Research Article

Fusion of PCA-Based and LDA-Based Similarity Measures for Face Verification

Mohammad T. Sadeghi,[1] Masoumeh Samiei,[1] and Josef Kittler[2]

[1] *Signal Processing Research Group, Department of Electrical and Computer Engineering, Yazd University,*
P.O. Box 89195-741, Yazd, Iran
[2] *Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

Correspondence should be addressed to Mohammad T. Sadeghi, m.sadeghi@yazduni.ac.ir

The problem of fusing similarity measure-based classifiers is considered in the context of face verification. The performance of face verification systems using different similarity measures in two well-known appearance-based representation spaces, namely Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is experimentally studied. The study is performed for both manually and automatically registered face images. The experimental results confirm that our optimised Gradient Direction (GD) metric within the LDA feature space outperforms the other adopted metrics. Different methods of selection and fusion of the similarity measure-based classifiers are then examined. The experimental results demonstrate that the combined classifiers outperform any individual verification algorithm. In our studies, the Support Vector Machines (SVMs) and Weighted Averaging of similarity measures appear to be the best fusion rules. Another interesting achievement of the work is that although features derived from the LDA approach lead to better results than those of the PCA algorithm for all the adopted scoring functions, fusing the PCA- and LDA-based scores improves the performance of the system.

1. Introduction

In spite of the rapid advances in machine learning, in many pattern recognition problems, the decision making is based on simple concepts such as distance from or similarity to some reference patterns. This type of approach is particularly relevant when the number of training samples available to model a class of objects is very limited. Examples of such situations include content-based retrieval from image or video databases, where the query image is the only sample at our disposal to define the object model, or biometrics where only one or a few biometric traits can be acquired during subject enrolment to create a reference template. In biometric identity verification, a similarity function measures the degree of similarity of an unknown pattern to the claimed identity template. If the degree exceeds a pre-specified threshold, the unknown pattern is accepted to be the same as the claimed identity. Otherwise, it is rejected.

Different similarity measures have been adopted in different machine vision applications. In [1], a number of commonly used similarity measures including the City-block, Euclidean, Normalised Correlation (NC), Chi-square (χ^2), and Chebyshev distance have been considered in an image retrieval system. The reported experimental results demonstrate that the City-block and Chi-square metrics are more efficient in terms of both retrieval accuracy and retrieval efficiency. In a similar comparative study, it has been shown that the Chi-square statistics measure outperforms the other similarity measures for remote sensing image retrieval [2]. In another study, the effect of 14 scoring functions such as the City-block, Euclidean, NC, Canberra, Chebyshev, and Distance based Correlation Coefficients has been studied in the context of the face recognition problem [3] in the PCA space. It has been shown that a simplified form of Mahalanobis distance outperforms the other metrics. In [4], four classical distance measures, City-block, Euclidean, Normalised Correlation, and Mahalanobis

distance have been compared in the PCA space. It has been shown that when the number of eigenvectors is relatively high, the Mahalanobis distance outperforms the other measures. Otherwise, a similar performance is achieved using different measures. It has been also propounded that no significant improvement is achieved by combining the distance measures.

A similarity score is computed in a suitable feature space. Commonly, similarity would be quantised in terms of a distance function, on the grounds that similar patterns will lie physically close to each other. Thus, the smaller the distance, the greater the similarity of two entities. The role of the feature space in similarity measurement is multifold. First of all, the feature space is selected so as to maximise the discriminatory information content of the data projected into the feature space and to remove any redundancy. However, additional benefits sought after from mapping the original pattern data into a feature space is to simplify the similarity measure deployed for decision making.

PCA and LDA are two classical tools widely used in the appearance-based approaches for dimensionality reduction and feature extraction. Many face recognition methods, such as eigenfaces [5] and fisherfaces [6], are built on these two techniques or their variants. Different researches show that in solving the pattern classification problems the LDA-based algorithms outperform the PCA-based ones, since the former take the between classes variations into account. The LDA is a powerful feature extraction tool for pattern recognition in general and for face recognition in particular. It was introduced to this application area by Belhumeur et al. in 1997 [6]. An important contributing factor in the performance of a face authentication system is the metric used for defining a matching score. Theoretically, Euclidean distance provides an optimal measure in the LDA space. In [7], it has been demonstrated that it is outperformed by the Normalised Correlation (NC) and Gradient Direction (GD). Also, in [8], the performance of the NC scoring function has been compared with the GD metric. The study has been performed on the BANCA database [9] using internationally agreed experimental protocols by applying a geometric face registration method based on manually or automatically annotated eyes positions. It has been concluded that overall the NC function is less sensitive to missregistration error but in certain conditions GD metric performs better. However, in [10], it has been further demonstrated that by optimising the GD metric, this metric almost always outperforms the NC metric for both manually and automatically registered data.

In this study, a variety of other metrics have been investigated, including Euclidean, City-block, Chebyshev, Canberra, Chi-square (χ^2), NC, GD, and Correlation coefficient-based distance. The experimental results in face verification confirm that, individually, other metrics on the whole do not perform as well as the NC and GD metrics in the LDA space. However, in different conditions, certain classifiers can deliver a better performance.

It is well known that a combination of many different classifiers can improve classification accuracy. Various schemes have been proposed for combining multiple classifiers. We concentrate on classifier combination at the decision-level, that is, combining similarity scores output by individual classifiers. Thus, the scores are treated as features, and a second-level classifier is constructed to fuse these scores.

Fusion rules can be divided into two main categories: fixed rules such as the sum, product, minimum, maximum, and median rule [11–13] and trained rules like the weighted averaging of classifiers outputs [14, 15], Support Vector Machines (SVM) [10], bagging, and boosting [16]. Overall, the fixed rules are most often used because of their simplicity and the fact that they do not require any training. Accordingly, equal weights are used for all the classifiers [11, 17].

However, in many studies it has been demonstrated that trained classifiers such as Support Vector Machines (SVMs) have the potential to outperform the simple fusion rules, especially when enough training data is available. In [18], AdaBoost has been adopted for combining unimodal features extracted from face and speech signals of individuals in multimodal biometrics. In [8] the fusion problem was solved by selecting the best classifier or a group of classifiers dynamically with the help of a gating function learnt for each similarity measure.

In summary, it is clear that it is still pertinent to ask which classifiers provide useful information and how the expert scores should be fused to achieve the best possible performance of the face verification system. In [19], considering a set of similarity measure-based classifiers within the LDA feature space, a sequential search algorithm was applied in order to find an optimum subset of similarity measures to be fused as a basis for decision making. The SVM classifier was used for fusing the selected classifiers.

In this paper, a variety of fixed and trained fusion rules are compared in the context of face authentication. Five fixed fusion rules (sum, min, max, median, and product) and two trained rules (the support vector machines and weighted averaging of scores) are considered. It is shown that a better performance is obtained by fusing the classifiers. Moreover, the adopted trained rules outperform the fixed rule. Although, the PCA-based classifiers perform nearly 3 times worse than the LDA-based one, an interesting finding of this paper compared to our previous work [19] is that the performance of the verification system can be further improved by fusing the LDA- and PCA-based classifiers. In [20], a similar study has been performed using Euclidean distance as the scoring function. In the training stage of the proposed algorithm, adopting a fixed reference as the central value of the decision making threshold, client specific weights are determined by calculating the average value of the Euclidean distance of all the patterns from each client template. The client specific weights are determined in both LDA and PCA spaces. The weights are then used within the framework of three simple untrained fusion rules. In the adopted experimental protocol, each subject images are divided into two parts as the training and test sets. The experimental study performed on the ORL and Yale data sets demonstrate that the combined classifier outperforms the individual PCA- and LDA-based classifiers [20]. Although the training and test images are different, since the same subjects are available within the training and test sets, the

weighting process is somehow biased so that the performance of the system in the presence of new impostors (not those used for training) could be worse.

The rest of the paper is organised as follows. In the next section, the adopted scoring functions are introduced. Fusion rules are reviewed in Section 3. A description of the experimental design including the face database used in the study, the experimental protocols, and the experimental setup are given in Section 4. The experimental results using the adopted scoring functions and the fusion results are presented and discussed in Section 5. Finally a summary of the main findings and conclusions can be found in Section 6.

2. Similarity Functions

In a similarity measure-based face verification system, a matching scheme measures the similarity or distance of the test sample, $\mathbf{x}$, to the template of the claimed identity, $\boldsymbol{\mu}_i$, both projected into an appropriate feature space. The general form of a group of similarity measures which is called *Minkowski Distance* or *power norm metrics* (L_p) is defined as

$$s_M = \left[\sum_{j=1}^{m} \left| \boldsymbol{\mu}_{ij} - \mathbf{x}_j \right|^p \right]^{1/p}, \tag{1}$$

where m is the dimensionality and j indexes the components of the two vectors.

The most commonly used similarity measures, *Manhattan* or *City-block* metric, *Euclidean* Distance (ED), and *Chebyshev* Distance are special cases of the Minkowski metric for $p = 1$, $p = 2$, and $p \to \infty$, respectively, that is, L_1, L_2, and L_∞ metrics:

$$s_{\text{City}} = \sum_{j=1}^{m} \left| \boldsymbol{\mu}_{ij} - \mathbf{x}_j \right|, \tag{2}$$

$$s_{\text{ED}} = \sqrt{\left(\mathbf{x} - \boldsymbol{\mu}_i\right)^T \left(\mathbf{x} - \boldsymbol{\mu}_i\right)}, \tag{3}$$

$$s_{\text{Cheby}} = \max_j \left| \boldsymbol{\mu}_{ij} - \mathbf{x}_j \right|. \tag{4}$$

The *Canberra* Distance is also given by

$$s_{\text{Canb}} = \sum_{j=1}^{m} \frac{\left| \boldsymbol{\mu}_{ij} - \mathbf{x}_j \right|}{\left| \boldsymbol{\mu}_{ij} \right| + \left| \mathbf{x}_j \right|}. \tag{5}$$

This can be considered as the normalised Manhattan Distance. The *Chi-squared* (χ^2) Distance is defined by

$$s_{\chi^2} = \sum_{j=1}^{m} \frac{\left(\boldsymbol{\mu}_{ij} - \mathbf{x}_j \right)^2}{\left| \boldsymbol{\mu}_{ij} \right| + \left| \mathbf{x}_j \right|} \tag{6}$$

which is basically a relative Euclidean squared distance and is usually meant for nonnegative variables only.

In [7], it has been demonstrated that a matching score based on *Normalised Correlation* (NC) scoring function, defined by the following equation, is more efficient:

$$s_N = \frac{\left\| \mathbf{x}^T \boldsymbol{\mu}_i \right\|}{\sqrt{\mathbf{x}^T \mathbf{x} \boldsymbol{\mu}_i^T \boldsymbol{\mu}_i}}. \tag{7}$$

Another similarity measure which is conceptually the same as the NC function is the *Correlation Coefficients*-based distance. For more details, the reader is referred to [3].

The *Gradient Direction* (GD) metric proposed in [7, 21] measures the distance between a probe image and a model in the gradient direction of the a posteriori probability function $P(i \mid \mathbf{x})$ associated with the hypothesised client identity i. A mixture of Gaussian distributions with isotropic covariance matrix has been assumed as the density function representing the anticlass (world population) estimated from the data provided by all the other users (for all $j \neq i$). The diagonal elements of the isotropic covariance matrix are assumed to have values related to the magnitude of variation of the image data in the feature space. It was demonstrated that in a face verification system, applying GD metric is even more efficient than the NC function. This matching score is defined as

$$s_O = \frac{\left\| \left(\mathbf{x} - \boldsymbol{\mu}_i\right)^T \nabla_O P(i \mid \mathbf{x}) \right\|}{\left\| \nabla_O P(i \mid \mathbf{x}) \right\|}, \tag{8}$$

where $\nabla_O P(i \mid \mathbf{x})$ refers to the gradient direction. For the isotropic structure of the covariance matrix, that is, $\Sigma = \sigma \mathbf{I}$, the optimal direction would be

$$\nabla_I P(i \mid \mathbf{x}) = \sum_{\substack{j=1 \\ j \neq i}}^{m} p(\mathbf{x} \mid j)\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i\right). \tag{9}$$

Note that the magnitude of σ will affect the gradient direction through the values of density $p(\mathbf{x} \mid j)(\mathbf{x}j)$.

3. Similarity Scores Fusion

One of the very promising research directions in the field of pattern recognition and computer vision is classifier fusion. It has been recognised that the classical approach to designing a pattern recognition system which focuses on finding the best classifier has a serious drawback. Any complementary discriminatory information that other classifiers may capture is not tapped. Multiple expert fusion aims to make use of many different designs to improve the classification performance. In the case considered here, as different metrics span the feature space in different ways, it seems reasonable to expect that a better performance could be obtained by combining the resulting classifiers.

Since the scores for different classifiers lie in different ranges, a normalisation process is required to transform these score to the same range before combining them [22]. The simplest normalisation technique is the min-max normalisation. The min-max normalisation is best suited for

the case where the bounds (maximum and minimum values) of the scores produced by a matcher are known. In this case, we can easily shift the minimum and maximum scores to 0 and 1, respectively. Given a set of scores for each classifier s_i, $i = 1, 2, \ldots, M$, where M is the number of samples, the normalised scores are given by

$$S_i = \frac{s_i - \min_i}{\max_i - \min_i}, \tag{10}$$

where s_i and S_i are, respectively, the original and normalised scores associated to the ith sample. $\min_i$ and $\max_i$ are the minimum and maximum scores determined from a training set.

As mentioned earlier, two main groups of fusion rules, untrained (fixed) and trained rules can be applied for classifiers fusion. The untrained methods such as Sum (or Average), Product, Min, Max, and Median are very well known approaches. For example, the Sum rule is defined as

$$S_{\text{new}} = \sum_{i=1}^{M} S_i, \tag{11}$$

where M is the number of classifiers. This is simply equivalent to averaging the normalised scores over the classifiers. A variety of trained fusion techniques such as neural network classifier, Bayesian classifier, and SVM have been suggested. It has been shown that the SVM classifier is among the best trained fusion rules. In [10], decision level fusion strategy using the SVMs has been adopted for combining the similarity measure-based classifiers. A very good performance has been reported using the adopted method.

Another promising trained rule involves a weighted averaging of similarity scores. Obviously, the technique used for determining the weight is an important factor in such a method.

3.1. Support Vector Machines.

A Support Vector Machine is a two-class classifier showing superior performance to other methods in terms of Structural Risk Minimisation [23]. For a given training sample $\{\mathbf{x}_i, y_i\}$, $i = 1, \ldots, N$, where $\mathbf{x}_i \in R^D$ is the object marked with a label $y_i \in \{-1, 1\}$, it is necessary to find the direction $\mathbf{w}$ along which the margin between objects of two classes is maximal. Once this direction is found the decision function is determined by threshold b:

$$y(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b). \tag{12}$$

The threshold is usually chosen to provide equal distance to the closest objects of the two classes from the discriminant hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$, which is called the optimal hyperplane. When the classes are linearly nonseparable some objects can be shifted by a value δ_i towards the right class. This converts the original problem into one which exhibits linear separation. The parameters of the optimal hyperplane

and the optimal shifts can be found by solving the following quadratic programming problem:

$$\text{minimise } \mathbf{w} \cdot \mathbf{w} + \mathcal{C} \sum_{i=1}^{N} \delta_i$$

$$\text{subject to: } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \delta_i, \tag{13}$$

$$\delta_i \geq 0, \quad i = 1, \ldots, N,$$

where parameter $\mathcal{C}$ defines the penalty for shifting the objects that would otherwise be misclassified in the case of linearly nonseparable classes.

The QP problem is usually solved in a dual formulation:

$$\text{minimise } \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

$$\text{subject to: } \sum_{i=1}^{N} \alpha_i y_i = 0, \tag{14}$$

$$0 \leq \alpha_i \leq C \quad i = 1, \ldots, N.$$

Those training objects $\mathbf{x}_i$ with $\alpha_i > 0$ are called Support Vectors, because only they determine direction $\mathbf{w}$:

$$\mathbf{w} = \sum_{i=1, \, \alpha_i > 0}^{N} \alpha_i y_i \mathbf{x}_i \tag{15}$$

The dual QP problem can be rapidly solved by the Sequential Minimal Optimisation method, proposed by Platt [24]. This method exploits the presence of linear constraints in (14). The QP problem is iteratively decomposed into a series of one variable optimisation problems which can be solved analytically.

For the face verification problem, the size of the training set for clients is usually less than the one for impostors. In such a case, the class of impostors is represented better. Therefore, it is necessary to shift the optimal hyperplane towards the better represented class. In this paper, the size of the shift is determined in the evaluation step based on the Equal Error Rate criterion.

3.2. Weighted Averaging of Similarity Measures.

Compare to the simple averaging rule, in the case of weighted averaging, different weights are considered for the scores achieved from different classifiers, that is,

$$S_{\text{new}} = \sum_{i=1}^{M} w_i S_i, \tag{16}$$

where w_i is the weight assigned to the ith classifier output.

In this study, three methods of weighted averaging are considered. In the first group, each classifier weight is determined based on the performance of the classifier in an evaluation step. The smaller the error rate, the greater the weight assigned to the classifier output, that is,

$$w_i = \frac{1}{\text{TERE}_i}, \quad i = 1, 2, \ldots, M, \tag{17}$$

where TERE$_i$ is the Total Error Rate of the ith classifier in the Evaluation stage.

The main idea behind the second adopted method is to minimise the correlation between classifier outputs. In practise, outputs of multiple classifiers are not uncorrelated, but some classifiers are more correlated than others. Therefore, it is reasonable to assign different weights to different classifiers according to their correlation. Principle Component Analysis, PCA, is one of the statistical techniques frequently used to decorrelate the data [25]. Denote by $\mathbf{S}$ the vector of scores delivered by the M classifiers, that is,

$$\mathbf{S} = [S_1 \ S_2 \cdots . \ S_N].\qquad(18)$$

Let λ_i and $\mathbf{v}_i, i = 1,\ldots,M$, be the eigenvalues and eigenvectors of the covariance matrix of the evaluation score vectors $\mathbf{S}$ retaining a certain proportion of the score variance. The eigenvectors are used as the bases of a new feature space. Applying the simple averaging rule (equation (11)) to the scores transformed to this feature space is equivalent to the weighted averaging of the original scores in (16) where w_i are determined using the following equation:

$$w_i = \sum_{j=1}^{M} \mathbf{v}_{ij}.\qquad(19)$$

As the third method of weighted averaging of the scores, the above mentioned idea can be extended by applying the LDA algorithm. In a face verification system, two groups of score vectors are considered: client scores and impostor scores. In the evaluation step, these classes of data can be used within the framework of the Linear Discriminant Analysis (LDA) for computing the feature space bases and the classifier weights.

4. Experimental Design

In this section, the face verification experiments carried out on images of the BANCA database are described. The BANCA database is briefly introduced first. The main specification of the experimental setup is then presented.

4.1. BANCA Database. The BANCA database has been designed in order to test multimodal identity verification systems deploying different cameras in different scenarios (Controlled, Degraded, and Adverse). The database has been recorded in several languages in different countries. Our experiments were performed on the English section of the database. Each section contains 52 subjects (26 males and 26 females).

Each subject participated to 12 recording sessions in different conditions and with different cameras. Sessions 1–4 contain data under *Controlled* conditions whereas sessions 5–8 and 9–12 contain *Degraded* and *Adverse* scenarios, respectively. In order to create more independent experiments, images in each session have been divided into two groups of 26 subjects (13 males and 13 females). Experiments can be performed on each group separately.

In the BANCA protocol, 7 different distinct experimental configurations have been specified, namely, Matched Controlled (Mc), Matched Degraded (Md), Matched Adverse (Ma), Unmatched Degraded (Ud), Unmatched Adverse (Ua), Pooled test (P), and Grand test (G). Table 1 describes the usage of the different sessions in each configuration. "T" refers to the client training while "C" and "I" depict client and impostor test sessions, respectively.

4.2. Experimental Setup. The performance of different decision making methods discussed in Section 2 is experimentally evaluated on the BANCA database using the configurations discussed in the previous section. The evaluation is performed in the LDA and PCA spaces. The original resolution of the image data is 720×576. The experiments were performed with a relatively low resolution face images, namely, 64×49. The results reported in this paper have been obtained by applying a geometric face normalisation based on the eyes positions. The eyes positions were localised either manually or automatically. A fast method of face detection and eyes localisation was used for the automatic localisation of eyes centre [26]. The XM2VTS database [27] was used for calculating the LDA and PCA projection matrices.

The thresholds in the decision making system have been determined based on the Equal Error Rate criterion, that is, by the operating point where the false rejection rate (FRR) is equal to the false acceptance rate (FAR). The thresholds are set either globally (*GT*) or using the client specific thresholding (*CST*) technique [21]. In the training sessions of the BANCA database 5 client images per person are available. In the case of global thresholding method, all these images are used for training the clients template. The other group data is then used to set the threshold. In the case of the client specific thresholding strategy, only two images are used for the template training and the other three along with the other group data are used to determine the thresholds. Moreover, in order to increase the number of data used for training and to take the errors of the geometric normalisation into account, 24 additional face images per each image were generated by perturbing the location of the eyes position around the annotated positions.

In the previous studies [21], it has been demonstrated that the Client Specific Thresholding (CST) technique is superior in the matched scenario (Mc, Md, Ma, and G) whereas the Global Thresholding (GT) method gives a better performance on the unmatched protocols. The results reported in the next section using thresholding have been acquired using this criterion.

5. Experimental Results and Discussion

As mentioned earlier, in the GD metric, the impostor distributions have been approximated by isotropic Gaussian functions with a standard deviation of σ, that is, $\Sigma = \sigma I$. The order of σ is related to the order of the standard deviation of the input data (grey level values in the LDA feature space). In the previous work [8], a fixed value equal to 10^4 has been used for σ. In this work, in order to optimise the metric

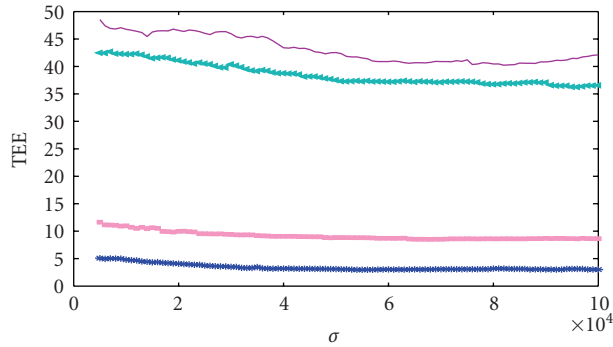TABLE 1: The usage of the different sessions in the BANCA experimental protocols.

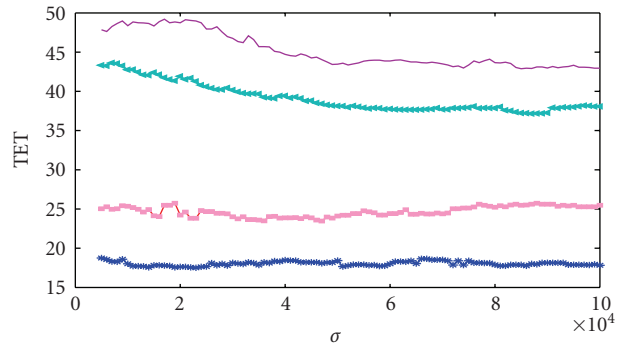|	1	2	3	4	5	6	7	8	9	10	11	12
Mc	TI	CI	CI	CI								
Md					TI	CI	CI	CI				
Ma									TI	CI	CI	CI
Ud	T				I	CI	CI	CI				
Ua	T								I	CI	CI	CI
P	TI	CI	CI	CI	I	CI	CI	CI	I	CI	CI	CI
G	TI	CI	CI	CI	TI	CI	CI	CI	TI	CI	CI	CI
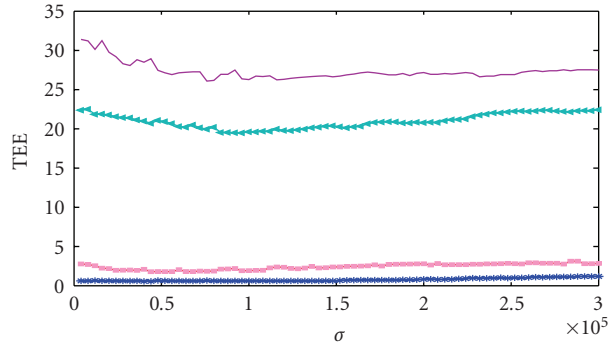
(a) Evaluation (Manual registration)
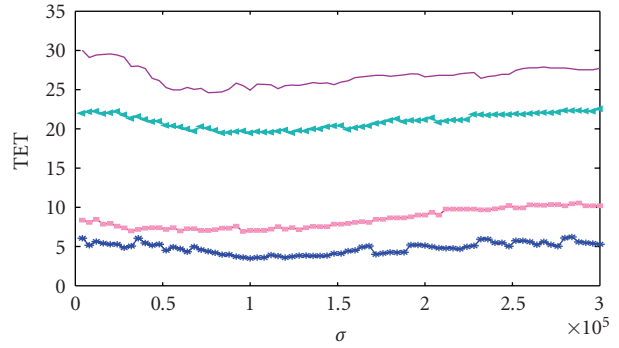
(b) Test (Manual registration)

(c) Evaluation (Automatic registration)

(d) Test (Automatic registration)

FIGURE 1: The performance of the GD metric versus the value of σ.

for dealing with different imaging conditions, the value of σ is adaptively determined in the evaluation step where the performance of the system for different values of σ is evaluated. As examples, Figure 1 contains plots of the Total Error rate versus the value of σ in the evaluation and test steps for the Mc, Md, Ud, and P protocols.

The evaluation plots show that by increasing the value of σ, the Total Error rate first rapidly decreases. Then, for larger values of σ, the TE rate remains relatively constant or increases gradually. From these plots, one can also see that the behaviour of the system in the evaluation and test phases is almost consistent. Therefore, the optimum σ can be found in the evaluation step by looking for the point after which the performance of the system is not significantly improved by increasing the value of σ. The associated value

of σ is then used in the test stage. Since, the effectiveness of a similarity measure depends on the adopted method of feature extraction, in the next subsection the experimental results using the PCA and LDA algorithms are reported. The fusion rules are presented in the sequel.

5.1. Experimental Results in the PCA and LDA Feature Spaces. Figure 2 contains the results obtained using the individual scoring functions on the evaluation and test data sets in the PCA and LDA spaces when manually annotated eyes position were used for the face geometric normalisation. The Total Error rates in the Evaluation (TEE) and Test (TET) stages have been used as performance measures in the plots. These results clearly demonstrate that among the adopted metrics, the GD metric is individually the outright winner.
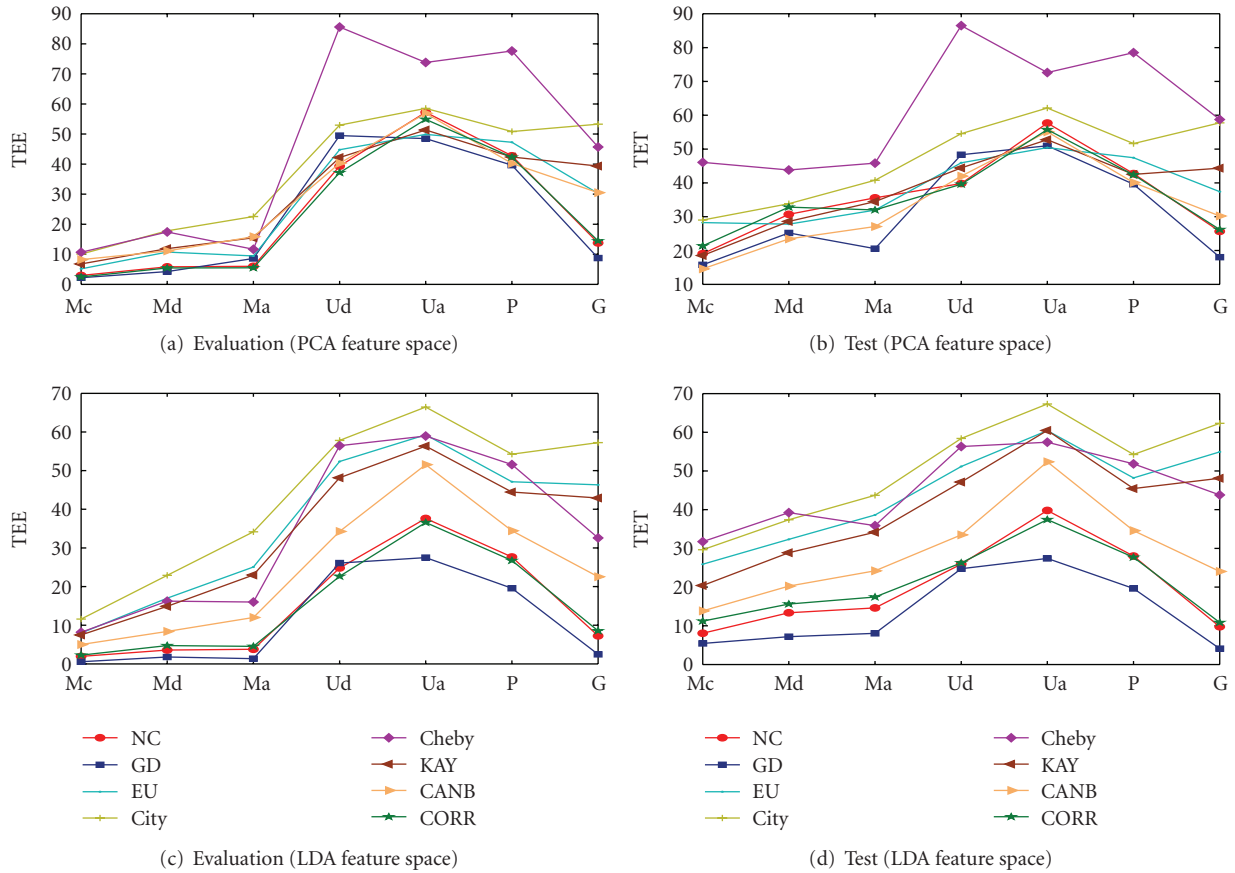
(a) Evaluation (PCA feature space)

(b) Test (PCA feature space)

(c) Evaluation (LDA feature space)

(d) Test (LDA feature space)

FIGURE 2: ID verification results using different scoring functions in the PCA and LDA feature spaces for the manually registered data.

(a) Evaluation (LDA feature space)
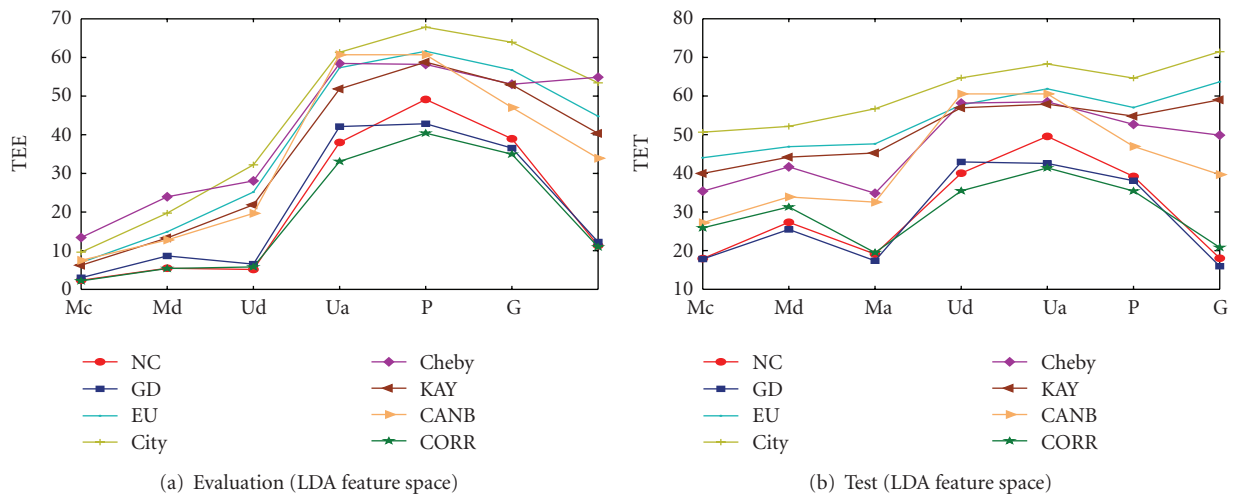
(b) Test (LDA feature space)

FIGURE 3: ID verification results using different scoring functions in the LDA feature space for automatically registered data.

For the sake of simplicity of comparison, Table 2 contains the evaluation and test results for the GD metric using the PCA and LDA spaces. These results demonstrate that a better performance can always be achieved using the LDA space.

Table 3 also contains a summary of the results obtained using the individual scoring functions on the evaluation and test sets when manually annotated eyes positions were used for the face geometric normalisation in the LDA space. The values in the table indicate the Total Error rates in the Evaluation (TEE) and Test (TET) stages, respectively.

The results of the similar experiments with automatically registered data in the LDA feature space demonstrate that in

TABLE 2: ID verification results using GD metric, LDA (left) and PCA (right). TEE: Total Error rate Evaluation; TET: Total Error rate Test.

| | Manual Registration | | | |
| | LDA | | PCA | |
	TEE	TET	TEE	TET
Mc	0.597	4.87	2.2	15.77
Md	1.77	7.18	4.26	25.19
Ma	1.56	8.03	8.6	20.54
Ud	26.09	24.74	49.49	48.32
Ua	27.5	27.4	48.49	50.96
P	19.56	19.64	39.64	39.6
G	2.43	4.12	8.74	18.04

TABLE 3: ID verification results using different similarity measures for the manual registered data in the LDA feature space.

| | Mc | | Md | | Ma | | Ud | | Ua | | P | | G | |
	TEE	TET	TEE	TET	TEE	TET	TEE	TET	TEE	TET	TEE	TET	TEE	TET
NC	1.93	8.08	3.57	13.36	3.79	14.61	24.81	25.93	37.63	38.81	27.69	28.01	7.26	9.75
GD	0.60	4.87	1.77	7.18	1.55	8.03	26.09	24.74	27.5	27.40	19.56	19.64	2.43	4.12
ED	7.97	25.89	17	32.34	25.06	38.62	52.37	51.15	59.26	60.42	47.12	48.22	46.33	54.93
City	11.6	29.65	22.9	37.4	34.17	43.71	57.82	58.4	66.44	67.3	54.25	54.25	57.24	62.26
Cheb	8.2	31.73	16.22	39.23	16	35.86	56.44	56.3	58.94	57.41	51.56	51.85	32.54	43.79
χ^2	7.49	20.41	14.88	28.88	22.99	34.17	48.17	47.15	56.35	60.48	44.46	45.45	42.91	48.12
Corr	2.25	11.22	4.74	15.6	4.54	17.43	22.66	26.25	36.57	37.44	34.44	34.54	8.02	10.85
Canb	5	13.85	8.69	20.25	12.01	24.2	34.26	33.5	51.54	52.37	26.74	27.69	22.54	24.04

TABLE 4: Fusion results for the different BANCA protocols using different fusion rules.

| | Sum | | WA1 | | WA2 | | WA3 | |
	TEE	TET	TEE	TET	TEE	TET	TEE	TET
Mc	2.51	10.61	.82	6.31	1.9	8.56	1.38	6.5
Md	5.98	16.28	2.93	10.67	5.35	15.93	2.5	9.33
Ma	7.54	18.75	2.24	11.06	6.29	17.14	2.55	10.32
Ud	30.03	30.54	26.16	25.38	29.45	30.48	19.55	21.79
Ua	40.41	41.19	36.47	37.95	40.35	41.4	26.96	29.61
P	29.65	29.8	25.3	24.87	18.51	28.57	18.33	19.94
G	15.5	19.12	3.8	5.66	14.92	18.47	3.32	4.55

this case the optimised GD function again delivers a better or at least comparable performance. The performance of other metrics, with the exception of NC, is much worse. These results are shown in Figure 3.

5.2. Fusion Results and Discussions. In the next step, we investigated the effect of fusing the classifiers employing the different similarity measures. In the first group of experiments, we compared the fixed combination rules (Sum, Product, Min, Max, and Median) in which all the classifiers are deemed to carry the same weight. The results obtained in the evaluation and test steps for both manually and automatically registered data are shown in Figure 4. These results clearly demonstrated that among the adopted fixed rules, the Sum rule outperforms the others for both manually and automatically registered data. For the sake of simplicity of comparison of the results using the untrained and trained rules, the fusion results using the Sum rule for manually registered data have been reported in Table 4.

In the second group of fusion experiments, different weighted averaging of the outputs of classifiers employing different similarity measures were examined. The results are presented in Table 4. In this table WA1, WA2, and WA3 represent the weighted averaging results for the error minimisation method, PCA, and LDA, respectively.

As can be seen, all the adopted weighted averaging methods give better results compared to the simple averaging (Sum) rule. Also, among the weighted averaging methods, a better performance is achieved using the LDA method.

(a) Evaluation (Manual registration)

(b) Test (Manual registration)

(c) Evaluation (Automatic registration)

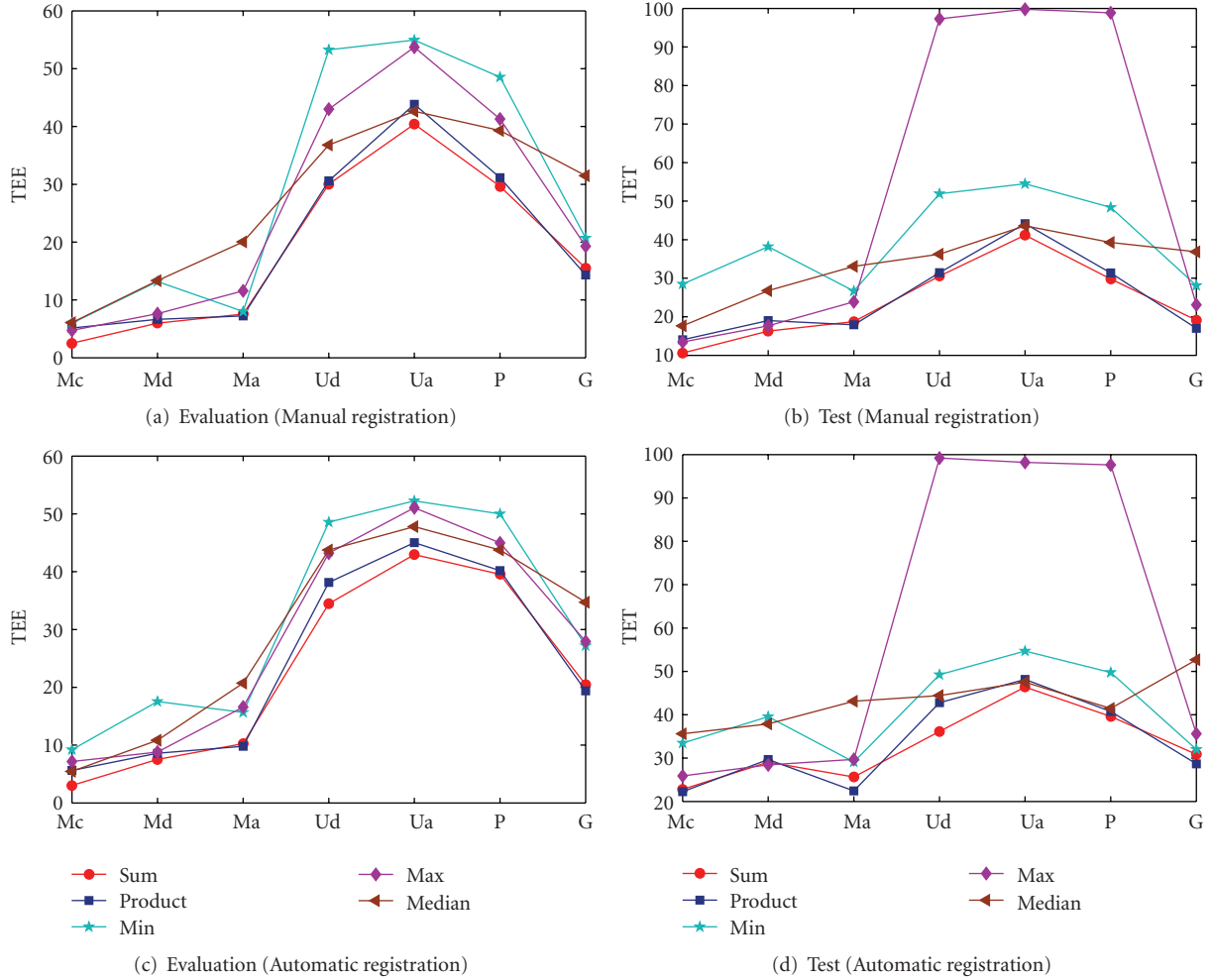(d) Test (Automatic registration)

FIGURE 4: Untrained fusion results in the evaluation and test steps for different BANCA experimental protocols.

TABLE 5: Fusion results on BANCA protocols with PCA and LDA space using SVM, manual registration (left), and automatic registration (right).

	Manual Registration						Automatic Registration					
	FARE	FRRE	TERE	FART	FRRT	TERT	FARE	FRRE	TERE	FART	FRRT	TERT
Mc	0.096	0.13	0.22	0.86	0.13	0.99	5.48	5.51	10.99	6.92	6.54	13.46
Md	0.96	1.02	1.98	1.06	2.18	3.24	2.88	2.95	5.83	21.83	6.41	28.24
Ma	1.44	1.54	2.98	0.38	3.72	4.1	0.86	0.9	1.76	0.86	7.56	8.42
Ud	10.19	10.13	20.32	9.14	14.61	23.75	10.48	10.38	20.86	9.81	15	24.81
Ua	10.77	10.9	21.67	11.83	10.51	22.34	15	14.87	29.87	26.15	22.44	48.59
P	7.6	7.52	15.12	7.92	9.83	17.75	14.87	14.82	29.6	12.08	17.52	29.6
G	1.31	1.33	2.64	1.15	1.7	2.85	6.35	6.41	12.76	9.87	8.93	18.8

Figure 5 contains comparative plots of the results using the Sum rule, LDA-based weighted averaging, and the SVMs. These plots demonstrate that the trained methods outperform the untrained (Sum) rule. In most of the cases, comparable results are obtained using LDA weighting and SVMs.

Since, the effectiveness of a similarity measure depends on the adopted method of feature extraction, in the next step, the merit of fusing the PCA- and LDA-based classifiers using SVM was investigated. Figure 6 contains the comparative plots of the Total Error rates obtained in the Evaluation (TEE) and Test (TET) stages for both manually

(a) Evaluation (Manual registration)

(b) Test (Manual registration)

(c) Evaluation (Automatic registration)

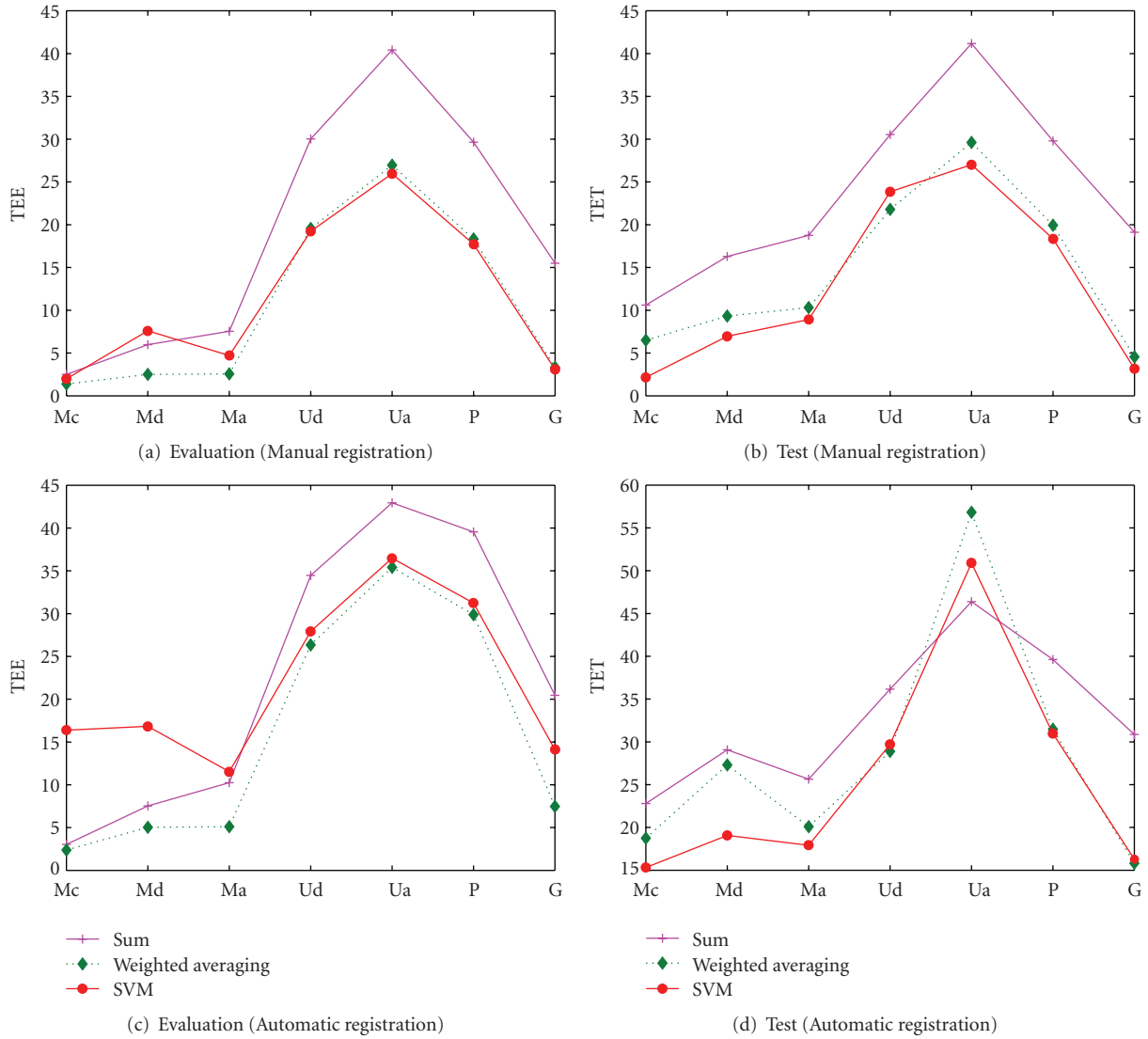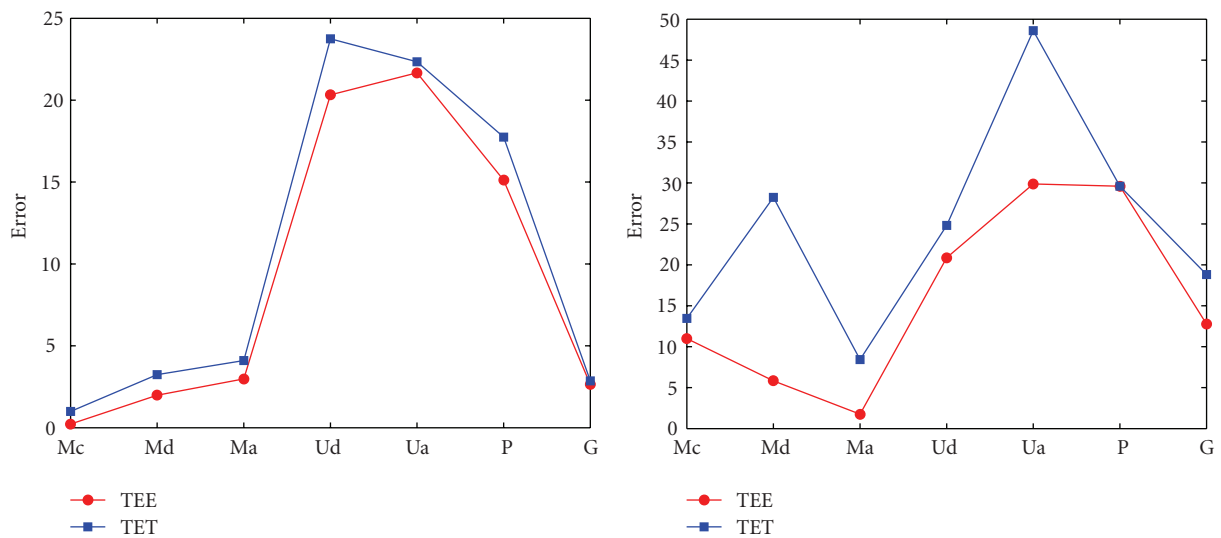(d) Test (Automatic registration)

FIGURE 5: Fusion using the Sum, LDA- based weighted averaging, and SVMs.

(a) (Manual registration)

(b) (Automatic registration)

FIGURE 6: Verification results by fusing the LDA- and PCA-based classifiers using SVMs.

and automatically registered data (see Table 5). These plots demonstrate that these methods outperform the other rules.

Overall, the results clearly demonstrate that the proposed similarity measure fusion considerably improves the performance of the face verification system.

6. Conclusions

The problem of fusing similarity measure-based classifiers in face verification was considered. First, the performance of face verification systems in PCA and LDA feature spaces with different similarity measure classifiers was experimentally evaluated. The study was performed for both manually and automatically registered face images. The experimental results confirm that our optimised Gradient Direction metric in the LDA feature space outperforms the other investigated metrics. Different methods for the selection and fusion of the various similarity measure-based classifiers were compared. The experimental results demonstrate that the combined classifiers outperform any individual verification algorithm. Moreover, the Support Vector Machines and Weighted Averaging of similarity measures have been shown to be the best fusion rules. It was also shown that although the features derived from the LDA approach lead to better results, than those of the PCA algorithm, fusing the PCA- and LDA-based scores improves the performance further. Based on our previous study within the LDA space [19], further improvement is also expected by adaptively selecting a subset of the LDA-based and PCA-based classifiers.

Acknowledgment

References

[1] D. Zhang and G. Lu, "Evaluation on similarity measurement for image retrieval," *Neural Network and Signal Processing*, vol. 2, pp. 228–231, 2003.

[2] Q. Bao and P. Guo, "Comparative studies on similarity measures for remote sensing image retrieval," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '04)*, pp. 1112–1116, October 2004.

[3] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognition Letters*, vol. 25, no. 6, pp. 711–724, 2004.

[4] W. S. Yambor, B. A. Draper, and J. R. Beveridge, "Analyzing PCA-based face recognition algorithm: eigenvector selection and distance measures," in *Empirical Evaluation Methods in Computer Vision*, H. Christensen and J. Phillips, Eds., World Scientific Press, Singapore, 2002.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[7] J. Kittler, Y. P. Li, and J. Matas, "On matching scores for LDA-based face verification," in *Proceedings of British Machine Vision Conference*, M. Mirmehdi and B. Thomas, Eds., pp. 42–51, 2000.

[8] M. T. Sadeghi and J. Kittler, "Confidence based gating of multiple face authentication experts," in *Proceedings of Joint IAPR International Workshops on Syntactical and Structural Pattern Recognition and Statistical Pattern Recognition (SSPR '06)*, vol. 4109 of *Lecture Notes in Computer Science*, pp. 667–676, Hong Kong, August 2006.

[9] E. Bailly-Bailliére, S. Bengio, F. Bimbot et al., "The BANCA database and evaluation protocol," in *Proceedings of International Conference on Audio and Video Based Person Anthentication*, vol. 2688, pp. 625–638, 2003.

[10] M. T. Sadeghi, M. Samiei, S. M. T. Almodarresi, and J. Kittler, "Similarity measures fusion using SVM classifier for face authentication," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications (VISAPP '08)*, vol. 2, pp. 105–110, Funchal, Madeira, Portugal, January 2008.

[11] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[12] J. Kittler and F. Roli, *Multiple Classifier Systems*, vol. 2096, Springer, Berlin, Germany, 2001.

[13] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[14] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combining of neural classifiers: a comparative study," *Pattern Recognition Letters*, vol. 20, pp. 429–444, 1999.

[15] F. Roli and G. Fumera, "Analysis of linear and order statistics combiners for fusion of imbalanced classifiers," in *Proceedings of the 3rd International Workshop on Multiple Classifier Systems*, pp. 252–261, Springer, Cagliari, Italy, June 2002.

[16] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[17] R. P.W. Duin, "The combining classifier: to train or not to train?" in *Proceedings of the International Conference on Pattern Recognition*, vol. 16, no. 2, pp. 765–770, 2002.

[18] K. Maghooli and M. S. Moin, "A new approach on multimodal biometrics based on combining neural networks using AdaBoost," in *Proceedings of the International ECCV Workshop on Biometric Authentication (BioAW '04)*, vol. 3087, pp. 332–341, Prague, Czech, May 2004.

[19] M. T. Sadeghi, M. Samiei, and J. Kittler, "Selection and fusion of similarity measure based classifiers using support vector machines," in *Proceedings of Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition (SSPR '08)*, vol. 5342 of *Lecture Notes in Computer Science*, pp. 479–488, 2008.

[20] G. L. Marcialis and F. Roli, "Fusion of LDA and PCA for face verification," in *Proceedings of the International ECCV Workshop on Biometric Authentication*, M. Marcialis and J. Bigun, Eds., vol. 2359 of *Lecture Notes in Computer Science*, pp. 30–37, 2002.

[21] M. T. Sadeghi and J. Kittler, "Decision making in the LDA space: generalised gradient direction metric," in *Proceedings of*

the 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 248–253, Seoul, Korea, May 2004.

[22] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[24] J. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," Tech. Rep. 98-14, Microsoft Research, Redmond, Wash, USA, April 1998.

[25] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

[26] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kälviäinen, and J. Matas, "Feature-based affine-invariant localization of faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1490–1495, 2005.

[27] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, " XM2VTSDB: the extended m2vts database," in *Proceedings of the 2nd International Conference on Audio and Video-based Biometric Person Authentication*, pp. 72–77, 1999.