The Corrected Normalized Correlation Coefficient: A Novel Way Of Matching Score Calculation for LDA-Based Face Verification*

Vitomir Štruc and Nikola Pavešić
Faculty of Electrical Engineering
University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia
vitomir.struc@fe.uni-lj.si, nikola.pavesic@fe.uni-lj.si

Abstract

The paper presents a novel way of matching score calculation for LDA-based face verification. Different from the classical matching schemes, where the decision regarding the identity of the user currently presented to the face verification system is made based on the similarity (or distance) between the "live" feature vector and the template of the claimed identity, we propose to employ a measure we named the corrected normalized correlation coefficient, which considers both the similarity with the template of the claimed identity as well as the similarity with all other templates stored in the database. The effectiveness of the proposed measure was assessed on the publicly available XM2VTS database where encouraging results were achieved.

1. Introduction

Over the last decades, automatic face recognition has become a highly active research area, mainly due to the countless application possibilities in both the private as well as the public sector. Research groups from around the world have directed their research efforts towards improving the recognition performance of automatic face recognition systems. However, most of this effort is directed at face localization and elaborate feature extraction techniques, while only a few studies take an interest in the matching scheme which also represents a vital part of the face recognition system.

Kittler et al. [4], for example, showed that when a face verification system employs linear discriminant analysis (LDA) for feature extraction the choice of an appropriate distance (or similarity) measure has a significant impact on the final performance of that system. They reported that better verification results can be obtained when the normalized correlation coefficient is used as the similarity measure rather than the Euclidian distance and that the performance can further be improved if an even more elaborate similarity measure is used for calculation of the matching score in the LDA subspace.

Recently, similar findings were presented by Liu [5], who proposed two novel similarity measures and also showed that when incorporated into the matching procedure the new measures significantly improved the performance of his face verification system.

In this paper we propose yet another measure, which opposed to the measures proposed by Liu [5] or Kittler et al. [4] in addition to the similarity with the template of the claimed identity also considers the similarity with all other templates in the system's database. The measure is derived based on some properties of the LDA subspace and evaluated on the XM2VTS face database. It's performance is assessed in conjunction with two methods that use LDA for feature extraction: the Fisherface method and Gabor-Fisher Classifier.

The rest of the paper is organized as follows: in Section 2 LDA is briefly reviewed and the basic concepts of the Fisherface and Gabor-Fisher Classifier methods are described. In Section 3 the novel corrected normalized correlation coefficient is introduced. The description of the face-image database and the associated experimental protocol is detailed in Section 4. The experimental results are presented in Section 5, while some final comments are given in Section 6.

2. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a powerful subspace projection technique which has been exten-

*Partially supported by the national research program P2-0250(C) Metrology and Biometric Systems, the bilateral project with the People's Republic of China Bi-CN/07-09-019, the bilateral project with the Bulgarian Academy of Sciences - Face and Signature Biometrics, the national project AvID M2-0210 and the EU-FP7 project 217762 Homeland security, biometric Identification and personal Detection Ethics (HIDE)

sively employed for feature extraction in many pattern recognition problems including face recognition. It derives a transformation matrix (i.e., the projection basis) which is used to project patterns, e.g., face images, into a subspace where between-class variations are maximized while within-class variations are minimized [8].

Given a set $\mathcal{X}$ of n d-dimensional training patterns i.e., $\mathcal{X} = \{\mathbf{x}_i : i = 1, 2, ..., n\}$, each belonging to one of N classes $C_1, C_2, ..., C_N$, one first computes the between-class and the within-class scatter matrices $\mathbf{S}_B = \sum_{i=1}^{N} n_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$ and $\mathbf{S}_W = \sum_{i=1}^{N} \sum_{\mathbf{x}_j \in C_i} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T$ and then derives the LDA transformation matrix $\mathbf{W}$ which maximizes Fisher's discriminant criterion [1, 8]:

$$J(\mathbf{W}) = arg \max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|}, \qquad (1)$$

where n_i denotes the number of samples in the i-th class, μ_i stands for the class conditional mean and μ represents the global mean of all training samples. Fisher's discriminant criterion is maximized when $\mathbf{W}$ is constructed by a simple concatenation of the $d' \leq N-1$ leading eigenvectors of the following eigenproblem:

$$\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{w}_i, \quad i = 1, 2, ..., d', \qquad (2)$$

that is $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_{d'}]$. Once the transformation matrix $\mathbf{W}$ is calculated, it can be used to project a test pattern $\mathbf{x}$ into the LDA subspace, thus reducing the pattern's dimension from d to d':

$$\mathbf{y} = \mathbf{W}^T (\mathbf{x} - \boldsymbol{\mu}), \qquad (3)$$

where $\mathbf{y}$ represents the d'-dimensional projection of the centered pattern $\mathbf{x}$.

Unfortunately, the number of training patterns available in the field of face recognition is commonly significantly smaller than the pattern's (e.g., face image's) dimension, making the within-class scatter matrix $\mathbf{S}_W$ singular and thus computation of the matrix $\mathbf{W}$ using the eigenproblem in (2) impossible. To overcome this problem, we used a modification of LDA, which first projects all patterns into the PCA subspace (to reduce their dimensionality and consequently ensure that the matrix $\mathbf{S}_W$ is invertible) and then performs LDA in the reduced space [1].

Several approaches have been presented in the literature which employ the described approach for facial-feature extraction, among them the Fisherface method and the Gabor-Fisher Classifier (GFC) are two of the most popular and have therefore also been used in our experiments. The Fisherface method applies the described LDA to the grey-scale distributions of the training images, while the GFC method first derives a vector of Gabor magnitude features for each of the training images and then performs LDA on the resulting set of Gabor feature vectors. A detailed description of both techniques can be found in [1] and [6].

3. The Corrected Normalized Correlation Coefficient

In a face recognition system the choice of an appropriate similarity measure which measures the similarity (or distance) between a given input feature vector and a user-template stored in the system's database is an important factor which has a great impact on the final recognition accuracy of the face recognition system. Traditionally, the same similarity measures are used for systems operating both in verification as well as identification mode. However, as we will show in this section, face recognition systems operating in verification mode require similarity measures which differ from those used for identification. Based on this finding we will present a novel measure, i.e., the corrected normalized correlation coefficient, which ensures enhanced face verification performance.

Let us for the time being confine ourselves to face recognition systems operating in identification mode. The identification problem may be stated as follows: given an input feature vector $\mathbf{y}$, determine the identity C_i, $i \in \{1, 2, ..., N, N+1\}$, where $C_1, C_2, ..., C_N$ denote the identities enrolled in the system and C_{N+1} represents the class of users for which no suitable identity can be found in the system's database [3]:

$$\mathbf{y} \in \begin{cases} C_i, & \text{if } \max_i \{\delta(\mathbf{y}, \overline{\mathbf{y}}_i)\} \geq t, \; i = 1, 2, ..., N \\ C_{N+1}, & \text{otherwise} \end{cases}.$$

Here δ denotes a function that measures the similarity between the "live" feature vector $\mathbf{y}$ and the user-template $\overline{\mathbf{y}}_i$ which corresponds to identity C_i, and t stands for a predefined threshold. We can see that the success of the identification procedure heavily depends on the employed similarity measure δ.

Recently C. Liu showed in [5] that some popular similarity measures (or distance measures) such as the Mahalanobis distance or the whitened cosine similarity measure can under certain assumptions be derived from the Bayes formula for conditional probabilities and thus the identification process based on these measures is related to the Bayes decision rule for minimum error. In the following, we will present a similar derivation which, however, considers some properties intrinsic to the LDA subspace. The Bayes formula for conditional probabilities is defined as [2]:

$$P(C_i|\mathbf{y}) = \frac{p(\mathbf{y}|C_i)P(C_i)}{\sum_{j=1}^{N} p(\mathbf{y}|C_j)P(C_j)}, \qquad (4)$$

where $p(\mathbf{y}|C_i)$ for $(i = 1, 2, ..., N)$ denote the conditional probability density functions, $P(C_i)$ for $(i = 1, 2, ..., N)$ represent the prior probabilities of the N classes and $P(C_i|\mathbf{y})$ stands for the a posteriori probability of the class C_i. In a face recognition system operating in identification mode where the goal is to assign an identity, i.e., class label C_i, to a given feature vector $\mathbf{y}$, the optimal decision rule is given by the Bayes rule for minimum error, i.e. [2, 5]:

$$P(C_i|\mathbf{y}) = \max_{j=1}^{N} P(C_j|\mathbf{y}) \rightarrow \mathbf{y} \in C_i, \tag{5}$$

The rule postulates that, if the a posteriori probability of class C_i is the largest among all N classes, than the feature vector $\mathbf{y}$ is assigned the identity, i.e., class label, C_i. However, in the identification scenario, the denominator of (4) for a given feature vector $\mathbf{y}$ is the same for all classes C_i $(i = 1, 2, ..., N)$ and can therefore be ignored. The Bayes decision rule for minimum error can now be written in it's equivalent form:

$$[p(\mathbf{y}|C_i)P(C_i)] = \max_{j=1}^{N}[p(\mathbf{y}|C_j)P(C_j)] \rightarrow \mathbf{y} \in C_i \tag{6}$$

or

$$ln[p(\mathbf{y}|C_i)P(C_i)] = \max_{j=1}^{N} ln[p(\mathbf{y}|C_j)P(C_j)] \rightarrow \mathbf{y} \in C_i. \tag{7}$$

If we assume that the conditional probability density functions $p(\mathbf{y}|C_j)$ in the d'-dimensional LDA subspace are modeled as multivariate Gaussian distributions with mean vectors $\overline{\mathbf{y}}_j$ and identity covariance matrices [4, 5], i.e.,

$$p(\mathbf{y}|C_j) = (2\pi)^{-\frac{d'}{2}} \exp[-\frac{1}{2}(\mathbf{y} - \overline{\mathbf{y}}_j)^T(\mathbf{y} - \overline{\mathbf{y}}_j)], \tag{8}$$

and we denote the discriminant function in (7) as $\delta(\mathbf{y})$, i.e., $\delta(\mathbf{y}) = ln[p(\mathbf{y}|C_j)P(C_j)]$ then we can write:

$$\delta(\mathbf{y}) = -\frac{1}{2}[(\mathbf{y} - \overline{\mathbf{y}}_j)^T(\mathbf{y} - \overline{\mathbf{y}}_j) + d' ln(2\pi)] + ln[P(C_j)], \tag{9}$$

which is equivalent to (7) under the Gaussian assumption.

Equation (9) can, however, be further simplified by dropping the constant terms $d' ln(2\pi)$ and $ln[P(C_j)]$ [1], i.e.,

$$\delta(\mathbf{y}) = -\frac{1}{2}[(\mathbf{y} - \overline{\mathbf{y}}_j)^T(\mathbf{y} - \overline{\mathbf{y}}_j)] \tag{10}$$

which turns the identification problem using the Bayes decision rule into the search for the smallest Euclidian distance, i.e., $\delta_E(\mathbf{y}, \overline{\mathbf{y}}_j) = [(\mathbf{y} - \overline{\mathbf{y}}_j)^T(\mathbf{y} - \overline{\mathbf{y}}_j)]^{\frac{1}{2}}$ between the given input vector $\mathbf{y}$ and the mean vectors $\overline{\mathbf{y}}_j$ of all classes $j = 1, 2, ..., N$:

$$\delta_E(\mathbf{y}, \overline{\mathbf{y}}_i) = \min_{j=1}^{N} \delta_E(\mathbf{y}, \overline{\mathbf{y}}_j) \rightarrow \mathbf{y} \in C_i. \tag{11}$$

Liu [5] further showed that if the feature and mean vectors $\mathbf{y}$ and $\overline{\mathbf{y}}_j$ are normalized to unit norm, the discriminant function of (10) is related to the cosine similarity measure:

$$\delta(\mathbf{y}) = -\frac{1}{2}[\|\mathbf{y}\|^2 + \|\overline{\mathbf{y}}_j\|^2 - 2\|\mathbf{y}\|\|\overline{\mathbf{y}}_j\|\delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_j)], \tag{12}$$

where

$$\delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_j) = \frac{\mathbf{y}^T\overline{\mathbf{y}}_j}{\|\mathbf{y}\|\|\overline{\mathbf{y}}_j\|} \tag{13}$$

denotes the cosine similarity measure based matching score of the vectors $\mathbf{y}$ and $\overline{\mathbf{y}}_j$ and $\|\cdot\|$ stands for the norm operator. Applying the unit norm assumption to (12), i.e., $\delta(\mathbf{y}) = \delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_j) - 1$, results in the cosine similarity decision rule:

$$\delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_i) = \max_{j=1}^{N} \delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_j) \rightarrow \mathbf{y} \in C_i. \tag{14}$$

In statistics, the expression in (13) is often referred to as the normalized correlation coefficient and is used for measuring the extent to which two samples, in our case the vectors $\mathbf{y}$ and $\overline{\mathbf{y}}_j$, are linearly related. When the absolute value of the normalized correlation coefficient equals one[2], then there exists a linear relation between the two samples, while on the other hand, when the value of the normalized correlation coefficient equals zero, then the two samples have no linear relation. Generally, the higher the absolute value of the coefficient, the stronger the linear relation between the two samples. Based on this fact the absolute value of the normalized correlation coefficient is commonly employed for computing the matching score between the input vector $\mathbf{y}$ in the client template $\overline{\mathbf{y}}_j$. If we replace the the cosine similarity measure in (14) with it's absolute value and we denote $\delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_j) = |\delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_j)|$, then the optimal decision rule can finally be written as:

$$\delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_i) = \max_{j=1}^{N} \delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_j) \rightarrow \mathbf{y} \in C_i. \tag{15}$$

The presented normalized correlation decision rule usually ensures better face recognition (identification and verification) performance than decision rules based on other similarity measures [4, 5].

Having the presented derivation in mind, let us now shift our attention to face recognition systems operating in verification mode. Different from the identification case, the problem statement for such systems can

[1]Note that it is assumed that the prior probabilities $P(C_j)$ are equal for all N classes and thus represent constant terms.

[2]Note that the range of the normalized correlation coefficient is given by the real valued interval $[-1, 1]$

be described as follows: given an input feature vector $\mathbf{y}$ and a claimed identity C_i, determine whether the claim of identity should be accepted or not. In case the claim is accepted assign the feature vector $\mathbf{y}$ to class w_1 (a genuine user), otherwise assign the feature vector $\mathbf{y}$ to class w_2 (an impostor). Typically, the acceptance of the claim is determined based upon the outcome of a matching procedure, in which the given input feature vector $\mathbf{y}$ is matched against the user-template $\overline{\mathbf{y}}_i$ that corresponds to the identity C_i [3], i.e.,

$$(C_i, \mathbf{y}) \in \begin{cases} w_1, & \text{if } \{\delta(\mathbf{y}, \overline{\mathbf{y}}_i)\} \geq t,\ i = 1, 2, ..., N \\ w_2, & \text{otherwise} \end{cases},$$

where t and δ again denote the predefined threshold and the function that measures the similarity between the vectors $\mathbf{y}$ and $\overline{\mathbf{y}}_i$, respectively.

Similarly to face recognition systems operating in identification mode, systems operating in verification mode also often employ the Euclidian distance or the normalized correlation coefficient to produce a matching score, based on which the input feature vector is assigned either to the class of clients or the class of impostors. However, as we have shown, these distances (or similarity measures) are derived (for the identification problem) from the Bayes formula for conditional probabilities under several assumptions which may not fully apply to the verification case. First of all, the distances (or similarity measures) were derived from the conditional density function $p(\mathbf{y}|C_i)$ rather than the a posteriori probability $P(C_i|\mathbf{y})$ as for the identification problem the denominator of (4) for a given $\mathbf{y}$ was equal for all classes and had therefore no impact on the maximum value of the scores across different classes. However, for the verification problem, only one matching score is computed for a given input feature vector $\mathbf{y}$, i.e., $\delta(\mathbf{y}, \overline{\mathbf{y}}_i)$, and this matching score is subsequently compared with the threshold t to decide to which of the two classes (w_1 or w_2) the feature vector $\mathbf{y}$ should be assigned to. In this case the denominator of (4) cannot simply be ignored as it represents a scaling factor which ensures that the a posteriori probabilities $P(C_1|\mathbf{y}), P(C_2|\mathbf{y}), ..., P(C_N|\mathbf{y})$ sum up to the value of one. From this point of view, the denominator of (4) serves as a normalization factor which ensures that the matching scores for different input feature vectors $\mathbf{y}$ are properly normalized and can be compared to a single global threshold t. Considering the presented derivation of the normalized correlation coefficient for the identification scenario, it is trivial to show that under similar assumptions, (i.e., multivariate normal distribution, equal priors, unit norms) the a posteriori probability $P(C_i|\mathbf{y})$ for the verification problem turns into a scaled form of the normalized correlation coefficient:

$$\delta_{SNC}(\mathbf{y}, \overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2, ..., \overline{\mathbf{y}}_N)_{C_i} = \frac{\delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_i)}{\sum_{j=1}^{N} \delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_j)}. \quad (16)$$

By closer examining the equation (16) we can see that the decision regarding the class membership of the given input feature vector $\mathbf{y}$ is no longer made solely on the distance (or similarity) between the vector $\mathbf{y}$ and the user-template $\overline{\mathbf{y}}_i$. The decision is now based on the distances (or similarities) between the feature vector $\mathbf{y}$ and all user-templates in database $\overline{\mathbf{y}}_j$ ($j = 1, 2, ..., N$). Furthermore, we can notice that equation (16) better reflects Fisher's criterion (1) as both inter-class and intra-class distances are incorporated into the scaled form of the normalized correlation coefficient.

As will be shown in the experimental section, the presented scaled form of normalized correlation coefficient already ensures superior face verification performance when compared to traditionally employed distances (or similarity measures) such as the Euclidian distance or the normalized correlation coefficient. However, we can not simply neglect the fact that the normalized correlation coefficient has ensured satisfactory verification performance in the past and therefore propose to combine both the classical normalized correlation coefficient and it's scaled form into a new measure which we call the corrected normalized correlation coefficient, i.e.,

$$\begin{aligned} &\delta_{CNC}(\mathbf{y}, \overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2, ..., \overline{\mathbf{y}}_N)_{C_i} = \\ &= \delta_{NC}(\mathbf{y}, \overline{\mathbf{y}}_i) + \gamma \cdot \delta_{SNC}(\mathbf{y}, \overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2, ..., \overline{\mathbf{y}}_N)_{C_i}. \end{aligned} \quad (17)$$

Here the scaled form of the normalized correlation coefficient is treated as a correction term which in addition to the intra-class distance (or similarity), i.e. nominator of (16), also accounts for the inter-class distances (or similarities), i.e., denominator of (16), of the given feature vector $\mathbf{y}$, while the classical normalized correlation coefficient accounts only for the intra-class distance (or similarity), i.e., the distance (or similarity) between the vector $\mathbf{y}$ and the user-template of the claimed identity $\overline{\mathbf{y}}_i$. γ represents the parameter which controls the amount of the correction and should be determined either through cross-validation or if available on data from an evaluation set.

4. Database and experimental protocol

The experiments presented in Section 5 were performed on the publicly available XM2VTS database [7] which contains frontal face images of 295 distinct subjects. The images were recorded in four photo sessions which were distributed over a time span of five months. At each session two images were recorded for each subject, resulting in a total of 2360 facial images which comprise the standard face image subset of the XM2VTS database.

Similar to other studies on face recognition (see for example [4, 6]) prior to feature extraction all of the face images from the XM2VTS database were subjected to a pre-processing procedure which comprised of:

- a conversion of the input color images to 8-bit grey-scale images,
- manual localization of the eyes and a subsequent normalization procedure which rotated and scaled the image in such a way that the centers of the eyes were located at predefined positions,
- a procedure that cropped the face region of the image to a standard size of 128×128 pixels and
- a normalization procedure which normalized the face region to zero mean and unit variance.

To allow for a direct comparison of our experimental results with other results presented in the literature the first configuration of the established experimental protocol, i.e., the Lausanne protocol - LP [7], associated with XM2VTS database was used to divide the subjects of the database into two groups - the group of clients (200 subjects) and the group of impostors (95 subjects). Images from these groups were then further partitioned into image sets which were used for training, evaluation and testing. The training set comprised of 600 images (3 images per subject) and was used to compute the LDA transformation matrices for both, the Fisherface method and the GFC technique and to construct the client models, i.e., mean feature vectors. The evaluation image set was comprised of 600 client images (3 per subject) and 200 impostor images (8 for each of the 25 subjects defined by the LP as evaluation impostors) and was employed to compute the decision threshold t as well as to determine the parameter γ which controls the amount of correction in the expression (17) for the corrected normalized correlation coefficient. Finally, the test set was comprised of 400 client images (2 per subject) and 560 impostor images (8 for each of the 70 subjects defined by the LP as test impostors) and was used exclusively for the final performance assessment.

In the verification experiments the feature vectors extracted from the client images were matched against the corresponding client templates, while the feature vectors belonging to the impostor images were matched against all client templates in the database. This setup resulted in a total of 600 client and $40,000$ impostor access trails during the evaluation phase and 400 client and $112,000$ access trails during the test phase. In each phase three error rates were measured to assess the performance of the proposed corrected normalized correlation coefficient, namely, the false acceptance rate (FAR) which measures the frequency with which an impostor is falsely accepted, the false rejection rate (FRR) which measures the frequency with which a client is falsely rejected and the total error rate (TER) which is defined as the sum of the FAR and FRR. However, as the values of FAR and FRR both depend on the employed decision threshold t an operating point had to be chosen. In accordance with the LP the equal error operating point, i.e., the point where FAR equals FRR, was selected for our experiments and therefore the threshold was set in a way, which ensured equal error rates in the evaluation phase.

5. Experiments

The goal of our first set of verification experiments was to determine the value of the correction parameter γ which would ensure the best possible verification performance (on the evaluation set) of the Fisherface and GFC methods in conjunction with the corrected normalized correlation coefficient. To this end, the value of the parameter was increased from 10 to 60 with a step size of 10, i.e., $\gamma = \{10, 20, 30, 40, 50, 60\}$. For each of the tested values of γ the values of the FAR, FRR and TER were computed for both methods (at the equal error operating point) and are presented in Tables 1 and 2.

Table 1. The error rates at the equal error operating point for the Fisherface method

γ	10	20	30	40	50	60
FAR(%)	2.18	1.77	1.81	1.76	1.79	1.94
FRR(%)	2.33	2.00	1.83	1.83	1.83	2.00
TER(%)	4.51	3.77	3.64	**3.59**	3.62	3.94

Table 2. The error rates at the equal error operating point for the GFC method

γ	10	20	30	40	50	60
FAR(%)	1.00	1.00	1.17	1.17	1.17	1.17
FRR(%)	1.00	0.98	0.91	1.11	1.10	1.13
TER(%)	2.00	**1.98**	2.08	2.28	2.27	2.30

We can see that the Fisherface method performed the best when the value of the parameter γ was set to 40. The GFC method, on the other hand, reached the lowest total error rate when γ was set to the value of 20. The experimental results suggest that these values should be used in the final comparative assessment.

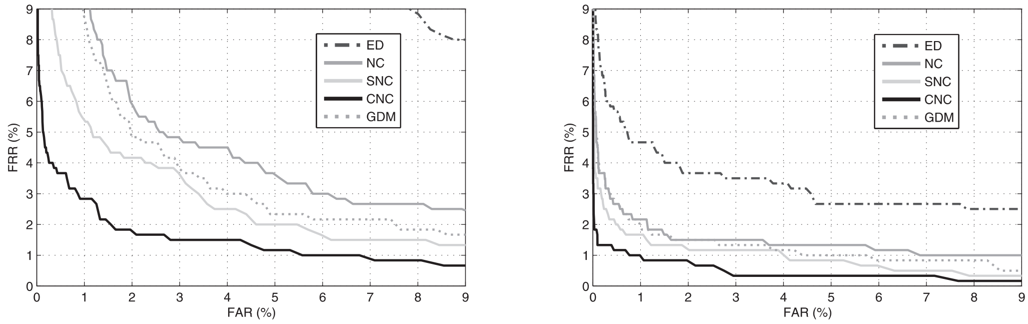

Figure 1. Comparison of the ROC curves of different distance (and similarity) measures for the Fisherface (left) and GFC (right) methods on the evaluation image set

In our second set of experiments we compared the performance of the Fisherface and GFC methods in conjunction with different distance (or similarity) measures. Specifically, the following distance (or similarity) measures were assessed in addition to the corrected normalized correlation coefficient (CNC): the Euclidian distance (ED), the normalized correlation coefficient (NC), the scaled form of the normalized correlation coefficient (SNC) and the gradient direction metric (GDM) that was presented in [4]. Note that the CNC was implemented with the values of γ that were determined in the previous set of experiments. The receiver operating characteristics (or ROC curve) of the experiments on the evaluation set which display the dependencies of the FAR and FRR at different operating points are shown in Figure 1. The final error rates (obtained on the test set) are given in Table 3.

Table 3. The error rates on the test set obtained with the threshold that ensured equal error rates on the evaluation set

Method	Fisherfaces			GFC		
Error (%)	FAR	FRR	TER	FAR	FRR	TER
ED	10.82	8.00	18.82	4.41	3.75	8.16
GDM	3.47	2.75	6.22	1.61	1.00	2.61
NC	4.51	3.50	8.01	1.67	1.00	2.67
SNC	2.95	2.50	5.45	1.15	0.75	1.90
CNC	1.73	1.75	**3.48**	0.99	0.25	**1.24**

We can notice that both methods performed the best when the corrected normalized correlation coefficient was used in the matching stage. Furthermore we can see that the relative ranking of the employed measures was the same for both methods, namely, the proposed measure performed the best, followed in order by the scaled form of the normalized correlation coefficient, the gradient direction metric, the normalized correlation coefficient and finally the Euclidian distance.

6. Conclusion

In this paper we have presented a novel measure for matching score calculation in the LDA subspace. The results of the verification experiments performed on the XM2VTS database show that the proposed measure reduces the total error rate of the Fisherface and GFC methods by more than 50% when compared to the commonly employed normalized correlation coefficient. As our future goal we plan to assess the performance of the proposed measure in conjunction with other feature extraction techniques that are not based on Fisher's discriminant criterion.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *Proceedings of the 4th ECCV*, pages 45–58, Cambridge, UK, April 1996.

[2] C. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, USA, 2006.

[3] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Curcuits and Systems for Video Technology*, 14(1):4–20, 2004.

[4] J. Kittler, Y. Li, and J. Matas. On matching scores for lda-based face verification. In *Proceedings of BMVC*, pages 42–51, Bristol, UK, 2000.

[5] C. Liu. The bayes decision rule induced similarity measures. *IEEE TPAMI*, 29(6):1086–1090, 2007.

[6] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[7] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. Xm2vtsdb: the extended m2vts database. In *Proceedings of AVBPA'99*, pages 72–77, Washington D.C., USA, March 1999.

[8] T. Savič and N. Pavešić. Personal recognition based on an image of the palmar surface of the hand. *Pattern Recognition*, 40(11):3152–3163, 2007.