Application of Principal Component Analysis and K-means Algorithm in Facial Biometric Recognition

Devyash Sanghai *(Author)*
Computer and Information Science
University of Florida
Gainesville, United States of America
devyashsanghai@gmail.com

***Abstract*— The goal of this paper is to implement and evaluate the performance of Principal Component Analysis and K-means Clustering algorithm in the context of problems faced during Facial Biometric Recognition and Soft Biometric Classification. Facial characteristics identification is a natural mode of identification and recognition in humans, however trying to replicate the same procedure using a Biometric Identifying System is not a simple task. One of the major problems faced by these identification system is the large amount of features in each test image(also called as 'Curse of dimensionality'). Further, problems faced by these systems are the large sample search space which increase the amount of time required to identify an individually correctly. This paper address these concerns by the application of PCA Algorithm and K-Means Clustering Algorithm.**

***Index Terms* —*Principal Component Analysis, K-means Clustering algorithm, Facial Biometric Recognition, Soft Biometric Classification, Curse of Dimensionality*.**

I. INTRODUCTION

A. Facial Biometric Recognition

Facial characteristics identification is a natural mode of identification and recognition in humans. It comes naturally to humans. Evolution has remarkably given us the ability to accurately identify faces irrespective of variations caused due to changes in expression or emotion, pose, illumination, makeup, ageing, hair growth etc. Therefore, face was also included in the set of biometric modalities. Systems which can identify or recognize individuals using their facial information were designed. [1]

The general face recognition system has the following main processes.

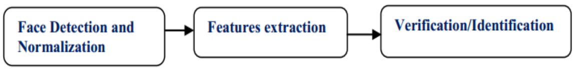


Fig 1.A.1: Face Recognition Process

Different techniques are existed for the above key elements, but the basic and more important are feature's extraction. The systems for recognition are getting rapidly advancement in their algorithms. Due to the advancement in these algorithms, the systems are able to recognize and detect other objects as well like car, humans and pedestrians. Because of these advancements, the use with these systems became popular in various fields include industrial manufacturing, security-related systems and medical field [9]. This area is become popular in public because of getting privacy and security. This still a complicated and challenging task for researchers due to human's face is very vigorous (strong) in nature. Changes in a human face can exist in short time (day to day) and long time (month or years) means due to age [10]. Real time face recognition is very important in any educational institutions now days that provide the facility of an automatic attendance system to save time. A researcher has still a big challenge to provide fast and accurate system [11]. The techniques of human face image processing mostly deal image as a two-dimensional signal there for the standard signal-processing technique is applied [12]. It is very difficult to construct a face-recognition model which is computationally less expensive because of the complexity of face. Therefore, in a face-recognition system, the feature extraction is very important for accurate recognition system. [13]

B. Multimodal Biometric System

In Multimodal biometric systems use multiple sensors or biometrics to overcome the limitations of unimodal biometric systems. For instance, iris recognition systems can be compromised by aging irises [11] and finger scanning systems by worn-out or cut fingerprints. While unimodal biometric systems are limited by the integrity of their identifier, it is unlikely that several unimodal systems will suffer from identical limitations. Multimodal biometric systems can obtain sets of information from the same marker (i.e., multiple images of an iris, or scans of the same finger) or information from different biometrics (requiring fingerprint scans and, using voice recognition, a spoken pass-code). [12][13]

C. Soft Biometrics

In Soft biometrics typically refer to attributes of people such as their gender, the shape of their head, the color of their hair, etc. There is growing interest in soft biometrics as a means of improving automated face recognition since they hold the promise of significantly reducing recognition errors, in part by ruling out illogical choices. In context of face recognition, empirical evidence suggests that significant gains using soft biometrics are hard to come by.[7]

II. PRINCIPAL COMPONENT ANALYSIS & DIMENSIONALITY REDUCTION

Principal component analysis is a statistical tool used to analyze data sets. The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of large number of interrelated variables, while retaining as much as possible of the variation present in the data set [4]. The mathematics behind principle component analysis is statistics and is hinged behind standard deviation, eigenvalues and eigenvectors. The entire subject of statistics is based around the idea that you have this big set of data, and you want to analyze that set in terms of the relationships between the individual points in that data set [5]. Images are technically data set whose component represents the image which we see. [7]

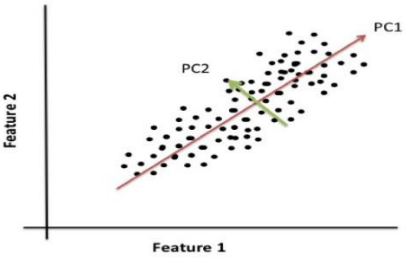


Fig II.1: Principal Component Analysis

III. K-MEANS CLUSTERING

Clustering is an unsupervised learning approach of partitioning the data set into clusters in the absence of class labels. The members of a cluster are more similar to each other than to the members of other clusters. One of the most fundamental and popular clustering techniques are KMeans [1] and Fuzzy K-Means [2] clustering algorithms. K-Means clustering technique uses the mean/centroid to represent the cluster. It divides the data set comprising of n data items into k clusters in such a way that each one of the n data items belongs to a cluster with nearest possible mean/centroid.

Procedure for K-Means Clustering:	
Input:	k: number of clusters D: the data set containing n items
Output:	A set of k clusters that minimizes the square-error function,
	Z = ∑ki=1 ∑ ||x-ci||2 Z: the sum of the squared error for all the n data items in the data set x: the data point in the space representing an item in cluster Ck ci: is the centroid/mean of cluster Ck
Steps:	1: Arbitrarily choose any k data items from D. These data items represent the initial k centroids/means. 2: Assign each of the remaining data items to the cluster that has the closest centroid. 3: Once all the data items are assigned to a cluster, recalculate the positions of the k centroids. 4: Reassign each data item to the closest cluster based on the mean value of the items in the cluster. 5: Repeat Steps 3 and 4 until the centroids no longer move.

This approach although very convenient to understand and implement has a major drawback. In case of extreme valued data items, the distribution of data will get uneven resulting in improper clustering. This makes K-Means clustering algorithm very sensitive to outliers and noise, thereby reducing its performance too. K-means is also does not work quite well in discovering clusters that have non-convex shapes or very different size. [3]

IV. METHODOLOGY

The first step is that Each 2-D facial image is represented as 1-D vector by concatenating each column (or row) into a long thin vector. So, the resulting vector should present as:

$$x^i = [x_1^i \ldots\ldots x_N^i]^T$$

The mean image is a column vector where each entry is the mean of all corresponding pixels of the training images. The preceding theory can be expressed in the following expression:

$$\bar{x}^i = x^i - m$$

Where:

$$m = \frac{1}{p}\sum_{i=1}^{p} x^i$$

Center data:

$$\bar{X} = [\bar{x}^1|\bar{x}^2|\ldots|\bar{x}^p||]$$

Covariance Matrix:

$$\Omega = \overline{XX}^T$$

Sorting the order of Eigen vectors

$$V = [V_1|V_2|\ldots|V_p||]$$

Projecting the images

$$\tilde{x}^i = V^T \bar{x}^i$$

Identifying the new images

$$\bar{y}^i = y^i - m$$

$$\tilde{y}^i = V^T \bar{y}^i$$

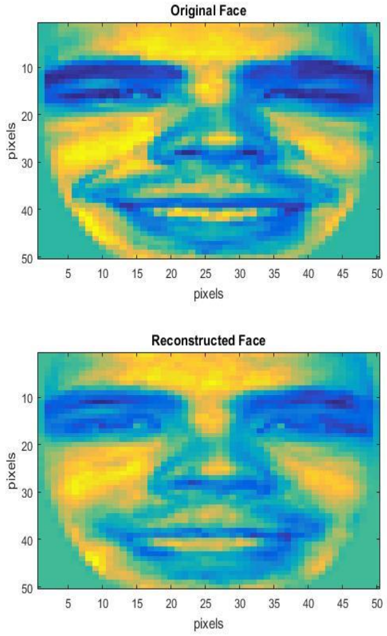
After using PCA, unwanted variations caused by the illumination, facial position and facial expression still retain.

Now Using Euclidean Distance as a distance measure we calculate the minimum or least distance a test image of ‘Probe Set’ is from the image of ‘Gallery Set.’

Formulae of Euclidian Distance

$$D(x_j, y_j) = \sqrt{\Sigma_{j=1}^{n}(x_j - y_j)^2}$$

After finding the minimum distance, we output that individual as the individual from the gallery set. Below results are calculated using the above assigned classes.

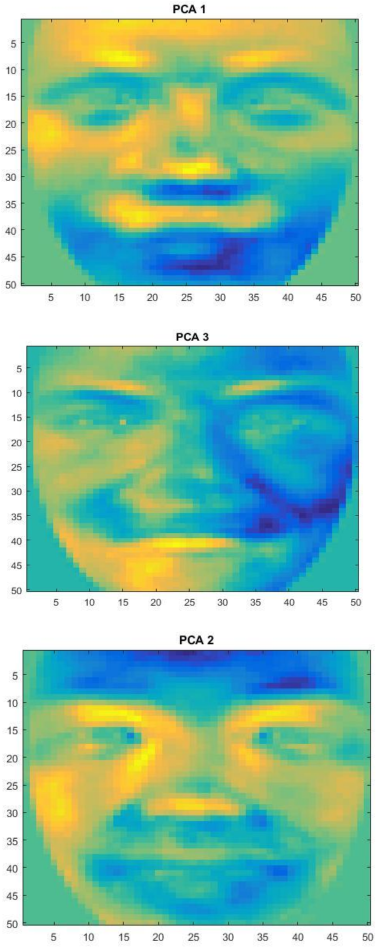

Original Face

Reconstructed Face

Fig IV.1 : The original Face vs the reconstructed face.

V. RESULTS AND ANALYSIS

A. PCA Projections or Eigen Faces


PCA 1

PCA 3

PCA 2

Fig V.A.1: First Three Principal Components (ordered from top to bottom)

Eigen faces is the name given to a set of eigenvectors when they are used in the computer vision problem of human face recognition. [14] The eigenvectors are derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images. The Eigen faces themselves form a basis set of all images used to construct the covariance matrix. This produces dimension reduction by allowing the smaller set of basis images to represent the original training images. Classification can be achieved by comparing how faces are represented by the basis set.

The largest principal component represents the direction of maximum variance in the images. The largest principal component also represents unwanted variations caused by the illumination, facial position and facial expression.

B. Recognition Performance with PCA

Recognition Rate is the accuracy of the correct Identification of the Biometric System.

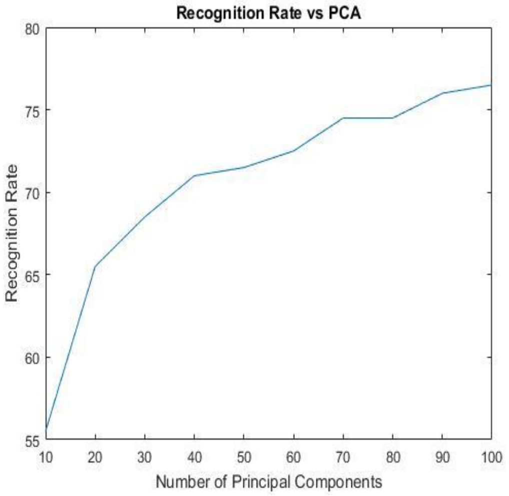


Fig V.B.1: Recognition rate after variating the number of principal components.

Here Recognition rate is calculated after the applying PCA. We variate the number of principal components used and plot the Correct Recognition rate. We observe that the recognition rate monotonically increases to the number of principal components used to project the data, and then remains constant after 100 principle components used.

We observe this as **the variance after 100 principal component has become negligible**, so the **biometric system is not getting any new information to enhance the identification of the individual.**

The **Maximum Recognition rate** that was observed was **76.5% using 100 principal components** in **0.056847 seconds.** Running the program on a personal computer on MATLAB.

C. Recognition Performance without P C A

Without reducing the dimension of the data given to us, if we calculated the Maximum Recognition rate, it was observed that, we were getting the same recognition rate as using just 100 features from the reduced dimensional data from PCA. This indeed gave significant boost in the time and memory required to recognize the individual. The time required to find the recognition rate was **0.495015 seconds**. Which is almost **10 times** more than time required using PCA.

This empirical result is significant as we just increased the **performance of the Biometric system 10 fold** as well as decreased the **memory requirement to just 8% of the initial memory requirement without PCA.**

D. Recognition Performance of K-Means Classification

Since K-Means, is unsupervised clustering algorithm, it gives random recognition performance, as the clustering is not only done using ‘male’, ‘female’ information., **K-Means is clustering the data based on their Euclidian distance**. **It is not compulsory that all ‘males’ have a similar face structure and all ‘females’ have a similar face structure. Hence, the recognition performance would be random.**

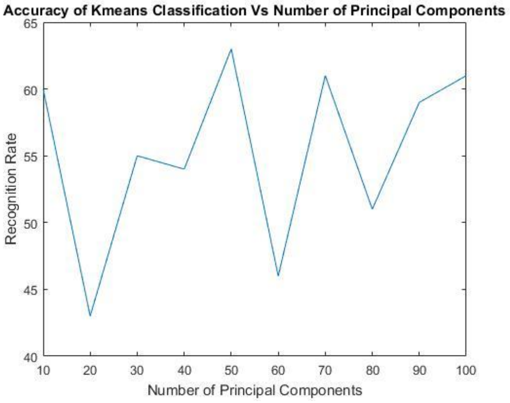


Fig V.D.1: Recognition rate after variating the number of principal components.

E. Internal K- Means Cluster Validity index- Davis Bouldin Index

In Davies–Bouldin index a lower value will mean that the clustering is better. It happens to be the average similarity between each cluster and its most similar one, averaged over all the clusters, where the similarity is defined as Si above. This affirms the idea that no cluster has to be similar to another, and hence the best clustering scheme essentially minimizes the Davies–Bouldin index[17]

The Below plot gives a **decreasing criterion value of Cluster size as 2. This verifies that clustering is better as the number of principal components are increased. Which basically means the more information is provided to the K-Means algorithm the better will it be able to cluster the data into different classes.**

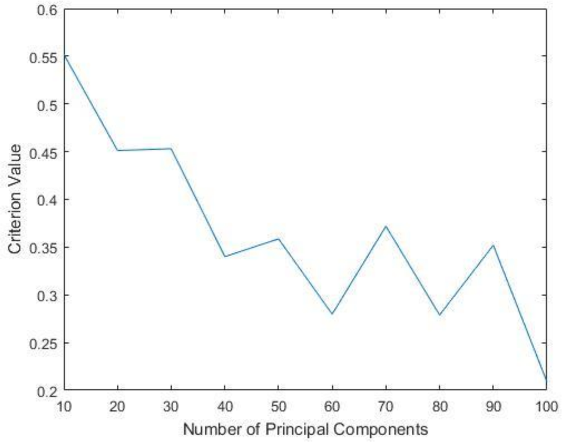


Fig V.E.1: Criterion value after variating the number of principal components.

F. External K- Means Cluster Validity index- F measure

F-measure The F-measure can be used to balance the contribution of false negatives by weighting recall through a parameter

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$

where P is the precision rate and R is the recall rate. We can calculate the F-measure by using the following formula: [16]

$$F_\beta = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 \cdot P + R}$$

Notice that when β , F_0=P . In other words, recall has no impact on the F-measure when β=0, and increasing βallocates an increasing amount of weight to recall in the final F-measure.

The F-Measure value is approximately =0.6712. F-measure has an intuitive meaning. **It tells you how precise your classifier is (how many instances it classifies correctly)**, as well as how robust it is (it does not miss a significant number of instances). With high precision but low recall, your classifier is extremely accurate, but it misses a significant number of instances that are difficult to classify. This is not very useful [18]

VI. Conclusion

We have effectively alleviated 2 major problems in Biometric Classification. We showed that the problem with large feature set could effectively reduce using PCA with little or almost nil empirically proved compromise. We also found that, after using PCA, unwanted variations caused by the illumination, facial position and facial expression still retain. However, the boost in memory and performance is significant. We showed that we can used soft Biometric as a means of classification to improve the large-scale facial recognition.

References

[1] Hartigan, J. A.; Wong, M. A. (1979), "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C 28 (1): 100–108

[2] Bezdek, James C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms.

[3] International Journal of Soft Computing, Mathematics and Control (IJSCMC), Vol. 3, No. 3, August 2014 DOI : 10.14810/ijscmc.2014.3301 1 K-MEDOIDS CLUSTERING USING PARTITIONING AROUND MEDOIDS FOR PERFORMING FACE RECOGNITION Aruna Bhat Department of Electrical Engineering, IIT Delhi, Hauz Khas, New Delhi

[4] I.T. Jolliffe, Prinicipal Component Analysis, 2nd Edition, Springer series in statistics 2002, page 1-3.

[5] Lindsay I Smith, A tutorial on Principal Components Analysis, February 26, 2002, page 2-8

[6] IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 2, November 2011 ISSN (Online): 1694-0814 www.IJCSI.org Principal Component Analysis-Linear Discriminant Analysis Feature Extractor for Pattern Recognition Aamir Khan1, Hasan Farooq

[7] SOFT BIOMETRICS ARE HARD Hao Zhang Department of Computer Science Colorado State University

[8] T. V. N. Rao, V. S. Aditya, S. Venkateshwarlu and B. Vasavi, "Partition Based Face Recognition System", Journal Of Global Research In Computer Science, vol. 2, no. 9, **(2011)** September.

[9] S. Singh, M. Sharma and N. S. Rao, "Robust & Accurate Face Recognition using Histograms", International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'2011), Bangkok, **(2011)** December

[10] H. Rady, "Face Recognition using Principle Component Analysis with Different Distance Classifiers", IJCSNS International Journal of Computer Science and Network Security, vol. 11, no. 10, **(2011)** October

[11] H. M. Hasan, W. A. Al Jouhar and M. A. Alwan, "Face Recognition Using Improved FFT Based Radon by PSO and PCA Techniques", International Journal of Image Processing (IJIP), vol. 6, no. 1, **(2012)**

[12] V. Vaidehi, S. Vasuhi, R. Kayalvizhi, K. Mariammal, M. B. Raghuraman, V. S. Raman, L. Meenakshi, V. Anupriyadharshini and T. Thangamani, "Person Authentication Using Face Detection", Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, San Francisco, USA, **(2008)** October 22-24.

[13] Analytical Study of Face Recognition Techniques Sabir Shah1, Faizanullah1, Sajid Ali khan1 and Naveed Riaz1

[14] M. Turk; A. Pentland (1991). "Face recognition using eigenfaces" (PDF). Proc. IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–591

[15] Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". Remote Sensing of Environment. 62 (1): 77–89. doi:10.1016/S0034-4257(97)00083-7

[16] Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hinrich. Introduction to Information Retrieval. Cambridge University Press. ISBN 978-0-521-86571-5.

[17] Davies, David L.; Bouldin, Donald W. (1979). "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224–227. doi:10.1109/TPAMI.1979.4766909

[18] How to interpret F-measure values? (n.d.). Retrieved November 28, 2016, from http://stats.stackexchange.com/questions/49226/how-to-interpret-f-measure-values