Indian Journal of Science and Technology, *Vol 8(35), DOI: 10.17485/ijst/2015/v8i35/77760, December 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Principal Component Analysis based Feature Vector Extraction

M. Murali*

Department of Computer Science and Engineering, SRM University, Kattankulathur - 603203, Tamil Nadu, India; murali.m@ktr.srmuniv.ac.in

Abstract

Objectives: A simple and efficient method is employed to extract feature vector from images and to reduce the dimension of data. Once the feature vectors are extracted, it can be used in face recognition module. **Methods/Analysis:** Principal Components Analysis (PCA) is used for face recognition technique for feature identification in large data sets and to highlight their similarities and differences is more essential step in face recognition. PCA is used as efficient tool in data analysis to reduce dimension and to obtain maximum variance of data. **Findings:** Experiment is conducted using Yale database B. The face images are formed with multiple factors on different lighting conditions, background interference, and face rotation etc. The experimental results on Yale database B are given to illustrate the proposed method. **Conclusion/Application:** A simple and effective feature extraction method is analyzed for face images and experimental results are shown.

Keywords: Eigenvalue, Eigenvector, Feature Vector Extraction, Face Recognition, Principal Component Analysis

1. Introduction

Face recognition is a challenging task. Feature vector extraction technique is important step in face recognition. A face in an input image must be registered in a standard size frame before being processed. Automatic detection of features is also an important. PCA is a simple, non-parametric method of extracting relevant information from large datasets. This system is described in terms of the preprocessing and some steps. During preprocessing, set of images are stored andthey are represented as a vector of pixel values. PCA can be used to find the maximum variance in the original space. The linear transformation maps the original d-dimensional space onto a s-dimensional feature subspace. Most discriminating features are identified as feature vectors and stored in face recognition module. Another advantage of using PCA is that the patterns found in the data can be compressed by reducing the number of dimensions, without loss of information. This technique may also be used in image compression.

There are various feature identification methods are available, Eigenface method[1,2] and Fisherface method[3–5] are more popular methods. Eigenface method based on Principal Component Analysis (PCA), is used to find the set of projection vectors. The Fisherface method, based on Linear Discriminant Analysis (LDA)[6,7], is used to find the optimal set of projection vectors. Other methods are Discrete Cosine Transform (DCT)[8], Independent Component Analysis (ICA)[9], Support Vector Machines (SVM)[10,11], and so on. In this paper, focus is given to extract feature vectors from images. The idea of using principal components to represent human faces was developed by Sirovich and Kirby[12] and used by Turk and Pentland[13] for face detection and recognition, PCA can be also called as Karhunen–Loeve transformation. SVM is used to solve binary classification problems[14] and in 2010, Naseem proposed a novel method for face recognition which is a Linear Regression Classification (LRC) algorithm[15].

2. Proposed Feature Vector Extraction Method

The feature vector extraction technique is heavily used in face recognition. In this approach, two dimensional

**Author for correspondence*

images are converted into one dimensional vectors. Since the vector space is too large, it is very difficult to calculate the covariance matrix. This can be done by PCA method, which is based on two dimensional image matrices. A two dimensional image transformation has major advantages in image feature vector extraction and dimensionality reduction. PCA is used to extract feature vector and reduce the dimensions of process data. Figure1 shows the procedure of our proposed method. Consider the face image database consisting of total N images. Each face image in the database is of the size m x n pixel. Here, the main task is to find out the features of images without many computations. The proposed method is used to extract feature vectors to identify the similarity between human faces are done by computing eigenvalues and eigenvectors.

There is a two dimensional input image, is to be compared with a set of data base images to find the match. It is assumed that the images in the database are all of same resolution. An input image with n pixels can be treated as a point in an d-dimensional space called the image space. The individual ordinates represent the intensity values of each pixel of the image and form a row vector. This vector

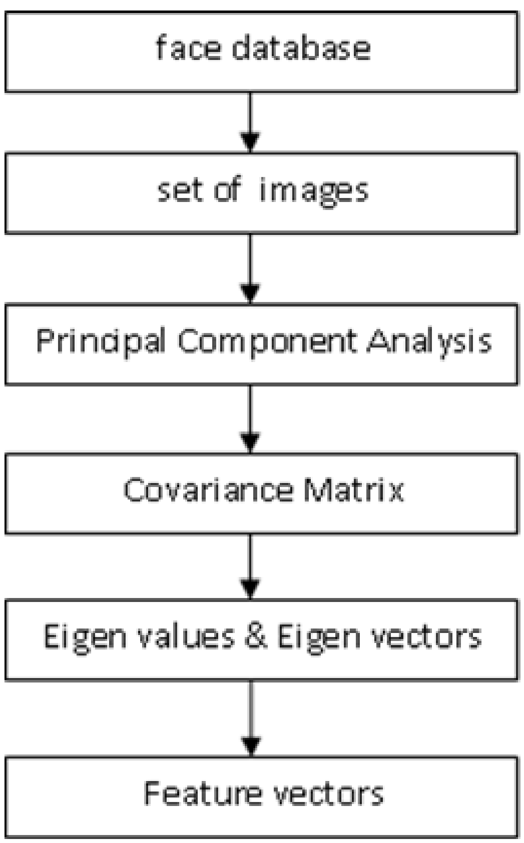

Figure 1. The flow chart of proposed method.

is constructed by concatenating each row of image pixels to a size of 128 by 128. PCA is a technique that can be used to simplify a dataset. It is a linear transformation that chooses a new coordinate system for the data set such that variance by any projection of the data set lie on the first axis (first principal component), the second variance on the second axis, and so on. PCA can be used for reducing dimensionality by eliminating the later principal components. Principal component analysis provides a roadmap for how to reduce a complex data set to a lower dimension. A training set istaken from the face database, TS = (img1, img2, img3 … imgM) with M images each with n pixels. Our data set is an M x n data matrix. Each row of data matrix represents one image of our data set. Since the mages are projected into this subspace the recognition of an image can be achieved by projecting a probe image into this subspace and the image closer to the probe image can be chosen.

2.1 Computing the Principal Components

Principal components can be identified by calculating the eigenvectors and eigenvalues of the data covariance matrix. This is equivalent to finding the axis system in which the co-variance matrix is diagonal. However many data sets have more than one dimension, and the aim of the statistical analysis of these data sets is usually to see if there is any relationship between the dimensions. Covariance is a measure to find the relationship between the dimensions among the datasets. Covariance is always measured between 2 dimensions. Each row corresponds to all measurements of a particular type X_i. Each column of X corresponds to a set of measurements from particular time instant. Since covariance is non-negative, the covariance matrix will be a diagonal matrix. PCA first selects a normalized direction in d-dimensional space along which the variance in X is maximized. Again it finds another direction, in which the variance is maximized. Thus m directions can be selected. The result set is the principal components.

$$COV(X,Y)=\frac{\sum_{n=1}^{n}(\mathrm{xi}-\bar{X})(\mathrm{yi}-\bar{Y})}{(\mathrm{n}-1)} \quad (1)$$

Covariance matrices represent the statistical interdependence structureand it is square matrix. The eigenvector with the largest eigenvalue is the direction of variation is considered as first principal component and the second largest eigenvalue with the next highest variation is the second principal component and so on. Let A be an nx nmatrix. The eigenvectors e1, e2,…en span on eigenspace, and they are known as orthonormal vectors. The vector e1 is called an eigenvector of A associated with the eigenvalue λ1, and vector e2 is called an eigenvector of A associated with the eigenvalue λ2 and so on. The next step is to order them by eigenvalue, highest to lowest, in order of significance. Now, no need to consider the components of lesser significance.

2.2 Face Image Database

The Yale database B is cropped to small size of 192 × 168 pixel and then used for training and testing. In the first experiment, the single light source images of 4 persons are selected for testing, producing a training set. It is shown in Figure 2. The feature vector extraction is carried out using proposed method considering for training and testing of the images. In the second experiment 6 images are selected for producing a training set. It is shown in Figure 3.

2.3 Experimental Results

In this paper, Yale database B is considered for extracting feature vectors. Figure 2 is showing 4 images. As per the proposed method, the images are read as a matrix. One single image is stored in a single row, likewise four images will be stored in four rows. An image size is 8.00Kb (8192 bytes) and it contains 192 rows and 168 columns (192 × 168). After adding all the four images in a single matrix, the size would become 4 × 32256. By applying Principal Component Analysis, the first step is finding the Covariance matrix. Afterwards, eigenvalues and the corresponding eigenvectors are calculated.

Figure 3 is showing 6 images. As per the proposed method, the images are read as a matrix. One single image is stored in a single row, likewise four images will be stored in four rows. All 6 images are same size and it contains 192 rows and 168 columns (192 × 168). After adding all the four images in a single matrix, the size would become 6 × 32256. By applying PCA, covariance matrix is found. Thereafter, eigenvalues and the corresponding eigenvectors are calculated.

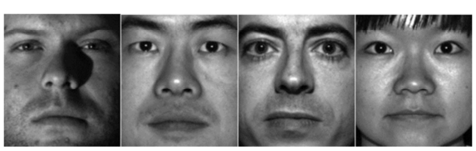

Figure 2. Images considered for experiment 1.

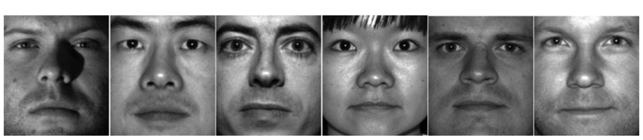

Figure 3. Images considered for experiment 2.

2.4 Choosing Components and Forming Feature Vector

Principal component analysis has been widely used in face recognition, primarily for reducing the number of variables. Feature vector extraction is an important step in face recognition. All images are of the same size and resolution. Each pixel is considered a variable it can be simplified by PCA. In our example, the eigenvector with the largest eigenvalue is identified. It is the most significant relationship between the data dimensions. Once eigenvectors are found from the covariance matrix, the next stepis to order them by eigenvalue, in descending order. This will give the components by their significance. Now we need to form a feature vector, which is a name for a matrix of vectors. This is done by taking the eigenvectors from the list of eigenvectors, and forming a matrix.

Feature Vector = (eig1, eig2, eig3 …eign)

Result of Experiment 1 using Figure 2:

The Eigenvalues are = [4.3282 2.0056 1.1181 0.6456]

The Feature vectors are

$$= \begin{bmatrix} 0.2003 & 0.9752 & -0.0757 & 0.0561 \\ 0.4513 & -0.1499 & -0.1004 & 0.8739 \\ 0.6377 & -0.0603 & 0.7238 & -0.2565 \\ 0.5912 & -0.1509 & -0.6785 & -0.4091 \end{bmatrix}$$

Result of Experiment 2 using Figure 3:

The Eigenvalues are = [5.4095 2.0080 1.2251 0.7189 0.5928 0.3518]

The Feature vectors are

$$= \begin{bmatrix} 0.1719 & -0.9752 & -0.1147 & 0.0684 & -0.0041 & 0.0398 \\ 0.3850 & 0.1479 & -0.2435 & 0.6682 & -0.5669 & 0.0516 \\ 0.5806 & 0.0599 & 0.5074 & 0.3088 & 0.5512 & -0.0518 \\ 0.5014 & 0.1486 & -0.7259 & -0.3261 & 0.3044 & -0.0243 \\ 0.1909 & -0.0337 & 0.1367 & -0.2233 & -0.2831 & -0.9021 \\ 0.4442 & 0.0177 & 0.3528 & -0.5452 & -0.4494 & 0.4228 \end{bmatrix}$$

3. Conclusion

In this paper, a simple feature vector extraction method is proposed which is based on PCA. PCA is one of the most important methods in pattern recognition. Therefore, PCA is used to extract feature and reduce the dimensions of process data. Afterwards, it can be used for face recognition. This proposed method effectively uses the PCA for feature vector extraction. The experimental results on Yale database B is presented to illustrate the feature vector extraction.

4. References

1. Turk M, Pentland A. Eigenfaces for recognition. J Cognitive Neurosci. 1991; 3(1):71–86.
2. Sharma UM. Hybrid feature based face verification and recognition system using principal component analysis. Indian Journal of Science and Technology. 2015 Jan; 8(S1):115–20.
3. Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell. 1997; 19(7):711–20.
4. Sivaram KS, et al. Object recognition by feature weighted matrix - A novel approach. Indian Journal of Science and Technology. 2015 Apr; 8(S7):278–91.
5. Sucharita V, Jyothi S, Mamatha DM. Texture feature extraction for the classification of panaeid prawn species using gabor filter. Indian Journal of Science and Technology. 2015 Aug; 8(17):1–4.
6. Lu GF, Zou J, Wang Y. Incremental complete LDA for face recognition. Pattern Recognition. 2012; 45(7):2510–21.
7. Kanchana S, Balakrishnan G. A novel Gaussian measure curvlet based feature segmentation and extraction for palmprint images. Indian Journal of Science and Technology. 2015 Jul; 8(15):1–7.
8. Zhong D, Defe I. DCT histogram optimization for image database retrieval. Pattern Recognition Lettters. 2005; 26(14):2272–81.
9. Bartlett MS, Sejnowski TJ. Independent components of face images: A representation for face recognition. Proceedings of the 4th Annual Joint Symposium on Neural Computation; Pasadena, CA. 1997 May. p. 3–10.
10. Deniz O, Castrillon M, Hernandez M. Face recognition using independent component analysis and support vector machines. Pattern Recognition Lettters. 2003; 24(13):2153–7.
11. Jayasri T, Hemalatha M. Categorization of respiratory signal using ANN and SVM based feature extraction algorithm. Indian Journal of Science and Technology. 2013 Sep; 6(9):5195–200.
12. Kirby M, Sirovich L. Application of the KL procedure for the characterization of human faces. IEEE Trans Pattern Mach Intell. 1990; 12(1):103–8.
13. Turk MA, Pentland AP. Eigenfaces for recognition. J Cogn Neuroscience. 1991; 3(1):71–86.
14. Chi WH, Chi JL. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks. 2009; 13(2):415–25.
15. Togneri IRN, Bennamoun M. Linear regression for face recognition. IEEE Trans Pattern Anal Mach Intelligence. 2010; 32(11):2106–12.