A new Face Database and Evaluation of Face Recognition Techniques

D. ALEXIADIS¹, V. SYRRIS², A. PAPASTERGIOU³, A. HATZIGAIDAS⁴, L. MARIUTA
Technological Educational Institute of Thessaloniki
Department of Electronics
Thessaloniki, P.O. Box 141, 57400 GREECE
¹dalexiad@mri.ee.auth.gr, ²vsyrris@auth.gr, ³natp@el.teithe.gr, ⁴sakis@el.teithe.gr

Abstract: This paper introduces ELA5, a new image database that is suitable for experimentation within the face recognition domain. Its design attempts to cover a variety of scenarios encompassing pose and illumination variations, different facial expressions and occlusion. In addition, established computational techniques such as Principal Component Analysis (PCA) and Multilinear PCA, combined with Fisher/Linear Discriminant Analysis (LDA) are evaluated and comparative results display their strengths and weaknesses in settings that simulate real-world conditions.

Key–Words: Face Database; Face Recognition; Biometric Identification; Image Processing

1 Introduction

In computer vision, the term face recognition refers to the automated process of human face identification or verification, based on input data derived from still images or sequences of images (videos). The processing chain is depicted at Fig. 1. As one can see, the processes of face detection and feature extraction constitute prerequisite steps for successful face recognition. Nevertheless, there are applications such as face tracking and monitoring of human body parts where these two processing stages may stand alone.

In face detection, the system checks the existence of objects that resemble human faces. It locates the image regions that seem to contain the human faces. Most of the algorithms in this category employ either feature-based approaches such as skin-color [1], or image classification approaches such as neural network-based detection [2] and support vector-based detection [3]. Some indicative survey papers are [4,5]. A very fast face-detection approach, that is used in the current paper, can be found in [6–8].

Face recognition is a typical classification task. As such, an abstract representation of the face is required that assists a classifier to execute partial matching of the faces. This representation usually takes the form of a set of distinctive features that are extracted from the face and the respective processing stage is named feature extraction. Depending on the extraction process and the face matching, the techniques are divided into a) appearance-based approaches, such as eigenfaces [9–14], feature-based [15], wavelet-based [16], component-based [17] and b) model-based approaches, such as elastic bunch graph matching [18], 2D modeling [19] and 3D modeling [20]. Face description and feature extraction constitute an active research area: the definition of quantities which characterize uniquely a person’s face and remain unvaried in different shooting conditions is proved to be a very intricate problem.

Depending on the specific application requirements, the final processing stage is either the face identification or the face verification (authentication). The recognition takes place by using one classifier [21,22] or an ensemble of classifiers [23] which learn to match images of faces through training that is based on the extracted features. The recognition rates depend heavily on the robustness of the chosen features. An analytical survey on this field can be found in [24].

Face recognition is an ill-defined problem due to several reasons: intense variations in illumination, shape scaling and head pose, camera distortion and noise, different makeup and hair style, aging problem, multiple facial expressions, complex background and occlusions produced by objects such as glasses, hat, earrings, ribbon, etc. The aforementioned factors can severely mislead the classifiers.

In this paper we introduce a new image database which aims to provide a testing bed for the evaluation of face recognition techniques. It simulates the basic real-world camera shooting scenarios such as different views and illumination, facial...
expressions and occlusion (glasses and hat). The database is created for research purposes at the Department of Electronics, Technological Institute of Thessaloniki and it is publicly available at http://mycad.c5lab.el.teithe.gr/ELA5DBASE.rar. We also provide indicative results derived from the experimentation with classical techniques: the Principal Component Analysis (PCA) and the Multilinear PCA, combined with Fisher/Linear Discriminant Analysis [11]. The last section contains results, discussion and conclusions.

2 The new ELA5 database

Several image databases have been developed in order to play an objective benchmark platform for the evaluation of face recognition techniques. A complete list with analytic description can be found at http://www.face-rec.org/databases. Our motivation for the creation of a new face database was to simulate in a high degree the shooting conditions that exist in real-world situations.

2.1 Database creation setup

The face of the persons was recorded in a physical mode (natural movement of the head, with different illumination, etc.) without attempting to maintain scenery similarities in an artificial way. The set consists of 1260 images, obtained for ten persons (six males and four females) in the following five different scenarios:

- 14 images with physical light and viewing angles from -75° to +75°.
- 14 images with physical light and the person wearing glasses (-75° to +75°).
- 14 images with physical light and the person wearing a hat (-75° to +75°).
- 14 images with dim light, provided by a lamp at the right side of the person (-75° to +75°).
- 10 images with dim light, provided by the lamp and the person wearing glasses (-40° to +40°).

The database creation setup is illustrated in Fig. 2. Some example images for the described scenarios are given in Figs 3-5.

2.2 Face detection and cropping

Given that any face recognition system performs face detection before classification, the software we developed was programmed to perform face detection and cropping, during image capturing. Both the original and the cropped images are available. The software detects the face rectangle, calculates its center and crops the surrounding image area of size 140 × 180, as shown in the right of Fig. 6. The Open Computer Vision (OpenCV) library v.2.0.0a (http://sourceforge.net/projects/opencvlibrary) was exploited for face detection. Some false detections (although few) are possible to be produced by the algorithm. Consequently, sometimes more than one “potential” faces are found. Therefore, the face detection outputs were further processed exploiting an eye detection strategy, which uses the OpenCV algorithm. Any candidate detected face that does not contain at least one eye is rejected.

The algorithm implemented in the OpenCV library is based on the works of [6–8] and operates extremely rapidly, achieving high detection rates. A set of haar-like features is extracted from an image, using 14 prototype convolution masks, which are scaled in vertical and horizontal directions in order to generate a rich set of features. The term “haar-like” is employed because the prototype functions are “black and white”, similar to the haar basis functions. The applied prototype functions can extract edge-like, line-like and center-surround features. The computational cost is largely decreased by introducing the notion of the integral image [7]. Furthermore, based on the ex-
extracted features, a cascade of binary (true/false) classifiers is used to classify an image rectangle as a face or a non-face. The classifier of each stage was trained to detect almost all “potential” faces, while rejecting a small fraction of the non-faces patterns. This means that it produces very high hit-rates but also relatively high false-positive rates. The cascade of many such classifiers results in a “strong” classifier. The cascade is trained using the Discrete Adaboost algorithm. For a detailed analysis the reader is referred to [6–8].

3 Employed face recognition approaches

3.1 Principal Component Analysis (PCA)

3.1.1 Lexicographic reshaping

In order to apply the standard 1-D PCA methodology, the 2-D face image representation $I(n_x, n_y)$ is simplified to an 1-D representation:

$$I(n) \leftarrow I(n_x, n_y), \quad n = 0, 1, \ldots, N - 1,$$

$$n_y = \lfloor \frac{n}{N_x} \rfloor, \quad n_x = n - n_y \cdot N_x, \quad (1)$$

where $N = N_x \cdot N_y$ and $\lfloor \cdot \rfloor$ represents the floor operation. Namely, the lexicographic ordering shown in Fig. 7, is employed. The $m$-th face image is represented by the vector $I_m(n)$ of length $N$, i.e. each image is represented as a point in $\mathbb{R}^N$.

3.1.2 1-D PCA analysis

Let $A$ represent the $N \times M$ matrix with its $M$ rows being equal to the vector images $I_m$. According to PCA analysis, the covariance matrix $C = AA^T$ of the image data is calculated. Its singular value decomposition (SVD) follows, such that $C$ is written as:

$$C = UV^T$$  \hspace{1cm} (2)

where $U$ and $V$ are orthogonal. When $M = N$, i.e. $C$ is square matrix, the columns $V_j, j = 1, \ldots, N$ of $V$ are equal to the eigenvectors of $C$ and $U = V^{-1}$. The matrix $A$ is a diagonal matrix containing the corresponding eigenvalues. The $j$-th eigenvalue corresponds to the variance of the projection of the training set to the $j$-th eigenvector. If $M < N$, there will be only $M$ meaningful eigenvectors (the remaining eigenvalues will have associated eigenvalues equal to zero).

Assuming that the eigenvalues are sorted in descending order and according to PCA, each vector image $I_m$ is represented by its projection to the only $M_1 << M$ first eigenvectors $V_j, j = 1, \ldots, M_1$ (principal components). The set of $V_j, j = 1, \ldots, M_1$ constitute the best undercomplete ($M < N$) orthonormal basis (with $M_1$ basis functions) for the space of the training vector images, in the least square sense.

In this way, the dimensionality of the vector images is much reduced. Furthermore, it is commonly assumed that the eigenvectors corresponding to small eigenvalues correspond to noise in the training dataset. For a detailed analysis, refer to [9, 10].
3.2 Multilinear PCA (MPCA)

3.2.1 The problems with lexicographic reshaping

As described, the naive application of PCA to a set of images requires their reshaping into vectors, using lexicographic ordering. In most of the cases, these vectors’ length is very high (e.g. for an image of size $256 \times 256$ the corresponding lexicographic vector is of length $N = 2^{16}$), obviously resulting in high processing cost (curse of the dimensionality). Beyond implementation issues, it is well understood that reshaping breaks the natural structure and correlation in the original data, removing the pixel dependencies along the $y$-direction. Therefore, it results potentially in a less compact or useful representation than those that can be obtained using the images in their original form. Reshaping, as PCA preprocessing, ignores the fact that images are naturally 2-D objects. Therefore, it is natural and desirable to use a PCA approach that exploits directly the images in their native matrix representation.

3.2.2 The MPCA approach in a nutshell

We use and evaluate versus the classic PCA, the Multilinear PCA approach, described in [13]. The authors of [13] propose a general multidimensional PCA-like framework, where the multi-dimensional objects, called tensor objects (i.e. gray images are 2nd-order tensors, gray image sequences are 3rd-order tensors, etc.), are treated in their native multi-dimensional form. The authors present a general framework and an iterative algorithm for an arbitrary order of the tensor objects. We have to highlight that for second order tensor objects, the MPCA algorithm is very similar to the matrix-rank reduction technique of [14]. Thus, we briefly present the related theory using matrix notations, similarly to [14]. For a detailed analysis, a curious reader is referred to [13].

Here, each image $I_m(n_x, n_y)$ is represented by a matrix $I_m \in \mathbb{R}^{N_x \times N_y}$ and the collection of data is represented by a sequence of such matrices. The problem is the approximation by a sequence of matrices $D_m \in \mathbb{R}^{p_1 \times p_2}$ of lower rank. The aim is to compute two orthogonal matrices $U_L \in \mathbb{R}^{N_x \times p_1}$ and $U_R \in \mathbb{R}^{N_y \times p_2}$, with $p_1 < N_x$ and $p_2 < N_y$, such that the sum of the norms:

$$\sum_{m=1}^{M} ||U_L \cdot D_m \cdot U_R^T - I_m||^2$$

is minimized. In the above equation, $D_m = U_L^T \cdot I_m \cdot U_R$ is the feature vector of the $m$-th image datum $I_m$. The projection matrices $U_L$ and $U_R$ constitute a two-sided linear transformation on the data. Two iterative algorithms for the estimation of $U_L$ and $U_R$ are presented in [14] and [13], which are almost identical, except from the initialization procedure and the termination criterion. In our experiments we use the one of [13].

3.2.3 Selection of features

**MPCA1:** With the MPCA approach described so far, a set of features with reduced dimensionality ($p_1 \times p_2 < N_x \times N_y$) is obtained. Within a simple unsupervised framework, the features are sorted in descending order, with respect to their variance for the training set and only the first $P$ features are utilized by the classifier. Namely, this approach does not take into account the class labels (ground truth). In this way, the variation captured in the projected subspace includes both the intra-class variation and the inter-class variation.

**MPCA2:** In a classification task, a large inter-class variation and a small intra-class variation are desirable. Therefore, the authors in [13] propose a feature selection strategy with respect to their class discrimination power, which is calculated as the ratio of the inter-class scatter over the intra-class scatter.

**MPCA+LDA:** Instead of using directly the extracted features for classification, a Linear Discrimination Analysis (LDA) [11–13] can be utilized. LDA seeks a projection matrix $V$ that linearly projects the feature space onto a new space where the discrimination (the ratio of the inter-class over the intra-class scatter) is maximized. We experimented also with this approach.
4 Experiments and results

4.1 Classifier and distance metrics

After the data are projected in the feature space by one of the described approaches, the simple nearest-neighbor classifier is used. Any distance metric can be utilized. We experimented with a) the city block ($L_1$), b) the Euclidean ($L_2$) and c) the cosine distance metrics. They produced very similar results, with the $L_1$ metric performing slightly better. Therefore, we present the results obtained using the $L_1$ metric.

4.2 Datasets and evaluation procedure

As described, the database contains two sub-datasets: One with normal light conditions for three scenarios (42 images per person) and one with directional dim light for two scenarios (24 images per person). Each face image was horizontally flipped, so that the total number of available images per person becomes 132.

In order to evaluate the performance of the described feature extraction techniques, we realized two series of experiments:

1. For each person, we select randomly 45 out of the 132 available images per person. These constitute the training set. Namely, the training set contains examples from all angle views, light conditions and scenarios. The remaining images constitute the evaluation set. Since the selection is random, each experiment is realized 20 times and the mean classification accuracy is calculated and presented.

2. For the training set, we select only the small angle views (-30° to +30°) for normal light conditions. This leads to a training set with $2 \times 3 \times 8 = 48$ images per person. The remaining images constitute the evaluation set. With this type of experiments we study a) The capability of the used feature extraction methods in “extrapolating” and generalizing for larger angle views and b) their robustness against light condition changes.

4.3 Results and discussion

The results are given with respect to the number of the features utilized by the classifier. The number of used features varies in the interval [9,33]. The results are presented in the diagrams of Figs. 8 and 9 for the two experimental settings, respectively. As one can notice from both diagrams, the MPCA-based approaches outperform significantly the PCA-based one, as expected.

![Figure 8: Results for the 1st series of experiments.](image1)

![Figure 9: Results for the 2nd series of experiments.](image2)

**First setting:** When the training examples are selected randomly from all angle views and light conditions available images, the MPCA-based approaches can achieve relatively high correct recognition rates, even with a small number of used features. Especially, if the MPCA features are selected according to their discrimination power, or when MPCA is combined with LDA, the recognition rates reach 95%. The PCA-based approach works relatively worse. These results indicate that the described methodologies, when given an adequately large training set, can “interpolate” for angles in-between those present in the training set.

**Second setting:** When the training examples refer only to a small interval of viewing angles (-30° to +30°) and normal light conditions, the performance of all approaches diminishes significantly. This result highlights the fact that conventional PCA-like techniques
cannot “extrapolate” to larger viewing angles. Also, they are sensitive to illumination conditions changes. Additionally, according to the diagrams the use of LDA worsens the situation, instead of improving it. Therefore, for such scenarios, it would be reasonable to search for feature extraction methodologies and classifiers that are less sensitive to illumination changes and viewing angles.

5 Conclusion

We introduced a new face database, designed to cover a variety of scenarios encompassing pose and illumination variations, different facial expressions and occlusion. The proposed ELA5 database is continuously growing. In addition, established techniques such as PCA and MPCA, combined with Fisher LDA were evaluated. The results indicate their strengths but also their weaknesses in settings that simulate real-world conditions. In order to further study the face recognition problem in real-world situations, as a future work we plan to develop a real-time face-recognition system on a Graphical Processing Unit (GPU).

References: