Integrating affinity propagation clustering method with linear discriminant analysis for face recognition

Chunhua Du,[a] Jie Yang,[a] Qiang Wu,[b] and Feng Li[c]
[a]Shanghai Jiao Tong University, Institute of Image Processing and Pattern Recognition, Shanghai, China
[b]University of Technology Sydney, Department of Computer Systems, Sydney, Australia
[c]University of Delaware, Department of Computer and Information Sciences, Newark, Delaware

Abstract. The Fisherface method suffers from the problem of using all training face images to recognize a test face image. To tackle this problem, we propose combining a novel clustering method, affinity propagation (AP), recently reported in the journal *Science*, with linear discriminant analysis (LDA) to form a new method, AP-LDA, for face recognition. By using AP, a representative face image for each subject can be obtained. Our AP-LDA method uses only these representative face images rather than all training images for recognition. Thus, it is more computationally efficient than Fisherface. Experimental results on several benchmark face databases also show that AP-LDA outperforms Fisherface in terms of recognition rate. © *2007 Society of Photo-Optical Instrumentation Engineers.*
[DOI: 10.1117/1.2801735]

1 Introduction

Face recognition (FR) has recently received extensive attention due to its broad applications. Many techniques have been developed over the past decades for FR. Subspace analysis is one of the most efficient techniques. The most popular subspace techniques are principal component analysis (PCA),[1] linear discriminant analysis (LDA),[2] and independent component analysis (ICA).[3] Besides directly processing image appearance, subspace methods can also be combined with the Gabor feature[4] to derive the Gabor-based methods.[5–7] However, the recognition time of these methods depends heavily on the size of the training set because a test image is compared to all training images. Such a recognition scheme is not efficient, especially when the training set is too large. Thus, we propose integrating a novel clustering method, affinity propagation (AP),[8] with LDA to form AP-LDA for FR. By using AP on the low-dimensional features obtained from LDA, a representative face image for each subject can be achieved. Thus, AP-LDA needs to use only representative face images rather than all training face images for recognition.

We combine AP with LDA for the following reasons: AP can cluster data points into different clusters and detect a representative example for each cluster. By using AP, we intend to achieve a representative face image for each subject. If so, we need to use only representative face images for recognition. However, directly using AP on gray pixel values is computationally expensive and inefficient. For example, the difference between two pixels on the same positions in two images of the same subject is obvious when the illumination is different. However, they should be considered to be similar to each other in fact. Thus, more discriminating and efficient low-dimensional features should be extracted before using AP for clustering. Therefore, LDA is adopted, as it can not only be used for dimensionality reduction but also extract discriminative features.

2 Review of the AP Clustering Method and the LDA Method

2.1 AP Clustering Method

AP[8] first builds a similarity matrix s, in which $s(i,k)$ between data points x_i and x_k is their negative Euclidean distance $[s(i,k)=-\|x_i-x_k\|^2]$. Before clustering, each data point also needs to be assigned a number $P(B)$, which describes the *a priori* knowledge of how good point B is as a representative example. The data points with larger values of $P(B)$ are more likely to be chosen as representative examples. These values are referred to as *preferences*. In fact, the probability of each point being the representative example is the same; thus, the preferences should be set to the same value, which can be varied to produce different numbers of clusters. Generally, such a value takes the median of the s. After the construction of a similarity matrix and the setting of preferences, two kinds of messages (responsibility and availability) are passed between data points. The responsibility $r(i,k)$, sent from x_i to candidate representative example x_k, reflects the accumulated evidence for how proper it would be for x_k to serve as the representative example for x_i. It is updated using the rule:

$$r(r,k) \leftarrow s(i,k) - \max_{k' \, s.t. k' \neq k}[a(i,k') + s(i,k')]. \quad (1)$$

The availability $a(i,k)$, sent from candidate representative example x_k to x_i, reflects the accumulated evidence for how well-suited x_i is to choose x_k as its representative example. It is computed by the rule:

$$a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' \, s.t. i' \notin \{i,k\}} [0, r(i',k)]\right\}. \quad (2)$$

It is clear that availabilities and responsibilities can be combined to recognize representative examples at any time. For x_i, the k that maximizes $a(i,k)+(i,k)$ indicates that x_k serves as the representative example for x_i.

2.2 LDA Method

Suppose there is a set of N d-dimensional samples $\{x_1, x_2, \ldots, x_N\}$ belonging to c classes $\{X_1, X_2, \ldots, X_N\}$. LDA[2] aims to find W_{opt} that maximizes the ratio of the between-class scatter matrix to the within-class scatter matrix, i.e.,

Table 1 Comparative results on Yale database.

FR method	Number of training face images per person					
	3	4	5	6	7	8
Fisherface	60.17%	69.33%	72.00%	77.33%	79.33%	84.89%
AP-LDA	55.33%	69.90%	76.89%	78.13%	83.33%	87.56%

$$W_{opt} = \arg \max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} = (w_1, w_2, \ldots, w_m), \qquad (3)$$

where S_B and S_W are the between-class matrix and the within-class scatter matrix:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T, \qquad (4)$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T, \qquad (5)$$

where μ_i is the mean of class X_i, μ is the total sample mean, and N_i is the size of the data points in X_i. $(w_1, w_2, \ldots, w_m)$ are generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $(\lambda_i | i = 1, 2, \ldots, m)$, i.e.,

$$S_B w_i = \lambda_i S_W w_i \quad i = 1, 2, \ldots, m. \qquad (6)$$

Note that there are at most $c-1$ nonzero generalized eigenvalues; thus, the dimension of the reduced space is $c-1$. At the same time, to make the within-class scatter matrix S_W nonsingular, PCA is first used to reduce the dimension of the feature to $N-c$, and then LDA is applied to reduce the dimension to $c-1$.

3 Summary of AP-LDA

Suppose that there is a face data set of N face images belonging to c different subjects. First, n face images per person (hence, $n \times c$ in total) are selected for training using LDA, obtaining $n \times c$ corresponding $c-1$ dimensional features. Second, AP is used to cluster these features into different clusters and obtain a representative feature for each cluster. Note that the number of clusters obtained by using AP may not equal the number of subjects, as the former is influenced by the values of preference,[8] while the latter is

fixed. To solve this problem, we repeatedly vary the value of the preferences until two such numbers equal to each other. Last, each test image is converted to a low-dimensional feature, which is then compared to c representative features and identified using a nearest-neighbor classifier. Since AP-LDA uses only c representative features for recognition, the recognition time of AP-LDA depends mainly on the number of subjects. It increases linearly with the c, while the recognition time of the Fisherface method increases with the training image size ($n \times c$). Obviously, our AP-LDA method is more computationally efficient than the Fisherface method.

4 Experiments and Discussion

In this section, experiments on three benchmark face data sets [Yale[9]; extended Yale[10]; and Pose, Illumination, and Expression (PIE)[11]] are carried out to show the effectiveness of our AP-LDA method and also to compare it to Fisherface. All face images are cropped based on the centers of eyes such that facial areas contain only the face. All cropped images are then normalized to the size of 32×32, with 256 gray levels per pixel. In the following experiments, different numbers of images per subject were randomly selected for training, and the rest were used for testing. To minimize the possible misleading results, the final recognition rates were obtained by averaging the results over five random splits.

The Yale database contains 165 images with 11 different images for each of the 15 distinct subjects. Comparative results are summarized in Table 1. It is clear that AP-LDA obviously outperforms Fisherface when the number of training images per person is more than 3.

The Extended Yale database is extended from the Yale database. It contains approximately 64 near-frontal images for each of 38 distinct subjects. We randomly selected 20 images per person for our experiments. Table 2 presents the results. We can see that AP-LDA remarkably outperforms Fisherface in all cases.

The CMU Pose, Illumination, and Expression (PIE) database contains 41,368 facial images of 68 individuals. We randomly chose 30 images per person for our experiments. The results are shown in Table 3. Again, AP-LDA consistently outperforms Fisherface.

Comparative experimental results show that:

1. The recognition rates of AP-LDA increase with the increase in the number of training face images per person because the more training images per person, the better the representative face images that can be detected.

Table 2 Comparative results on subset from Extended Yale database.

FR method	Number of training face images per person						
	4	5	6	7	8	9	10
Fisherface	59.61%	63.30%	70.08%	71.01%	74.08%	77.18%	80.32%
AP-LDA	65.03%	70.49%	74.40%	76.88%	80.53%	81.24%	84.21%

Table 3 Comparative results on subset from CMU PIE database.

FR method	Number of training face images per person						
	4	5	6	7	8	9	10
Fisherface	50.84%	58.09%	63.39%	64.59%	68.69%	68.43%	70.66%
AP-LDA	59.24%	65.96%	70.92%	71.71%	73.68%	72.61%	74.15%

2. Although AP-LDA performs better than Fisherface on the Yale face data sets in most cases, the improvement is not significant. Moreover, it performs not as well as Fisherface when the number of training images per person is 3. This is because the AP method could not find good representative face images in these cases.
3. AP-LDA remarkably outperforms Fisherface on the Extended Yale and CMU PIE databases. In fact, for these two databases, we chose only a fraction of images for our experiments. We found that our AP-LDA method performed significantly better than Fisherface when all images were used. However, such improved performance was achieved at the expense of increase of computational cost.

5 Conclusions

We have introduced in this letter a novel AP-LDA method for face recognition. Unlike Fisherface, which uses all training face images for recognition, our AP-LDA method uses only representative face images. This makes AP-LDA more efficient. Experiments also indicate that AP-LDA outperforms Fisherface in terms of recognition rate.

References

1. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, Proc. CVPR 91*, IEEE Computer Society Conference, pp. 586–591 (1991).
2. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
3. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.* **13**(6), 1450–1464 (2002).
4. L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997).
5. C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(5), 572–581 (2004).
6. C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. Neural Netw.* **14**(4), 919–928 (2003).
7. C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.* **11**(4), 467–476 (2002).
8. B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science* **315**(5814), 972–976 (2007).
9. P. N. Belhumeur and D. J. Kriegman, "The Yale face database," http://cvc.yale.edu/projects/yalefaces/yalefaces.html(1997).
10. K. C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 684–698 (2005).
11. T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. Fifth IEEE Int. Conf. in Automatic Face and Gesture Recognition*, pp. 46–51 (2002).