

Stable and orthogonal local discriminant embedding using trace ratio criterion for dimensionality reduction

Xiaojun Yang² · Gang Liu³ · Qiang Yu³ · Rong Wang¹

Received: 24 March 2017 / Revised: 1 July 2017 / Accepted: 6 July 2017 /
Published online: 18 July 2017
© Springer Science+Business Media, LLC 2017

Abstract Stable orthogonal local discriminant embedding (SOLDE) is a recently proposed dimensionality reduction method, in which the similarity, diversity and interclass separability of the data samples are well utilized to obtain a set of orthogonal projection vectors. By combining multiple features of data, it outperforms many prevalent dimensionality reduction methods. However, the orthogonal projection vectors are obtained by a step-by-step procedure, which makes it computationally expensive. By generalizing the objective function of the SOLDE to a trace ratio problem, we propose a stable and orthogonal local discriminant embedding using trace ratio criterion (SOLDE-TR) for dimensionality reduction. An iterative procedure is provided to solve the trace ratio problem, due to which the SOLDE-TR method is always faster than the SOLDE. The projection vectors of the SOLDE-TR will always converge to a global solution, and the performances are always better than that of the SOLDE. Experimental results on two public image databases demonstrate the effectiveness and advantages of the proposed method.

Keywords Trace ratio criterion · Manifold learning · Dimensionality reduction · Diversity

✉ Rong Wang
wangrong07@tsinghua.org.cn

¹ The Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

² The School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China

³ The Xi'an Research Institute of Hi-Tech, Xi'an 710025, China

1 Introduction

Recently, the techniques for dimensionality reduction have received a lot of attentions in the fields of computer vision and pattern recognition [9, 10, 12, 18, 24]. Real data are usually depicted in high dimensions, a common way to deal with the high-dimensional data adequately and avoid the curse of dimensionality is to use the dimensionality reduction techniques [8, 23]. Principal component analysis (PCA) [3, 16, 19], linear discriminant analysis (LDA) [1, 20] and locality preserving projections (LPP) [6, 14] are the most widely used techniques for reduced dimensionality. The high-dimensional data information in the real world tend to lie on a smooth nonlinear low-dimensional manifold [2, 23]. However, both PCA and LDA see only the global Euclidean structure and fail to discover the underlying manifold structure [4, 23]. The LPP is proposed to uncover the essential manifold structures by preserving the local structure of image samples [26].

Based on LPP, many related dimensionality reduction approaches including orthogonal locality preserving projections (OLPP) [4], marginal fisher analysis (MFA) [7, 24] and SOLDE [12], and Fast and Orthogonal Locality Preserving Projections (FOLPP) [22] have been developed. In MFA, the intraclass compactness (similarity) of data was characterized by minimizing the distance among nearby data belonging to same classes, and at the same time the interclass separability of data was characterized by maximizing the distance among nearby data belonging to different classes. Both the similarity and the interclass separability are important for improving the algorithmic discriminating ability. However, the projection vectors obtained by LPP and MFA are mutually nonorthogonal. Some researchers [4, 7, 25] have pointed out that enforcing an orthogonality relationship between the projection directions is more effective for preserving local geometry of data and improving the discriminating ability. Thus, the OLPP was proposed by characterizing the similarity of data and enforcing an orthogonality relationship between projection vectors, simultaneously.

Moreover, both MFA and OLPP characterize the similarity of data by minimizing the sum of the distance among nearby data from the same class. Following this idea, nearby data from the same class can be mapped to a single data point in the reduced space. In this way, the intraclass separability (diversity) of data was completely ignored [12], which is also important for preserving local geometry of data. Combining above analysis, we can see that the above mentioned discriminant approaches ignore the diversity of data resulting in instable intrinsic structure representation [12]. Recently, the stable orthogonal local discriminant embedding (SOLDE) was proposed, which takes the similarity, diversity, interclass separability and orthogonal constraint together into account, to preserve local geometry of data and improve the algorithmic discriminating ability. And it shows a good performance for face recognition compared with many prevalent approaches including LPP, OLPP, MFA, locality sensitive discriminant analysis (LSDA) [5] and maximum margin criterion (MMC) [17]. However, it adopted a step-by-step procedure to obtain the orthogonal projection vectors sequentially, due to which the algorithmic computation burden is seriously increased [4, 11].

We solve the problem in SOLDE using the trace ratio criterion. Due to the fact that there is no closed form solution for solving the trace ratio problem, Wang et al. [21] proposed an efficient algorithm, called iterative trace ratio (ITR) algorithm, to find the optimal solution based on an iterative procedure. It is faster than the prevalent approach proposed by Guo et al. [13] and will converge to global solutions. In this paper, we propose a stable and orthogonal local discriminant embedding using trace ratio criterion (SOLDE-TR), which is much faster than SOLDE. Specially, we first generate the objective function of SOLDE to a trace ratio maximization problem. Then, we adopt the ITR algorithm to solve the trace ratio maximization problem iteratively [15, 21]. By this procedure, the orthogonal projection vectors

were optimized simultaneously, and will always converge to a global solution. Especially, the solutions are always better than that of SOLDE.

The rest of this paper is organized as follows: Section 2 gives a brief review of SOLDE. The SOLDE-TR is proposed in Section 3. Section 4 reports all experimental results, and conclusion are drawn in Section 5.

2 Brief review of SOLDE

Given N training data $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in \mathbb{R}^d$ from C classes. Denote the training data matrix by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$. Then, three adjacency graphs $\mathbf{G}_{g-s} = \{\mathbf{X}, \mathbf{S}\}$, $\mathbf{G}_{g-v} = \{\mathbf{X}, \mathbf{H}\}$ and $\mathbf{G}_d = \{\mathbf{X}, \mathbf{F}\}$ are constructed over the training data to model the local similarity, diversity and interclass separability. The elements $S_{ij}, i, j = 1, \dots, N$ in weight matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ are defined as

$$S_{ij} = \begin{cases} K(\mathbf{x}_i, \mathbf{x}_j), & \text{if } \mathbf{x}_i \in N_s(\mathbf{x}_j) \text{ or} \\ & \mathbf{x}_j \in N_s(\mathbf{x}_i), \text{ and } \tau_i = \tau_j, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma)$, $\sigma > 0$ is a proper parameter, $\|\cdot\|$ denotes the ℓ_2 -norm of a vector, $N_s(\mathbf{x}_i)$ stands for the set of s nearest neighbors of \mathbf{x}_i and τ_i is the class label of \mathbf{x}_i . This weight matrix \mathbf{S} mainly characterizes the important similarity of patterns, when the nearby points lie on a compact region.

Moreover, when the nearby points lie on a sparse region, a weight matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$ is defined to characterize the important diversity of patterns. The elements of \mathbf{H} are defined as

$$H_{ij} = \begin{cases} 1 - K(\mathbf{x}_i, \mathbf{x}_j), & \text{if } \mathbf{x}_i \in N_s(\mathbf{x}_j) \text{ or} \\ & \mathbf{x}_j \in N_s(\mathbf{x}_i), \text{ and } \tau_i = \tau_j, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In order to characterize the interclass separability of data in low dimensional representation, a weight matrix $\mathbf{F} \in \mathbb{R}^{N \times N}$ is defined, of which the elements are defined as

$$F_{ij} = \begin{cases} K(\mathbf{x}_i, \mathbf{x}_j), & \text{if } \mathbf{x}_i \in N_s(\mathbf{x}_j) \text{ or} \\ & \mathbf{x}_j \in N_s(\mathbf{x}_i), \text{ and } \tau_i \neq \tau_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Then, three objective functions of SOLDE are constructed to map the original high-dimensional data to a line so that the nearby points with small distance in a same classes can be mapped close in the reduced space, the nearby points with large distance in a same classes can be mapped far, and at the same time the nearby points with small distance in the different classes can be mapped far. The objective functions of SOLDE are defined as

$$\min \sum_{ij} (y_i - y_j)^2 S_{ij} = 2\mathbf{w}^T \mathbf{X} \mathbf{L}_s \mathbf{X}^T \mathbf{w}, \tag{4}$$

$$\max \sum_{ij} (y_i - y_j)^2 H_{ij} = 2\mathbf{w}^T \mathbf{X} \mathbf{L}_v \mathbf{X}^T \mathbf{w}, \tag{5}$$

$$\max \sum_{ij} (y_i - y_j)^2 F_{ij} = 2\mathbf{w}^T \mathbf{X} \mathbf{L}_m \mathbf{X}^T \mathbf{w}, \tag{6}$$

where $y_i = \mathbf{w}^T \mathbf{x}_i$ is the one-dimensional representation of \mathbf{x}_i , \mathbf{w} is a projection vector, $\mathbf{L}_s = \mathbf{D}^s - \mathbf{S}$, \mathbf{D}^s is a diagonal matrix whose elements on diagonal are column sum of \mathbf{S} ,

$L_v = D^v - H$, D^v is a diagonal matrix whose elements on diagonal are column sum of H , $L_m = D^m - F$, D^m is a diagonal matrix whose elements on diagonal are column sum of F . After some simple algebraic steps, the optimal objective functions of (4), (5) and (6) are integrated into the following objective function

$$w^* = \arg \max \frac{w^T X L_m X^T w}{w^T X L_d X^T w}, \tag{7}$$

where $L_d = aL_s - (1 - a)L_v$, a is a suitable constant and $0.5 \leq a \leq 1$.

Aiming to find a set of orthogonal projection vectors, the objective function of (7) can be written as

$$w_1 = \arg \max_w \frac{w^T X L_m X^T w}{w^T X L_d X^T w}, \tag{8}$$

and

$$w_k = \arg \max_w \frac{w^T X L_m X^T w}{w^T X L_d X^T w}, \tag{9}$$

$$w_k^T w_1 = w_k^T w_2 = \dots = w_k^T w_{k-1} = 0.$$

Motivated by [4, 11], the orthogonal projection vectors were obtained by adopting a step-by-step procedure, because of which the algorithmic computation burden was seriously increased. This step-by-step procedure is written as follows:

1. Compute w_1 as the eigenvector of $(X L_d X^T)^{-1} X L_m X^T$ associated with the largest eigenvalue.
2. Compute w_k as the eigenvector of

$$M^{(k)} = \{I - (X L_d X^T)^{-1} W^{k-1} [B^{k-1}]^{-1} [W^{k-1}]^T \cdot (X L_d X^T)^{-1} X L_m X^T, \tag{10}$$

associated with the largest eigenvalue, where $\{w_1, w_2, \dots, w_{k-1}\}$ are the first $k - 1$ projection vectors,

$$W^{k-1} = [w_1, w_2, \dots, w_{k-1}] \text{ and } B^{k-1} = [W^{k-1}]^T (X L_d X^T)^{-1} W^{k-1}.$$

3 Stable and orthogonal local discriminant embedding using trace ratio criterion (SOLDE-TR)

In order to obtain a set of orthogonal projection vectors based on the adjacency graphs of $G_{g-s} = \{X, S\}$, $G_{g-v} = \{X, H\}$ and $G_d = \{X, F\}$ with a relatively less computation burden, we introduce a fast stable and orthogonal local discriminant analysis algorithm, termed as SOLDE-TR. The objection functions of (8) and (9) can be generalized to the following trace ratio maximization problem:

$$W^* = \arg \max_{W^T W = I} \frac{\text{tr}(W^T S_p W)}{\text{tr}(W^T S_d W)}, \tag{11}$$

and the optimum trace ratio value

$$\lambda^* = \frac{\text{tr}(W^T S_p W)}{\text{tr}(W^T S_d W)}, \tag{12}$$

where $W = [w_1, w_2, \dots, w_m]$, m is the desired lower feature dimension, $S_p = X L_m X^T$ and $S_d = X L_d X^T$. We firstly assume that the data $x_i, i = 1, \dots, N$ have been projected

into the PCA subspace such that the matrix S_d is nonsingular. Thus, the denominator of the objective function (11) is always positive for non-zero W , that is, S_d is positive definite.

In order to solve the trace ratio maximization problem (11), we resort to solve the following trace different problem iteratively [21]

$$W^* = \arg \max_{W^T W = I} \text{tr}[W^T (S_p - \lambda_t S_d) W], \tag{13}$$

where λ_t is the trace ratio value calculated from the projection matrix W_{t-1} of the previous step, i.e.,

$$\lambda_t = \frac{\text{tr}(W_{t-1}^T S_p W_{t-1})}{\text{tr}(W_{t-1}^T S_d W_{t-1})}. \tag{14}$$

Then, we propose our fast stable and orthogonal local discriminant analysis method to solve the trace ratio maximization problem (11). The algorithmic procedure of the proposed algorithm is stated as follows:

1. PCA Projection: Project the data $x_i, i = 1, \dots, N$ into the PCA subspace such that the matrix $S_d = X L_d X^T$ is nonsingular. And denote the transformation matrix of PCA by W_{PCA} .
2. Construct the Adjacency Graphs and Choose Weights: Construct the adjacency graphs of $G_{g-s} = \{X, S\}$, $G_{g-v} = \{X, H\}$ and $G_d = \{X, F\}$. Compute the weight matrices S, H , and F according to (1), (2) and (3), respectively.
3. Compute the Orthogonal Projection Vectors:

- (a) Initialize W_0 as an arbitrary matrix whose columns are orthogonal and normalized, $t = 1$;
- (b) Compute the trace ratio value λ_t from the projection matrix W_{t-1} according to

$$\lambda_t = \frac{\text{tr}(W_{t-1}^T S_p W_{t-1})}{\text{tr}(W_{t-1}^T S_d W_{t-1})}. \tag{15}$$

- (c) Compute the eigenvalue decomposition of $(S_p - \lambda_t S_d)$ as:

$$(S_p - \lambda_t S_d) w_k^t = \rho_k^t w_k^t, \tag{16}$$

where ρ_k^t is the k th largest eigenvalue of $(S_p - \lambda_t S_d)$ with the corresponding eigenvector w_k^t .

- (d) Set $W_t = [w_1^t, w_2^t, \dots, w_m^t]$, where m is the desired lower feature dimensions.
 - (e) If $|\lambda_{t+1} - \lambda_t| < \varepsilon$ ($0 < \varepsilon < 10^{-6}$), go to (f); else set $t = t + 1$, go to (b).
 - (f) Output $W_{SOLDE-TR} = W_t$.
4. Orthogonal Embedding: Thus, the embedding is given as follows:

$$x \rightarrow y = W^T x \quad W = W_{PCA} W_{SOLDE-TR},$$

where y is a m dimensional representation of the data x , and W is the transformation matrix.

In each step, a trace difference problem $\arg \max_W \text{tr}[W^T (S_p - \lambda_t S_d) W]$ is solved with λ_t being the trace ratio value computed from the previous step. The projection matrix of our proposed approach will converge to a global solution [21]. And the trace ratio value

will monotonic increase, which have been proved based on point-to-set map theories [21]. Due to the fact that the intrinsic structure of real data is complex, and the testing data are usually different from the training data for a same subject, only similarity or diversity is not sufficient to guarantee the algorithmic generalization capability and stableness. Thus, the proposed SOLDE-TR is more stable and robust than the prevalent OLPP and MFA for testing data.

4 Experimental results

In this section, several experiments were carried out to show the effectiveness of the proposed method on the ORL database and the AR database. More details of the two databases include:

1. ORL Database: ORL consists of 400 facial images from 40 subjects. For each subject, the images were taken at different times. The facial expressions and occlusion also vary. The images were taken with a tolerance for fitting and rotation up to 20 degrees. Each subject contributes 10 different images. All images are grayscale and normalized to the size of 64×64 .
2. AR Database: AR consists of over 4000 color face images of 126 people (70 men and 56 women). The images of the most persons (65 men and 55 women) are taken in two sessions (separated by two weeks). Each session contained 13 gray scale images for each subject, which have been normalized to the size of 50×40 . The images from each session consist of three images with illumination variations, three images with wearing scarf, three images with wearing sun glasses, and four images with expression variations.

We compare the proposed method with the LPP, OLPP and SOLDE for face recognition problem. Without loss generality, a nearest neighbor classifier was used for classification [4]. Likewise, the data vectors was firstly normalized to unit length [4]. In all experiments, the parameter σ was set to the mean distance of the normalized data vectors and the parameters α to 0.9. The weight matrix S was computed according to the fixed σ for each database. In the following experiments, the best recognition accuracies in Tables 1, 2 and the training time of the SOLDE and SOLDE-TR in Table 3 have been emphasized with the bold numbers.

4.1 Face recognition on the ORL database

For the ORL database, four random subsets with ($l = 2, 3, 4, 5$) images per individual were taken with labels to form the training sets. The rest of the database were considered to be the testing sets.

Table 1 Best recognition accuracies (%) on the ORL database

	LPP	OLPP	SOLDE	SOLDE-TR
2 Train	62.34	68.44	71.75	76.69
3 Train	69.52	78.84	83.45	86.77
4 Train	80.44	83.75	90.90	91.29
5 Train	84.93	87.83	95.03	95.03

Table 2 Best recognition accuracies (%) on the AR database

	LPP	OLPP	SOLDE	SOLDE-TR
3 Train	58.63	80.83	75.83	80.83
7 Train	76.43	78.21	78.93	80.71
13 Train	73.35	75.68	75.58	75.96

First, we compare the best recognition accuracies of the LPP, OLPP, SOLDE and our proposed SOLDE-TR on the ORL database. Table 1 summarises the best recognition accuracies of the four methods with the subsets of ($l = 2, 3, 4, 5$). The projection dimension was set to 100, except for the training set ($l = 2$), in which the projection dimension was set to 50. For each given l , we averaged the results over 20 random splits. As can be seen, the SOLDE-TR performs the best in most training sets ($l = 2, 3, 4$). As to the last training set ($l = 5$), SOLDE-TR performs an equal level with SOLDE.

Then, we concentrate on the convergence speed of the SOLDE-TR on the ORL database. Figure 1 shows the increasing trend of the trace ratio value in (15) vs. the iterative numbers. The projection dimension was set to 100 for the ORL training set ($l = 5$). As can be seen, the trace ratio value approximated the maximum after 3th iteration. And our proposed method empirically converged from 6 to 9 iterations.

Next, we show the recognition accuracies versus dimensions on the ORL database. Figure 2 shows the classification performance of the SOLDE, LPP, OLPP and SOLDE-TR on the subset ($l = 3$) of the ORL database. We averaged the results over 20 random splits. As can be seen, the SOLDE-TR outperformed the other methods distinctly.

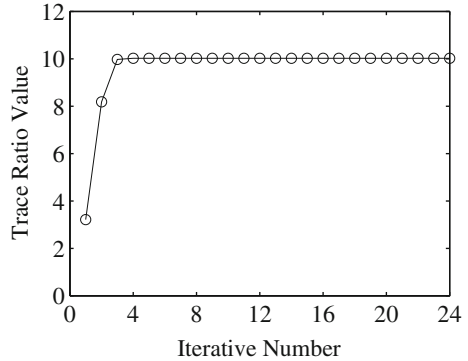
4.2 Face recognition on the AR database

For the AR database, we randomly selected 20 men and 20 women that had participated in two sessions with each individual having 13 images in each session. Then, three subsets with ($l = 3, 7, 13$) images per individual from the first session were selected for the training subsets. The first subset ($l = 3$) was composed of the images with merely illumination variations. The second subset ($l = 7$) was composed of the images with merely expression variations and the images of the first subset. The third subset ($l = 13$) was composed of the total images from the first session. And the corresponding images from the second session of each individual composed the testing sets.

Table 3 Training time in seconds required by LPP, OLPP, SOLDE and SOLDE-TR on the AR database

Dimension	LPP	OLPP	SOLDE	SOLDE-TR
100	9.1 s	33.2 s	74.5 s	35.7 s
200	35.3 s	267.1 s	502.4 s	98.8 s
300	59.3 s	838.1s	1582.3 s	180.9 s
400	111.7 s	2757.5 s	4690.5 s	346.4 s
500	179.8 s	7764.4 s	9949.3 s	602.1 s

Fig. 1 Trace ratio value vs. iterative numbers on ORL database



First, we compare the best recognition accuracies of the LPP, OLPP, SOLDE and our proposed SOLDE-TR on the AR database. Table 2 summarizes the best recognition accuracies of the four methods with the subsets of ($l = 3, 7, 13$). It is shown that the SOLDE-TR performs the best in most training sets ($l = 7, 13$). As to the first training set ($l = 3$), SOLDE-TR performs an equal level with OLPP.

Then, we show the recognition accuracies versus dimensions on the AR database. Figure 3 shows recognition accuracies of the SOLDE, LPP, OLPP and SOLDE-TR on the subset ($l = 3$) of the AR database. We averaged the results over 20 random splits. As can be seen, the SOLDE-TR outperformed the other methods in most cases.

Next, we concentrate on the time consumption of the LPP, OLPP, SOLDE and SOLDE-TR on the AR database. Table 3 summarises the average CPU time consumption of the training procedure with projection dimensions ranging from 100 to 500, measured in seconds, required by LPP, OLPP, SOLDE and SOLDE-TR method on the subset ($l = 13$) of the AR database. All algorithms have been implemented on Matlab R2014a and a computer with Intel I7 2600 CPU (3.5Ghz) and 8 GB RAM. As can be seen, the SOLDE-TR outperforms the SOLDE distinctly. Thus, the computation burden of SOLDE-TR is evidently alleviated compared with SOLDE.

Fig. 2 Recognition accuracy vs. dimensions on the ORL database

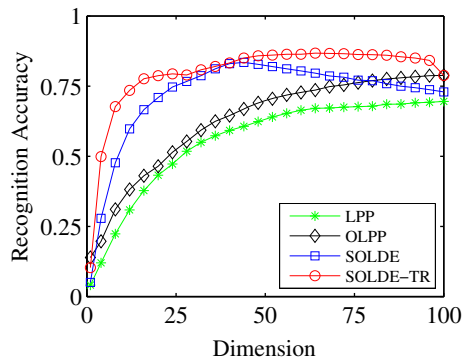
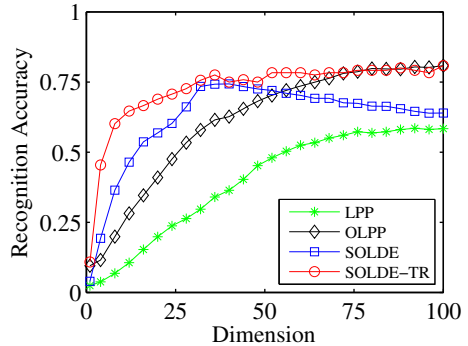


Fig. 3 Recognition accuracy vs. dimensions on the AR database



5 Conclusion

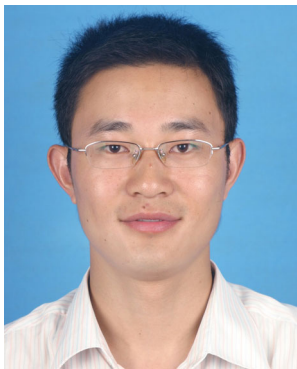
Stable orthogonal local discriminant embedding (SOLDE) is a recently proposed dimensionality reduction method. However, it solves the orthogonal projection vectors through a step-by-step procedure, and is thus computationally expensive. In this paper, a stable and orthogonal local discriminant embedding using trace ratio criterion (SOLDE-TR) method was proposed for dimensionality reduction. The objective function of SOLDE was firstly generalized to a trace ratio maximization problem. Then, the ITR algorithm was provided to optimize the orthogonal projection vectors simultaneously resulting in global solutions. Therefore, the new SOLDE-TR method usually leads to a better solution for face recognition in contrast to the conventional LPP, OLPP and SOLDE. Especially, the solutions of our proposed method are always faster than that of SOLDE. Experimental results on the ORL databases and the AR databases demonstrate the effectiveness and advantages of the proposed method.

Acknowledgements This paper is supported by National Natural Science Foundation of China (No.61401471 and No.61501471); General Financial from the China Postdoctoral Science Foundation (No.2014M552589) and Special Financial from the China Postdoctoral Science Foundation (No.2015T81114).

References

1. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
2. Benmokhtar R, Delhumeau J, Gosselin PH (2013) Efficient supervised dimensionality reduction for image categorization. In: *Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp 2425–2428
3. Boughnim N, Marot J, Fossati C, Bourennane S, Guerault F (2013) Fast and improved hand classification using dimensionality reduction and test set reduction. In: *Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp 1971–1975
4. Cai D, He X, Han J, Zhang H (2006) Orthogonal laplacianfaces for face recognition. *IEEE Trans Image Process* 15(11):3608–3614
5. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis. In: *Proceedings of the 20th international joint conference on artificial intelligence*, pp 708–713
6. Chang X, Ma Z, Yang Y, Zeng Z, Hauptmann AG (2016) Bi-level semantic representation analysis for multimedia event detection. *IEEE Trans Cybern* 1–18. doi:[10.1109/TCYB.2016.2539546](https://doi.org/10.1109/TCYB.2016.2539546)

7. Chang X, Nie F, Wang S, Yang Y, Zhou X, Zhang C (2016) Compound rank- k projections for bilinear analysis. *IEEE Trans Neural Netw Learn Syst* 27(7):1502–1513
8. Chang X, Yang Y (2016) Semisupervised feature analysis by mining correlations among multiple tasks. *IEEE Trans Neural Netw Learn Syst* 1–12. doi:[10.1109/TNNLS.2016.2582746](https://doi.org/10.1109/TNNLS.2016.2582746)
9. Chang X, Yang Y, Long G, Zhang C, Hauptmann AG (2016) Dynamic concept composition for zero-example event detection. In: AAAI
10. Chang X, Yu YL, Yang Y, Xing EP (2016) Semantic pooling for complex event analysis in untrimmed videos. *IEEE Trans Pattern Anal Mach Intell* (99). doi:[10.1109/TCYB.2015.2479645](https://doi.org/10.1109/TCYB.2015.2479645)
11. Duchene J, Leclercq S (1988) An optimal transformation for discriminant and principal component analysis. *IEEE Trans Pattern Anal Mach Intell* 10(6):978–983
12. Gao Q, Ma J, Zhang H, Gao X, Liu Y (2013) Stable orthogonal local discriminant embedding for linear dimensionality reduction. *IEEE Trans Image Process* 22(7):2521–2531
13. Guo YF, Li SJ, Yang JY, Shu TT, Wu LD (2003) A generalized foley–sammon transform based on generalized fisher discriminant criterion and its application to face recognition. *Pattern Recogn Lett* 24(1):147–158
14. He X, Yan S, Hu Y, Niyogi P, Zhang H (2005) Face recognition using laplacianfaces. *IEEE Trans Pattern Anal Mach Intell* 27(3):328–340
15. Jia Y, Nie F, Zhang C (2009) Trace ratio problem revisited. *IEEE Trans Neural Netw* 20(4):729–735
16. Li BN, Yu Q, Wang R, Xiang K, Wang M, Li X (2016) Block principal component analysis with nongreedy ℓ_1 -norm maximization. *IEEE Trans Cybern* 46(11):2543–2547
17. Li H, Jiang T, Zhang K (2006) Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans Neural Netw* 17(1):157–165
18. Luo M, Chang X, Nie L, Yang Y, Hauptmann AG, Zheng Q (2017) An adaptive semisupervised feature analysis for video semantic recognition. *IEEE Trans Cybern*. doi:[10.1109/TCYB.2017.2647904](https://doi.org/10.1109/TCYB.2017.2647904)
19. Parisotto E, Ghassabeh YA, Freydoonjad S, Rudzicz F (2015) Eeg dimensionality reduction in automatic identification of synonymy. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 847–851
20. Wang H, Lu X, Hu Z, Zheng W (2014) Fisher discriminant analysis with L1-norm. *IEEE Trans Cybern* 44(6):828–842
21. Wang H, Yan S, Xu D, Tang X, Huang T (2007) Trace ratio vs. ratio trace for dimensionality reduction. In: Proceedings of the 2007 IEEE conference on computer vision and pattern recognition, pp 1–8
22. Wang R, Nie F, Hong R, Chang X, Yang X, Yu W (2017) Fast and orthogonal locality preserving projections for dimensionality reduction. *IEEE Trans Image Process* PP(99):1–1. ISSN 1057-7149 doi:[10.1109/TIP.2017.2726188](https://doi.org/10.1109/TIP.2017.2726188)
23. Wang S, Chen H, Peng X, Zhou C (2011) Exponential locality preserving projections for small sample size problem. *Neurocomputing* 74(17):3654–3662
24. Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S (2007) Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans Pattern Anal Mach Intell* 29(1):40–51
25. Zhang T, Huang K, Li X, Yang J, Tao D (2010) Discriminative orthogonal neighborhood-preserving projections for classification. *IEEE Trans Syst Man Cyberns Part B: Cybern* 40(1):253–263
26. Zoidi O, Nikolaidis N, Pitas I (2014) Semi-supervised dimensionality reduction on data with multiple representations for label propagation on facial images. In: Proceedings of the 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 6019–6023



Xiaojun Yang receive his M.S. and Ph.D. degrees from High-Tech Institute of Xi'an, Xi'an, China in 2006 and 2010, respectively. His research interests include passive location, nonlinear filter, signal processing and data fusion.



Gang Liu is the professor in High-Tech Institute of Xi'an and received his Ph.D. degree from Northwestern Polytechnical University in 1999. His current research interests include signal processing, navigation and control.



Qiang Yu received his B.S. and M.S. degrees from High-Tech Institute of Xi'an, Xi'an, China in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree in High-Tech Institute of Xi'an, Xi'an, China. His research interests focus on machine learning and its applications.



Rong Wang received his Ph.D. degree from High-Tech Institute of Xi'an, Xi'an, China in 2013. His current research interests include signal processing and machine learning.