olivetti database. Similar results are obtained. Owing to the limited space, the figures cannot be included.

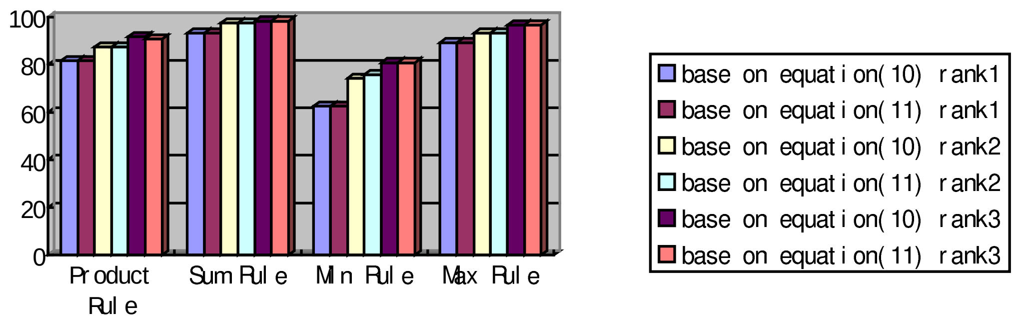

Figure1: Based on equation (10), (11)

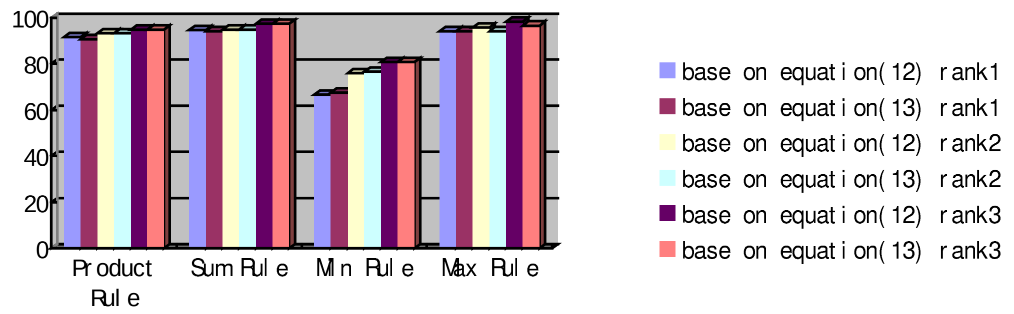

Figure 2: Based on equation (12), (13)

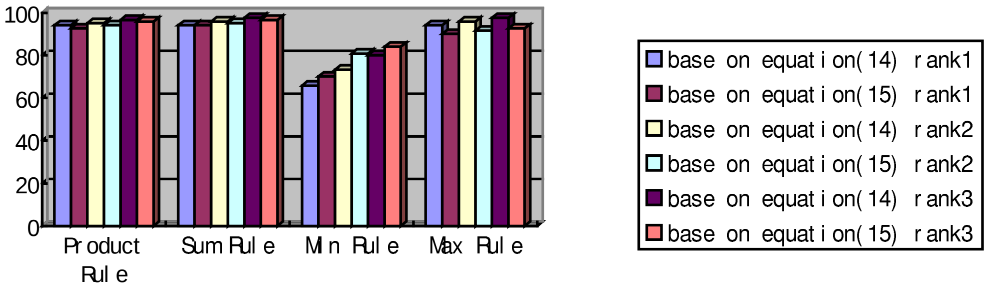

Figure 3: Based on equation (14), (15)

5. Conclusions

In this paper, we have proposed three normalization methods for combining classifiers. We have also demonstrated that combining local and global features performs better than either local or global feature is used. Moreover, the following two conclusions can be drawn in this paper,

1. Exponential normalization method and linear exponential normalization method have better performance than linear normalization method for all combining strategies.
2. Global and local features have been tested with various kinds of combining strategy. Either exponential normalization method or linear exponential normalization method with Sum Rule to combining global and local features gives the best performance.

Acknowledgement

This project was supported by the Science Faculty Research Grant, Hong Kong Baptist University. We would like to thank Yale University for providing the face image databases.

References

1. R Chellappa, C L Wilson and S Sirohey, "Human and machines recognition of faces: A survey", *Proceeding of The IEEE*, Vol. 83, No.5, pp. 705-740, 1995.
2. L Wiskott, J Fellous, and C von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 775-779, 1997.
3. J Kittler, M Hataf, R Duin, and J Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp.226- 239, 1998.
4. J H Lai, P C Yuen and G C Feng, "Face recognition using holistic Fourier invariant features", *Pattern Recognition*, Vol.34, No.1, pp.95-109, 2001.
5. G C Feng, P C Yuen and D Q Dai, "Human face recognition using PCA on Wavelet subband", *SPIE Journal of Electronic Imaging*, Vol. 9, No.2 pp.226-233, 2000.
6. B Ackermann, H Bunke, "Combination of Classifiers on the Decision Level for Face Recognition", *Technical Report IAM-96-002*, Institut für Informatik und angewandte Mathematik, Universität Bern, 1996.
7. P C Yuen and J H Lai, "Face representation using independent component analysis", *Pattern Recognition*, accepted for publication, March 2001.