EasyChair Preprint

№ 4931

A Study for Face Recognition Using Techniques PCA and KNN

Prachi Sasankar and Usha Kosarkar

January 25, 2021

A study for Face Recognition using techniques PCA and KNN.

Prof. Prachi Sasankar
Department of Computer Science
Sadabai Raisoni Women's College,
Nagpur, India.
Email:sasankar. prachi @gmail.com

Prof. Usha Kosarkar
Department of Computer Science
G.H.Raisoni Institute of IT,
Nagpur, India.
Email: usha.kosarkar@raisoni.net

Abstract- Face recognition has now become one of the interesting fields of research and has received a substantial attention of researchers from all over the world. Face recognition techniques has been mostly used in the discipline of image analysis, image processing, etc. This paper provides various techniques which are often used for face recognition in face recognition systems. In this paper performance of face recognition with two well-known image recognition methods such as Principal Component Analysis (PCA) and K-Nearest Neighbour (KNN). The effectiveness of color information plays an important role when face images are taken under strong variation in illumination, as well as low spatial resolution. In this paper, we improve a face recognition system using Principal Component Analysis (PCA) to extract features from the face images and reduce the dimensionality of each image and K nearest neighbor to classify data.

Keywords: Face Recognition System, Principal Component Analysis (PCA), K-Nearest Neighbour (KNN).

I. INTRODUCTION

Face recognition techniques is used in wide variety of face recognition systems. A face recognition technique involves identification of faces and then comparing it with the images in the database. The idea of face recognition is to give a computer system the ability of finding and recognizing human faces fast and precisely in images or videos. Numerous algorithms and techniques have been developed for improving the performance of face recognition. The face recognition is an integral part of biometrics. In biometrics, basic traits of human are matched to the existing data. Facial features are extracted and implemented through algorithms, which are efficient and some modifications are done to improve the existing algorithm models. The face recognition system generally involves two stages:

Face Detection- where the input image is searched to find any face, then image processing cleans up the facial image for easier recognition.

Face Recognition- where the detected and processed face is compared to the database of known faces to decide who that person is.

The difference between face detection and recognition is that in detection we just need to determine if there is some face in the image, but in recognition we want to determine whose face it is. Features extracted from a face present in the database. In general, face recognition techniques can be divided into two groups:

Face representation techniques-these techniques use holistic texture features and are applied to either whole face or specific regions in a face image.

Features-based techniques-these techniques use geometric facial features (mouth, eyes, brows, etc.), and geometric relationship between them.

II. Face Recognition System

A facial recognition system is a technology capable of identifying or verifying a person from a digital image or a video frame from a video source. There are multiple methods in which facial recognition system work, but in general by comparing selected facial features from given image with faces within a database. Face recognition system basically consists of following blocks:

1. Input image
2. Facial detection
3. Image processing
4. Facial recognition-features extraction
5. Verification/validation.

A. Input image

A physical or behavioural sample of image is captured from any physical image capturing device, which serves as an input the system.

B. Facial detection

It basically means locating a face in a given image.

C. Image processing

Images are cropped to avoid facial image remains and images are converted into black and white and then to gray image.

D. Facial recognition

It means to extract the features of a face (eyes, nose, etc) and make a template of it.

E. Verification

System recognizes a given individual and gives a decision.

F. Validation

In this system compares a given individual face with faces in the database to identify a given face.

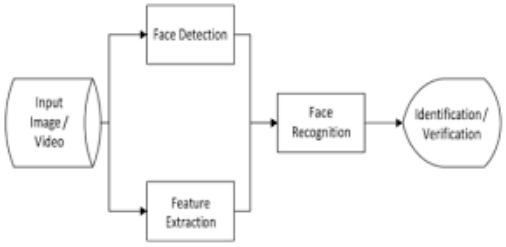

Fig 1. Block diagram of face recognition system

Figure 1 shows the block diagram of face recognition system. Nevertheless ample work has been done on face recognition problem, but still not works is up to the mark for the implementation point of view.

The images will be scanned by scanner and stored into the database. Again the image of the same candidate and stored into the database. Now two images of the same candidate will be stored into the database. The first step is to select desired images from the database then for comparisons the next step is to detect faces from each image. Final step is to recognize that images as of the same candidate or not.

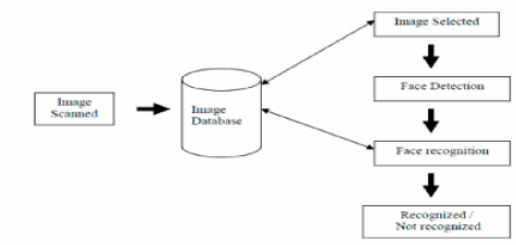

Fig 2. Structure of face recognition system.

III Face Recognition using Principal Component Analysis (PCA)

Principal Component Analysis is the simplest of the true eigenvector-based multivariate analysis. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance.

If a multivariate dataset (e.g. set of images) visualized as a set of coordinates in a high dimensional data space(1 axis per variable).

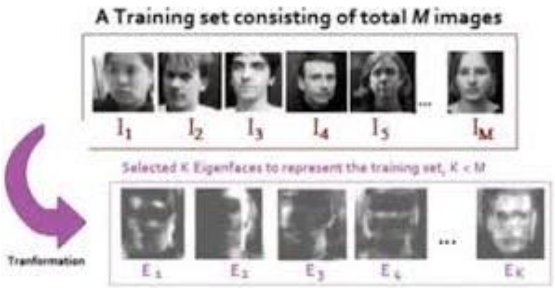

Fig 3. Face images to Eigen faces

The PCA can supply the user with a lower dimensional picture, “shadow” of the object when viewed from its informative viewpoint. Principal Component Analysis is a of possibly correlated M face images into a swt of values of uncorrelated variables called eigenfaces.

The number of principal compunents is always less than or equal to the number of original variables.$i.e\ K < M$.

This transformation is defined in such a way that the first principal component shows the most dominant “direction”/ “features” of the dataset and each succeeding component in turn shows the next most possible dominant “direction/features”. Under the constraints that it is uncorrelated the preceding components. To reduce the calculations needed for finding these principal components, the dimensionality of the original dataset is reduced before they are calculated.

Since Principal Components show the “directions” of data and each proceedings components shows less “directions” and more “noise” , Only few first Principal Components(Sat K) are selected whereas the rest of the last components is discarded.

The K principal components can safely represent the whole original dataset because they depict the major features/directions that make up the dataset.

Figure4. Selected k useful Eigenfaces

Each image in the original dataset can be represented as a linear combination (weighted sum) of Eigenfaces.

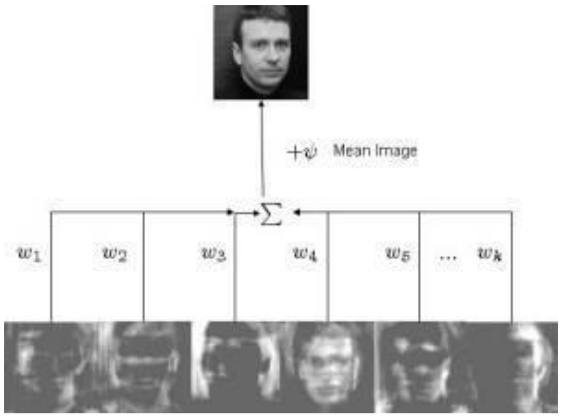

Figure5. Weighted sum of Eigen faces

So, PCA reduces the number of values (from M to K) needed to recognize a face.

IV K – Nearest NeighbourClassifier

In this method, nearest neighbour is searched, and if maximum number of neighbour belongs to one class, we can classify test image to that class. With increase k value, the classification rate decreases.

In pattern recognition, the K-nearest neighbour algorithm (KNN) is a widely used classifier for classifying objects based on closest training examples in the feature space. The k-nearest neighbour algorithm is the simplest classifier of all machine learning algorithms. In this classifier image is classified by a majority vote of its neighbours. In KNN classifier the Euclidean distance between the testing image feature and each training image feature is determined to form a distance matrix. The summation value of distance matrix is estimated and stored in increasing order. The first K elements are selected and majority class value is determined for classifying the image accurately.

The K- Nearest neighbour classifier is by far the simplest machine learning / image classification algorithm. Inside this algorithm simply relies on the distance between feature vectors. Simply put, the K-NN algorithm classifies unknown data points by finding the most common class among the k- closest examples. Each data point in the K-closest examples casts a vote and the category with the most votes wins.

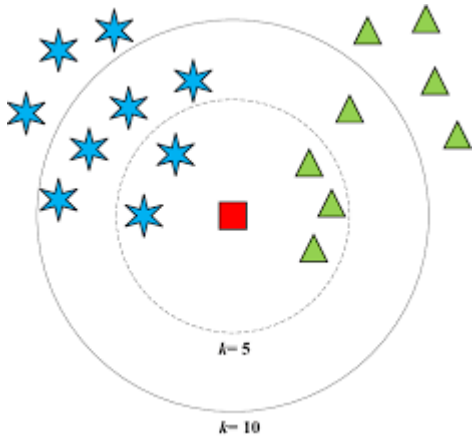

Fig.6: The principle diagram of the K-NN classification algorithm.

As shown in fig6., the red square represents the sample to be classified. It needs to be classified into blue stars or green triangle. It is obvious that it is classified to green triangle while k is set to 5, since the probability of classifying it into green triangle is 60%, which is higher than that of classifying it to blue star (40%). While k is set to 10, the red square is classified into blue star, since the probability of classifying it into blue star is 60%, higher than the probability of classifying it into green triangle (40%). In order to apply the k-nearest Neighbour classification , we need to define a distance metric or similarity function. Common choices include the Euclidean distance:

$$d(p,q) = \sqrt{\sum_{i=1}^{N}(q_i - p_i)^2}$$

Where N is the number of variables, and $q_i and p_i$ are the values of the i^{th} variable at points p and q respectively.

Other distance metrics/ similarity functions can be used depending on the type of data (the chi – squared) distance is often used for distributions.

V Results and Discussions

The proposed method different persons face images are used to test the performance of the system. For each person the database contains different facial expression images. Face images features are considered as training feature. All the images are transforming into lower dimensional images. After transformation last component of the images can be discarded.

The KNN classifier is used to classify the different face images. The Euclidean distance between the testing image feature and the training image feature is determined by finding the difference between the testing and the training feature and a distance matrix is created. In the distance matrix first k values are considered and the majority label of the k value is considered as the correct label of the given testing image. The performance of the system is measured in terms of accuracy. The accuracy is given by

$$Accuracy = \frac{correctly\ detected\ face\ images}{total\ number\ of\ face\ images}$$

VI Conclusion

The research in face recognition has been an exciting area and it will keep attracting many engineers and scientists for many years in the field. In this paper different techniques that can be used in face recognition as discussed. It also gives the whole concept of face recognition system. Image pre-processing is necessary to reduce the noise from the images and to give better recognition rate. PCA works better for features extraction. K nearest neighbour techniques and Distance-wise classification methods are used for further recognition and classification. Combining this two techniques, a face recognition application or system can be developed. The accuracy obtained in this method is much better than other results available in the literature. In future optimal set of features is found out from the space for increasing the accuracy.

VI References.

[1] P. Shih and C. Liu, “Improving the face recognition grand challenge baseline performance using color configurations across color spaces,” in Proc. IEEE Int. Conf. Image Process., 2006, pp. 1001–1004.
[2] H. M. Z. Hewa Majeed Zangana, “A New Algorithm for Human Face Detection Using Skin Color Tone,” IOSR Journal of Computer Engineering, vol. 11, no. 6, pp. 31–38, 2013.
[3] M. N. S. Zainudin et al., “Face Recognition using Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA),” no. 05, 2012.
[4] A. Amine, S. Ghouzali, and M. Rziza, “Face Detection in Still Color Images Using Skin Color Information,” Network, vol. 24, no. July 2014, p. 4, 2002.