Comparison of Eigenface-Based Feature Vectors under Different Impairments

Aristodemos Pnevmatikakis, Lazaros Polymenakos
Athens Information Technology Institute
Autonomic Computing Laboratory
Peania 19002, Athens, Greece
apne@ait.gr, lcp@ait.gr

Abstract

We study the performance of a new eigenface-based method for face recognition. Specifically, we perform DCT preprocessing followed by the PCA-LDA combination. We compare the new method to existing ones (PCA, PCA-LDA, DCT-PCA) under impairments like changes in brightness, direction-of-illumination, hairstyle, clothing, expression, head orientation, and added noise.

1. Introduction

Face recognition has an ever growing number of applications; for example, pilot systems are evaluated at public transportation terminals. If performance could be boosted to that achieved by other biometric systems, face recognition systems would rank among the most desirable, because of the non-intrusiveness and the simplicity of the infrastructure required.

Face recognition techniques have been classified in [1] into feature-based and appearance-based. We focus on appearance-based methods that are global: they treat the face as a single entity and transform it into a feature vector. After training, the feature vectors are classified to represent different persons. Common feature extraction techniques for appearance-based face recognition are: eigenfaces [2], Gabor wavelets [3] and DCT [4].

Eigenfaces use Principal Component Analysis (PCA) [5] to reduce the dimension of the feature vector (vectorized image). PCA maximizes the total scatter of the training vectors while reducing their dimensions. The resulting feature vectors are robust to noise and minor head rotations, but performance suffers from changes in the direction-of-illumination [4,6].

Since its introduction in 1991, the eigenface technique has seen many modifications [6,7]. It was realized that successful recognition relies on minimizing the within-class scatter. This is achieved by Linear Discriminant Analysis (LDA) [5]. A PCA-LDA scheme was proposed in [6] which proved robust under direction-of-illumination changes.

The robustness of the DCT-based feature vectors to direction-of-illumination changes inspired a new method [4] with DCT preprocessing of the images followed by the eigenface technique. The method was reported [4] to achieve robustness both for noise and for direction-of-illumination changes.

We experimented on DCT preprocessing followed by PCA-LDA, and realized that performance analysis of four eigenface-based techniques is required: (i) PCA, (ii) DCT-PCA, (iii) PCA-LDA and (iv) DCT-PCA-LDA.

In this paper, we briefly introduce the DCT-PCA-LDA method and provide exhaustive performance comparison of the four eigenface-based methods, under different types of impairments: noise, direction-of-illumination changes, variations in clothing and hair style, and medium changes of head orientation (tilts up-down, left-right, slight rotations). Our motivation is to evaluate the new method, establish which methods work best under which impairment(s) and, if possible, identify a method that is robust for all cases. The results are obtained using two different face databases: the Aberdeen database from University of Stirling [8], and the ORL database from University of Cambridge/AT&T [9]. These databases suffer from different impairments. We also added noise and artificial direction-of-illumination changes. The use of both databases enables the assessment of the suitability of each feature extraction technique to the different inherent impairments.

The paper is organized as follows: In section 2 the four feature extraction methods are outlined and in section 3 the two databases are presented and compared. In section 4 the performance of the four feature extraction methods is evaluated under all the mentioned impairments, followed by the conclusions in section 5.

2. Eigenface-based features

According to the eigenface feature extraction method [2], the \( N \) images of the training set are converted to \( N \) column vectors \( \mathbf{x} \), of length \( D \) equal to the number of pixels in the images. If their mean is \( \mathbf{u} \), then the total scatter is given by the \( D \times D \) matrix...
The eigenvectors of $S_T$ (normalized to unity norm) that correspond to the $D_{PCA} \leq D$ largest eigenvalues form the linear transformation matrix $W_{PCA}$ that projects the face vectors into the space where the transformed total scatter matrix $W_{PCA}^T S_T W_{PCA}$ has maximum determinant, i.e. the total scatter is maximized. This is the PCA technique [5]. In practice, $D_{PCA} \ll D$ is chosen and PCA achieves feature space dimension reduction. PCA maximizes the total scatter, hence it is a good technique for representation. It does not guarantee improvement of the discrimination of the classes; indeed, in many cases the classes are smeared, hindering classification.

The way around the discrimination problem of PCA is LDA [5]. Here the ratio of the between-class scatter over the within-class scatter is maximized. The between-class scatter matrix $S_B$ is defined as

$$S_B = \sum_{i=1}^{K} n_i (\mu_i - \mu)(\mu_i - \mu)^T$$

(2)

where $K$ is the number of classes, $\mu_i$ the mean vectors of each class and $n_i$ is the number of training faces per class. The within-class scatter matrix $S_W$ is defined as

$$S_W = \sum_{i=1}^{K} \sum_{k=1}^{n_i} (x_i^{(k)} - \mu_i)(x_i^{(k)} - \mu_i)^T$$

(3)

where $x_i^{(k)}$, $i=1,\ldots,n_i$ are the training vectors of the $k$-th class. Thus, LDA tries to simultaneously increase the total scatter and decrease the scatter of each class. Therefore classification is enhanced. The transformation matrix that achieves this is $W_{LDA}$, and has the eigenvectors of $S_w^{-1} S_B$ as columns. Only the largest $K-1$ eigenvalues of $S_w^{-1} S_B$ are non-zero. The rank of $S_w$ is less than $N-K$ and in general $N \ll D$. Hence $S_w$ is singular and this technique is not directly applicable to face recognition [6].

The way around the singularity problem of $S_w$ is to first apply PCA to reduce the dimension of $x_i$ from $D$ to $N-K$ and then to apply LDA to further reduce the dimension to $K-1$ [6]. $S_w$ is now a $(N-K) \times (N-K)$ non-singular matrix.

PCA and/or LDA can be applied on the values of the DCT coefficients of the images, instead the images themselves [4]. To do so, DCT is applied on 8 by 8 blocks with 50% overlapping. The coefficients are zig-zag scanned as in JPEG. The first three coefficients of each block are replaced by the three horizontal and three vertical Delta coefficients, as proposed in [4]. We characterize the DCT features by $6d+nDCT$, where $n \in [0,61]$ is the number of the retained coefficients after discarding the three first coefficients.

3. Aberdeen and ORL databases

The Aberdeen database comprises 469 frontal facial images of 29 different persons. There are from 11 up to 19 images of each person. Their size is 56 by 48 pixels, taken during different times, so there are changes in hairstyle and upper-torso clothing. They also suffer from large variations in brightness. On the other hand they are frontal, with minor in-plane rotations and mostly neutral expressions. Example images of a single person from this database are shown in Figure 1.

![Figure 1. A person from the Aberdeen database](image)

The ORL database comprises 400 frontal facial images of 40 different persons. There are 10 images of each person. They are 112 by 92 pixels and vary in head orientation (tilts up-down, left-right, slight rotations) and expression. On the other hand, they have uniform within-class brightness. They are more closely cropped, so hairstyle and clothing are irrelevant. Example images of a single person from this database are shown in Figure 2.

![Figure 2. A person from the ORL database](image)

The impairments of the two databases are summarized in Table 1.

Table 1. Within-class impairments exhibited by the two databases.

<table>
<thead>
<tr>
<th>Impairment</th>
<th>Aberdeen</th>
<th>ORL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hairstyle</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Torso</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Clothing</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Brightness</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Illumination</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Rotations &amp; tilts</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

We use both of them to assess the performance of the four feature extraction methods under different impairments. Also, both can be contaminated with noise and can have
artificial direction-of-illumination changes added to them, using a variation of the technique in [4].

4. Comparison of the Methods

The probability of misclassification (PMC) of all four feature extraction techniques on both databases is first determined. We performed 200 runs with random selection of the training images per class and the averaged results are shown in Table 2. The performance of the non-LDA features is grossly different for the two databases. For the Aberdeen database, PCA alone is much worse - by a factor of 3 - than PCA-LDA showing that PCA is not capable of handling the types of impairments found in the Aberdeen database: changes in hairstyle, in upper-torso clothing and in brightness. On the other hand, PCA alone is just 29% worse than PCA-LDA for the ORL database, indicating that PCA alone can handle more successfully the types of impairments found in the ORL database (expression changes and out-of-plane head rotations).

Next, we equalized the within-class brightness variation of the Aberdeen database, in order to determine if it plays detrimental role to PCA. This improves the PCA performance by 50% and the PCA-LDA performance by 102%. Hence, the PCA performance loss on the Aberdeen database is not due to brightness but mainly due to the changes in hairstyle and upper-torso clothing.

A second observation is the pronounced difference in the performance of the DCT preprocessing when applied to the two databases. For the Aberdeen database, DCT improves PCA performance by 38% and PCA-LDA performance marginally. However, for the ORL database, DCT worsens PCA performance by 57% and PCA-LDA performance by 44%. This performance loss is attributed to the six delta elements of the DCT since if we repeat the experiment using DCT without delta coefficients there is no loss in performance. We conclude that delta features cannot be used when there are expression changes and out-of-plane head rotations. We also note that brightness equalization marginally changes the performance of the DCT preprocessed features.

Table 2. PMC of the four features on both databases without any additional impairment. 7 training images per class are used.

<table>
<thead>
<tr>
<th>Feature extraction</th>
<th>ORL</th>
<th>Aberdeen</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>6.81</td>
<td>24.6</td>
</tr>
<tr>
<td>6d+26DCT PCA</td>
<td>15.9</td>
<td>17.8</td>
</tr>
<tr>
<td>PCA-LDA</td>
<td>5.28</td>
<td>6.26</td>
</tr>
<tr>
<td>6d+26DCT PCA-LDA</td>
<td>9.45</td>
<td>6.22</td>
</tr>
</tbody>
</table>

The dependence of the performance of the PCA and the PCA-LDA to the number of training images per class is also studied. The results are shown in Figure 3. The PCA features are less sensitive to the number of training images per class. This is expected since LDA requires more parameters, i.e., a robust estimate of both within-class and between-class scatter. Regarding the ORL database (where for large training sets the performance of the PCA is comparable to PCA-LDA) PCA outperforms PCA-LDA for training sets of 3 or less images per class.

The use of DCT preprocessing allows keeping less than 64 values per DCT block to form the input vectors for the PCA. After zig-zag scanning the DCT coefficients of each block, only the first \( n \) are kept. The first 3 of these are replaced with 6 delta coefficients. Varying \( n \) results to different PMC and feature space dimension for the following PCA. In Figure 4, the PMC of the DCT preprocessing followed by PCA-LDA relative to that of plain PCA-LDA is plotted as a function of the feature space dimension relative to that of no DCT preprocessing. Down to 85% of the feature space (6 Delta and 10 DCT coefficients) there is no performance loss.
than the PCA-LDA features. Similar results are obtained for the ORL database. At low SNR values (less than 6 dB for the Aberdeen, 4 dB for the ORL) PCA is better than PCA-LDA.

![Figure 5. PMC of the four features as a function of the SNR. 10 training images per class are used.](image)

Finally, the direction-of-illumination is changed artificially. Instead of using a linear brightness gradient as in [4], we add Gaussian gradients from left-to-right, right-to-left and centre-to-edges. The strength of the artificial illumination is determined by the maximum value of the gradient, $b$. As a result, the size of the databases is quadrupled; to obtain averaged results, 800 runs are used. The results on the Aberdeen database for different values of $b$ are given in Table 3. We observe that PCA cannot cope with direction-of-illumination changes. DCT preprocessing helps PCA considerably, but the combination still remains far worse than PCA-LDA. The latter does not improve with DCT preprocessing.

<table>
<thead>
<tr>
<th>Feature extraction</th>
<th>$b=0$</th>
<th>$b=0.3$</th>
<th>$b=0.4$</th>
<th>$b=0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>23.4</td>
<td>61.4</td>
<td>72.7</td>
<td>83.4</td>
</tr>
<tr>
<td>6d+26DCT PCA</td>
<td>15.9</td>
<td>19.9</td>
<td>30.2</td>
<td>45.5</td>
</tr>
<tr>
<td>PCA-LDA</td>
<td>3.78</td>
<td>3.54</td>
<td>3.91</td>
<td>4.98</td>
</tr>
<tr>
<td>6d+26DCT PCA-LDA</td>
<td>3.87</td>
<td>3.65</td>
<td>4.05</td>
<td>5.19</td>
</tr>
</tbody>
</table>

Table 3. PMC of the four features on the artificially illuminated Aberdeen database. 10 training images per class are used.

5. Conclusions

In this paper the performance of two modifications to the eigenface feature extraction method are evaluated alone and in combination, resulting to four different feature extraction techniques.

PCA alone performs well under minor head rotations, expression changes and noise. Different hairstyles and clothing degrade its performance, while direction-of-illumination changes destroy it completely.

The PCA-LDA combination performance is superior to just PCA under minor head rotations and expression changes, but its superiority is more evident under different hairstyle and clothing and direction-of-illumination changes. Under extreme noise its performance degrades faster than PCA alone, and requires more training.

The DCT preprocessing prior to PCA is helpful under different hairstyle and clothing. It is mandatory under direction-of-illumination changes. Unfortunately, it cannot be used under minor head rotations and expression changes, as it degrades performance.

Finally the DCT preprocessing prior to the combination of PCA-LDA provides minor improvements under some impairments, except for head rotations and expression changes, where it degrades performance.

Since the use of DCT preprocessing does not always yield successful feature vectors, and when it does, they are only marginally better than the PCA-LDA combination alone, it should not be used in general. The most successful feature vectors are obtained using the combination of PCA-LDA. DCT pre-processing is only useful under controlled conditions, e.g. for verification systems, where the face images are frontal and the expressions neutral.

References


