Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions

C. Kotropoulos*, A. Tefas, I. Pitas

Department of Informatics, Aristotle University of Thessaloniki, Box 451, Thessaloniki 540 06, Greece

Received 11 February 1999; accepted 12 July 1999

Abstract

In this paper, morphological elastic graph matching is applied to frontal face authentication on databases ranging from small to large multimedia ones collected under either well-controlled or real-world conditions. It is shown that the morphological elastic graph matching achieves a very low equal error rate on databases collected under well-controlled conditions. However, its performance deteriorates when it is applied to databases recorded in real-world scenarios. The compensation for variable recording conditions such as changes in illumination, scale differences and varying face position prior to the application of morphological elastic matching is proposed.

Keywords: Frontal face verification; Morphological elastic graph matching; Large databases; Face normalization; Receiver operating characteristics

1. Introduction

The interest and the research activities in automatic face recognition have increased significantly over the past few years. The growth is mainly driven by application demands, such as identification for law enforcement and authentication for remote banking and access-control applications. A recent survey on face recognition can be found in Chellapa et al. [1]. Machine recognition of faces yields problems of the following two categories:

- **Face recognition.** Given a test face and a set of reference faces in a database find the $N$ closest reference faces to a test one.
- **Face authentication.** Given a test face and a reference one, decide if the test face is identical to the reference one.

The just mentioned problems are conceptually different. On the one hand, a face recognition system usually assists a human face-recognition expert to determine the identity of the test face by computing all similarity scores between the test face and each human face stored in the system database and by ranking them. On the other hand, a face authentication system should decide itself if the face is a **client** (i.e., he or she claims his/her own identity) or is an **impostor** (i.e., he or she pretends to be someone else).

The evaluation criteria for face recognition systems are different from those applied to face authentication systems. The performance of face recognition systems is quantified in terms of the percentage of correctly identified faces within the $N$ best matches. By varying the rank $N$ of the match, the curve of cumulative match score versus rank is obtained [2]. The performance of face authentication systems is measured in terms of the **false rejection rate** (FRR) achieved at a fixed **false acceptance rate** (FAR) or vice versa. By varying FAR, the **receiver operating characteristic** (ROC) curve is obtained. If a scaler figure of merit is used to judge the performance of an authentication algorithm, then we usually choose the operating point having $FAR = FRR$, the so-called **equal error rate** (EER). However, a more detailed comparison should also include the values $FRR_{FAR=0}$ and $FAR_{FRR=0}$.

*Corresponding author. Tel.: +30-31-998225; fax: +30-31-998419.

E-mail addresses: costas@zeus.csd.auth.gr (C. Kotropoulos), tefas@zeus.csd.auth.gr (A. Tefas), pitas@zeus.csd.auth.gr (I. Pitas).
A third difference is in the requirements needed when face recognition/authentication systems are trained. Face recognition systems are usually trained on sets having one frontal image per person. For example, in face recognition experiments conducted on FERET database, the fa images are used to train the system, while fb images are used to test the system. Face authentication systems usually need more images per individual for training to capture intra-class variability (i.e., to model the variations of the face images corresponding to the same individual). The requirements increase dramatically when linear discriminant analysis is employed to accomplish feature selection [11]. Although the algorithms employed in both face recognition and authentication systems are of common origin (for example, the dynamic link architecture [16]), there is no a priori guarantee that the same algorithm would enjoy the same performance level in both cases.

Face recognition has received more attention than face authentication by the scientific community. There are not many published works on face authentication. Even for face recognition only a handful of algorithms have been tested on databases greater than 150 persons. For a comparative study, the interested reader may consult Wiskott et al. [3]. Moreover, the conditions under which face databases are collected affect seriously the performance of a face recognition system. Recently, a comparative study has been performed for three well-known face recognition techniques, namely, the eigenfaces, the auto-association and classification neural networks, and the elastic graph matching [4]. It has been found that the eigenfaces work well when the face images have relatively small lighting and moderate expression variations. Their performance deteriorates significantly as lighting variation increases. On the contrary, elastic graph matching is found relatively insensitive to variations in lighting, face position and expressions. Zhang et al. attributed the robustness of elastic graph matching to the use of Gabor filters and to the rigid and deformable matching stages of the recognition algorithm. Recognition errors close to 20% are the best rates reported for elastic graph matching when scale variations are present. Auto-association and classification networks actually break down (i.e., their recognition errors are \( \approx 60\% \)) in the presence of scale variations.

In the closely related work, Adini et al. [5] presented an empirical study that evaluates the sensitivity of several image representations (e.g. edge maps, directional and non-directional derivatives of Gaussian filters, images convolved by 2-D Gabor-like filters) in changes of the illumination conditions. It has been found that all the aforementioned image representations are insufficient to overcome variations due to changes in illumination direction, viewpoint and expressions.

Motivated by the studies [2,4,5], first we test the performance of morphological elastic graph matching to databases ranging from small ones to large ones collected under either well-controlled or real conditions. Secondly, by evaluating the sensitivity of face authentication systems employing the morphological elastic graph matching and other competitive algorithms developed within the framework of EU-ACTS project M2VTS to changes in illumination, face size and position and facial expressions, we propose simple and powerful face normalization techniques that compensate for varying recording conditions. Accordingly, the paper complements the previously published works of Zhang et al. [4] and Adini et al. [5].

More specifically, several frontal-face authentication algorithms have been developed within M2VTS project. Among others we mention the morphological dynamic link architecture (MDLA), a sort of morphological elastic graph matching [8,11], the morphological signal decomposition-dynamic link architecture [9,10], the elastic graph matching (residual matching and local discriminants) [12], the optimized robust correlation [13] and the grey level frontal face matching [14]. All algorithms were initially tested on the M2VTS database [15] which contains video data of 37 persons' in 4 shots. A brief comparative study of their performance is included in Section 3.1.

The paper is focused on the morphological dynamic link architecture and its application to frontal face authentication on four databases. Section 2 briefly describes the morphological elastic graph matching algorithm (abbreviated as MDLA throughout the paper). The first two databases are multimedia ones collected under well-controlled conditions ranging from small (37 persons) to large ones (295 persons). In Section 3, we demonstrate that the morphological elastic graph matching achieves a very low equal error rate on databases collected under well-controlled conditions. The remaining two databases are small galleries recorded during real-world tests, such as access-control to buildings and cash dispenser services or access-control to tele-services via Internet in a typical office environment. Several types of degradation are present in the latter databases, such as, varying face size (i.e., scale differences), varying face position, changes in lighting as well as facial expression variations. More details can be found in Section 4. Although, MDLA achieves a very low equal error rate under well-controlled conditions, its performance deteriorates when it is applied to databases recorded during real-world tests. The compensation for variable recording conditions such as changes in illumination, scale differences and varying face position is addressed next. The use of simple and powerful pre-processing techniques aiming at compensating for the aforementioned conditions prior to the application of morphological elastic matching is proposed. Section 5 outlines the face normalization techniques. Experimental results that quantify the success of the techniques are included in Section 5 as well. The
results obtained indicate that the proposed normalization technique overcomes the image variations and stabilizes the performance of the authentication algorithm.

2. Morphological elastic graph matching

An alternative to linear techniques for generating an information pyramid is the scale-space morphological techniques. In the following, a brief description of morphological dynamic link architecture (MDLA) is given. In MDLA, we substitute the image representation part of elastic graph matching [16] that is based on Gabor wavelets by the multiscale morphological dilation-erosion.

Let $\mathcal{R}$ and $\mathcal{D}$ denote the set of real and integer numbers, respectively. Given an image $f(x) : \mathcal{D} \subseteq \mathbb{R}^2 \rightarrow \mathcal{R}$ and a structuring function $g(x)$, the dilation of the image $f(x)$ by $g(x)$ is defined by [6,7]

$$ (f \ominus g)(x) = \max_{x \in \mathcal{R}, \ x \in \mathcal{D}} \{ f(x-z) + g(z) \}.$$  \hspace{1cm} (1)

Its complementary operation, the erosion is given by

$$ (f \oslash g)(x) = \min_{x \in \mathcal{R}, \ x \in \mathcal{D}} \{ f(x+z) - g(z) \}.$$  \hspace{1cm} (2)

The scaled hemisphere is employed as a structuring function. The multiscale dilation-erosion of the image $f(x)$ by $g_{\sigma}(x)$ is defined by [17]

$$ (f \ominus g_{\sigma})(x) = \begin{cases} (f \ominus g_{\sigma})(x) & \text{if } \sigma > 0, \\ f(x) & \text{if } \sigma = 0, \\ (f \ominus g_{\sigma})(x) & \text{if } \sigma < 0. \end{cases}$$  \hspace{1cm} (3)

The outputs of multiscale dilation-erosion for $\sigma = -9, \ldots, 9$ form the feature vector located at the grid node $x$, i.e., $j(x) = (f \ominus g_{\sigma})(x), \ldots, f \ominus g_{\sigma})(x)$.

A 8 x 8 sparse grid has been created by measuring the feature vectors $j(x)$ at equally spaced nodes over the output of the face detection algorithm described in Kotropoulos et al. [25].

Let the superscripts $t$ and $r$ denote a test and a reference person (or grid), respectively. The $L_2$ norm of the difference between the feature vectors at the $i$th grid node has been used as a (signal) similarity measure, i.e.,

$$ S_t(j(x_i^t), j(x_i^r)) = \| j(x_i^t) - j(x_i^r) \|.$$  \hspace{1cm} (4)

As in DLA [16], the quality of a match is evaluated by taking into account the grid deformation as well. Let us denote by $\mathcal{V}$ the set of grid nodes. The grid nodes are simply the vertices of a graph. Also let $\mathcal{N}(i)$ denote the neighborhood of vertex $i$. A four-connected neighborhood has been used. An additional cost function:

$$ C_t(i,j) = S_t(d_i^j, d_{i,j}^r) = \| d_i^j - d_{i,j}^r \| \ \forall i \in \mathcal{V} ; j \in \mathcal{N}(i) \hspace{1cm} (5)$$

with $d_i^j = (x_i - x_j)$ can be used to penalize grid deformations. The penalty (5) can be incorporated into the cost function:

$$ C_t(\{x_i^t\}) = \sum_{i \in \mathcal{V}} \left\{ S_t(j(x_i^t), j(x_i^r)) + \lambda \sum_{j \in \mathcal{N}(i)} S_t(i, j) \right\}. $$  \hspace{1cm} (6)

The cost function (6) is actually a matching error that defines a distance measure between two persons. Lades et al. argue that a two-stage coarse-to-fine optimization procedure suffices for the minimization of Eq. (6) [16]. In our experiments, the above-mentioned approach is proved inadequate. Accordingly, we propose: (i) to exploit the face detection results that are provided by the hierarchical rule-based system [25] for initializing the minimization of the cost function, and (ii) to replace the two stage optimization procedure by a probabilistic hill climbing algorithm (i.e., a simulated annealing algorithm) that is reminiscent of the Algorithm 1.4 [18] that does not make a distinction between coarse and fine matching.

The design of the face authentication system based on morphological elastic graph matching spans several issues that have been solved separately. Linear projection algorithms for feature selection could be used to increase the efficiency of the method [10]. Automatic weighting of nodes according to their discriminatory power has also been treated [11].

3. Performance assessment on small and large multimedia databases collected under well-controlled conditions

In this section, we assess the performance of morphological elastic graph matching when it is applied for face authentication to two multimedia databases collected under well-controlled conditions. The databases are the M2VTS database [15] and the extended M2VTS database (XM2VTSDB) [19]. First, we briefly review the performance achieved by the morphological elastic graph matching on the M2VTS database. We proceed then to its performance assessment on the XM2VTSDB.

3.1. Face authentication from M2VTS database images

The morphological dynamic link architecture has been initially tested on the M2VTS database [15]. The database contains both sound and image information. Four recordings (i.e., shots) of 37 persons have been collected. In our experiments, the sequences of rotated heads have been considered by using only the luminance information at a resolution of $286 \times 350$ pixels. We have used only one frontal image from the image sequences of each person that has been chosen based on symmetry considerations. Four experimental sessions have been implemented by employing the “leave one out” principle
thus producing 5328 client claims and another 5328 imposter claims. Details on the experimental protocol used in the performance evaluation as well as on the computation of thresholds that discriminate each person from the remaining persons in the database can be found in Kotropoulos et al. [8]. Fig. 1 shows the receiver operating characteristics of morphological dynamic link architecture with and without any linear projections as well as when discriminatory power coefficients are employed.

Table 1 summarizes the equal error rates achieved by the face authentication algorithm under study and compares them to those achieved by the other frontal authentication algorithms on M2VTS database. It is seen that the optimized MDLA with discriminatory power coefficients is ranked as the first method in terms of the equal error rate.

### 3.2. Face authentication from extended M2VTS database images

The extended M2VTS database contains video data of 295 persons’ which include speech and image sequences of rotated heads [19]. Eight recordings of 295 persons have been collected under well-controlled recording conditions. Morphological elastic graph matching has been tested on the XM2VTSDB according to the two configurations defined in Luettin et al. [20]. A subset of 200 persons (i.e., training clients) has been used for training. Let us denote by $\mathcal{S}$ the set of training clients. The design of a person authentication system based on MDLA is split into three procedures, namely the training, evaluation and test procedure. In the following, we describe the design of such a system in detail for the first configuration.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic graph matching (residual matching) [12]</td>
<td>11.8</td>
</tr>
<tr>
<td>Morphological dynamic link architecture (without linear projection) [8]</td>
<td>9.3</td>
</tr>
<tr>
<td>Grey-level frontal face matching [14]</td>
<td>8.5</td>
</tr>
<tr>
<td>Elastic graph matching (local discriminants) [12]</td>
<td>6.1</td>
</tr>
<tr>
<td>Morphological dynamic link architecture (with linear projections of the feature vectors) [10]</td>
<td>5.4</td>
</tr>
<tr>
<td>Optimized robust correlation [13]</td>
<td>4.8</td>
</tr>
<tr>
<td>Morphological dynamic link architecture with discriminatory power coefficients [11]</td>
<td>3.7</td>
</tr>
</tbody>
</table>

In the first configuration three frontal images have been used for each client. The objective of the training procedure is to determine a threshold, $T(X; Q)$, for each training client as follows. Let $i, j$ be the recording indices. For $i, j = 1, 2, 3$ and $j \neq i$:

1. Compute the intra-class distances $d(X_i, X_j)$. Let $D(X, X) = \min_{i,j} \{d(X_i, X_j)\}$.
2. Compute the inter-class distances:
   $$d(Y_i, X_j), X \in \mathcal{S}_1, Y \in (\mathcal{S}_1 - \{X\})$$
   
   Let $D(Y, X) = \min_{i,j} \{d(Y_i, X_j)\}$.
3. Let $OS_{Q}(\mathcal{A})$ denote the $Q$-th order statistic of a set $\mathcal{A}$. For $Q = 0, 1, \ldots$ compute a threshold by
   $$T(X; Q) = OS_{Q+1}(D(Y, X) Y \neq X).$$
It is evident that the training procedure results in 200 client claims and 39,800 impostor claims. For several choices of $Q$, different pairs of false acceptance and false rejection rates are found. Accordingly, the receiver operating characteristics on the training set is obtained which is plotted in Fig. 2(a). It is seen that an EER $\approx 3.0\%$ is achieved.

A second subset of 25 persons has been selected to evaluate the performance of the authentication algorithm. These persons act as evaluation impostors. To implement the evaluation procedure, another three frontal images have been used for each client and eight frontal images have been employed for each evaluation impostor. The evaluation client images and the evaluation impostor images build the evaluation set. By using the just-mentioned set, 600 client claims and 40,000 impostor ones are produced. Although there is a possibility to adapt the thresholds determined previously to the evaluation set and to use the updated thresholds on the test set, we decided not to do so in order to keep the evaluation claims disjoint to the claims that are employed in the subsequent test procedure. Then the evaluation claims can be exploited to train a fusion manager. The estimation of FAR and FRR parallels the steps described subsequently in the test procedure. The ROC on the evaluation set is overlaid in Fig. 2(a). It can easily be seen that an EER $\approx 8.0\%$ is obtained on the evaluation set.

A third subset of 70 persons has been selected to constitute the set of test impostors. Two new frontal images for each client have been used to implement client accesses. Eight frontal images for each test have been used for each test impostor. The test client images and the test impostor images form the test set. By using the test set, 400 client claims and 112,000 impostor ones can be produced. The objective of the test procedure is to provide estimates of FAR and FRR by applying the previously determined thresholds on the test set. Let us denote the new frontal images of person $X$ by $X_i$, $i = 1, 2$. For each $X \in \mathcal{S}_1$, the number of false rejections is obtained as follows:

1. Compute the distances $d(X, X_i)$ for $i = 1, 2, 3$.
2. Normalize the distances to the range $[0, 1]$:
   
   \[ z(X, X_i) = f(d(X, X_i), T(X, Q)) \]

3. Compute $D(X, X) = \min_i \{z(X, X_i)\}$.
4. Count a false rejection if $D(X, X) < 0.5$.

Let us denote the set of test impostor images by $\mathcal{S}_3$. For each $X \in \mathcal{S}_1$ and $Y \in \mathcal{S}_3$ the number of false acceptances is obtained as follows:

1. Compute the distances $d(Y, X_i)$ for $i = 1, 2, 3$.
2. Normalize the distances to the range $[0, 1]$:
   
   \[ z(Y, X_i) = f(d(Y, X_i), T(X, Q)) \]

3. Compute $D(Y, X) = \min_i \{z(Y, X_i)\}$.
4. Count a false acceptance if $D(Y, X) \geq 0.5$.

By varying the parameter $Q$, different thresholds are recalled from the training procedure and the receiver operating characteristic on the test set is created. The test ROC is shown overlaid in Fig. 2(a). The inspection of Fig. 2(a) reveals that an EER $\approx 6.57\%$ is achieved.
Table 2
Rates at $FAR \approx 0$, $FAR \approx FRR$, and $FRR \approx 0$ on XM2VTSDB in the two configurations of the experimental protocol. All rates are in %

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Operating point</th>
<th>Evaluation set FAR</th>
<th>Test set FAR</th>
<th>Evaluation set FRR</th>
<th>Test set FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$FAR \approx 0$</td>
<td>0.5</td>
<td>0.46</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>$FAR \approx FRR$</td>
<td>8.11</td>
<td>8.23</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$FRR \approx 0$</td>
<td>48.42</td>
<td>46.63</td>
<td>1.66</td>
<td>0.75</td>
</tr>
<tr>
<td>II</td>
<td>$FAR \approx 0$</td>
<td>0.46</td>
<td>0.46</td>
<td>18.75</td>
<td>16.25</td>
</tr>
<tr>
<td></td>
<td>$FAR \approx FRR$</td>
<td>6.04</td>
<td>6.17</td>
<td>6.5</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>$FRR \approx 0$</td>
<td>36.85</td>
<td>34.67</td>
<td>1.25</td>
<td>0.75</td>
</tr>
</tbody>
</table>

A second configuration is also tested. It employs four training images for each client, two images for evaluation and another two images for testing. This configuration results in 400 evaluation client claims whereas the number of evaluation impostor claims, test client claims and test impostor claims are left intact. By adapting the just described training and evaluation procedures and by applying the previously presented test procedure, the ROCs plotted in Fig. 2(b) are obtained. It is seen that the EER is approximately 1.5% on the training set, 6.2% on the evaluation set and 5% on the test set.

Table 2 summarizes the performance of morphological elastic graph matching in terms of the FAR and FRR achieved at three operating points defined on the evaluation set, namely, $FAR \approx 0$, $FAR \approx FRR$, and $FRR \approx 0$.

The false rejection rates at approximately the same false acceptance rates in the test procedure for morphological elastic graph matching, standard elastic graph matching and optimized robust correlation. Both configurations are considered. By inspecting the entries in Table 3, it is seen that MDLA outperforms the elastic graph matching by 2.5 and 4% in Configurations I and II, respectively. The corresponding gain in FRR with respect to the optimized robust correlation is 1 and 1.75%, respectively.

4. Problems occurred in real-world tests

It is well-known that the conditions under which face databases are collected affect drastically the performance of face recognition/authentication algorithms [4,5]. A key to the successful development of a general face authentication system is to systematically account for the different acquisition parameters, i.e., the varying lighting, background, pose and scale of the face, etc. Accordingly, the evaluation of any authentication algorithm on a database collected under well-controlled conditions is inadequate when the algorithm is to be integrated in a platform for commercial exploitation. In such a case, it is of utmost importance to evaluate the authentication performance of the algorithm on real-world tests. Towards this goal two small galleries are collected under “real operating conditions”, namely, the MATRANORTEL database and the IBERMATIC database.

The experiments reported on MATRANORTEL database have been conducted on 21 persons. However, the size of the database progressively increases (51 persons are currently included in the database). Several sources of degradation are modeled in the database:

(a) Face size and position. In practice, it is very difficult to control the position of the subject with respect to the camera.
(b) Changes in illumination. If a spotlight is not used, lighting variations occur. For example close to a window, the lighting depends strongly on the day-time and the weather.
(c) Facial expressions. In practice, it is almost impossible to control the mood of the subject. The smile causes probably the largest image variation among all facial expressions.

In addition to images belonging to the just-mentioned cases, the database contains one set of training images (1 image per person) and one set of test images (2 images per person) recorded under well-controlled conditions. That is, a uniform white background exists in the images, uniform lighting conditions are used, the face is of neutral expression and is located at the center of the image. All images in the MATRANORTEL database

<table>
<thead>
<tr>
<th>Method</th>
<th>Configuration I</th>
<th>Configuration II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAR FRR FAR FRR</td>
<td></td>
</tr>
<tr>
<td>Elastic graph matching (residual matching) [21]</td>
<td>8.08 8.50 7.67 7.25</td>
<td></td>
</tr>
<tr>
<td>Morphological dynamic link architecture (without linear projections)</td>
<td>8.23 6 7.94 3.25</td>
<td></td>
</tr>
<tr>
<td>Optimized robust correlation [22]</td>
<td>7.76 7.25 5.71 5.75</td>
<td></td>
</tr>
<tr>
<td>Morphological dynamic link architecture (without linear projections)</td>
<td>7.61 6.25 5.52 4.00</td>
<td></td>
</tr>
</tbody>
</table>

1 It was collected by MATRANORTEL Communications, France.
2 It was collected by Ibermática, Spain.
are recorded in 256 grey levels and they are of dimensions 144 × 192 pixels. Sample images demonstrating the variable recording conditions are shown in Fig. 3.

A set of experiments has been implemented so that the performance of the morphological elastic matching is correlated to each degradation source. More specifically, 441 impostor claims and another 21 client claims have been tested under each degradation source. In addition to these claims, 781 impostor claims and 38 client claims have been tested under the just-mentioned well-controlled conditions. This is tantamount to 2545 impostor claims and 122 client claims in total. The fourth column in Table 4 summarizes the performance of morphological dynamic link architecture for each source of degradation separately and in the general case that encompasses all varying conditions. The next two columns describe the performance of the elastic graph matching [12] and the optimized robust correlation [13] for each source of degradation [23].

A set of similar experiments has been conducted on the IBERMATIC database. The experiments reported here have been conducted on 11 persons. It is worth noting that the size of the database progressively increases (17 persons are currently included in the database). For each person two training sample images are stored in the database. A large number of test images has also been recorded. In particular, more than 20 test images of each user have been recorded when he or she claims the correct identity (client images). In addition, a couple of test images per individual have been recorded when he or she pretends that he or she is someone else (impostor images). In total, 514 client accesses have been tested against 56 impostor ones. The three types of degradation previously described are present in this database as well. All images in the IBERMATIC database are recorded in 256 grey levels and they are of dimensions 320 × 240. Sample images are shown in Fig. 4. The performance of the elastic graph matching [12] and the morphological elastic graph matching has been evaluated on the IBERMATIC database. The EERs obtained using the just-mentioned techniques are shown in Table 5. From the inspection of Tables 4 and 5, it becomes evident that the performance of the algorithms developed within the M2VTS project depends strongly on the variable

![Sample images](image_url)

Fig. 3. Sample images from MATRANORTEL database demonstrating the recording of images under (a) lighting variations, (b) scale differences (c) varying face position, and (d) when facial expressions are not neutral.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Number of claims</th>
<th>EER (%)</th>
<th>MDLA</th>
<th>EGM</th>
<th>ORC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Client</td>
<td>Impostor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well controlled</td>
<td>38</td>
<td>781</td>
<td>12</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Lighting variations</td>
<td>21</td>
<td>441</td>
<td>33</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>Face size variations</td>
<td>21</td>
<td>441</td>
<td>28</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Varying face position</td>
<td>21</td>
<td>441</td>
<td>23</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Variable facial expressions</td>
<td>21</td>
<td>441</td>
<td>17</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>General</td>
<td>122</td>
<td>2545</td>
<td>22</td>
<td>21</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 4

Equal error rates achieved by elastic graph matching (EGM), morphological elastic graph matching (MDLA) and optimized robust correlation (ORC) when they are applied to MATRANORTEL database.
Table 5
Equal errors rates achieved by elastic graph matching and morphological elastic graph matching when they are applied to IBERMATIC database

<table>
<thead>
<tr>
<th>Method</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic graph matching [12,24]</td>
<td>25</td>
</tr>
<tr>
<td>Morphological elastic graph matching (MDLA)</td>
<td>35</td>
</tr>
<tr>
<td>Normalization + morphological elastic graph matching</td>
<td>20</td>
</tr>
</tbody>
</table>

recording conditions. To alleviate such a dependency, a compensation for the different conditions is needed. To do so, the use of simple and powerful face normalization techniques prior to the application of any authentication algorithm is proposed in Section 5.

5. Face normalization techniques

The proposed techniques are based on the detection of the facial region in the image and its splitting in two segments, the left segment and the right one. We assume that: (1) the background of the images is uniform, and (2) only one person appears in the scene. Fig. 5 demonstrates pictorially the steps of the algorithm that are described in detail below.

Step 1: The oval shape of a face can be approximated by an ellipse. Therefore, the detection of the facial area in an image can be performed by detecting an object of elliptical shape. To do so, first we have to discard the image background. The starting point is an edge detection algorithm. By thresholding the resulting image after edge detection, a zero value is assigned to the background. Since the background is uniform and does not contain any edges, it can easily be discarded by employing a grass-fire algorithm. Accordingly, the image is segmented into two regions, one of which contains the facial region and the other contains the background.

Step 2: The next step is to model the face-like region by an ellipse using moment-based features [26,27]. Let us denote the face-like area by \( C \) and the best-fit ellipse by \( \mathcal{E} \). An ellipse is defined by its center \((x_0, y_0)\), its orientation \( \theta \) and the length \( a \) and \( b \) of its semi-major and semi-minor axes. The center of the ellipse is estimated by the center of mass of the region \( C \). The orientation of the ellipse is computed by determining the angle between the axis of the least moment of inertia and the horizontal axis of the coordination system, i.e.,

\[
\theta = \frac{1}{2} \arctan \left( \frac{2\rho_{1,1}}{\rho_{2,0} - \rho_{0,2}} \right),
\]

where \( \rho_{i,j} \) denotes the \((i,j)\)-central moment of the region \( C \). The length of the semi-major axis \( a \) and the length of the semi-minor axis \( b \) can be computed by evaluating the least and the greatest moment of inertia, respectively. The least and the greatest moment of inertia of an ellipse with orientation \( \theta \), \( I_{\min} \) and \( I_{\max} \), are given by

\[
I_{\min} = \sum_{(x,y) \in C} [(y - y_0) \cos \theta - (x - x_0) \sin \theta]^2,
\]

\[
I_{\max} = \sum_{(x,y) \in C} [(y - y_0) \sin \theta + (x - x_0) \cos \theta]^2,
\]

where \( x, y \) denote the horizontal and vertical coordinate of a pixel. Accordingly, \( a \) and \( b \) are given by

\[
a = \left( \frac{4}{\pi} \right)^{1/4} \left[ \frac{I_{\max}}{I_{\min}} \right]^{3/8},
\]

\[
b = \left( \frac{4}{\pi} \right)^{1/4} \left[ \frac{I_{\min}}{I_{\max}} \right]^{3/8},
\]

respectively. To find the ellipse that models the given region best, we iteratively maximize the measure

\[
\mathcal{M} = \sum_{(x,y) \in \mathcal{E} \cap C} 1 - \sum_{(x,y) \in \mathcal{E} \cap \mathcal{C}} 1,
\]

where \( \mathcal{C} \) denotes the complement of the region \( C \) (i.e., the background). The maximization of Eq. (13) corresponds to the maximization of the number of correctly modeled pixels (i.e., \((x, y) \in \mathcal{E} \cap C \)) and the minimization of the number of incorrectly modeled pixels (i.e., \((x, y) \in \mathcal{E} \cap \mathcal{C} \)). In order to find a maximum of \( \mathcal{M} \) the following recursive procedure is proposed.

Step (a). Calculate the parameters of the initial ellipse \( \mathcal{E}_0 \) and the measure \( \mathcal{M}_0 \) for the initial region \( C_0 \).
Step (b). Find the new region \( C_i = C_{i-1} \cap \mathcal{E}_{i-1} \).
Step (c). Calculate the parameters of the ellipse \( \mathcal{E}_i \) that fits \( C_i \) and the measure \( \mathcal{M}_i \).

Fig. 4. Sample images from IBERMATIC database exhibiting scale and lighting differences.
Fig. 5. Pictorial description of the proposed face normalization algorithm.

Step (d). If \( M_i > M_{i-1} \) go the Step (b). Otherwise, the best ellipse is \( E_{i-1} \) that has already been found.

By using this iterative algorithm the ellipse fitting becomes more robust to noise, i.e., to pixels that correspond to clothes, hair, etc.

Step 3: The first ellipse found is a coarse approximation of the facial area, because the hair, and in some cases, parts of the clothes are included in it. Thus, the skin region can be segmented. To overcome this problem the ellipse is subdivided into its left and right segments with respect to the vertical axis. Moreover, the subdivision aids to compensate for the different lighting conditions which effect unevenly the two parts of the face.

Step 4: The next step is to apply a clustering algorithm to each segment of the ellipse, separately. By choosing \( K = 2 \), a \( K \)-means algorithm hopefully succeeds to relate the skin-like area with a single cluster in each segment. Thus, the skin region is detected accurately. The mean intensity values of each face side are also estimated. They provide information for the lighting conditions during the recording procedure.

Step 5: The union of clusters in the left and right segments that correspond to skin-like areas is modeled by an ellipse using the algorithm described in Step 2. The
quality of fit is measured again by Eq. (13). By applying Steps (a)–(d), a finer approximation is obtained. The ellipse found at the last iteration is the best-fit ellipse we searched for.

In addition to the use of the just-described technique in facial area detection, the method provides: (i) an estimate of the face center that can be used in compensating for face translation, (ii) an estimate of the face width which is related to the length of the minor axis of the best-fit ellipse, and (iii) an estimate of the main intensity value of each segment of the face (left and right) which is related to the lighting conditions of the recording procedure.

5.1. Illumination normalization

Lighting may cause uneven illuminations of the right and the left face segments. We assume that the illumination conditions in the left and right face segments are uniform. To compensate for the aforementioned effect, the mean intensities of both face segments should be equalized. Let \( m_L \), \( m_R \) be the main intensity values in the left and the right segment, respectively. The initial image \( I(x, y) \) is transformed so that the left and right segments of the normalized image, \( I_N(x, y) \), have the same (desired) mean intensity \( I_d \):

\[
I_N(x, y) = I_d \left( \frac{1/m_R - 1/m_L}{1 + \exp \left( \frac{(x_0 - x)/\lambda}{1/m_L} \right)} + \frac{1/m_L}{1 + \exp \left( \frac{(x_0 - x)/\lambda}{1/m_L} \right)} \right) I(x, y),
\]

where \( \lambda \) controls the slope of the sigmoidal function that appears in the denominator of the first fractional term inside parentheses. Let \( L \) be the image region in which \( \exp \left( \frac{(x_0 - x)/\lambda}{1/m_L} \right) \to \infty \) and \( R \) be the image region in which \( \exp \left( \frac{(x_0 - x)/\lambda}{1/m_L} \right) \approx 0 \). These regions correspond to the left and the right segments of the best-fit ellipse modeling the face, respectively. It can be easily proven that:

\[
E[I_N(x, y), (x, y) \in L] = E[I_N(x, y), (x, y) \in R] = I_d,
\]

which satisfies our objective. Fig. 6 demonstrates the illumination normalization achieved when images from MATRANORTEL database are used.

5.2. Face position and size

Varying face position and size are easily compensated for, if the face is accurately approximated by an ellipse. The problem of varying face position can be solved by translating the initial image so that the center of the ellipse always coincides with the image plane center. The width of the face can be approximated by the length of the minor axis \( 2b \). Size (i.e., scale) normalization can be achieved by resizing the image with a horizontal scale factor \( W_d/2b \), where \( W_d \) is the desired width of the normalized face. Image resizing is achieved by linear interpolation. Fig. 7 shows the compensation for scale variations in images from MATRANORTEL database.

The successful compensation for varying face position in images from MATRANORTEL database is demonstrated in Fig. 8. Examples of normalized face images from IBERMATICA database are depicted in Fig. 9.

5.3. Impact of face normalization on the authentication performance of morphological elastic graph matching

To quantify the success of the face normalization technique we evaluate the ROC when the proposed normalization technique is incorporated into the morphological elastic graph matching on the MATRANORTEL and the IBERMATICA databases. The ROCs of morphological elastic graph matching with and without face normalization for the three degradation sources, namely, the changes in lighting, the varying face size and position, are plotted in Fig. 10. In each subplot, first the ROC without applying the proposed face normalization technique is given. The ROC, when we compensate for the specific source of degradation, is plotted next. Finally, the
Fig. 7. Facial images possessing scale differences. The best ellipses determined by the algorithm are drawn on the original images. The normalized images are shown in the second and fourth columns.

Fig. 8. Facial images exhibiting differences in face position. The best ellipses determined by the algorithm are drawn on the original images. The normalized images are shown in the second and fourth columns.

Fig. 9. Normalized images corresponding to the sample images of Fig. 4 from IBERMATICA database.

ROC when we simultaneously compensate for the three sources of degradation is also shown. The gain in EER when the proposed face normalization technique is employed is shown in Table 6 for each degradation source separately as well as in the general case. A significant drop 7.3% in EER (which amounts to a 33% relative drop) is achieved in the general case. For comparison purposes, the EERs achieved without normalization are also included.

The ROC, when we compensate for the three types of degradation on the IBERMATICA database is plotted in Fig. 11. The ROC without applying the proposed face normalization technique is overlaid in the same plot as well. The EER drops to 20% when the face normalization technique is employed. That is, a significant drop of 15% in EER (which amounts to a 43% relative drop) is achieved.

6. Conclusions

A thorough assessment of morphological elastic graph matching as a frontal face authentication algorithm on four databases has been presented. The databases range from small galleries to large multimedia ones collected under either well-controlled or real-world conditions. It
has been shown that the morphological elastic graph matching achieves a very low equal error rate on databases recorded under well-controlled conditions (3.7–5%). However, its performance deteriorates when it is applied to real-world databases. A face normalization method that succeeds to compensate for lighting changes, varying face size and position has been proposed. It guarantees an almost stable authentication performance of morphological elastic graph matching in field tests under any recording conditions. As a by-product image variations attributed to different facial expressions are compensated for as well.

7. Summary

In this paper, morphological elastic graph matching is applied to frontal face authentication on four databases.
Two of them are multimedia databases collected under controlled conditions. The first one contains audio and video data of 37 persons in four shots. The second database is extended to audio and video data of 295 persons recorded at eight time instants. The remaining databases are extended to audio and video data of 295 persons recorded at eight time instances. The remaining two databases are small galleries recorded during real-world tests, such as access-control to buildings and cash dispenser services or access-control to tele-services via Internet in a typical office environment. It is demonstrated that the morphological elastic graph matching achieves a very low equal error rate on databases collected under well-controlled conditions. However, its performance deteriorates when it is applied to databases recorded during field tests. The compensation for image variations attributed to variable recording conditions, i.e., changes in illumination, face size differences and varying face position, is addressed next. The use of simple and powerful pre-processing techniques aiming at compensating for the aforementioned variations prior to the application of morphological elastic graph matching is proposed. The results obtained indicate that such an approach overcomes the image variations and stabilizes the performance of the authentication algorithm.

References