

FACE RECOGNITION VIA ADAPTIVE DISCRIMINANT CLUSTERING

Marios Kyperountas, Anastasios Tefas, Ioannis Pitas

Aristotle University of Thessaloniki, Department of Informatics
Artificial Intelligence and Information Analysis Laboratory
Box 451, GR-54124 Thessaloniki, Greece

ABSTRACT

This paper presents a methodology that tackles the face recognition problem by accommodating multiple clustering steps. At each clustering step, the test and training faces are projected to a discriminant space and the projected training data are partitioned into clusters using the k-means algorithm. Then a subset of the training data clusters is selected, based on how similar the faces in these clusters are to the test face. In the clustering step that follows a new discriminant space is defined by processing this subset and both the test and training data are projected to this space. This process is repeated until one final cluster is selected and the most similar, to the test face, face class contained is set as the identity match. The UMIST and XM2VTS face databases have been used to evaluate the algorithm and results indicate that the proposed framework provides a promising solution to the face recognition problem.

Index Terms — *adaptive classification, face recognition, discriminant analysis*

1. INTRODUCTION

In recent years, developing face recognition (FR) technology has received great attention. For the face recognition problem, the true match to a test face, out of a number of N different training faces stored in a database, is sought. Unfortunately, when large in size databases are considered, the performance of many state-of-the-art FR methods deteriorates rapidly [1,2]. Differences in viewpoint, illumination and facial expression incapacitate the ability of methods that use linear criteria to generalize for all the introduced variations. In addition, problems such as over-fitting, computational complexity and difficulties in optimizing the involved non-linear parameters often appear [1].

Recently, various methods have attempted to solve the aforementioned problems. A widely used principle that

has been used is the ‘divide and conquer’, which decomposes a database into smaller sets in order to piecewise learn the complex distribution by a mixture of local linear models. In [1] a linear discriminant analysis (LDA)-like separability criterion is used to partition a training set from a large database into a set of smaller maximal separability clusters (MSCs). Then, a hierarchical classification framework consisting of two levels of nearest neighbour classifiers is applied to the MSCs and the match is found. The work in [3] concentrates on the hierarchical partitioning of the feature spaces. A space-tessellation tree is generated by employing principal component analysis (PCA) followed by LDA, at each level of the tree. By deriving a recursively better-fitted set of features for each of the recursively subdivided sets of training samples, the limitation that stems from using global linear features is avoided. In general, the use of hierarchical trees has been extensive for pattern recognition purposes. In [4], an owner-specific LDA-subspace is developed in order to create a personalized face verification system. The training set is partitioned into a number of clusters and only a single cluster, which contains face data that is most similar to the owner face, is retained. The system assigns the owner training images to this particular cluster and this new data set is used to determine an LDA subspace that is used to compute the verification thresholds and matching score when a test face claims the identify of the owner. Rather than using the LDA space created by processing the full training set, the authors show that verification performance is enhanced when an owner-specific subspace is utilized.

This paper presents a novel framework which employs person-specific adaptive discriminant clustering (ADC) with clustering and discriminant analysis parameters that are heavily affected by the characteristics of the test face. This methodology is not restricted to face recognition, but is able to deal with any problem that fits into the same formalism. At this point, it is imperative that two terms that are frequently used in this paper are defined: ‘class’ refers to a set of face images from the same person, whereas ‘cluster’ refers to a set of classes. The i^{th} face

class is denoted by \mathcal{Y}_i whereas the i^{th} cluster by C_i . It should be mentioned that a case exists such that the face images of a person may be partitioned into more than one cluster; thus, each of these clusters will contain a class from that particular person.

Initially, the training and test face vectors are projected onto a LDA-space by employing Fisher's criterion [5], thus producing the most discriminant features (MDF). Subsequently, k-means is used to partition the training data into a set of K discriminant clusters C_i and the distance of the test face from the cluster centroids is used to collect a subset of K' clusters that are closest to the test face. The training data that reside in these K' clusters are merged and a new MDF-space of the merged face classes is found by applying LDA and k-means is once again used to partition the data into a set of discriminant clusters. This process is repeated in as many iterations as are necessary until a single cluster is selected. Then, discriminant analysis is performed on this cluster, by using the data that reside in this cluster to produce the MDF-space, and the face class that is most similar to the test face is set as its identity match.

2. THE ADAPTIVE DISCRIMINANT CLUSTERING ALGORITHM

The ADC algorithm is an iterative process that at each iteration uses an adaptive MDF space that is closely related to the characteristics of the test face. More specifically, the set of clusters to be included in the training process that will define the future MDF space are selected based on how close they are to the test face in the current MDF space. Let us assume that an image \mathbf{X} of a test face is to be assigned to one of the Y distinct classes \mathcal{Y}_i that lie in the training set space \mathcal{T} . In addition, assume that each i^{th} class in \mathcal{T} is represented by $N_{\mathcal{Y}_i}$ images and the total number of training images is $N_{\mathcal{Y}}$. Thus, the face images that comprise the training set \mathcal{T} can be represented by $\{\mathbf{Y}_n\}_{n=1}^{N_{\mathcal{Y}}}$.

2.1. Linear Discriminant Analysis

We would like to linearly transform the face vectors such that they become separable, by projecting them to the MDF space. Let S_w and S_B be within-class and between-class scatter matrices [6] of the training set \mathbf{Y} . The most known and plausible criterion is to find a projection that maximizes the ratio of the between-class scatter against the within-class scatter (Fisher's criterion):

$$J(\mathbf{W}) = \frac{\mathbf{W}^T S_B \mathbf{W}}{\mathbf{W}^T S_w \mathbf{W}}. \quad (1)$$

Therefore, LDA is applied on \mathbf{Y} and the discriminant matrix \mathbf{W} of (1) is found. The training and test feature vectors are then projected to the MDF-space by

$$\mathbf{y}'_n = \mathbf{W}^T \mathbf{y}_n, \quad n = 1, \dots, N_{\mathcal{Y}} \quad (2)$$

and

$$\mathbf{x}' = \mathbf{W}^T \mathbf{x}. \quad (3)$$

where \mathbf{y}_n and \mathbf{x} are the training and test images in the form of vectors. Each training feature vector \mathbf{y}'_n is stored in a column of \mathbf{Y}' .

2.2. Clustering using k-means

The k-means algorithm is then employed in an effort to partition the training data into the Y distinct face classes. Given a set of N data vectors, realized by $\{\mathbf{y}_n\}_{n=1}^N$, in the d -dimensional space, k-means is used to determine a set of K vectors in \mathfrak{R}^d , called cluster centroids, so as to minimize the mean squared distance from each data vector to its nearest centroid. The objective function of k-means [7] that is used in this paper employs the squared Euclidean distance and is defined as

$$J = \sum_{i=1}^K \sum_{n=1}^N z_{in} \|\mathbf{y}_n - \boldsymbol{\mu}_i\|^2, \quad (4)$$

where z_{in} is defined as

$$z_{in} = \begin{cases} 0 & \text{if } \mathbf{y}_n \in C_i \\ 1 & \text{otherwise} \end{cases}, \quad (5)$$

where C_i and $\boldsymbol{\mu}_i$ denote the i^{th} cluster and centroid vector, respectively.

The k-means clustering method uses a two-phase iterative algorithm [8]. During each iteration of the first phase, vectors are reassigned to their nearest cluster centroid, all at once, followed by recalculation of cluster centroids. In the second phase, the vectors are individually reassigned if doing so will reduce the sum of distances, and cluster centroids are recomputed after each reassignment. Each iteration during this second phase consists of one pass through all the data vectors.

In order to incorporate k-means clustering, we set $\mathbf{y}_n = \mathbf{y}'_n$, $K = Y$ and $N = N_{\mathcal{Y}}$ in (4). The Y centroid vectors $\{\boldsymbol{\mu}_i\}_{i=1}^Y$ are found by

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{n=1}^N z_{in} \mathbf{y}'_n, \quad (6)$$

where

$$N_i = \sum_{n=1}^N z_{in}. \quad (7)$$

The distance between each training feature vector and the Y centroids is found by employing the Euclidean distance measure:

$$D_i^n(\mathbf{y}'_n, \boldsymbol{\mu}_i) = \|\mathbf{y}'_n - \boldsymbol{\mu}_i\|, \quad i = 1, \dots, Y. \quad (8)$$

The training vector \mathbf{y}'_n is then assigned to the cluster with the minimum vector-to-cluster-centroid distance, among the Y distances that are calculated.

2.3. Adaptive MDF-spaces

At this point we would like to redefine the original classification problem to a simpler one by discarding part of the training data and applying discriminant analysis on the new subset. The training feature vector data in these K' clusters, corresponding to the K' minimum vector-to-cluster-centroid distances, are collected. Let's assume that the Y' classes \mathcal{Y}'_i are contained in the subset that is selected and that each i^{th} class in \mathcal{T}' is represented by $N_{\mathcal{Y}'_i}$ images. Now, the total number of training feature vectors is $N_{\mathcal{Y}'}$ and these vectors are stored as columns in $\mathbf{Y}'_s \in \mathcal{T}'$. The value of K' , i.e. the number of clusters that are retained, is selected bearing in mind that after a few iterations one final cluster should be retained, and is calculated by the following simple function that converges to unity:

$$(K^j)_j = \text{round}\left(K^{\frac{1}{2^j}}\right), \quad j = 0, \dots, K_{\text{final}} \quad (9)$$

where j denotes the clustering step and K_{final} the clustering step for which K' becomes one for the first time.

In the new MDF-space that will be created using the face data from the K' clusters, LDA will attempt to discriminate the different classes found in each of the K' clusters. This enables the algorithm to formulate a clustering process that considers possible large variations in the set of images that each face class is represented by. For example, a portion of the set of images that corresponds to the i^{th} training person may present this person having facial hair, whereas others as not having facial hair. If these variations are larger than identity-related variations, then they are clustered into disjoint clusters. As a result, the match with the subset of the training images of class i whose appearance is most similar to the test face is considered, so the best match can be found.

3. EXPERIMENTAL RESULTS

In this section, the classification ability of ADC is investigated by observing FR experiments using data from the XM2VTS and UMIST databases [9]. Essentially, as in most FR applications, the classification experiments that are carried out fall under the small sample size (SSS) problem where the dimensionality of the samples is larger than the number of available training samples per subject [10]. The performance of ADC is presented for various degrees of how severe the SSS problem is. This is done by providing recognition rates for experiments where each face class \mathcal{Y}_i is represented by the smallest to the largest possible number of training samples, N_T . Since ADC employs discriminant analysis, the smallest possible value is 2. The largest possible value of training samples for each face class \mathcal{Y}_i is determined by the number of available images in this class, $N_{\mathcal{Y}_i}$, and by considering that at least one of these samples needs to be excluded in order to be able to test the recognition performance for that particular class. The remaining images that do not comprise the training set are used to test the performance of ADC, thus, they constitute the test set. The training and test sets are created by random selection on each set of the $N_{\mathcal{Y}_i}$ images of each face class. To give statistical significance to our experiments, this random selection is repeated N_R times and N_R recognition rates are accumulated and then averaged in order to calculate the final recognition rate R_{rec} .

The UMIST database consists of $K = 20$ different face classes, each of which is represented by at least $N_{\mathcal{Y}_i} = 19$ images. Consequently, 17 recognition rates were derived for training sets that contained $N_T = 2, \dots, 18$ images from each of the 20 face classes. Each corresponding rate was the average out of $N_R = 10$ repetitions. The XM2VTS database consists of $K = 200$ different face classes, each of which is represented by $N_{\mathcal{Y}_i} = 8$ images.

The number of clusters K' that are retained at each clustering level is selected by using (9). As a result, for the UMIST database experiments four clustering steps are required with each one forming 20, 4, 2 and 1 clusters respectively. For the XM2VTS experiments, five clustering steps are required with each one forming 200, 14, 4, 2 and 1 clusters, respectively. The face classes residing in the final cluster are projected to the MDF-space that is created by processing only this specific set of data. The face class that is closest to the test face in this

TABLE I
MEAN RECOGNITION RATES FOR VARIOUS NUMBERS OF TRAINING SAMPLES PER SUBJECT

UMIST				XM2VTS	
Training Samples N_T	R_{rec} (%)	Training Samples N_T	R_{rec} (%)	Training Samples N_T	R_{rec} (%)
2	58.5	11	95.7	2	32.1
3	80.3	12	95.9	3	91.4
4	89.1	13	96.4	4	94.2
5	90.3	14	96.7	5	96.5
6	91.5	15	97.3	6	96.9
7	93.4	16	97.6	7	98.1
8	93.8	17	98.0		
9	94.2	18	98.7		
10	94.9				

MDF-space is selected as the true match in identity. Table I reports the mean recognition rates, R_{rec} , obtained for FR experiments carried out on both face databases, for $N_R = 10$ independent runs.

4. CONCLUSION

A FR methodology that employs person-specific adaptive discriminant clustering is proposed and its performance is evaluated. The ADC algorithm adapts the coordinates of the MDF-space with respect to the characteristics of the test face and the training faces that are more similar to the test face. Thus, the FR problem is broken down to multiple easier classification tasks, in terms of achieving linear separability. This method was tested on the UMIST and XM2VTS face databases and results show that the proposed framework provides a promising solution to the face recognition problem.

5. REFERENCES

- [1] J. Lu and K.N. Plataniotis, "Boosting face recognition on a large-scale database", in *Proc. IEEE Int. Conf. on Image Processing*, Rochester, New York, USA, September 22-25, 2002.
- [2] G.D. Guo, H.J. Zhang, and S.Z. Li, "Pairwise face recognition", in *Proc. 8th IEEE Int. Conf. on Computer Vision*, vol. 2, pp. 282-287, Vancouver, Canada, July 2001.
- [3] D.L. Swets and J. Weng, "Hierarchical discriminant analysis for image retrieval", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 386-401, May 1999.
- [4] H.-C. Liu, C.-H. Su, Y.-H. Chiang; Y.-P. Hung, "Personalized face verification system using owner-specific cluster-dependent LDA-subspace", in *Proc. IEEE Int. Conf. on Pattern Recognition*, vol. 4, pp. 344-347, Aug. 23-26, 2004.
- [5] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using LDA based algorithms", *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 195-200, Jan. 2003.
- [6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York, NY: Academic Press, 1990.
- [7] Y. Lee, S. Choi, "Minimum entropy, k-means, spectral clustering", in *Proc. IEEE Int. Joint Conf. on Neural Networks*, vol. 1, pp. 117-122, Budapest, July 25-29, 2004.
- [8] G.A.F. Seber, *Multivariate Observations*, Wiley, New York, 1984.
- [9] M. Kyperountas, A. Tefas, and I. Pitas, "Weighted piecewise LDA for solving the small sample size problem in face verification", *IEEE Trans. on Neural Networks*, vol. 18, no. 2, pp. 506-519, March 2007.
- [10] M. Kyperountas, A. Tefas, and I. Pitas, "Face verification using locally linear discriminant models", in *Proc. IEEE Int. Conf. on Image Processing*, vol. 4, pp. 469-472, San Antonio, 16-19 Sep., 2007.