ABSTRACT:

Like all biometrics solutions, face recognition technology measures and matches the unique characteristics for the purposes of identification or authentication. Often leveraging a digital or connected camera, facial recognition software can detect faces in images, quantify their features, and then match them against stored templates in a database. Facial recognition doesn’t just deal with hard identities, but also has the ability to gather demographic data on crowds. This has made face biometrics solutions much sought after in the retail marketing industry. In this paper, we presented a comprehensive review of Face recognition system and the face datasets used in the face biometric.

Keywords: Face Recognition, PCA, ICA, LDA, Kernel, EBGM, EP, Trace transform, AAM, 3D

[1] INTRODUCTION

Facial recognition is a type of biometric procedure that can identify a specific individual in a digital image by analysing and comparing patterns. So many computer users are notoriously apt to use poor, easily guessed credentials (such as password), resulting in break-ins where intruders can guess another user’s credentials and gain unauthorized access to a digital system. The term biometrics can resolve all type of credential problems by requiring an additional credential something associated with the person’s own body. The primary intention of this application is to prevent the frauds from accessing secured resources. This Biometrics offers reliable solution
through some technologies. As part of biometrics, the face recognition plays a very important role. A facial recognition system is a computer application capable of identifying or verifying a person from a digital image or a video frame from a video source. In this article, we discussed so many state of the art face recognition procedures with their performance.

[2] FACE RECOGNITION METHODOLOGIES

A. Principal Component Analysis (PCA)

It is one of the statistical technique for reducing the dimensionality of the dataset, and possibly to feature selection. PCA based algorithm have been the base of the several research projects in computer vision. Basically PCA is the unsupervised learning technique used for the producing the optimal linear least square decomposition of a datasets. PCA is a typical de correlation repetition in present research, one originates an orthogonal projection basis which directly leads to dimensionality reduction for classification problems.

B. Independent Component Analysis (ICA)

There are a number of algorithms for performing ICA [11], [13], [14], [25]. Author chose the infomax algorithm proposed by Bell and Sejnowski [11], which was derived from the principle of optimal information transfer in neurons with sigmoidal transfer functions [27]. The algorithm is motivated as follows: Let X be an -dimensional (n-D) random vector representing a distribution of inputs in the environment. (Here, boldface capitals denote random variables, whereas plain text capitals denote matrices). Let W be an n×n invertible matrix, U=WX and Y=f(U) an n-D random variable representing the outputs of n-neurons. Each component of f = (f1,….,fn) is an invertible squashing function, mapping real numbers into the interval. Typically, the logistic function is used

\[
    f_i(u) = \frac{1}{1+e^{-u}}.
\]

The U1,...,Un variables are linear combinations of inputs and can be interpreted as presynaptic activations of -neurons. The Y1,...,Yn variables can be interpreted as postsynaptic activation rates and are bounded by the interval . The goal in Bell and Sejnowski’s algorithm is to maximize the mutual informa-tion between the environment X and the output of the neural network Y . This is achieved by performing gradient ascent on the entropy of the output with respect to the weight matrix . The gradient update rule for the weight matrix, is as follows:

\[
    \Delta W \propto \nabla_w H(Y) = (W^T)^{-1} + E(YY^T)
\]

of the nonlinear where transfer function is the same as the cumulative density functions of the underlying ICs (up to scaling and translation) it can be shown that maximizing the joint entropy of the outputs in also minimizes the mutual information be-tween the individual outputs in [12], [42]. In practice, the logistic transfer function has been found sufficient to separate mixtures of natural signals with sparse distributions including sound sources [11]. 

The algorithm is speeded up by including a “sphering” step prior to learning [12]. The row means of are subtracted, and then is passed through the whitening matrix, which is twice the inverse square root2 of the covariance matrix
This removes the first and the second-order statistics of the data; both the mean and covariances are set to zero and the variances are equalized. When the inputs to ICA are the “sphered” data, the full transform matrix is the product of the sphering ma-trix and the matrix learned by ICA.

C. Linear Discriminant Analysis (LDA):

Both PCA and LDA have been used for face recognition [5, 6, 7, 8, 15, 16, 11]. With PCA, the input face images usually needed to be warped to a standard face because of the large within-class variance [6, 7]. This processing stage reduces the within-class variance dramatically, thus improving the recognition rate. Author first built a simple system based on pure LDA [8], but the performance was not satisfactory on a large dataset of person not present in the training set. The idea of combining PCA and LDA has been previously explored by Weng et al [15]. Although the pure LDA algorithm does not have any problem discriminating the trained samples, Author have observed that it does not perform very well for the following three cases:

- When the testing samples are from persons not in the training set
- When markedly different samples of trained classes are presented
- Samples with different background are presented
- Basically this is a generalization problem since the pure LDA based system is very much tuned to the specific training set, which has the same number of classes as persons, with 2 or 4 samples per class!

Combining PCA and LDA, Author obtain a linear projection which maps the input image $x$ first into the face-subspace $y$, and then into the classification space $z$:

$$y = \mathbf{T}x$$  \hspace{1cm} (6) $$z = \mathbf{Wy}^T y$$

$$z = Wx^T x$$  \hspace{1cm} (8)

where $\mathbf{T}$ is the PCA transform, $\mathbf{W}_y$ is the best linear discriminating transform on PCA feature space, and $\mathbf{W}_x$ is the composite linear projection from the original image space to the classification space. After this composite linear projection, recognition is performed in the classification space based on some distance measure criterion.

One thing need to be pointed out is that for the im-plementation of PCA, in many cases, even though the covariance matrix is a full-rank matrix, the large con-dition number will create a numerical problem. One way around this is to compute the eigenvalues and eigenvectors for $\mathbf{C}+\mathbf{I}$ instead of $\mathbf{C}$, where $\mathbf{I}$ is a positive number. This is based on the following lemma:Lemma 1 Matrices $\mathbf{C}$ and $\mathbf{C}+\mathbf{I}$ have same eigen-vectors but different eigenvalues with the relationship: $\mathbf{C}+\mathbf{I} = \mathbf{C} \ + I$ as long as $I$ is not equal to zero.

Performance improvement of this method over pure LDA based method is demonstrated through our own experiments and FERET test. Author believe that by combining PCA and LDA, using PCA to construct a task-specific sub-space and then applying LDA on that subspace, other image recognition systems such as fingerprint, optical character recognition can be improved. Author will study the subspace-LDA approach in detail and explore the possible applications in future work.
D. Eigenspace-based adaptive approach (EP):

The task of EP is to search for a face basis through the rotated axes defined in a properly whitened PCA space. Evolution is driven by a fitness function defined in terms of performance accuracy and class separation (scatter index). Accuracy indicates the extent to which learning has been successful so far, while the scatter index gives an indication of the expected fitness on future trials. Together, the accuracy and the scatter index give an indication of the overall performance ability. In analogy to the statistical learning theory [43], the scatter index is the conceptual analog for the capacity of the classifier and its use is to prevent overfitting. By combining these two terms together (with proper weights), GA can evolve balanced results and yield good recognition performance and generalization abilities.

One should also point out that just using more principal components (PCs) does not necessarily lead to better performance, since some PCs might capture the within class scatter which is unwanted for the purpose of face recognition [25],[27]. In proposed method, one can search the 20- and 30-dimensional whitened PCA spaces corresponding to the leading eigenvalues, since it is in those spaces that most of the variations characteristic of human faces occur.

In analogy to pursuit methods, EP seeks to learn an optimal basis for the dual purpose of data compression and pattern classification. The challenge for EP is to increase the generalization ability of the learning machine as a result of seeking the trade-off between minimizing the empirical risk encountered during training and narrowing the confidence interval for reducing the guaranteed risk for future testing on unseen images. Towards that end, EP implements strategies characteristic of genetic algorithms (GAs) for searching the space of possible solutions and determining an optimal basis. Within the face recognition framework, EP seeks an optimal basis for face projections suitable for compact and efficient face encoding in terms of both present and future recognition ability. Experimental results, using a large and varied subset from the FERET facial database, show that the EP method compares favorably against two popular methods for face recognition Eigenfaces and Fisherfaces.

E. Elastic Bunch Graph Matching (EBGM):

A first set of graphs is generated manually. Nodes are located at fiducial points and edges between the nodes as well as correspondences between nodes of different poses are defined. Once the system has an FBG (possibly consisting of only one manually defined model), graphs for new images can be generated automatically by elastic bunch graph matching. Initially, when the FBG contains only few faces, it is necessary to review and correct the resulting matches, but once the FBG is rich enough (approximately 70 graphs) one can rely on the matching and generate large galleries of model graphs automatically. Matching a FBG on a new image is done by maximizing a graph similarity between an image graph and the FBG of identical pose. It depends on the jet similarities and a topography term, which takes into account the distortion of the image grid relative to the FBG grid.
Fig. 1. The Face Bunch Graph (FBG) serves as a general representation of faces. Each stack of discs represents a jet. From a bunch of jets attached to a single node only the best fitting one is selected for a match, indicated by gray shading.

Fig. 2. Object-adapted grids for different poses. The nodes are positioned automatically by elastic bunch graph matching.

The system presented is general and flexible. It is designed for an in-class recognition task, i.e., for recognizing members of a known class of objects. Authors have applied it to face recognition but the system is in no way specialized to faces and author assume that it can be directly applied to other in-class recognition tasks, such as recognizing individuals of a given animal species, given the same level of standardization of the images. In contrast to many neural network systems, no extensive training for new faces or new object classes is required. Only a moderate number of typical examples have to be inspected to build up a bunch graph, and individuals can then be recognized after storing a single image.

Author tested the system with respect to rotation in depth and differences in facial expression. Some experiments included mirror reflection. Author did not investigate robustness to other variations, such as illumination changes or structured background. The performance is high on faces of same
pose. Author also showed robustness against rotation in depth up to about 22°. For large rotation angles the performance degrades significantly. Our system performs well compared to other systems. Results of a blind test of different systems on the FERET database were published in [6] and [7].

The Trace transform, a generalization of the Radon transform, is a new tool for image processing which can be used for recognizing objects under transformations, e.g. rotation, translation and scaling. To produce the Trace transform one computes a functional along tracing lines of an image. Different Trace transforms can be produced from an image using different trace functionals.

Each line is characterized by two parameters, namely its distance \( p \) from the centre of the axes and the orientation \( \theta \) the normal to the line has with respect to the reference direction. In addition, Author define parameter \( t \) along the line with its origin at the foot of the normal. The definitions of these three parameters are shown in figure 1. The image is transformed to another image with the Trace transform which is a 2-D function depending on parameters \( (\Omega, p) \). Different Trace transforms can be produced from an image using different trace functionals. An example of the Trace transform is shown in figure 2. It is shown that the image space in the \( x \) and \( y \) directions is transformed to the Trace transform space in the \( \Omega \) and \( p \) directions.

![Fig. 3. Tracing line on an image with parameters $\Omega$, $p$ and $t$.](image)

![Fig. 4. An image and its Trace transform.](image)
One of the key properties of the Trace transform is that it can be used to construct features invariant to rotation, translation and scaling. Author should point out that invariance to rotation and scaling is harder to achieve than invariance to translation. Let us assume that an object is subjected to linear distortions, i.e. rotation, translation and scaling. It is equivalent to saying that the image remains the same but viewed from a linearly distorted coordinate system. Consider scanning an image with lines in all directions. Let us denoted the set of all these lines with $\square$. The Trace transform is a function $g$ defined on $\square$ with the help of $T$ which is some functional of the image function when it is considered as a function of variable $t$. $T$ is called the trace functional. 

$$g(\square, p) \equiv T F(\square, p, t),$$ (1) where $F(\square, p, t)$ stands for the values of the image function along the chosen line. Parameter $t$ is eliminated after taking the trace functional. The result is therefore a 2-D function of parameters $\square$ and $p$ and can be interpreted as another image defined on $\square$.

F. Active Appearance Model (AAM):

An Active Appearance Model (AAM) is an integrated statistical model which combines a model of shape variation with a model of the appearance variations in a shape-normalized frame. An AAM contains a statistical model of the shape and gray-level appearance of the object of interest which can generalize to almost any valid example. Matching to an image involves finding model parameters which minimize the difference between the image and a synthesized model example projected into the image.

G. 3-D Face Recognition:

Human face is a surface lying in the 3-D space intrinsically. Therefore the 3-D model should be better for representing faces, especially to handle facial variations, such as pose, illumination etc. Blantz et al. proposed a method based on a 3-D morphable face model that encodes shape and texture in terms of model parameters, and algorithm that recovers these parameters from a single image of a face.

The main novelty of this approach is the ability to compare surfaces independent of natural deformations resulting from facial expressions. First, the range image and the texture of the face are acquired. Next, the range image is preprocessed by removing certain parts such as hair, which can complicate the recognition process. Finally, a canonical form of the facial surface is computed. Such a representation is insensitive to head orientations and facial expressions, thus significantly simplifying the recognition procedure. The recognition itself is performed on the canonical surfaces.

H. Bayesian Framework:

A probabilistic similarity measure based on Bayesian belief that the image intensity differences are characteristic of typical variations in appearance of an individual. Two classes of facial image variations are defined: intrapersonal variations and extrapersonal variations. Similarity among faces is measures using Bayesian rule.
CURRENT TRENDS ON FACE RECOGNITION METHODS IN FACE BIOMETRICS

All of the face recognition systems cited above (indeed the majority of the face recognition system published in the open literature) relay on similarity matrices which are invariably based on Euclidean distance or normalized correlation, thus corresponding to standard “template-matching”, i.e., nearest-neighbour based recognition. For example, simplest form the similarity measure $S(I_1, I_2)$ between two facial images $I_1$ and $I_2$ can be set to be inversely proportional to the norm $\| I_1 \triangle I_2 \|$

I. Support Vector Machine (SVM) :

Given a set of points belonging to two classes, a Support Vector Machine (SVM) finds the hyperplane that separates the largest possible fraction of points of the same class on the same side, while maximizing the distance from either class to the hyperplane. PCA is first used to extract features of face images and then discrimination functions between each pair of images are learned by SVMs.

J. Hidden Markov Models (HMM) :

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. HMM consists of two interrelated processes: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density functions associated with each state.

K. Boosting & Ensemble Solutions:

The idea behind Boosting is to sequentially employ a weak learner on a weighted version of a given training sample set to generalize a set of classifiers of its kind. Although any individual classifier may perform slightly better than random guessing, the formed ensemble can provide a very accurate (strong) classifier. Viola and Jones build the first real-time face detection system by using AdaBoost, which is considered a dramatic breakthrough in the face detection research. On the other hand, papers by Guo et al. are the first approaches on face recognition using the AdaBoost methods.

[3] DATA SETS

In biometrics when benchmarking a procedure it is suggested to use a standard test data set for researchers to be able to directly compare the results. In the state of the art literature there are numerous datasets in use at present, the choice of an appropriate database to be used should be made based on the task given (aging, expressions, lighting etc). The other method to choose the data set specific to the property to be tested (e.g. how algorithm behaves when given images with lighting deviations or images with diverse facial expressions). Some of the benchmark datasets are given below:

The Color FERET Database, USA
SCface - Surveillance Cameras Face Database
SCfaceDB Landmarks
Multi-PIE
The Yale Face Database
The Yale Face Database B
PIE Database, CMU
Project - Face In Action (FIA) Face Video Database, AMP, CMU
AT&T "The Database of Faces" (formerly "The ORL Database of Faces")
Cohn-Kanade AU Coded Facial Expression Database
MIT-CBCL Face Recognition Database
Image Database of Facial Actions and Expressions - Expression Image Database
Face Recognition Data, University of Essex, UK
NIST Mugshot Identification Database
NLPR Face Database
M2VTS Multimodal Face Database (Release 1.00)
The Extended M2VTS Database, University of Surrey, UK
The AR Face Database, The Ohio State University, USA
The University of Oulu Physics-Based Face Database
CAS-PEAL Face Database
Japanese Female Facial Expression (JAFFE) Database
BioID Face DB - HumanScan AG, Switzerland
Psychological Image Collection at Stirling (PICS)
The Sheffield Face Database (previously: The UMIST Face Database)
Face Video Database of the Max Planck Institute for Biological Cybernetics
Caltech Faces
EQUINOX HID Face Database
VALID Database
The UCD Colour Face Image Database for Face Detection
Georgia Tech Face Database
Indian Face Database
VidTIMIT Database
Labeled Faces in the Wild
The LFWcrop Database
Labeled Faces in the Wild-a (LFW-a)
3D_RMA database
GavabDB: 3D face database, GAVAB research group, Universidad Rey Juan Carlos, Spain
FRAV2D Database
FRAV3D Database
BJUT-3D Chinese Face Database
The Bosphorus Database
PUT Face Database
The Basel Face Model (BFM)
Plastic Surgery Face Database
The Iranian Face Database (IFDB)
The Hong Kong Polytechnic University NIR Face Database
The Hong Kong Polytechnic University Hyperspectral Face Database (PolyU-HSFD)
MOBIO - Mobile Biometry Face and Speech Database
Texas 3D Face Recognition Database (Texas 3DFRD)
Natural Visible and Infrared facial Expression database (USTC-NVIE)
FEI Face Database
ChokePoint
UMB database of 3D occluded faces
VADANA: Vims Appearance Dataset for facial ANAlysis
MORPH Database (Craniofacial Longitudinal Morphological Face Database)
Long Distance Heterogeneous Face Database (LDHF-DB)
PhotoFace: Face recognition using photometric stereo
The EURECOM Kinect Face Dataset (EURECOM KFD)
YouTube Faces Database
YMU (YouTube Makeup) Dataset
VMU (Virtual Makeup) Dataset
MIW (Makeup in the "wild") Dataset
3D Mask Attack Database (3DMAD)
Senthilkumar Face Database (Version 1.0)
McGill Real-world Face Video Database
SiblingsDB Database
FaceScrub - A Dataset With Over 100,000 Face Images of 530 People
LFW3D and Adience3D sets
Indian Movie Face database (IMFDB)
Labeled Wikipedia Faces (LWF)
10k US Adult Faces Database
Denver Intensity of Spontaneous Facial Action (DISFA) Database
BU-3DFE Database (Static Data)
BU-4DFE Database (Dynamic Data)
BP4D-Spontaneous Database
CAFE - The Child Affective Face Set
UFI - Unconstrained Facial Images
Senthil IRTT Face Database Version 1.1
Senthil IRTT Face Database Version 1.2
Senthil IRTT Video Face Database 1.0
VT-AAST Bench-marking Dataset
SEAS-FR-DB (School of Engineering & Applied Science - Face Video Database).

[4] CONCLUSION

Face recognition is a necessity of the modern age as the need for identification of individual has increased with the globalization of the world. Personal authentication through face has been under research since last two decades. The performance of the face recognition system has been enhanced using various algorithms. A generic facial authentication method contains three major steps i.e. face detection, facial features segmentation and face recognition. There are many commonly used algorithms used for this purpose. This paper provides an overview of different face recognition approaches. These approaches are categorized into four classes in this paper. These are holistic based approach, model based approach, hybrid based approach and feature based approach. Various techniques introduced in each of these categories are discussed.
REFERENCES

CURRENT TRENDS ON FACE RECOGNITION METHODS IN FACE BIOMETRICS


[33] A. Bronstein, M. Bronstein, R. Kimmel, and A. Spira. 3D face recognition without facial surface reconstruction, in Proceedings of ECCV 2004, Prague, Czech Republic, May 11-14, 2004


