ISBN 978-93-5156-328-0

An Overview of Face Recognition Schemes

Nisha Patel [a], Shailendra Kumar Dewangan [b]

[a] Department of Electronics and Telecommunication Engineering, Chhatrapati Shivaji Institute of Technology, Durg, Chhattisgarh, India

[b] Department of Electronics & Instrumentation Engineering, Chhatrapati Shivaji Institute of Technology, Durg, Chhattisgarh, India

Article Info

Article history:
Received 9 January 2015
Received in revised form
15 January 2015
Accepted 22 January 2015
Available online 31 January 2015

Keywords

Face Recognition,
PCA,
ICA,
LDA

Abstract

A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. It is typically used in security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems. The aim of this paper is to present an independent, comparative study of three most popular appearance-based face recognition projection methods (PCA, ICA, and LDA) in completely equal working conditions regarding preprocessing and algorithm implementation.

1. Introduction

The face plays a major role in our social intercourse in conveying identity and emotions. Face recognition can be applied for a wide variety of problems like image and film processing, human-computer interaction, criminal identification etc. This has motivated researchers to develop computational models to identify the faces, which are relatively simple and easy to implement. Computational models of faces have been an active area of research since late 1980s, however, developing a computational model of face recognition is quite difficult, because faces are complex, multidimensional, and subject to change over time.

From the sequence of images captured by camera, the goal is to find best match with a given image. Using a pre-stored image database, the face recognition system should be able to identify or verify one or more persons in the scene. Before face recognition is performed, the system should determine whether or not there is a face in a given image, a sequence of images. This process is called face detection. Once a face is detected, face region should be isolated from the scene for the face recognition. The face detection and face extraction are often performed simultaneously. [1]

Typical Applications of Face Recognition System are : Human-Robot-interaction, Human-Computer-interaction, Driver’s license, Smart cards, National ID, Passports, Voter registration Personal device logon, Desktop logon, Information security, Database security, Intranet security, Internet access, Medical records Video surveillance, CCTV control and Suspect tracking and investigation.

1.1 Principal Component Analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

Corresponding Author,
E-mail address: shailendra.dev@gmail.com

This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. [2] Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Eigen-face method based on Principal Component Analysis (PCA), is simple, fast and accurate in constrained environments. Our goal is to implement the model for a particular face and distinguish it from a large number of stored faces with some real-time variations as well [2].

The k^{th} component can be found by subtracting the first $k - 1$ principal component from X:

$$\hat{\mathbf{X}}_{k-1} = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X}\mathbf{w}_{(s)}\mathbf{w}_{(s)}^{\mathrm{T}}$$

and then finding the loading vector which extracts the maximum variance from this new data matrix. It turns out that this gives the remaining eigenvectors of X^TX, with the maximum values for the quantity in brackets given by their corresponding eigen values.

$$\mathbf{w}_{(k)} = \underset{\|\mathbf{w}\|=1}{\arg\max} \left\{ \|\hat{\mathbf{X}}_{k-1}\mathbf{w}\|^2 \right\} = \arg\max \left\{ \frac{\mathbf{w}^T \hat{\mathbf{X}}_{k-1}^T \hat{\mathbf{X}}_{k-1}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \right\}$$

The *k*th principal component of a data vector $\mathbf{x}_{(i)}$ can therefore be given as a score $t_{k(i)} = \mathbf{x}_{(i)} \cdot \mathbf{w}_{(k)}$ in the transformed co-ordinates, or as the corresponding vector in the space of the original variables, $\{\mathbf{x}_{(i)} \cdot \mathbf{w}_{(k)}\}\, \mathbf{w}_{(k)}$, where $\mathbf{w}_{(k)}$ is the *k*th eigenvector of $\mathbf{X}^T\mathbf{X}$.

The full principal components decomposition of X can therefore be given as

$$\mathbf{T} = \mathbf{X}\mathbf{W}$$

Where W is a p-by-p matrix whose columns are the eigenvectors of X^TX

1.2 Independent Component Analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents by assuming that the subcomponents are non-Gaussian signals and that they are all statistically independent from each other. ICA is a special case of blind source separation. As an example, sound is usually a signal that is composed of the numerical addition, at each time t, of signals from several sources. The question then is whether it is possible to separate these contributing sources from the observed total signal. When the statistical independence assumption is correct, blind ICA separation of a mixed signal gives very good results. It is also used for signals that are not supposed to be generated by a mixing for analysis purposes. A simple application of ICA is the "cocktail party problem", where the underlying speech signals are separated from a sample data consisting of people talking simultaneously in a room. [3] Usually the problem is simplified by assuming no time delays or echoes. An important note to consider is that if N sources are present, at least N observations (e.g. microphones) are needed to recover the original signals. This constitutes the square case (J = D, where D is the input dimension of the data and J is the dimension of the model). Other cases of underdetermined (J > D) and over determined (J < D) have been investigated that the ICA separation of mixed signals gives very good results are based on two assumptions and three effects of mixing source signals. Two assumptions:

1. The source signals are independent of each other.
2. The distribution of the values in each source signals is non-Gaussian.

Three effects of mixing source signals:

1. Independence: As what we assume, the source signals are independent, however, their signal mixtures are not. That is because the signal mixtures share the same source signals.
2. Normality: Based on the Central Limit Theorem, the distribution of a sum of independent random variables tends towards a Gaussian distribution. Loosely speaking, a sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original variables. Here we consider the value of each signal as the random variable.
3. Complexity: The temporal complexity of any signal mixture is greater than that of its simplest constituent source signal.

Those principles contribute to the basic establishment of ICA. If the signals we happen to extract from a set of mixtures are independent like sources signals, or have non-Gaussian histograms like source signals, or have low complexity like source signals, then they must be source signals.

1.3 Linear Discriminant Analysis (LDA) method is used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

LDA is closely related to ANOVA (analysis of variance) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. However, ANOVA uses categorical independent variables and a continuous dependent variable, whereas discriminant analysis has continuous independent variables and a categorical dependent variable (i.e. the class label). Logistic regression and probit regression are more similar to LDA, as they also explain a categorical variable by the values of continuous independent variables. [4] These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the LDA method.

LDA is also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made.

LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.

In the case where there are more than two classes, the analysis used in the derivation of the Fisher discriminant can be extended to find a subspace which appears to contain all of the class variability. Suppose that each of C classes has a mean μ_i and the same covariance Σ. Then the between class variability may be defined by the sample covariance of the class means

$$\Sigma_b = \frac{1}{C}\sum_{i=1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T$$

Where μ is the mean of the class means. The class separation in a direction $\vec{w}$ in this case will be given by

$$S = \frac{\vec{w}^T \Sigma_b \vec{w}}{\vec{w}^T \Sigma \vec{w}}$$

This means that when $\vec{w}$ is an eigenvector of $\Sigma^{-1}\Sigma_b$ the separation will be equal to the corresponding eigen value

2. Methodology

In this work, the focus is on image-based face recognition. Given a picture taken from a digital camera towards this goal the entire process of face recognition can generally be separated into three steps: Face Detection, Feature Extraction, and Face Recognition.

2.1 Face Detection

The main function of this step is to determine (1) whether human faces appear in a given image, and (2) where these faces are located at. The expected outputs of this step are patches containing each face in the input image. In order to make further face recognition system more robust and easy to design, face alignment are per-formed to justify the scales and orientations of these patches. Besides serving as the pre-processing for face recognition, face detection could be used for region of interest detection, retargeting, video and image classification, etc.

2.2 Feature Extraction

After the face detection step, human-face patches are extracted from images. Directly using these patches for face recognition have some disadvantages, first, each patch usually contains over 1000 pixels, which are too large to build a robust recognition system1. Second, face patches may be taken from different camera alignments, with different face expressions, illuminations, and may suffer from occlusion and clutter. To overcome these drawbacks, feature extractions are performed to do in-formation packing, dimension reduction, salience extraction, and noise cleaning. After this step, a face patch is usually transformed into a vector with fixed dimension or a set of fiducial points and their corresponding locations. We will talk more detailed about this step in Section 2. In some literatures, feature extraction is either included in face detection or face recognition.

2.3 Face Recognition

After formulizing the representation of each face, the last step is to recognize the identities of these faces. In order to achieve automatic recognition, a face database is required to build. For each person, several images are taken and their features are extracted and stored in the database. Then when an input face image comes in, we perform face detection and feature extraction, and compare its feature to each face class stored in the database.

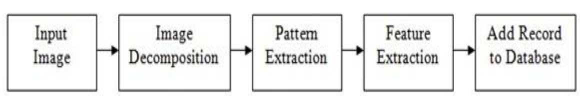

Fig: 1. Procedure for Facial Feature Extraction

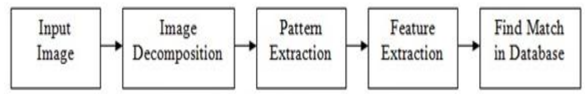

Fig: 2. Procedure for Face Recognition

There have been many researches and algorithms proposed to deal with this classification problem, and we'll discuss them in later sections. There are two general applications of face recognition, one is called identification and another one is called verification. Face identification means given a face image, we want the system to tell who he / she is or the most probable identification; while in face verification, given a face image and a guess of the identification, we want the system to tell true or false about the guess.

3. Conclusion

Face recognition is a challenging problem in the field of image analysis and computer vision that has received a great deal of attention over the last few years because of its many applications in various domains. Research has been conducted vigorously in this area for the past four decades or so, and though huge progress has been made, encouraging results have been obtained and current face recognition systems have reached a certain degree of maturity when operating under constrained conditions; however, they are far from achieving the ideal of being able to perform adequately in all the various situations that are commonly encountered by applications utilizing these techniques in practical life. The ultimate goal of researchers in this area is to enable computers to emulate the human vision system. Strong and coordinated effort between the computer vision, signal processing, and psychophysics and neurosciences communities is needed to attain the objective of face recognition.

References

[1] S. L. Wijaya, M. Savvides, B. V. K. V. Kumar, Illumination Tolerant Face Verification of Low Bit Rate JPEG2000 Wavelet Images With Advanced Correlation Filters for Handheld Devices, Applied Optics, 44, 2005, 655-665

[2] E. Acosta, L. Torres, A. Albiol, E. J. Delp, An Automatic Face Detection & Recognition System for Video Indexing Applications, IEEE International Conference on Acoustics, Speech and Signal Processing, Orlando, Florida, 4, 2002, 3644-3647

[3] J. N. K. Liu, M. Wang, B. Feng, iBotGuard : An Internet-Based Intelligent Robot Security System Using Invariant Face Recognition Against Intruder, IEEE Transactions on Systems Man & Cybernetics, 35, 2005, 97-105

[4] B. Moghaddam, M. H. Yang, Learning Gender with Support Faces, IEEE Transactions on Pattern Analysis & Machine Intelligence, 24, 2002, 707-711

[5] A. Colmenarez, B. J. Frey, T. S. Huang, A Probabilistic Framework for Embedded Face & Facial Expression Recognition, IEEE Conference on Computer Vision and Pattern Recognition, 1, Ft. Collins, CO, USA, 1999, 1592-1597

[6] M. M. Rahman, R. Hartley, S. Ishikawa, A Passive & Multimodal Biometric System for Personal Identification, International Conference on Visualization, Imaging & Image Processing. Spain, 2005, 89-92