MULTI-DISCRIMINANT CLASSIFICATION ALGORITHM FOR FACE VERIFICATION

Cheng-Ho Huang and Jhing-Fa Wang
Dept. of Electrical Engineering, National Cheng Kung University, No. 1 University Rd., Tainan City, Taiwan

Keywords: Linear discriminant analysis, face verification, multi-discriminant classification.

Abstract: Linear discriminant analysis (LDA) is a conventional approach for face verification. For computing large amounts of data collected for a given face verification system, this study proposes a multi-discriminant classification algorithm to classify and verify voluminous facial images. In the training phase, the algorithm extracts all discriminant features of the training data, and classifies them as the clients' multi-discriminant sets. The algorithm verifies a claim to the client's multi-discriminant set, and then determines whether the claimant is the client. Comparative results demonstrate that the proposed algorithm reduces the false acceptance rate in face verification.

1 INTRODUCTION

Two primary applications of face recognition are face identification and face verification. Face identification identifies two similar faces between unknown user and genuine users; face verification compares an unknown user to a genuine user, and decides whether the two are the same. Therefore, impostors present a problem in face verification. In particular, impostors are greater in number than clients. Eigenface (Turk and Pentland, 1991) and Fisherface (Belhumeur et al., 1997) are two of the best known methods that adopt feature transformation in order to discriminate differences in facial features for the purpose of face verification. However, the performance of Eigenface method is not ideal when numbers of the sample sets are voluminous. Fisherface, an implementation of linear discriminant analysis (LDA) (Martinez and Kak, 2001), is often utilized for face verification. It employs both the PCA and Fisher criterion to extract discriminant information from a set of training data. Many methods (Liu and Wechsler, 1998; Loog et al., 2001; Wang and Tang, 2004) have been proposed to enhance the performance and stability of LDA. Both classical and modified LDA methods are efficient for face recognition.

Although improved LDA approaches are superior to classical LDA approaches, they still do not provide adequate discriminant information to permit accurate discrimination of the highly complex and voluminous data of facial images. Main reason for this limitation is given below.

The voluminous data of facial images are not true Gaussian distributions. Consequently, the classical linear transform of the "between-class" and the "within-class" cannot effectively extract the differential features from the classes.

Therefore, classical LDA is not appropriate for direct analysis of complex and numerous data. As the amount of data increases, computational loading of LDA also increases, and the time required for calculation grows longer, making the method less practical. To reduce the computations of numerous data, k-nearest neighbor (KNN) and k-means algorithm are adopted to classify data into small units. However, KNN is sensitive to feature mapping; if the feature mapping is not well distribution, KNN does not obtain robust classifications. K-means , which is an unsupervised classification algorithm, has problems with initial centroids and specifying the number of clusters. Otherwise, if the selected threshold value of the algorithm is unsuitable, then the false acceptance rate (FAR) and false rejection rate (FRR) increase; in particular, the algorithm cannot effectively tune the threshold parameters for FAR and FRR.

Due to these above-mentioned problems, in order to avoid the resulting decrease in efficiency of the overall performance caused by the large amounts of complex data, this study proposes a verification algorithm without setting any threshold value to separate complex data into simple units and verify face images. This algorithm splits all of the training data,

Huang C. and Wang J. (2008).
MULTI-DISCRIMINANT CLASSIFICATION ALGORITHM FOR FACE VERIFICATION.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 299-304
DOI: 10.5220/0001082202990304

enabling each individual's features to be distinguished and yield subsets of distinguishable features for each person. Combining the results obtained by separately discriminating these subsets is synonymous with verifying whether an unknown user is the genuine user. Thus, as evidenced from volumes of face verification, this study has achieved good efficiency to avoid impostors, and increased the overall robustness of the method.

The paper is organized as follows. Section 2 presents the multi-discriminant classification algorithm on volumes of face verification. Experiments and final conclusions are provided in Sections 3 and 4, respectively.

2 THE PROPOSED MULTI-DISCRIMINANT CLASSIFICATION ALGORITHM FOR FACE VERIFICATION

The proposed multi-discriminant classification algorithm (MDCA) consists of two modules, multi-discriminant classifier and evaluator. Figure 1 shows the entire framework. Each module is discussed below.

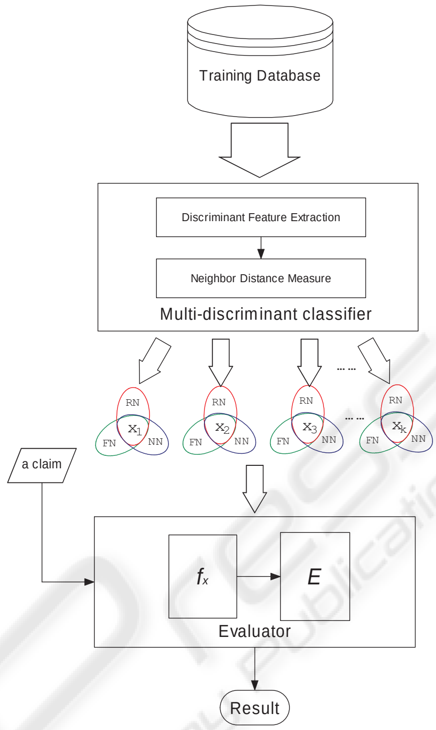

Figure 1: Framework of the proposed algorithm.

2.1 Multi-discriminant Classifier

The proposed approach using generalized singular value decomposition LDA (GSVD/LDA) (Howland and Park, 2004) constructs multi-discriminant sets (MDS) and performs discriminant analysis to verify a claimant in the client's MDS.

Suppose that m-dimensional patterns $\mathbf{A} = \{x_i\}_{i=1,\ldots,n}$ belong to c different classes $\{C_i\}_{i=1,\ldots,c}$. $\mathbf{A} \in \Re^{m\times n}$. Let n_k denote the number of patterns in class k; thus, $\sum_{k=1}^{c} n_k = n$.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

$$\mu_k = \frac{1}{n_k}\sum_{x\in C_k} x_k, \tag{2}$$

where μ denotes the average of ensemble facial features and μ_k denotes the mean of class C_k. The between-class scatter matrix $\mathbf{S}_B$ is defined as

$$\mathbf{S}_B = \sum_{k=1}^{c} n_k(\mu_k-\mu)(\mu_k-\mu)^T. \tag{3}$$

The within-class scatter matrix $\mathbf{S}_W$ is defined as

$$\mathbf{S}_W = \sum_{k=1}^{c}\sum_{j\in C_k}(x_j-\mu_k)(x_j-\mu_k)^T, \tag{4}$$

and

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W. \tag{5}$$

The transformation matrix $\mathbf{G}^T \in \Re^{l\times m}$ reduces vector x_i of $\mathbf{A}$ to vector y_i in the $l-$dimensional space:

$$y_i = \mathbf{G}^T\mathbf{A} \in \Re^{l\times n}, l \ll m. \tag{6}$$

The maximum ratio of the between-class to within-class scatter is obtained by the determinant of the objective function of the scatter matrices and is defined as

$$J(\mathbf{G}) = trace((\mathbf{G}^T\mathbf{S}_T\mathbf{G})^{\dagger}(\mathbf{G}^T\mathbf{S}_B\mathbf{G})). \tag{7}$$

When $\mathbf{T}$ is full rank,

case 1: $l = n$
$\mathbf{T}^{\dagger} = (\mathbf{T})^{-1}$,
case 2: $l < n$

$\mathbf{T}^{\dagger} = \mathbf{T}^T(\mathbf{T}\mathbf{T}^T)^{-1}$,

case 3: $l > n$

$\mathbf{T}^{\dagger} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$,

where $\mathbf{T} = \mathbf{G}^T(\mathbf{S}_T)\mathbf{G}$ and $\mathbf{T}^{\dagger}$ is the Moore-Penrose pseudo-inverse can be obtained by GSVD.

The columns of an optimal G comprise the generalized eigenvectors corresponding to the l largest eigenvalues in

$$\mathbf{S}_B^y \mathbf{g}_i = \lambda_i(\mathbf{S}_T^y), i = 1, 2, ..., l \tag{8}$$

where $\mathbf{S}_B^y$ and $\mathbf{S}_W^y$ are chosen from $\mathbf{S}_B$ and $\mathbf{S}_W$, respectively; and $\mathbf{g}_i$ is the set of generalized eigenvectors of $\mathbf{S}_B^y$ and $\mathbf{S}_W^y$ corresponding to the l largest generalized eigenvalues λ.

The distance measure is derived from the differences in features between average face (μ) and everyone's ($\acute{\mu}$) , and then classifing facial images into the clients' MDS. In this case, $\acute{\mu} = (\mu + \frac{1}{n_k}\sum_{i=1}^{n_k} x_i)/2$. $\mathbf{G}^T$ is obtained from μ and $\acute{\mu}$ in Eq. (7), and the discriminant feature D is then defined as follows:

if $\mathbf{G}^T\acute{\mu} < \mathbf{G}^T\mu$

$$D = -(\left\|\mathbf{G}^T\acute{\mu}\right\| + \left\|\mathbf{G}^T\mu\right\|) \tag{9}$$

if $\mathbf{G}^T\acute{\mu} > \mathbf{G}^T\mu$

$$D = \left\|\mathbf{G}^T\acute{\mu}\right\| + \left\|\mathbf{G}^T\mu\right\| \tag{10}$$

Algorithm 1 illustrates the pseudocode for extracting discriminant features. The clients' MDS are constructed using these differences after extracting the discriminant features of all faces.

Algorithm 1 The pseudocode of discriminant feature extraction.

for i=1 to all **do**
  Calculate $\mathbf{G}_i^T$ by Eq. (7).
  if $\mathbf{G}_i^T\acute{\mu} < \mathbf{G}_i^T\mu$ **then**
    $D_i \leftarrow -(\left\|\mathbf{G}_i^T\acute{\mu}_i\right\| + \left\|\mathbf{G}_i^T\mu\right\|)$,
  else
    $D_i \leftarrow \left\|\mathbf{G}_i^T\acute{\mu}_i\right\| + \left\|\mathbf{G}_i^T\mu\right\|$.
  end if
end for

Consider a certain personal set $\mathbf{P}_{client}^S$ with x members. There exist three special subsets NN= $\{x_i | i = P_{client}...P_{ns}\}$, FN= $\{x_j | j = P_{client}...P_{fs}\}$ and RN= $\{x_m | m = P_{client}...P_{rs}\}$, $S = \{NN, FN, RN\}$; where NN denotes a nearest neighbor subset; FN denotes a farthest neighbor subset and RN denotes a random neighbor subset; ns and fs selections of people are similar and non-similar to p_{client}, respectively, and rs denoted the random selections of people to P_{client}. Algorithm 2 illustrates the pseudocode of multi-discriminant classifier. A subset of t members is chosen as a subset, where $10 \leq t \leq 20$.

Algorithm 2 The pseudocode of multi-discriminant classifier.

for $i = 1$ to all **do**
  if $D_i \sim D_{P_{client}}$ and $ns \leq t$ **then**
    Select x_i into an $NN_{P^S_{client}}$.
    $ns = ns + 1$.
  end if
  if $D_i \nsim D_{P_{client}}$ and $fs \leq t$ **then**
    Select x_i into an $FN_{P^S_{client}}$.
    $fs = fs + 1$.
  end if
end for
for $rs = 1$ to t **do**
  Randomly select x into a $RN_{P^S_{client}}$.
end for

2.2 The Evaluator of Face Verification

The evaluator determines whether a claim is the client by an evaluation function on the results of discriminations from NN, FN and RN.

Equation (11) is a similar description described by the following expression:

$$f_x = \begin{cases} 1, & \begin{aligned} &if\ dist(x_{claim}, x_{P_{client}}) \\ &\quad = \min(dist(x_{claim}, x)) \end{aligned} \\ 0, & \begin{aligned} &if\ dist(x_{claim}, x_{P_{client}}) \\ &\quad > \min(dist(x_{claim}, x)) \end{aligned} \end{cases} \tag{11}$$

where $dist$ is the distance measure function. If x_{claim} is similar to $x_{P_{client}}$, then $f_x = 1$ or 0.

Equation (12) is an evaluation function of MDCA, and is defined as follows:

$$\begin{aligned} E(x_{claim}, x_{P_{client}}) = &(f^{NN} \bullet f^{FN} + f^{RN}) \\ &+ (f^{NN} \bullet f^{RN} + f^{FN}) \\ &+ (f^{FN} \bullet f^{RN} + f^{NN}), \end{aligned} \tag{12}$$

where E denotes an evaluator; and $\bullet$ and $+$ are AND and OR Boolean operators, respectively. If x_{claim} is similar to $x_{P_{client}}$ for two out of the three discriminated results of subsets NN, FN, and RN, then x_{claim} indicates the genuine user $x_{P_{client}}$. If E is equal to 1, the result is an acceptance, or a rejection.

Thus, the face verification problem can be depicted by a multi-identification problem. The evaluation algorithm is illustrated in Alg. 3.

Algorithm 3 The pseudocode of the evaluator.

```
Calculate dist(x_claim, x).
if dist(x_claim, x_Pclient) = min(dist(x_claim, x)) then
    f_{x_Pclient} ← 1,
else
    f_x ← 0.
end if
if E(x_claiim, x_Pclient) = 1 then
    Accept,
else
    Reject.
end if
```

For instance, the statuses of MDS which owns ten members are described in the Table 1, Table 2 and Table 3,respectively. Eq. (12) is used to evaluate x_{claim} and $x_{P_{client}}$, and then obtains $E = 1$. Therefore, the result of verification is an acceptance.

Table 1: Select top ten nearest neighbors of D into an NN.

Member	D	$dist()$	f
$P2$	282	251	0
$P4$	247	342	0
$P6$	217	221	0
$P7$	371	175	0
$P10$	391	232	0
$P11$	182	172	0
$P12$	389	119	0
$P15$	193	120	0
$P20$	387	149	0
P_{client}	120	76	1

Table 2: Select top ten farthest neighbors of D into an FN.

Member	D	$dist()$	f
$P1$	891	240	0
$P9$	777	310	0
$P13$	909	130	0
$P14$	1111	171	0
$P17$	877	234	0
$P18$	976	140	0
$P21$	701	127	0
$P24$	761	134	0
$P25$	865	182	0
P_{client}	120	91	1

Table 3: Select ten random neighbors of D into a RN.

Member	D	$dist()$	f
$P3$	617	123	0
$P5$	489	141	0
$P8$	412	231	0
$P13$	909	211	0
$P15$	193	435	0
$P16$	435	156	0
$P19$	430	176	0
$P22$	600	183	0
$P23$	533	145	0
P_{client}	120	83	1

3 EXPERIMENTS

The experiments were carried out on the FERET (Rizvi et al., 1998), XM2VTS (Messer et al., 1999) and UNDBD-B (Bowyer and Flynn, 2003) face databases. FERET is a well-known face database provided by the NIST. The FERET database contains 994 people and over 11,000 face images, including profiles, frontal faces, expressions, and poses. The XM2VTSDB contains 2560 frontal images, which are four recordings of 295 people taken over a period of four months. The UNDBD-B database contains 33,247 visible frontal images of 749 people. This study adopted only the frontal face images as training faces, and adopted the other types and frontal images together as test data.

Therefore, the XM2VTS and UNDBD-B were adopted as the training and testing databases in Experiments (1) and (2), respectively. In Experiment (3), the FERET database was considered as outside data to test proposed algorithm. Two evaluations were adopted to evaluate the system performance:

- False acceptance rate (FAR): the ratio of the number of false acceptances to that of impostor accesses.

- False rejection rate (FRR): the ratio of the number of false rejections to that of authentic accesses.

The experimental results of the proposed face verification using the MDCA are presented below. In this evaluation, the sizes of the multi-discriminant sets were 10 and 20. The MDCA was adopted to verify these cases in the multi-discriminant sets. The results in Table 4 indicate that as the FAR and FRR of NN, FN and RN are decreased as the size of an multi-discriminant set increases from 10 to 20. In Experiment (1), the optimum value of FAR was 0.34%, while that of FRR was 4.04%, while the results of

Table 4: Comparison results of FAR and FRR between MDCA and LDA with KNN.

Experiment	E	Evaluation	**MDCA** size 10	**MDCA** size 20	LDA with KNN (K=10)	
		FAR	6.25%	5.86%		
	only NN	FRR	4.12%	4.03%		
		FAR	7.56%	6.91%	FAR	12.5%
XM2VTS	only FN	FRR	4.78%	4.60%		
		FAR	6.70%	6.30%		
(1)	only RN	FRR	4.71%	4.31%		
		FAR	0.73%	0.34%	FRR	8.7%
	NN, FN, RN	FRR	4.44%	4.04%		
		FAR	6.36%	5.91%		
	only NN	FRR	4.33%	4.00%		
		FAR	7.31%	6.89%	FAR	14.7%
UNDBD-B	only FN	FRR	4.96%	4.36%		
		FAR	7.23%	6.20%		
(2)	only RN	FRR	5.51%	4.24%		
		FAR	0.69%	0.23%	FRR	11.6%
	NN, FN, RN	FRR	4.91%	4.11%		
		FAR	6.41%	5.94%		
	only NN	FRR	4.43%	4.08%		
XM2VTS		FAR	8.32%	7.38%	FAR	16.8%
+UNDBD-B	only FN	FRR	5.11%	4.21%		
+FERET$_{impostors}$		FRR	6.64%	6.26%		
	only RN	FRR	4.88%	4.18%		
(3)		FAR	0.72%	0.31%	FRR	13.1%
	NN, FN, RN	FRR	5.08%	4.18%		

the NN, FN and RN intersected together. The FAR and FRR were 0.23% and 4.11%, respectively in Experiment (2), and 0.31% and 4.18%, respectively in Experiment (3). Regardless of the results of the NN, FN and RN, their intersection demonstrated the best performance in each experiment. The proposed performed better overall than LDA with KNN (Lin et al., 2005).

4 CONCLUSIONS

This study proposes an algorithm to enhance the face verification performance in numerous databases by using multi-discriminant classification. Experimental results indicate that proposed algorithm elevates the performance of face verification. Moreover, the proposed method does not require the construction of any miscellaneous thresholding rule and can actively solve the verified problem of face verification. The experimental results reveal that FAR can be decreased from 8.32% to 0.31% when utilizing evaluation function E with three discriminant subsets.

REFERENCES

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711–720.

Bowyer, K. and Flynn, P. (2003). University of notre dame biometrics database-b. http://www.nd.edu/ cvrl/UNDBiometricsDatabase.html.

Howland, P. and Park, H. (2004). Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(8):995–1006.

Lin, D., Yan, S., and Tang, X. (2005). Feedback-based dynamic generalized lda for face recognition. *Int. Conf. on Image Processing*, 2:922–925.

Liu, C. and Wechsler, H. (1998). Enhanced fisher linear discriminant models for face recognition. *Proc. of the 14th Int. Conf. on Pattern Recognition*, 2:1368.

Loog, M., Duin, R. P. W., and Haeb-Umbach, R. (2001). Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(7):762–766.

Martinez, A. M. and Kak, A. C. (2001). Pca versus lda. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):228–233.

Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. (1999). XM2VTSDB: The Extended M2VTS Database. *Proc. 2nd International Conference on Audio- and Video-based Biometric Person Authentication*.

Rizvi, S. A., Phillips, P. J., and Moon, H. (1998). The feret verification testing protocol for face recognition algorithms. *Proc. of the 3rd. Int. Conf. on Face & Gesture Recognition*, page 48.

Turk, M. A. and Pentland, A. P. (1991). Face recognition using eigenfaces. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 586–591.

Wang, X. and Tang, X. (2004). A unified framework for subspace face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(9):1222–1228.