K-MEDOIDS CLUSTERING USING PARTITIONING AROUND MEDOIDS FOR PERFORMING FACE RECOGNITION

Aruna Bhat

Department of Electrical Engineering, IIT Delhi, Hauz Khas, New Delhi

ABSTRACT

Face recognition is one of the most unobtrusive biometric techniques that can be used for access control as well as surveillance purposes. Various methods for implementing face recognition have been proposed with varying degrees of performance in different scenarios. The most common issue with effective facial biometric systems is high susceptibility of variations in the face owing to different factors like changes in pose, varying illumination, different expression, presence of outliers, noise etc. This paper explores a novel technique for face recognition by performing classification of the face images using unsupervised learning approach through K-Medoids clustering. Partitioning Around Medoids algorithm (PAM) has been used for performing K-Medoids clustering of the data. The results are suggestive of increased robustness to noise and outliers in comparison to other clustering methods. Therefore the technique can also be used to increase the overall robustness of a face recognition system and thereby increase its invariance and make it a reliably usable biometric modality.

KEYWORDS:

MEDOIDS, CLUSTERING, K-MEANS, K-MEDOIDS, PARTITIONING AROUND MEDOIDS

1. INTRODUCTION

Face is a natural mode of identification and recognition in humans. It comes intuitively to people for recognising others. They have a remarkable 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 recognise individuals using their facial information were designed [1]. One of the most useful advantages of having face as a morphological trait for recognition purpose was its non invasiveness. It was beneficial both in terms of cost, time and efforts to record the data for the biometric system. It altogether removed the need of having expensive scanners which were vital for other biometric systems like fingerprint, iris etc. It could also be used even without the knowledge of the user and immediately found its application in surveillance.

There have been different approaches for developing an efficient face recognition system. Various techniques for face detection, feature extraction and classification have been designed. Viola and Jones developed the method of face detection using a combination of filters to accurately identify face in an image [2]. The method was further enhanced by R. Lienhart and J. Maydt [3]. However detecting a face which is the first step in face recognition is far more challenging in uncontrolled environments. The detection is followed by image processing and

DOI : 10.14810/ijscmc.2014.3301

feature extraction. The method of using Principal component Analysis (PCA) [4] was proposed by M.A.Turk & A.P.Pentland [5] [6]. The face images are converted into Eigen faces which represent the principal components of the former in the form of eigen vectors. M. Kirby & L. Sirovich developed the Karhunen-Lo`eve procedure for the characterization of human faces [7]. It was followed by a new system which used Fisher Linear Discriminant Analysis [8] instead of PCA and generated Fisher faces [9] for recognition. Several variations of the methodology have been developed based on Kernel PCA, Probabilistic and Independent Component Analysis [10]. These methods have been enhanced further and utilised in dynamic face recognition [11]. Reasonable success has also been achieved with the use of unsupervised learning methods. Variants of clustering techniques for face recognition have been proposed. Mu-Chun Su and Chien-Hsing Chou suggested a modified Version of the K-Means Algorithm for face recognition [12]. Statistical Face Recognition using K-Means Iterative algorithm was proposed by Cifarelli, Manfredi and Nieddu [13]. K-Means algorithm also has been applied in recognising variations in faces like expression and emotions [14]. Face recognition using Fuzzy c-Means clustering and sub-NNs was developed by Lu J, Yuan X and Yahagi T [15]. Clustering has also found its use in dynamic and 3D face recognition applications too [16] [17]. There are many other methodologies which have been proposed for efficient face recognition and are being improvised incessantly [18].

The method discussed in this paper describes a novel approach of K-Medoids clustering [21] for face recognition. The choice of using this algorithm comes from its robustness as it is not affected by the presence of outliers or noise or extremes unlike clustering techniques based on K-Means [19] [20]. This advantage clearly leads towards the development of sturdy face recognition system which is invariant to the changes in pose, gait, expressions, illumination etc. Such a robust and unobtrusive biometric system can surely be applied in real life scenarios for authentication and surveillance.

The rest of this paper is organized as follows: Section 2 discusses the conceptual details about K-Medoids Clustering and Partitioning Around Medoids (PAM) [21]. The proposed methodology of applying K-Medoids Clustering to face recognition is discussed in Section 3. Experimental results are provided in Section 4. Section 5 discusses the conclusion and the future scope.

2. OVERVIEW OF METHODOLOGY

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 K-Means [19] and Fuzzy K-Means [20] 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:

<u>Input:</u>

k: number of clusters

D: the data set containing n items

<u>Output:</u>

A set of k clusters that minimizes the square-error function,

$$\mathbf{Z = \sum_{i=1}^{k} \sum ||x - c_i||^2} \quad \mathbf{(1)}$$

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 C_k
c_i: is the centroid/mean of cluster C_k

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. This calls for another approach to clustering that is based on similar lines, yet is robust to outliers and noise which are bound to occur in realistic uncontrolled environment.

K-Medoids clustering [21] is one such algorithm. Rather than using conventional mean/centroid, it uses medoids to represent the clusters. The medoid is a statistic which represents that data member of a data set whose average dissimilarity to all the other members of the set is minimal. Therefore a medoid unlike mean is always a member of the data set. It represents the most centrally located data item of the data set.

The working of K-Medoids clustering [21] algorithm is similar to K-Means clustering [19]. It also begins with randomly selecting k data items as initial medoids to represent the k clusters. All the other remaining items are included in a cluster which has its medoid closest to them. Thereafter a new medoid is determined which can represent the cluster better. All the remaining data items are yet again assigned to the clusters having closest medoid. In each iteration, the medoids alter their location. The method minimizes the sum of the dissimilarities between each data item and its corresponding medoid. This cycle is repeated till no medoid changes its placement. This marks the end of the process and we have the resultant final clusters with their medoids defined. K clusters are formed which are centred around the medoids and all the data members are placed in the appropriate cluster based on nearest medoid.

Procedure for K-Medoid Clustering:

Input:

k: number of clusters
D: the data set containing n items

Output:

A set of k clusters that minimizes the sum of the dissimilarities of all the objects to their nearest medoids.

$$\mathbf{Z = \sum^{k}_{i-1} \sum |x-m_i|} \quad \textbf{(2)}$$

Z: Sum of absolute error for all items in the data set
x: the data point in the space representing a data item
m_i: is the medoid of cluster Ci

Steps:

1: Arbitrarily choose k data items as the initial medoids.
2: Assign each remaining data item to a cluster with the nearest medoid.
3. Randomly select a non-medoid data item and compute the total cost of swapping old medoid data item with the currently selected non-medoid data item.
4. If the total cost of swapping is less than zero, then perform the swap operation to generate the new set of k-medoids.
5. Repeat steps 2, 3 and 4 till the medoids stabilize their locations.

There are various approaches for performing K-Medoid Clustering. Some of them are listed below:

I. PAM (Partitioning Around Medoids):
It was proposed in 1987 by Kaufman and Rousseeuw [21]. The above K-Medoid clustering algorithm is based on this method.

It starts from an initial set of medoids and iteratively replaces one of the medoids by one of the non-medoids if it improves the total distance of the resultant clustering. It selects k representative medoid data items arbitrarily. For each pair of non-medoid data item x and selected medoid m, the total swapping cost S is calculated. If S< 0, m is replaced by x. Thereafter each remaining data item is assigned to cluster based on the most similar representative medoid. This process is repeated until there is no change in medoids.

Algorithm:

1. Use the real data items in the data set to represent the clusters.
2. Select k representative objects as medoids arbitrarily.
3. For each pair of non-medoid item $\mathbf{x_i}$ and selected medoid $\mathbf{m_k}$, calculate the total swapping cost $S(x_i m_k)$. For each pair of $\mathbf{x_i}$ and $\mathbf{m_k}$
   If $S < 0$, $\mathbf{m_k}$ is replaced by $\mathbf{x_i}$
   Assign each data item to the cluster with most similar representative item i.e. medoid.
4. Repeat steps 2-3 until there is no change in the medoids.

II. CLARA (CLustering LARge Applications) [23] was also developed by Kaufmann & Rousseeuw in 1990. It draws multiple samples of the data set and then applies PAM on each sample giving a better resultant clustering. It is able to deal more efficiently with larger data sets than PAM method.

CLARA applies sampling approach to handle large data sets. Rather than finding medoids for the entire data set D, CLARA first draws a small sample from the data set and then applies the PAM algorithm to generate an optimal set of medoids for the sample. The quality of resulting medoids is measured by the average dissimilarity between every item in the entire data space D and the medoid of its cluster. The cost function is defined as follows:

$Cost(m_d,D) = \sum^{n}_{i-1} d(x_i, rpst(m_d, x_i)) / n$
where, m_d is a set of selected medoids, d(a, b) is the dissimilarity between items a and b and $rpst(m_d, x_i)$ returns a medoid in m_d which is closest to x_i.

The sampling and clustering processes are repeated a pre-defined number of times. The clustering that yields the set of medoids with the minimal cost is selected.

III. CLARANS (Randomized CLARA) was designed by Ng & Han [24]. CLARANS draws sample of neighbours dynamically. This clustering technique mimics the graph search problem wherein every node is a potential solution, here, a set of k medoids. If the local optimum is found, search for a new local optimum is done with new randomly selected node. It is more efficient and scalable than both PAM and CLARA.

3. Proposed Methodology

3.1. Pre-processing and Feature Extraction

Open source data sets were used to evaluate the performance of the technique. JAFFE database [25] contains the face images with varying expressions and emotions. The face images were segmented and processed as these preliminary steps directly impact the final results in recognition. After determining the ROI, the features were extracted. Viola-Jones object detection algorithm [2] was used to detect the frontal faces as well the features like eyes, nose and lips in the respective ROI images of the faces. Viola-Jones object detection framework is a robust technique and is known to perform accurate detection even in real time scenarios. It has been used extensively to detect faces and face parts in the images. Therefore this algorithm was used to extract the faces from the images and features from the former. For n face images, a two dimensional feature vector D was created such that n rows represent each of the n faces and p columns represent the complete feature information of every face.

3.2. Classification through Clustering

The information thus obtained from the facial images in the data set was clustered using K-Medoid Clustering. Partitioning Around Medoids (PAM) [21] technique was used to perform the clustering of the data space D.

The number of clusters is decided by the number of classes we have in the data set i.e. the number of individuals whose face images are present in the data set. To find the k medoids from the feature data space D, PAM begins with an arbitrary selection of k objects. It is followed by a swap between a selected object Ts and a non-selected object Tn, if and only if this swap would result in an improvement of the quality of the clustering. To measure the effect of such a swap between Ts and Tn, PAM computes costs $TCost_{jsn}$ for all non-selected objects Tj. Let d(a,b) represent the dissimilarity between items a and b. The cost $TCost_{jsn}$ is determined as follows depending on which of the cases Tj is in:

I. Tj currently belongs to the cluster represented by C_i and Tj is more similar to Tj' than Tn.
d(Tj, Tn) >= d(Tj, Tj') where Tj' is the second most similar medoid to Tj.
So if Ts is replaced by Tn as a medoid, Tj would belong to the cluster which is represented by Tj'.
The cost of the swap is:

$TCost_{jsn} = d(Tj, Tj') – d(Tj, Ts)$

(3)

This always gives a non-negative $TCost_{jsn}$ indicating that there is a non-negative cost incurred in replacing Ts with Tn.

II. Tj currently belongs to the cluster represented by C_i and Tj is less similar to Tj' than Tn, d(Tj,Tn) < d(Tj, Tj')

So if Ts is replaced by Tn as a medoid, Tj would belong to the cluster represented by Tn.

Thus, the cost for Tj is:

$TCost_{jsn} = d(Tj', Tn) – d(Tj, Ts)$

(4)

Here, $TCost_{jsn}$ can be positive or negative, based on whether Tj is more similar to Ts or to Tn.

III. Tj currently belongs to a cluster other than the one represented by Ts and Tj is more similar to Tj' than Tn,

Let Tj' be the medoid of that cluster. Then even if Ts is replaced by Tn, Tj would stay in the cluster represented by Tj'.

Thus, the cost is zero:

$TCost_{jsn} = 0$

(5)

IV. Tj currently belongs to a cluster represented by Tj' and Tj is less similar to Tj' than Tn.

Replacing Ts with Tn would cause Tn to shift to the cluster of Tn from that of Tj'.

Thus the cost involved is:

$TCost_{jsn} = d(Tj, Tn) – d(Tj, Tj')$

(6)

This cost is always negative.

Combining the above four cases, the total cost for replacing Ts with Tn is given by:

$TotalCost_{jsn} = TCost_{jsn}$

(7)

<u>Algorithm:</u>

1. From the data space D, select k representative objects randomly and mark these as medoids.
2. Remaining data items are non-medoids.
3. Repeat till medoids stabilise/converge

```
    for all medoid items Ts
        for all non medoid items Tn
            calculate the cost of swapping TCost_jsn
        end
    end
    Select s_min and n_min such that TCost_s_min,n_min = Min TCost_jsn
        if TCost_s_min,n_min < 0,
            mark s_min as non medoid and n_min as medoid item.
        end
```

4. Generate k clusters C_1………….. C_k.

Once the clustering is done, all the face images are assigned to a particular cluster based on the extent of similarity to the medoid data item of that cluster.

Thus resultant clustering also classifies the face images (in different expressions/pose/emotions/illumination) to correct individual classes.

4. EXPERIMENTAL RESULTS

JAFFE database [25] was used to experimentally evaluate the efficacy of the proposed method. After performing processing and feature extraction of the face images in the data set, K-Medoid clustering using Partitioning Around Medoids (PAM) [21] was done over the data space D which represented the feature information from n face images.

The standard K-Means clustering [19] technique was also evaluated in order to compare the performance of face recognition by K-Medoid and K-Means clustering.

The results with K-Medoid clustering were more robust to the outliers. It was also effective in classifying the images reasonably even in presence of variations in the image owing to changes in expressions and emotions. The algorithm is effective in terms of accuracy and time with medium sized data sets and performs better than K-Means clustering technique in case of noise and outliers. The classification precision of K-Medoid clustering using PAM was observed to be more than that of K-Means clustering for all the data sets where outliers and noise was present. The PAM algorithm recognised the faces with different expressions more accurately as summarised in Table 1.

However as the data set size increases and the extent of noise and outliers are reduced, the performance is nearly similar to standard K-Means with comparably higher computation cost involved.

Table 1. Comparison of results K-Medoids vs K-Means Face Recognition

Data Set					**Clustering** (Accuracy %)	
					K-Means	**K-Medoids**
	Size	images/individual (varying expressions)	Noise	Outliers		
Set I	100	4	Y	Y	65	78
	100	4	Y	N	70	78
	100	4	N	Y	72	77
	100	4	N	N	81	79
Set II	150	5	Y	Y	66	78
	150	5	Y	N	70	77
	150	5	N	Y	72	76
	150	5	N	N	82	79
Set III	200	20	Y	Y	68	77
	200	20	Y	N	72	76
	200	20	N	Y	72	76
	200	20	N	N	82	77

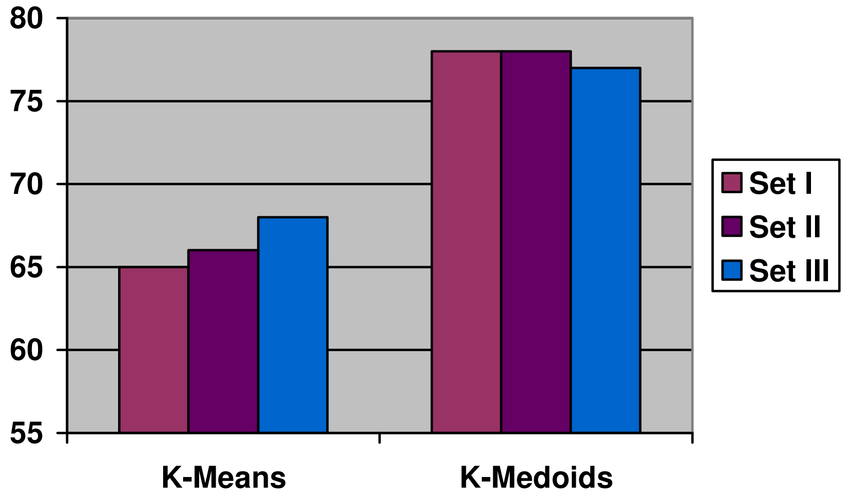

Figure 1. Comparison of results K-Means vs. K-Medoids (Noise: Y, Outliers: Y)

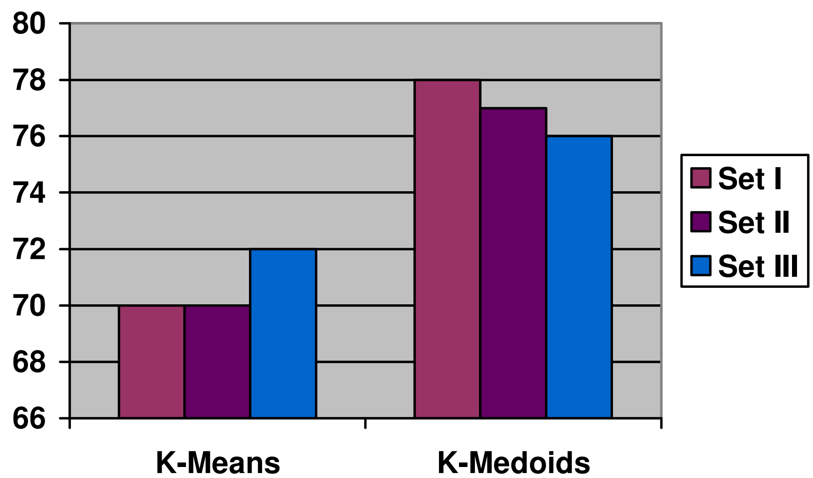

Figure 2. Comparison of results K-Means vs. K-Medoids (Noise: Y, Outliers: N)

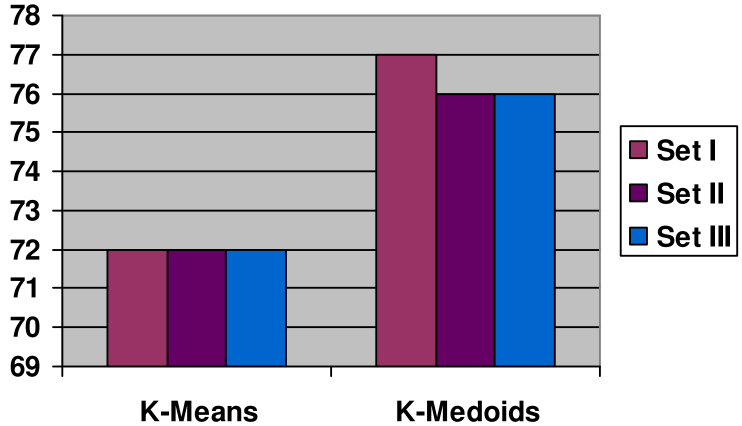

Figure 3. Comparison of results K-Means vs. K-Medoids (Noise: N, Outliers: Y)

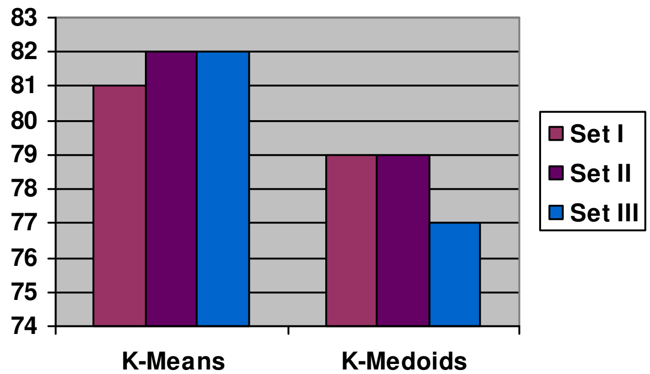

Figure 4. Comparison of results K-Means vs. K-Medoids (Noise: N, Outliers: N)

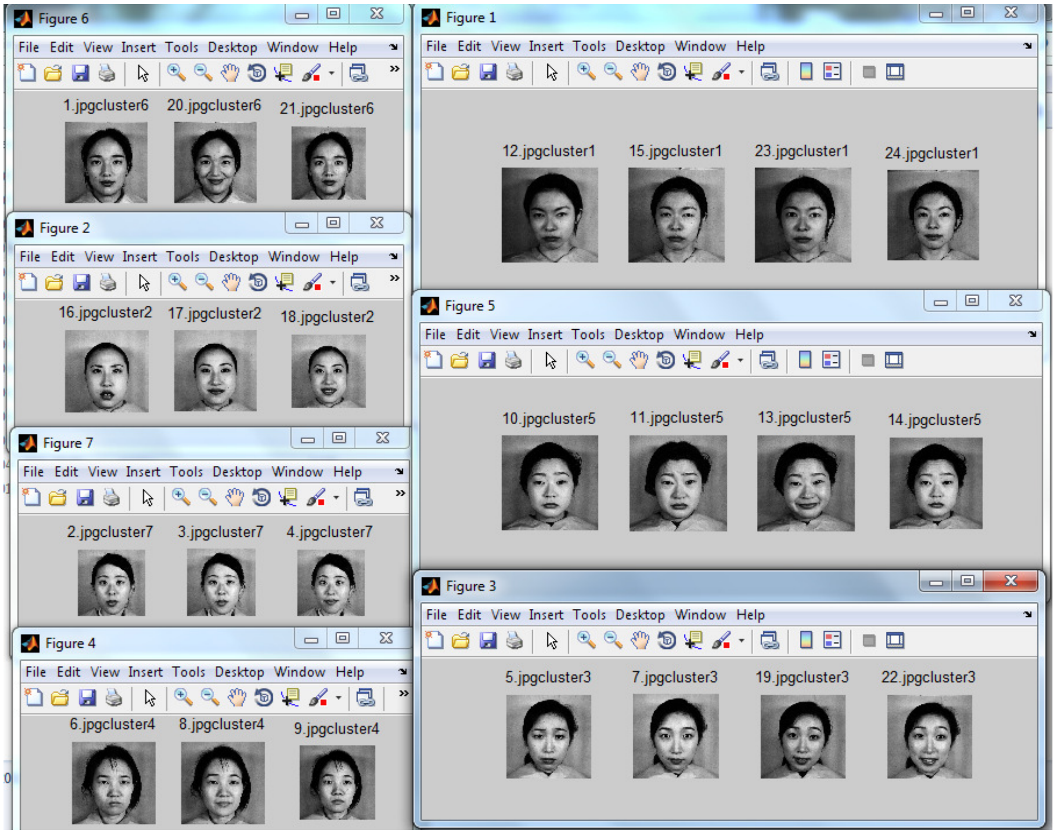

Figure 5. An instance from the K-Medoid clustering results

The results obtained clearly depict an increase in the recognition accuracy in spite of the presence of noise and/or outliers in the face images when the classification is done using Partioning Around Medoids method of K-Medoid clustering. The resultant system is also reasonably invariant to the changes in expressions as well. In comparison to K-Means clustering, the proposed method is far more robust and usable in real scenarios where noise and outliers are bound to be present.

5. Conclusions and Future Scope

As observed through the experimental analysis, the K-Medoids clustering [21] technique using Partitioning Around Medoids (PAM) has performance comparable to that of K-Means clustering technique in absence of noise and outliers. However its remarkable efficiency in the presence of extreme values or outliers in the data in the data set makes it unique.

It also showcased robustness to noise and variations of expressions and emotions in the face images. The recognition accuracy stays high without any adverse impact by aberrations that are caused by noise and outliers.

Therefore K-Medoid clustering technique can help in designing sturdy face recognition systems which are invariant to the changes in pose, illumination, expression, emotions, facial distractions like make up and hair growth etc. The real time uncontrolled environment will always have some noise factor or variations in face. The ability of this algorithm to deal with these unavoidable distractions in the data set encourages its use in designing robust face recognition systems.

The higher computation cost of K-Medoid clustering technique using PAM in comparison to K-Means clustering is a concern for its application to bigger data sets. However many variants to PAM have been developed now which are computationally as favourable as K-Means algorithm and perform better than both PAM and K-means. We can apply such K-Medoid algorithms to face recognition and evaluate their performance.

We may also use different facial feature extraction techniques like Eigen faces and Fisher faces [22] etc and perform K-Medoid clustering over that data. It could also be used with various feature detectors and descriptors like SIFT [26] and SURF [27].

ACKNOWLEDGEMENTS

The author would like to thank the faculty and staff at the Department of Electrical Engineering, IIT Delhi for the immense guidance and help they have graciously provided.

REFERENCES

[1] V. Bruce, P.J.B. Hancock and A.M. Burton, (Apr 1998) “Comparisons between human and computer recognition of faces”, Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, Vol., Iss., 14-16 Pages:408-413.

[2] P. Viola and M. Jones., (2001) “Rapid object detection using a boosted cascade of simple features”. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001.

[3] R. Lienhart and J. Maydt, (Sep 2002) “An extended Set of Haar-like Features for Rapid Object Detection”, IEEE ICIP 2002, Vol 1, pp. 900-903.

[4] P.J.B. Hancock, A.M. Burton & V. Bruce, (1996) “Face processing: Human perception and principal components analysis”, Memory and Cognition, 24(1):26-40.

[5] M.Turk & A.Pentland, (1991) “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991a.

[6] M. Turk & A. Pentland, (1991) “Face Recognition Using Eigenfaces”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591.

[7] M. Kirby & L. Sirovich, (1990) "Application of the Karhunen-Lo`eve procedure for the characterization of human faces", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, N1:103-108.

[8] W.Zhao, R. Chellappa & A. Krishnaswamy, (1998) “Discriminant Analysis of Principal Components for face Recognition”, Proc. Third Int’l Conf. Automatic Face and Gesture Recognition, pp. 336-341.

[9] P.N. Belhumeur, J.P. Hespanha & D.J. Kriegman, (July 1997) “Eigenfaces vs. Fisherfaces: Recognition Using Class-Specific Linear Projection”, PAMI(19), No. 7, pp. 711-720.

[10] Marian Stewart Bartlelt, Javier R. Movellan & and Ferrence J. Sejnowski, (Nov’2002) “Face Recognition by Independent Component Analysis”, IEEE Trans. on Neural Networks, vol.13, no. 6.

[11] A. Pavan Kumar & V. Kamakoti & Sukhendu Das (2004), An Architecture For Real Time Face Recognition Using WMPCA, ICVGIP 2004.

[12] Mu-Chun Su and Chien-Hsing Chou, (June 2001)A Modified Version of the K-Means Algorithm with a Distance Based on Cluster Symmetry, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 23, No. 6.

[13] Cifarelli, C. Manfredi, G. Nieddu, L. (2008), "Statistical Face Recognition via a k-Means Iterative Algorithm", Machine Learning and Applications, ICMLA '08.

[14] Zeki, A.M., bt Serda Ali, R., Appalasamy, P. (2012), "K-means approach to facial expressions recognition", Information Technology and e-Services (ICITeS).

[15] Lu J, Yuan X, Yahagi T., (2007 Jan)A method of face recognition based on fuzzy c-means clustering and associated sub-NNs, IEEE Trans Neural Netw,18(1):150-60.

[16] Marios Kyperountasa, Author Vitae, Anastasios Tefasa, Author Vitae, Ioannis Pitas, (March 2008) Dynamic training using multistage clustering for face recognition Pattern Recognition , Elsevier, Volume 41, Issue 3, Pages 894–905.

[17] Zheng Zhang, Zheng Zhao, Gang Bai , (2009) 3D Representative Face and Clustering Based Illumination Estimation for Face Recognition and Expression Recognition, Advances in Neural Networks – ISNN 2009 Lecture Notes in Computer Science Volume 5553, pp 416-422.

[18] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, “Face Recognition: A Literature Survey”.

[19] 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.

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

[21] Kaufman, L. and Rousseeuw, P.J. (1987), Clustering by means of Medoids, in Statistical Data Analysis Based on the L1 - Norm and Related Methods, edited by Y. Dodge, North- Holland, 405–416.

[22] M.Turk, W. Zhao, R chellapa, (2005), “Eigenfaces and Beyond”, Academic Press.

[23] Kaufman, L. and Rousseeuw, P. J. (2005). Finding groups in data: An introduction to cluster analysis. John Wiley and Sons.

[24] Ng, R.T., Jiawei Han, (2002) "CLARANS: a method for clustering objects for spatial data mining", Knowledge and Data Engineering, IEEE Transactions on (Volume:14, Issue:5), Pg 1003 - 1016, ISSN: 1041-4347.

[25] JAFFE database, Coding Facial Expressions with Gabor Wavelets Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi & Jiro Gyoba Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.

[26] M. Bicego, A. Lagorio, E. Grosso, M. Tistarelli, (2006) "On the use of SIFT features for face authentication," Proc. of IEEE Int Workshop on Biometrics, in association with CVPR, NY.

[27] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, (2008) "Speeded-up robust features(SURF)," Comput. Vis. Image Underst.,110(3), 346-359.

Authors

The author is a research scholar at Indian Institute of Technology, Hauz Khas, Delhi.