ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Face Recognition and Facial Expression Identification Using PCA

Yapa Ashok
M.Tech (Digital Electronics and Communication Systems),
Siddhartha Institute of Engineering and Technology.

Dr.Dasari Subba Rao, Ph.D
HOD,
Siddhartha Institute of Engineering and Technology.

ABSTRACT:

The face being the primary focus of attention in social interaction plays a major role in conveying identity and emotion. 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. The main aim of this paper is to analyse the method of Principal Component Analysis (PCA) and its performance when applied to face recognition. This algorithm creates a subspace (face space) where the faces in a database are represented using a reduced number of features called feature vectors. The PCA technique has also been used to identify various facial expressions such as happy, sad, neutral, anger, disgust, fear etc. Experimental results that follow show that PCA based methods provide better face recognition with reasonably low error rates. From the paper, we conclude that PCA is a good technique for face recognition as it is able to identify faces fairly well with varying illuminations, facial expressions etc.

Principal Component Analysis:

Principal component analysis (PCA) is a mathematical 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. 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 be orthogonal to (i.e., uncorrelated with) the preceding components.

Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT),the Hotelling transform or pro per orthogonal decomposition (POD).

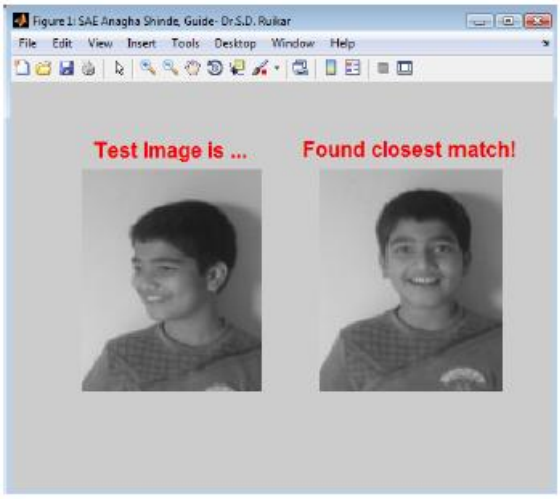

Figure 9:Test image and recognized image from the training base.

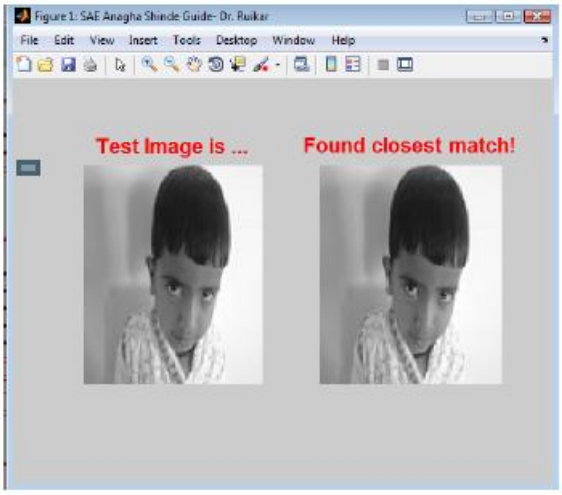

Figure 9:Test image and recognized image from the training base.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

PCA was invented in 1901 by Karl Pearson. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigen value 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. The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), 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. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data.

If a multivariate dataset is visualised 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, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.PCA is closely related to factor analysis; indeed, some statistical packages deliberately conflate the techniques. True factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.

Details:

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Define a data matrix, X^T, with zero empirical mean (the empirical (sample) mean of the distribution has been subtracted from the data set), where each of the n rows represents a different repetition of the experiment, and each of the m columns gives a particular kind of datum (say, the results from a particular probe). (Note that X^T is defined here and not X itself, and what we are calling X^T is often alternatively denoted as X itself.) The singular value decomposition of X is $X = W\Sigma V^T$, where the m × m matrix W is the matrix of eigenvectors of the covariance matrix XX^T, the matrix Σ is an m × n rectangular diagonal matrix with nonnegative real numbers on the diagonal, and the n × n matrix V is the matrix of eigenvectors of X^TX. The PCA transformation that preserves dimensionality (that is, gives the same number of principal components as original variables) is then given by:

$$\begin{aligned}\mathbf{Y}^\top &= \mathbf{X}^\top \mathbf{W} \\ &= \mathbf{V}\boldsymbol{\Sigma}^\top \mathbf{W}^\top \mathbf{W} \\ &= \mathbf{V}\boldsymbol{\Sigma}^\top\end{aligned}$$

V is not uniquely defined in the usual case when $m < n - 1$, but Y will usually still be uniquely defined. Since W (by definition of the SVD of a real matrix) is an orthogonal matrix, each row of Y^T is simply a rotation of the corresponding row of X^T. The first column of Y^T is made up of the "scores" of the cases with respect to the "principal" component, the next column has the scores with respect to the "second principal" component, and so on. If we want a reduced-dimensionality representation, we can project X down into the reduced space defined by only the first L singular vectors, W_L:

$$\mathbf{Y} = \mathbf{W_L}^\top \mathbf{X} = \boldsymbol{\Sigma_L} \mathbf{V}^\top$$

Where $\boldsymbol{\Sigma_L} = \mathbf{I}_{L\times m}\boldsymbol{\Sigma}$ with $\mathbf{I}_{L\times m}$ the $L \times m$ rectangular identity matrix. The matrix W of singular vectors of X is equivalently the matrix W of eigenvectors of the matrix of observed covariance's $C = X X^T$,

$$\mathbf{X}\mathbf{X}^\top = \mathbf{W}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{W}^\top$$

Given a set of points in Euclidean space, the first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted from the

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

points. The singular values (in Σ) are the square roots of the eigen values of the matrix XX^T. Each eigen value is proportional to the portion of the "variance" (more correctly of the sum of the squared distances of the points from their multidimensional mean) that is correlated with each eigenvector. The sum of all the eigen values is equal to the sum of the squared distances of the points from their multidimensional mean. PCA essentially rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information.

PCA is often used in this manner for dimensionality reduction. PCA has the distinction of being the optimal orthogonal transformation for keeping the subspace that has largest "variance" (as defined above). This advantage, however, comes at the price of greater computational requirements if compared, for example and when applicable, to the discrete cosine transform. Nonlinear dimensionality reduction techniques tend to be more computationally demanding than PCA. PCA is sensitive to the scaling of the variables. If we have just two variables and they have the same sample variance and are positively correlated, then the PCA will entail a rotation by 45° and the "loadings" for the two variables with respect to the principal component will be equal.

But if we multiply all values of the first variable by 100, then the principal component will be almost the same as that variable, with a small contribution from the other variable, whereas the second component will be almost aligned with the second original variable. This means that whenever the different variables have different units (like temperature and mass), PCA is a somewhat arbitrary method of analysis. (Different results would be obtained if one used Fahrenheit rather than Celsius for example.) Note that Pearson's original paper was entitled "On Lines and Planes of Closest Fit to Systems of Points in Space" – "in space" implies physical Euclidean space where such concerns do not arise. One way of making the PCA less arbitrary is to use variables scaled so as to have unit variance.

Discussion:

Mean subtraction (a.k.a. "mean centering") is necessary for performing PCA to ensure that the first principal component describes the direction of maximum variance. If mean subtraction is not performed, the first principal component might instead correspond more or less to the mean of the data. A mean of zero is needed for finding a basis that minimizes the mean square error of the approximation of the data. Assuming zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), the principal component w_1 of a data set X can be defined as:

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \operatorname{Var}\{\mathbf{w}^\top \mathbf{X}\} = \arg\max_{\|\mathbf{w}\|=1} E\left\{\left(\mathbf{w}^\top \mathbf{X}\right)^2\right\}$$

(See arg max for the notation.) With the first $k-1$ component, the kth component can be found by subtracting the first $k-1$ principal components from X:

$$\hat{\mathbf{X}}_{k-1} = \mathbf{X} - \sum_{i=1}^{k-1} \mathbf{w}_i \mathbf{w}_i^\top \mathbf{X}$$

and by substituting this as the new data set to find a principal component in

$$\mathbf{w}_k = \arg\max_{\|\mathbf{w}\|=1} E\left\{\left(\mathbf{w}^\top \hat{\mathbf{X}}_{k-1}\right)^2\right\}.$$

PCA is equivalent to empirical orthogonal functions (EOF), a name which is used in meteorology. An auto encoder neural network with a linear hidden layer is similar to PCA. Upon convergence, the weight vectors of the K neurons in the hidden layer will form a basis for the space spanned by the first K principal components. Unlike PCA, this technique will not necessarily produce orthogonal vectors. PCA is a popular primary technique in pattern recognition. It is not, however, optimized for class separability.[6] An alternative is the linear discriminant analysis, which does take this into account.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

III. RECOGNITION SYSTEM:

Automatic systems for facial expression recognition usually take the form of a sequential configuration of processing steps, which adheres to a classical pattern recognition model. The main steps to proceed:

A. Image Acquisition:

Images used for facial expression recognition are static images or image sequences. An image sequence contains potentially more information than a still image, because the former also depicts the temporal dimensionality of input images, 2-D monochrome (grey-scale) facial image sequences are the most popular type of pictures used for automatic expression recognition. However, colour images could become prevalent in future, owing to the increasing availability of low-cost colour image acquisition equipment, and the ability of colour images to convey emotional cues such as blushing.

- **IMAGE READ:**

A= imread (filename,fmt)

It reads a gray scale or color image from the file specified by the string filename. If the file is not in the current directory, or in a directory on the MATLAB path, specify the full pathname.

- **DISPLAY IMAGE:**

B= imshow (I)

It display the gray scale image I.

B. Pre-Processing:

Image pre-processing often takes the form of signal conditioning (such as noise removal, and normalization against the variation of pixel position or brightness), together with segmentation, location, or tracking of the face or its parts. Expression representation can be sensitive to translation, scaling, and rotation of the head in an image. To combat the effect of these unwanted transformations, the facial image may be geometrically standardized prior to classification. This normalization is usually based on references provided by the eyes or nostrils.

Segmentation is concerned with the demarcation of image portions conveying relevant facial information. Face segmentation is often anchored on the shape, motion, color, texture, and spatial configuration of the face or its Components. The face location process yields the position and spatial extent of faces in an image; it is typically based on segmentation results. A variety of face detection techniques have been developed. However, robust detection of faces or their constituents is difficult to attain in many real-world Settings. Tracking is often implemented as location, of the face or its parts, within an image sequence, whereby previously determined location is typically used for estimating location in subsequent image frames.

- **STANDARD SIZING:**

B= imresize (A, [mrowsncols])

This instruction returns image B that has the number of rows and columns specified by [mrowsncols]. Either NUMROWS or NUMCOLS, in which case imresize computer the number of rows or column automatically to preserve the image aspect ratio.

- **EDGE FINDING:**

BW=edge (I)

It takes a gray scale or a binary image I as its input, and returns a binary image BW of the same size as I, with 1‟s where the function finds edges in 1 and 0‟s elsewhere. By default, edge uses the Sobal method to detect edges but the following provides a complete list of all the edge finding methods supported by this functin:

1. Sobel method
2. Prewitt method
3. Roberts method

These methods find edges using the sobel, prewittor roberts approximation to the derivative. It returns edges .

LIGHTING COMPENSATION:

J= imadjust(I,[low_in;high_in],[low_out; high_out])

This instruction maps the values in I to new values in J such values between low_in and high_in map to values

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

between low_out and high_out. Values below low_in and above high_in are clipped; that is, values below low_in map to low_out, and those above high_in map to high_out.

C. Feature Extraction:

Feature extraction converts pixel data into a higher-level representation of shape, motion, color, texture, and spatial configuration of the face or its components. The extracted representation is used for subsequent expression categorization. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure should (ideally) retain essential information possessing high discrimination power and high stability. Such dimensionality reduction may mitigate the „curse of dimensionality". Geometric, kinetic, and statistical- or spectral-transform-based features are often used as alternative representation of the facial expression prior to classification.

- **PERFORMING PCA:**

[Coeff, Score, latent, tsquare] = princomp (X)

X is n b p data matrix. Rows of X correspond to observations and columns to variables.

- **Coeff:** Coeff is a p-by-p matrix, each column containing coefficients for one principal component. The columns are in order of decreasing component variance.
- **Score:** Representation of X is principal comp. Space rows of score correspond to observation, columns to components.
- **Latent:** Eigen values of the covariance matrix of X. It is the variance of Score.

D. Classification:

Expression categorization is performed by a classifier, which often consists of models of pattern distribution, coupled to a decision procedure. A wide range of classifiers, covering parametric as well as nonparametric techniques, has been applied to the automatic expression recognition problem. The two main types of classes used in facial expression recognition are action units (AUs), and the prototypic facial expressions defined by Ekman.

The 6 prototypic expressions relate to the emotional states of happiness, sadness, surprise, anger, fear, and disgust. However, it has been noted that the variation in complexity and meaning of expressions covers far more than these six expression categories.

E. Post-processing:

Post-processing aims to improve recognition accuracy, by exploiting domain knowledge to correct classification errors, or by coupling together several levels of a classification hierarchy, for example.

CONCLUSION:

Face recognition method using eigen faces is proposed. We used database of face images which contains 190 images of 38 different persons (5 images per person). From the results, it can be concluded that, for recognition, it is sufficient to take about 10% eigen faces with the highest eigen values. It is also clear that the recognition rate increases with the number of training images per person. It is obvious that if the minimum distance between the test image and other images is zero, the test image entirely matches the image from the training base.

REFERENCES:

[1]Sunita Kumari, Pankaj K. Sa, and Banshidhar Majhi. Genderclassi_cation by principal component analysis and support vector machine. In ACM

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

International Conference on Communication, Computing & Security, ICCCS 2011, pages 339 { 342, Rourkela, India February 2011.

[2] M. A. Turk and A. P. Pentland. „Face recognition using eigenfaces.‟ In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 91, pages 586 , 591, 1991.

[3]P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigen faces vs. Fisher faces: Recognition using class specific linear projection,” IEEE Trans. Patt. Anal. Mach. Intell., vol. 19, no. 7, 1997.

[4]A. M. Martinez and A. C. Kak, “PCA versus LDA,” IEEE Trans. Patt. Anal. Mach. Intell., vol. 23, pp. 228–233, 2001.

[5]P. C. Yuen and J. H. Lai, “Face representation using independent component analysis,” Pattern Recognit., vol. 35, no. 6, pp. 1247–1257,2002.

[6] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” IEEE Trans. Neural Netw., vol.13, no. 6, pp. 1450–1464, 2002.

[7]M. H. Yang, “Kernel eigne faces vs. Kernel fisher faces: Face recognition using kernel methods,” in Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition, May 2002, pp. 215–220.

[8]G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach,” Neural Comput., vol. 12, pp. 2385–2404, 2000.

[9]J. W. Lu, K. Plataniotis, and A. N. Venetsanopoulos, “Face recognition using kernel direct discriminate analysis algorithms,” IEEE Trans. Neural Netw., vol. 14, no. 1, pp. 117–126, Jan. 2003.

[10]J. Huang, P. C. Yuen, W. Chen, and J. Lai, “Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations,” IEEE Trans. Syst., Man, Cybern. B, pp.847–862, 2007.

Author’s Details:

Yapa Ashok

M.Tech (Digital Electronics and Communication Systems), Siddhartha Institute of Engineering and Technology.

Dr.D.Subba Rao,

is a proficient Ph.D person in the research area of Image Processing from Vel-Tech University, Chennai along with initial degrees of Bachelor of Technology in Electronics and Communication Engineering (ECE) from Dr. S G I E T, Markapur and Master of Technology in Embedded Systems from SRM University, Chennai. He has 13 years of teaching experience and has published 12 Papers in International Journals, 2 Papers in National Journals and has been noted under 4 International Conferences. He has a fellowship of The Institution of Electronics and Telecommunication Engineers (IETE) along with a Life time membership of Indian Society for Technical Education (ISTE). He is currently bounded as an Associate Professor and is being chaired as Head of the Department for Electronics and Communication Engineering discipline at Siddhartha Institute of Engineering and Technology, Ibrahimpatnam, Hyderabad.