

# Non-recursive method for Deriving uncorrelated Discriminants

Mianshu Chen

Communication Engineering College  
Nanling Campus, Jilin University  
Changchun, Jilin, 130025, China  
chenmianshu@sohu.com

Qiwei Zhang

Communication Research Group  
School of Electronics and Computer Science  
University of Southampton  
Highfield, Southampton  
SO17 1BJ, UK  
qz403@ecs.soton.ac.uk

Hexin Chen

Communication Engineering College  
Nanling Campus, Jilin University  
Changchun, Jilin, 130025, China  
chx@jlu.edu.cn

Wei Liu

Communication Engineering College  
Nanling Campus, Jilin University  
Changchun, Jilin, 130025, China  
davidiee@yahoo.com.cn

## Abstract

*Recently recursive algorithm has been used in getting Fisher criterion based linear discriminants through which the features obtained are statistically uncorrelated. This paper investigates a simpler and faster approach to compute uncorrelated discriminants. This is done by first mapping the input data onto an uncorrelated feature space and computing Fisher criterion based discriminants in that space. By avoiding recursive procedure, our algorithm is  $(C - 1)/2$  times faster than recursive algorithm when the number of classes  $C$  is large. Theoretical formulation and real world data confirmed the effectiveness of our approach.*

*Index Terms—Uncorrelated discriminants, Pattern Recognition, Feature extraction, Uncorrelated space*

## 1. Introduction

Recent years have shown a strong trend in the computer vision and pattern recognition community away from geometry towards statistical and appearance based models [1, 2, 3]. Among the widely used approaches, Fisher linear discriminant analysis can be interpreted as a benchmark both for dimensionality reduction and classification method [4, 5]. To generalize Fisher LDA from two classes to multiclass application, Liu [6] introduced the recursive algorithm that can obtain optimal linear discriminants in multiclass observations. Although Liu's discriminants are

optimal in the sense of Fisher criterion at each recursion, the features drawn from these discriminants are statistically correlated. While in feature extraction field, we always hope to extract uncorrelated features to characterize different patterns accurately. To deal with this problem, Jin [7] presented an improved algorithm that can get uncorrelated discriminants. This is done by adding an un-correlation constraint at each recursion. Jin's feature extraction algorithm achieved high scores in pattern classification experiment. However, the problem remains that the time cost of this approach is prohibitive considering it still computes discriminant vectors in a recursive way.

Along this line, our paper aims at obtaining optimal uncorrelated discriminants in a simpler and faster way. The basic idea behind our approach is: First we construct a convenient uncorrelated feature space and map the input data into that space. Then in the uncorrelated space we conduct the second mapping based on Fisher criterion. The combination of two mapping amounts to getting optimal discriminants in original input space through which we get uncorrelated features.

In next section, we first give the formulation of our Uncorrelated Discriminants on Feature Space (UDFS), then investigate its relationship with Jin [7]'s uncorrelated Discriminants in Recursive Algorithm (UDRA). Section 3 is devoted to the experiments confirming the effects of the new method, and the conclusions are given in the last section.

## 2. Discriminants on uncorrelated feature space

We first fix some notations used in the following formulation. Let  $\omega_1, \omega_2, \dots, \omega_C$  be  $C$  known patterns, and  $\mathbf{x}$  be a sample in  $n$  dimensional space.  $\mathbf{m}_i$  and  $P_i (i = 1, 2, \dots, C)$  are the mean and prior probability of  $\omega_i$  respectively. Then the between-class scatter matrix  $\mathbf{S}_b$ , within-class scatter matrix  $\mathbf{S}_w$ , and total population scatter matrix  $\mathbf{S}_t$  can be defined as:

$$\mathbf{S}_b = \sum_{i=1}^C P_i [\mathbf{m}_i - E(\mathbf{x})][\mathbf{m}_i - E(\mathbf{x})]^\top \quad (1)$$

$$\mathbf{S}_w = \sum_{i=1}^C P_i E \left[ (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^\top | \omega_i \right] \quad (2)$$

$$\begin{aligned} \mathbf{S}_t &= E \left\{ [\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^\top \right\} \\ &= \mathbf{S}_b + \mathbf{S}_w \end{aligned} \quad (3)$$

The optimal uncorrelated discriminants are such vectors: All the features obtained by mapping input data on these discriminants vectors are uncorrelated. Meanwhile, on every discriminants the ratio of the between-class distance to the within-class distance is maximum.

**Definition 1.** For any training sample  $\mathbf{x}$ , assume  $\Phi = [\phi_1, \phi_2, \dots, \phi_k]$  to be discriminant vectors. Then the linear transformation of  $R^n \rightarrow R^k$  is defined as:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_k^\top \end{bmatrix} \mathbf{x} \quad (4)$$

If  $E[(y_i - Ey_i)(y_j - Ey_j)] = \phi_j^\top \mathbf{S}_t \phi_i = 0 (j \neq i)$ , and  $\Phi$  maximize the ratio of between-class distance to the within-class distance, then  $\Phi = [\phi_1, \dots, \phi_k]$  are uncorrelated discriminants.

### 2.1. Uncorrelated feature space

**Theorem 1.** Any orthogonal transformation of an identity matrix leads to an identity matrix.

This conclusion is obvious. From theorem 1 we know that in first mapping if we transform  $\mathbf{S}_t$  into an identity matrix with a column orthogonal matrix  $\mathbf{V}$ , we can get an uncorrelated feature space  $\mathcal{F}$ , in which the total population scatter matrix becomes identity matrix. Then we can always get  $\phi_j^\top \mathbf{S}_t \phi_i = 0 (j \neq i)$  and  $\phi_j^\top \mathbf{S}_t \phi_i = 1 (j = i)$  as long as the projection directions  $\phi_i$  and  $\phi_j$  used for the second mapping are orthogonal. According to definition 1,  $\phi_i$  is the uncorrelated vectors. We can obtain discriminants based on Fisher criterion in  $\mathcal{F}$ .

The matrix  $\mathbf{V}$  used in the first mapping can be derived from  $\mathbf{S}_t$ . Suppose  $\mathbf{S}_t$  to be non-singular matrix with eigenvectors  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$  and eigenvalues  $[\lambda_1, \lambda_2, \dots, \lambda_n]$ . Then,

$$\mathbf{U}^\top \mathbf{S}_t \mathbf{U} = \Lambda \quad (5)$$

where  $\Lambda$  is diagonal matrix with eigenvalues  $\lambda_i$ .

Let  $\mathbf{V} = \mathbf{U}\Lambda^{-1/2}$ , there exists

$$\mathbf{V}^\top \mathbf{S}_t \mathbf{V} = \mathbf{I} \quad (6)$$

where  $\mathbf{I}$  is identity matrix.

Then the  $\mathbf{V}$  in (6) is the first projection matrix we want. Therefore to get discriminants we merely need find out a set of orthogonal projection vectors in  $\mathcal{F}$ . Section 2.2 would demonstrate that the discriminants drawn from fisher criterion fulfill all the requirements.

### 2.2. Fisher discriminants on uncorrelated feature space

Since we have obtained an uncorrelated feature space  $\mathcal{F}$ , the next task is to find a set of orthogonal discriminant vectors maximizing Fisher coefficient.

**Theorem 2.** In a  $n$ -dimensional space  $R^n, \mathbf{x} \in R^n, f(\mathbf{x}) \geq 0, g(\mathbf{x}) > 0$ . Let  $h_1(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$ ,  $h_2(\mathbf{x}) = f(\mathbf{x})/[f(\mathbf{x}) + g(\mathbf{x})]$ . Then  $\mathbf{x}$  maximize  $h_1(\mathbf{x})$  if and only if it maximize  $h_2(\mathbf{x})$ .

The proving procedure is easy and can be referred to Jin [7].

The Fisher coefficient in feature space  $\mathcal{F}$  is defined as:

$$J_F(\phi) = \frac{\phi^\top \mathbf{S}_b^u \phi}{\phi^\top \mathbf{S}_w^u \phi} \quad (7)$$

where  $\mathbf{S}_b^u$  and  $\mathbf{S}_w^u$  denote the between-class scatter matrix and within-class scatter matrix of samples in  $\mathcal{F}$ . According to theorem 2, the above function can be rewritten in a slightly different form:

$$J_F(\phi) = \frac{\phi^\top \mathbf{S}_b^u \phi}{\phi^\top \mathbf{S}_t^u \phi} \quad (8)$$

where  $\mathbf{S}_t^u$  is the total population scatter matrix in the uncorrelated feature space  $\mathcal{F}$ . Considering that we perform the second projection in  $\mathcal{F}$ , and any unit vector  $\phi$  in  $\mathcal{F}$  fulfill  $\phi^\top \mathbf{S}_t^u \phi = 1$ , we can simplify the Fisher criterion as:

$$J_F(\phi) = \phi^\top \mathbf{S}_b^u \phi \quad (9)$$

In this eigenvalues problem the eigenvectors corresponding to the  $k (k \leq n)$  largest eigenvalues of  $\mathbf{S}_b^u$  can be chosen as the  $k$  projection directions. We reach this conclusion not only because these eigenvectors maximize  $J_F(\phi)$  but also because they are orthogonal vectors, which is a premise in theorem 1. In this way, we get optimal discriminant vectors on which all the features projected are uncorrelated.

### 2.3. Solution of uncorrelated discriminants on feature Space (UDFS)

After the theoretical analysis, the procedure of obtaining uncorrelated discriminant vectors can be decomposed into the following several steps:

- Compute the between-class scatter matrix  $\mathbf{S}_b$  and the total population scatter matrix  $\mathbf{S}_t$ .
- Compute the eigenvalues and eigenvectors of  $\mathbf{S}_t$ , then obtain matrix  $\mathbf{\Lambda}$  and  $\mathbf{U}$  in (5).
- Implement the first mapping by use of the eigenvectors of  $\mathbf{S}_t$  as projection directions, and get the between-class scatter matrix in  $\mathcal{F}$ :  $\mathbf{S}_b^u = \mathbf{V}^\top \mathbf{S}_b \mathbf{V}$ .
- Compute the first  $k$  largest eigenvalues and corresponding eigenvectors of  $\mathbf{S}_b^u$  (usually we fix  $k$  to  $C-1$ , with  $C$  the number of classes). Thus we get  $n \times k$  matrix  $\mathbf{\Phi} = [\phi_1, \dots, \phi_k]$  used in the second projection.
- Merge two projection matrixes into one,  $\mathbf{W} = \mathbf{V}\mathbf{\Phi}$ , and normalize the column vectors of  $\mathbf{W}$ . Finally the column vectors of  $\mathbf{W}$  are the expected uncorrelated discriminants.

### 2.4. Relationship with Jin's algorithm (UDRA)

If  $[\lambda_1, \dots, \lambda_n]$  are the eigenvalues of  $S_b^u$  in descending size, the corresponding eigenvectors are  $\mathbf{\Phi} = [\phi_1, \dots, \phi_n]$ , and corresponding vectors in original space  $R^n$  are  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n] = [\mathbf{V}\phi_1, \dots, \mathbf{V}\phi_n]$

Suppose  $\mathbf{w}_i$  ( $i = 1, \dots, r-1$ ) to be the first  $(r-1)$  uncorrelated discriminants that have been verified, i.e.,

$$\tilde{\mathbf{w}}_i = \mathbf{w}_i \quad (i = 1, \dots, r-1). \quad (10)$$

As to the  $r$ th discriminant  $\tilde{\mathbf{w}}_r$ , we define a  $\mathbf{w}$  in  $R^n$  that satisfies

$$\mathbf{w}^\top \mathbf{S}_t \tilde{\mathbf{w}}_i = 0 \quad (i = 1, \dots, r-1). \quad (11)$$

Let  $\mathbf{Q}_{r-1} = \text{span}\{\mathbf{V}\phi_1, \dots, \mathbf{V}\phi_{r-1}\}$  and  $\bar{\mathbf{Q}}_{r-1} = \text{span}\{\mathbf{V}\phi_r, \dots, \mathbf{V}\phi_n\}$ .  $\bar{\mathbf{Q}}_{r-1}$  is the complementary subspace of  $\mathbf{Q}_{r-1}$ .

Obviously  $\mathbf{w} \in \bar{\mathbf{Q}}_{r-1}$ , and we can get an expansion for  $\mathbf{w}$  in terms of linearly uncorrelated  $\mathbf{w}_i$ , i.e.,

$$\mathbf{w} = \sum_{i=r}^n \alpha_i \mathbf{w}_i \quad (12)$$

Then,

$$\begin{aligned} \mathbf{w}^\top \mathbf{S}_b \mathbf{w} &= \sum_{i=r}^n \sum_{j=r}^n \alpha_i \alpha_j \mathbf{w}_i^\top \mathbf{S}_b \mathbf{w}_j \\ &= \sum_{i=r}^n \sum_{j=r}^n \alpha_i \alpha_j \phi_i^\top \mathbf{V}^\top \mathbf{S}_b \mathbf{V} \phi_j \\ &= \sum_{i=r}^n \sum_{j=r}^n \alpha_i \alpha_j \phi_i^\top S_b^u \phi_j = \sum_{i=r}^n \alpha_i^2 \lambda_i \end{aligned} \quad (13)$$

Analogous to (13),

$$\mathbf{w}^\top \mathbf{S}_t \mathbf{w} = \sum_{i=r}^n \alpha_i^2 \quad (14)$$

$$\frac{\mathbf{w}^\top \mathbf{S}_b \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_t \mathbf{w}} = \frac{\sum_{i=r}^n \alpha_i^2 \lambda_i}{\sum_{i=r}^n \alpha_i^2} \leq \lambda_r = \frac{\mathbf{w}_r^\top \mathbf{S}_b \mathbf{w}_r}{\mathbf{w}_r^\top \mathbf{S}_t \mathbf{w}_r} \quad (15)$$

As a result, among all the possible  $\mathbf{w}$  in  $\bar{\mathbf{Q}}_{r-1}$ , only  $\mathbf{w}_r$  maximizes Fisher coefficient. Hence  $\tilde{\mathbf{w}}_r = \mathbf{w}_r$ . Since UDIA seeks to find optimal discriminant that maximizes Fisher coefficient in  $\bar{\mathbf{Q}}_{r-1}$  at each recursion, the  $r$ th discriminant in our method is identical with Jin's.

## 3. Experiment performances and analysis

In this section we conducted two experiments to corroborate the formulation and test the performance of our method. The experiments were based on a widely used pattern recognition benchmark database. As face recognition is an active area in pattern recognition and also a tough problem, the ORL face dataset was adopted in our experiments. This dataset include forty distinct subjects and each subject has ten images with resolution of  $112 \times 92$ . All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position. The variation in scale is up to 10%. The dataset is available at <http://www.uk.research.att.com/facedatabase.html>.

In the initial experiment we chose the first 8 classes of faces in ORL dataset. In each class, we utilize the first 5 of 10 as training samples. All the sample images were resized to  $7 \times 6$  and for comparison, the 42 dimension feature vectors were compressed to 8 dimension using the method in Jin [7]. In the following feature extraction procedure, both our method and Jin's were used for obtaining uncorrelated discriminants, and the results were shown in table 1 and table 2.

From the tables, it is easy to see that discriminants  $\mathbf{w}_1$ ,  $\mathbf{w}_3$ ,  $\mathbf{w}_5$  in two methods were exactly identical, and  $\mathbf{w}_2$ ,  $\mathbf{w}_4$ ,  $\mathbf{w}_6$ ,  $\mathbf{w}_7$  have same values but opposite directions. As we know the direction has no influence on the computation of distances. Thus all the discriminants obtained in the above two methods are identical. This interesting result confirmed the analysis we did in section 2.4.

In the second experiment, we aim at comparing the time cost of our method and Jin's. Respectively we chose the first 30, 32, 34, 36, 38, 40 classes in the dataset. We also used first 5 image in each class to compute  $\mathbf{S}_b$  and  $\mathbf{S}_t$ . But this time we compressed the  $7 \times 6 = 42$  dimensional images to  $C$  dimension ( $C$  is the number of classes). Then two algorithms were used in feature extraction phase. All the com-

**Table 1. Uncorrelated discriminants in our algorithm (UDFS)**

$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	$w_7$
0.2402	0.3216	0.6219	-0.2110	-0.8737	0.4070	0.0755
-0.4337	0.4198	0.3397	0.5778	0.3889	0.4812	0.3737
0.1100	-0.2711	0.5327	-0.5481	-0.0815	-0.6925	0.8783
-0.6231	0.5339	-0.1934	-0.1619	0.0182	-0.1060	0.1611
-0.0435	0.2123	-0.3655	0.2738	-0.1943	-0.0336	0.0311
0.4838	-0.1501	-0.1999	-0.0892	-0.0519	0.1688	0.2340
0.2984	0.5382	0.0552	0.0938	0.1061	-0.1140	-0.0384
-0.1700	-0.0667	-0.0029	0.4509	0.1636	-0.2636	-0.0057

**Table 2. Uncorrelated discriminants in Jin's algorithm (UDRA)**

$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	$w_7$
0.2402	-0.3216	0.6219	0.2110	-0.8737	-0.4070	-0.0755
-0.4337	-0.4198	0.3397	-0.5778	0.3889	-0.4812	-0.3737
0.1100	0.2711	0.5327	0.5481	-0.0815	0.6925	-0.8783
-0.6231	-0.5339	-0.1934	0.1619	0.0182	0.1060	-0.1611
-0.0435	-0.2123	-0.3655	-0.2738	-0.1943	0.0336	-0.0311
0.4838	0.1501	-0.1999	0.0892	-0.0519	-0.1688	-0.2340
0.2984	-0.5382	0.0552	-0.0938	0.1061	0.1140	0.0384
-0.1700	0.0667	-0.0029	-0.4509	0.1636	0.2636	0.0057

**Table 3. The time cost of two methods**

Number of classes	30	32	34	36	38	40
Time cost of UDFS	0.05	0.05	0.06	0.06	0.06	0.06
Time cost of UDRA	0.44	0.55	0.72	0.83	0.98	1.15

putations were conducted on a computer of PIII 366MHz. Table 3 gave the time cost of the two methods.

In table 3, we see that the time cost of our method is roughly constant, while Jin's increases significantly. The reason lies in our algorithm includes only two feature extraction procedure. Yet Jin's algorithm has to do  $(C - 1)$  recursion to extract  $(C - 1)$  discriminants, and the time cost is  $o(C - 1)$ . When  $C$  is large enough, the time cost between two algorithms tends to be  $2 : (C - 1)$ .

## 4. Conclusions

This paper presented a new method for deriving optimal uncorrelated discriminants. Complete theoretical analysis and experimental results demonstrate that our method can get identical discriminants as those in recursive algorithm. Besides, the new method is simpler and costs less time. This paper can also cast light on the optimization problems under some constraints: We may first construct a constrained space, where all the solutions fulfill the constraints. Then the optimization problems can be solved in that constrained space.

## References

- [1] P. N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, January 1999.
- [2] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K. Müller. Fisher Discriminant Analysis with Kernels. *Neural Networks for Signal Processing*, 99(9):41–48, 1999.
- [3] B. Moghaddam. Principal Manifolds and Probabilistic Subspaces for Visual Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(6): 780–788, June 2002.
- [4] Y. F. Guo, S. J. Li, J. Y. Yang, T. T. Shu. A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition. *Pattern Recognition Letters*, 24:147–158, 2003.
- [5] L. F. Chen, H. Y. Liao, M. T. Ko et al. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33:1713-1726, 2000.
- [6] K. Liu et al. . A generalized optimal set of discriminant vectors. *Pattern Recognition*, 25:731-739, 1992.
- [7] Z. Jin et al. Face Recognition based on the uncorrelated Discriminant transform. *Pattern Recognition*, 34:1405-1416, 2001.