Recognition of Image using Principal Component Analysis

Deven Ketkar
Department of Electronics Engineering
Shah and AnchorKutchhi Engineering College , Chembur, Maharashtra, India

Abstract- The proposed work in this paper basically deals with calculating features of training images using Principal Component Analysis(PCA) method. Suppose we are having 50 training images, then features corresponding to that images which are nothing but mean, deviation, Eigen vectors and Euclidean distance are calculated. For testing purpose,we are going to use one of the image among the training dataset. Features corresponding to test image which are mentioned above are calculated. Recognition of image takes place according to the smallest Euclidean distance between training and testing image.

Keywords –Principal Component Analysis(PCA),Mean, Deviation, Eigen Vectors, Euclidean distance.

I. INTRODUCTION

Here,we are going to consider points related to images such as mean, deviation, Eigen vectors and Euclidean distance. Suppose we are having four by four matrix, then row wise mean corresponding to that image i.e matrix is calculated. This can be done by addition of four elements in the first row divided by four. Mean for the remaining three rows are calculated in same fashion. The above mentioned is the simple four by four matrix. Practically we are going to deal with 256 by 256 images. Mean corresponding to that images are calculated according to the same procedure mentioned above. For calculating mean, following formula is used

(1)

The next feature we are going to calculate is deviation. This feature is related to mean since in order to find deviation we have to find the difference between mean and each value in the image matrix.An eigenvector of a square matrix A is a non-zero vector V that, when it is multiplied with A, yields a scalar multiple of itself; the scalar multiplier is often denoted by λ. That is:

$$AV = \lambda V \qquad (2)$$

This equation uses post-multiplication of the matrix A by the vector V, it describes a eigenvector.The number λ is called the eigenvalue of A corresponding to V.If two-dimensional space is viewed as a piece of cloth being stretched by the matrix, the eigenvectors would make up the line along the direction the cloth is stretched in and the line of cloth at the center of the stretching, whose direction isn't changed by the stretching either. The eigenvalues for the first line would give the scale to which the cloth is stretched, and for the second line the scale to which it is tightened. A reflection can be seemed as stretching a line to scale −1 while shrinking the axis of reflection to scale 1. The eigenvectors form the axis of rotation in 3-D case, and scale of the axis is unchanged by the rotation, their eigenvalues are all 1.

The next feature is nothing but the Euclidean distance. It is defined as the distance or ordinary distance present between the two points in Euclidean space. If we consider two points p and q, then Euclidean distance is simply given by the formula

$$\sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (3)$$

The position of a point in a Euclidean *n*-space is called a Euclidean vector. Hence, **p** and **q** are Euclidean vectors, starting from the origin and their tips indicate two points.

II.PROPOSED ALGORITHM

A. Principal Component Analysis(PCA)–

Principal component analysis (PCA) is a statistical procedure that uses an 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 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.

Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT) in signal processing.PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute.

PCA is one of the most successful techniques that have been used in image recognition and compression.The purpose of PCA is to reduce the large dimensionality of the data.The task of the MR image recognizer is to find the most similar feature vector among the training set to the feature vector of a given test image.Let T1 be a training image of image 1 which has a pixel resolution of M x N (M rows,N columns).In order to extract PCA features of T1, first convert the image into a pixel vector $\Phi 1$ by concatenating each of the M rows into a single vector. The length (or,dimensionality) of the vector $\Phi 1$ will be M x N. Here, the PCA algorithm is used as a dimensionality reduction technique which transforms the vector $\Phi 1$ to a vector $\omega 1$which has a dimensionality d where $d << M$xN.For each training image Ti, these feature vectors ωi are calculated and stored.In the testing phase, the feature vector ωj of the test image Tj is computed using PCA. In order to identify the test image Tj, the similarities between ωj and all of the feature vectors ωi's in the training set are computed.The similarity between feature vectors is computed using Euclidean distance. The identity of the most similar ωi is the output of the image recognizer. If i = j, it means that the MR image j has correctly identified, otherwise if $i \neq j$,it means that the MR image j has misclassified.

By using Principal Component Analysis(PCA),we can perform feature extraction corresponding to collected data set.

Features to be calculated-:
1. Mean
2. Deviation
3. Eigen values
4. Euclidean distance

The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores and loadings (the weight by which each standardized original variable should be multiplied to get the component score).

PCA is the simplest of the true eigenvector-based multivariate analyses. Its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture.This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced. It is also related to canonical correlation analysis (CCA). CCA defines coordinate systems that optimally describe the cross-covariance between two datasets while PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset.

PCA can be thought of as fitting an *n*-dimensional ellipsoid to the data, where each axis of the ellipsoid represents a principal component. If some axis of the ellipse is small, then the variance along that axis is also small, and by omitting that axis and its corresponding principal component from our representation of the dataset, we lose only a small amount of information.

To find the axes of the ellipse, we must first subtract the mean of each variable from the dataset to center the data around the origin.Then, we compute the covariance matrix of the data, and calculate the eigenvalues and their corresponding eigenvectors of this covariance matrix. Then, we must orthogonalize the set of eigenvectors, and normalize each to become unit vectors. Once this is done, each of the mutually orthogonal, unit eigenvectors can be interpreted as an axis of the ellipsoid fitted to the data.

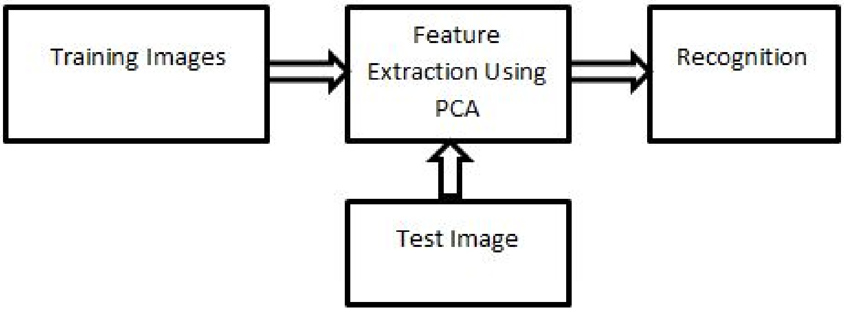

Figure 1. Block diagram

III.CONCLUSION

Here, we took 50 number of training images and features corresponding to that images which are mean, deviation, Eigen vectors and Euclidean distance are calculated using PCA. We took test image as one of the training images. PCA is applied to that image and above mentioned features are calculated. The smallest Euclidean distance between the testing and training images are used for image recognition.

REFERENCES

[1] Bruce poon1, m. ashraful amin2, hong yan1, "PCA BASED FACE RECOGNITION AND TESTING CRITERIA",Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009,pp-2945-2949.

[2] Vinod Kumar, Jainy Sachdeva, Indra Gupta,Niranjan Khandelwal, Chirag Kamal Ahuja"Classification of brain tumours using PCA-ANN" 2011 World Congress on Information and Communication Technologies.pp-1079-1083

[3] Xiaofei Zhou1.., Yong Shi1,2, Peng Zhang1, Guangli Nie1, Wenhan Jiang3" A New Classification Method for PCA-based Face Recognition", 2009 International Conference on Business Intelligence and Financial Engineering,pp-445-449.