Multilinear Tensor-Based Non-parametric Dimension Reduction for Gait Recognition

Changyou Chen, Junping Zhang*, and Rudolf Fleischer

Shanghai Key Lab of Intelligent Information Processing
School of Computer Science, Fudan University, Shanghai 200433, China
cchangyou@gmail.com, jpzhang@fudan.edu.cn, rudolf@cse.ust.hk

Abstract. The small sample size problem and the difficulty in determining the optimal reduced dimension limit the application of subspace learning methods in the gait recognition domain. To address the two issues, we propose a novel algorithm named *multi-linear tensor-based learning without tuning parameters* (MTP) for gait recognition. In MTP, we first employ a new method for automatic selection of the optimal reduced dimension. Then, to avoid the small sample size problem, we use multi-linear tensor projections in which the dimensions of all the subspaces are automatically tuned. Theoretical analysis of the algorithm shows that MTP converges. Experiments on the USF Human Gait Database show promising results of MTP compared to other gait recognition methods.

Keywords: Subspace learning, multi-linear tensor, small sample size problem, dimension reduction, gait recognition.

1 Introduction

Dimension reduction techniques play a crucial role in the discovery of the intrinsic low-dimensional structure in high-dimensional data such as gait sequences. Principal Component Analysis (PCA) attempts to find a subspace maximizing the total variance of the data and minimizing the mean squared errors. Linear Discriminant Analysis (LDA) constructs an optimal discriminative subspace for classification using label information. Furthermore, by incorporating spectral graph theory, algorithms such as LPP [1] and DNE [2] were developed for discovering the intrinsic low-dimensional manifolds in high-dimensional data. Recently, a unified framework of dimension reduction algorithms was proposed by Yan et al. [3] from the viewpoint of graph embedding.

The above-mentioned algorithms have in common that they are based on a vector representation. The disadvantages of using vector representation are 1) the small sample size problem for high-dimensional data, and 2) unfolding of natural data such as images into vector representation may lose some important information of the original data. To overcome these problems, tensor based learning algorithms [4] were proposed for extending vector based algorithms into tensor based counterparts. For example, He et al. [5] generalized LPP into tensor representation, while the LDE algorithm was extended by Dia et al. [6]. More results about tensor based learning can be found in the work of Yan et al. [7] and Xu et al. [8].

* Corresponding author.

Most of the subspace learning algorithms mentioned above have been applied for gait recognition. Wang et al. [9] used PCA to reduce the dimension of gait frames for gait recognition. Han et al. [10] enhanced recognition performance by the combination of LDA with some additional gait data. Furthermore, by integrating random transformations and LDA, Boulgouris et al. [11] proposed a new feature extraction process for gait recognition. While manifold learning methods such as LLE [12] and LPP in [13] had been introduced for gait recognition, the results were not very satisfactory. To obtain higher recognition rates, tensor-based methods such as MMFA [13], TLLE [12] and GTDA [14] were proposed.

It is quite obvious that the effectiveness of the algorithms above, which employ either vector representation or tensor representation, heavily depends on empirically tuned parameters. As a result, it is difficult to determine the intrinsic dimension and choose the corresponding subspace for a given problem instance. While DNE [2] tried to solve this problem by introducing a special Laplacian matrix, the nearest k -neighbor factor has to be predefined, and the automatic selection of the optimal projection subspace remains unanswered in [3].

To tackle the above mentioned problems, we now propose an extended tensor based parameter-free version of the DNE algorithm, named MTP (multi-linear tensor-based learning without tuning parameters). The advantage of DNE and MTP is that the intrinsic dimension can be automatically chosen by solving an eigenvalue problem. Unlike in DNE, we suggest that for the automatical selection of the optimal dimension of a subspace, the Laplacian matrix should be non-positive semi-definite, which was not required in DNE [2]. Furthermore, the relationship matrix is defined without the need of a predefined neighborhood factor. With these two improvements, we propose an objective function that preserves the relationship of pairwise data points in the automatically selected optimal subspace. Since the objective is to preserve pairwise relationships but not to achieve good discriminant ability, we enhance the classification ability by using the LDA technique without the singular matrix problem. We then generalize the proposed algorithm to a multi-linear tensor version and prove its convergence.

The rest of this paper is organized as follows. In Section 2 we present our new MTP algorithm together with some theoretical analysis. We report on some experiment in Section 3, and we conclude the paper in Section 4.

2 The MTP Algorithm

In this section, we first state an important property of the Laplacian matrix. Based on this property, we can define a relationship preserving projection on either vector data or tensor data, followed by LDA for enhancing its discriminant ability. Furthermore, we propose an iterative learning algorithm, and show its convergence. In the following, we denote the training data as $X \in \mathbb{R}^{k_1 \times k_2 \times \dots \times k_n \times N}$, where k_i is the dimension of the projection space in the i -th order of the tensor, m is the dimension of the projection space, and N is the number of training samples. Let $Y \in \mathbb{R}^{m \times N}$ be the projected vector data of each *core tensor* $\mathcal{Y}$, as defined in Section 2.2.

2.1 Non-positive Semi-definite Laplacian Matrix

Yan et al. [3] observed that most dimension reduction techniques can be unified in the graph embedding framework. Its objective is given by

$$F = \min_Y \{tr\{YLY^T\}\} \quad (1)$$

subject to some constraints (in the following, the constraints are usually omitted), where L is the Laplacian matrix that we want to find. However, it is not clear from Eq. (1) how to automatically obtain the optimal dimension of Y , i.e., the value of m . The following lemma addresses this problem.

Lemma 1. *We can calculate the optimal dimension of Y in Eq. (1) to find its minimum value if the Laplacian matrix L is not positive semi-definite, i.e., there exist some vectors $y \in \mathbb{R}^N$ such that $y^T Ly < 0$.*

Proof. $\implies$: Let $Y = (y_1^T, y_2^T, \dots, y_m^T)^T$. Then Eq. 1 can be rewritten as

$$F = \min \sum_{i=1}^m y_i Ly_i^T. \quad (2)$$

Note that L is not positive semi-definite and m is to be determined, we can choose those y_i such that $y_i Ly_i^T < 0$ to reconstruct Y . That is, $Y = (y_1^T, y_2^T, \dots, y_{m^*}^T)^T$ with all y_i ($0 \leq i \leq m^*$) satisfying $y_i Ly_i^T < 0$. In this way, Eq. (1) achieves its minimum value and we can automatically choose the optimal subspace.

In general, the Laplacian matrix L can be decomposed into two symmetric matrices, D and S , such that $L = D - S$ and $D(i, i) = \sum_j S(i, j)$ is a diagonal matrix. We call S the *relationship matrix*, because it reflects the relationship between data points, and the goal of the projection is to preserve these relations. According to Lemma 1, to construct a non-positive semi-definite Laplacian matrix L , we define the relationship matrix S as follows.

$$S(i, j) = \begin{cases} 1 & \text{if } x_i, x_j \text{ are in the same class} \\ -1 & \text{otherwise} \end{cases} \quad (3)$$

Obviously, the matrix L is then non-positive semi-definite, because Eq. (1) can be rewritten as

$$F = \min \sum_{i,j} \|y_i - y_j\|^2 S(i, j). \quad (4)$$

By analyzing the relationship matrix defined in Eq. (3), we can see that F is not always positive, and thus L is not positive semi-definite. According to Lemma 1, we can get an optimal subspace for the projection automatically. Note that this definition is different from the definition in Zhang et al. [2], where it is necessary to predefine a nearest neighbor factor which is not necessary in our new parameter free version.

2.2 Automatic Subspace Learning Using Multi-linear Tensor Projections

In this section, we employ multi-linear tensor projections in the framework proposed in Section 2.1 to learn an optimal subspace which can preserve the relationships of the original data pairs.

Tensors are an extension of vectors and matrices. We write an order- n tensor as $X \in \mathbb{R}^{N_1 \times N_2 \cdots \times N_n}$, which means an order- n tensor has n -ways varying dimensions. Following the SVD decomposition of a second order tensor (i.e., a matrix), we can decompose an order- n tensor into $X = \mathcal{Y} \times_1 U_1^T \times_2 \cdots \times_n U_n^T$, where $\mathcal{Y}$ is called the *core tensor* of X , the U_k are the order principal components of the k -th mode, and $\times_k$ means the k -mode product. A more detailed survey of tensors and tensor operations can be found in [15]. Intuitively, a vector is viewed as an order-1 tensor, a matrix as an order-2 tensor, and a video stream as an order-3 tensor, etc. Fig. 1 shows some visual descriptions from order-1 to order-4 tensors.

In the following, we assume that the data are given as order- n tensors. Our goal is to learn multi-linear projections U_k , for $k = 1, \dots, n$, such that the core tensors of the original tensors can achieve the objective defined in Section 2.1, namely to preserve the relationships between data point pairs. To obtain an optimal decomposition of the tensor, we can get the optimal order principal component U_k of the k -th mode of the data tensor X iteratively, similarly to the SVD decomposition of tensors. Suppose one order principal component U_f is unknown, while all other U_i , $i \neq f$, are known. To rewrite the objective function of Eq. (1) in multi-linear tensor projection form with the unknown order principal component U_f , we first project the data using the known order principal components U_i , $i \neq f$, onto a lower subspace, then unfold the projected data into a matrix on the f -mode, using the f -mode unfolding defined in [15].

$$\begin{aligned}
 F_t(U_f) &= \min \left\{ \sum_{i,j,i \neq j} \left\| x_i \prod_{k=1}^n \times_k U_k - x_j \prod_{k=1}^n \times_k U_k \right\|^2 S_{ij} \right\} \\
 &= \min \left\{ \text{tr} \left\{ U_f \left(\sum_{i,j} \left[\left(x_i \prod_{k \neq f} \times_k U_k \right)^f - \left(x_j \prod_{k \neq f} \times_k U_k \right)^f \right] \right. \right. \right. \\
 &\quad \left. \left. \left[\left(x_i \prod_{k \neq f} \times_k U_k \right)^f - \left(x_j \prod_{k \neq f} \times_k U_k \right)^f \right]^T S_{ij} U_f^T \right) \right\} \right\} \\
 &\quad \text{s.t. } U_f^T U_f = I
 \end{aligned} \tag{5}$$

where y^f means the f -mode unfolding of tensor y , and S_{ij} is the relationship matrix in Eq.3). Let

$$A_f = \sum_{i,j} \left[\left(x_i \prod_{k \neq f} \times_k U_k \right)^f - \left(x_j \prod_{k \neq f} \times_k U_k \right)^f \right] \left[\left(x_i \prod_{k \neq f} \times_k U_k \right)^f - \left(x_j \prod_{k \neq f} \times_k U_i \right)^f \right]^T S_{ij}. \quad (6)$$

Since the Laplacian matrix corresponding to the relationship matrix S is not positive semi-definite, it is easy to see by Lemma 1 that the optimal projection matrix U_f consists of the eigenvectors of the matrix A_f with negative eigenvalues. In this way, we can iteratively optimize the objective function defined in Eq. (5) to obtain the optimal projection matrix U_f given all other $n - 1$ projection matrices $U_i, i \neq f$. Note that when the data are in vector form, there is only a single projection matrix, so we can derive a closed-form solution immediately.

Fig. 1. An order-1 tensor, order-2 tensor, order-3 tensor, and order-4 tensor, from left to right

2.3 Convergence of MTP

To obtain all the projection matrices for tensor based learning, we use the following iterative procedure for searching an optimal solution.

1. We randomly initialize the projection matrices U_i^0 , for $i = 1, \dots, n$. In our experiments, we set $U_i^0 = E_{k_i}$, where E_{k_i} denotes the k_i -dimensional identity matrix.
2. For each U_f , by fixing the other $n - 1$ projection matrices $U_i, i \neq f$, we calculate the eigenvalues and eigenvectors of A_f defined in Section 2.2, and choose the eigenvectors corresponding to negative eigenvalues to form the projection matrix U_f .
3. Let ε be a small positive number. We stop if $\|(U_f^{t-1})^T U_f^t - E_{k_f}\|_F^2 < \varepsilon$ for all U_f , otherwise we continue with the next iteration in step (2).

The following theorem shows that the algorithm always converges. The experiments in Section 3 indicate that it usually converges in four to five iterations.

Theorem 2. *The subspace learning procedure of MTP converges to a local optimum.*

Proof. Let

$$g(U|_{l=1}^n) = \sum_{i,j} \|(x_i - x_j) \prod \times_k U_k\|^2 S_{ij} \quad (7)$$

Let U_k^t denote the k -th projected matrix learned in the t -th iteration, and g_k^t the minimum objective function value of Eq. 7 in the current state. Then,

$$g_k^t \triangleq \min g(U_l^t|_{l=1}^k, U_l^{(t-1)}|_{l=k+1}^n) \leq g_{k-1}^t \quad (8)$$

which means U_k gets updated with a better projection matrix. Thus,

$$g_0^0 \geq \dots \geq g_i^{t-1} \geq g_j^{t-1} \geq \dots \geq g_i^t \geq g_j^t \geq \dots, \quad (9)$$

where $0 < i < j < n$. Note that Eq. (7) has a lower bound because each x_i and U_i in Eq. (7) has finite norm, suggesting that all g_i^t have a lower bound. As a result, the iterative procedure will stop if $\|(U_l^t)^T U_l^{t-1} - I\| < \varepsilon$.

2.4 Improving MTP

Note that in Eq. (3), a relationship matrix S is defined to give us a non-positive semi-definite Laplacian matrix in the MTP algorithm. From the objective function in Eq. (5) and the relationship matrix S , we can conclude

Remark 3. The objective function in Eq. (5) with the relationship matrix S defined in Eq. (3) will result in data projection onto an optimal low dimensional subspace in which the discriminant ability is not strong.

This can easily be seen from the definition of the objective function in Eq. (5), since it only preserves the pairwise relationships of data points. The advantage is that MTP can automatically find an optimal lower dimensional subspace to represent the data. In this subspace, the traditional LDA technique can be used without the singular matrix problem. Of course, the time complexity is also much lower when we use LDA in the subspace. Thus, we can enhance the discriminant ability of MTP by adding a post-processing step, namely to use LDA in the learned subspace. Experiments show that this step is necessary for a satisfactory discriminant ability.

3 Experiments

3.1 The Gait Database

We evaluated MTP on the USF HumanID gait database [16] version 2.1. In this database, gait sequences of 122 subjects are sampled with varying view points, shoe type, surface type, briefcase and elapse time. Selecting sequences like ‘‘Grass, Shoe TypeA, Right Camera, No Briefcase, Time t_1 ’’ as the gallery set, Sarkar et al. [16] developed a set of 12 benchmark experiments under different conditions.

Since Gait Energy Image (GEI) [17] is insensitive to noise [10], we employ the same GEI techniques as in [16] and [10] for the representation of gait features. Then the distance between any two gait sequences using GEIs is calculated as

$$D_L(G_i, P) = \frac{1}{R_p} \sum_{k=1}^{R_p} \min_{j=1, \dots, R_i} d_L(g_{ij}, p_k), \quad (10)$$

where G_i represents the i -th gait sequence in the gallery set, and P is a specific probe sequence. Here, G_i and P contains R_i and R_p GEIs, respectively, that is, $G_i = \{g_{i_1}, \dots, g_{i_{R_i}}\}$, and $P = \{p_1, \dots, p_{R_p}\}$. Furthermore, $d_L(r_1, r_2)$ is the distance between a GEI r_1 in the gallery set and a GEI r_2 in the probe set as computed by our algorithm. Fig. 2 shows some examples of the GEIs of this database. More details can be found in [16,17].

Fig. 2. The GEIs of one individual in the USF gait database. The leftmost one is in the gallery set, while the others are in the probe sets A-L, correspondingly.

3.2 Experimental Results

To evaluate the performance of MTP, we compared it with five different methods: 1) a baseline approach proposed by Sarkar et al. [16] which represents the similarities between frames using the ratio of the number of pixels in their intersection to the number of pixels in their union; 2) two recently developed and popularly used methods, LDA¹ and LDA² [10]. LDA¹ refers to performing LDA on the original gait data, while LDA² fuses the results of LDA on the original gait data and some generated virtual gait data using LDA; 3) the DNE algorithm by Zhang et al. [2]; 4) 2DLDA [18]. Note that there exist some tensor based dimension reduction algorithms such as [14] whose performance depends on the manual tuning of some parameters. For this reason, we don't compare such kinds of algorithms with our proposed algorithm.

Theoretically, MTP does not need to tune any parameters to obtain an optimal learned projection dimension. However, in the experiments, we observed that some negative eigenvalues of the corresponding matrix were close to zero, which we did not expect to happen. To eliminate this noise, we selected only the prominent negative eigenvalues. More exactly, the number of negative eigenvalues is the minimal m satisfying the following criterion:

$$\gamma = \frac{\sum_{i=1}^m |\lambda_i|}{\sum_{i=1}^n |\lambda_i|} > 0.98, \quad (11)$$

where n is the total number of negative eigenvalues, and λ_i is the i -th smallest eigenvalue of the corresponding matrix.

Furthermore, we used MTP for gait recognition using both vector data and tensor data (order-2 tensors). We denote the algorithm dealing with vector data as MTP_V while the one working on tensor data as MTP_T. When we compared it to 2DLDA [18], we set the dimension of their projection matrices to be the same as MTP's dimension for a fair comparison. Tables 1 and 2 show the results of Rank-1 and Rank-5 performances of MTP compared with other methods as in [16].

We can see in Tables 1 and 2 that MTP improves on the average the recognition rate compared to other benchmark algorithms. Because MTP does not need any pre-defined parameters, it actually improves greatly the performance of other tensor learning algorithms.

Table 1. Comparison of Rank-1 performances of different algorithms on the USF gait database version 2.1, with the best performance highlighted by boldface. The projection dimension automatically learned by MTP was 20×17 .

Experiments	Rank-1 Performance						
	baseline [16]	LDA ¹ [10]	LDA ² [10]	DNE[2]	2DLDA[18]	MTP _V	MTP _T
A	73%	86%	88%	88%	78%	90%	90%
B	78%	90%	94%	87%	89%	89%	91%
C	48%	79%	83%	74%	69%	83%	83%
D	32%	29%	31%	21%	23%	35%	37%
E	22%	33%	35%	25%	30%	42%	43%
F	17%	17%	22%	14%	9%	22%	23%
G	17%	23%	26%	18%	15%	28%	25%
H	61%	58%	55%	60%	74%	60%	56%
I	57%	51%	61%	58%	67%	60%	59%
J	36%	53%	48%	40%	53%	56%	59%
K	3%	3%	6%	9%	3%	9%	9%
L	3%	9%	12%	0%	6%	6%	6%
Average	40.95%	47.05%	49.15%	44.16%	45.85%	51.30%	51.57%

Table 2. Comparison of Rank-5 performances of different algorithms on the USF gait database version 2.1, with the best performance highlighted by boldface. The projected dimension automatically learned by MTP was 20×17 .

Experiments	Rank-5 Performance						
	baseline [16]	LDA ¹ [10]	LDA ² [10]	DNE[2]	2DLDA[18]	MTP _V	MTP _T
A	88%	94%	94%	93%	88%	94%	94%
B	93%	94%	94%	93%	94%	94%	93%
C	78%	88%	90%	87%	81%	93%	91%
D	66%	57%	53%	51%	54%	64%	64%
E	55%	58%	60%	57%	52%	67%	68%
F	42%	37%	42%	36%	30%	47%	51%
G	38%	43%	50%	37%	48%	52%	52%
H	85%	85%	86%	82%	92%	89%	88%
I	78%	85%	85%	80%	85%	85%	83%
J	62%	77%	75%	75%	73%	78%	82%
K	12%	15%	15%	30%	12%	21%	18%
L	15%	18%	18%	18%	18%	21%	15%
Average	64.54%	66.71%	67.54%	64.7%	64.46%	71.28%	71.38%

In Theorem 2 we saw that MTP will converge. To experimentally justify this claim, we evaluated the recognition rates of experiments A, B, H and J with increasing iteration numbers from 0 to 10 in Fig. 3(a). Fig. 3(b) and (c) show the differences of the two projection matrices between consecutive iterations, which are defined as $diff = \|(U_i^t)^T U_i^{t-1} - I\|$, for $i = 1, 2, \dots$. In Fig. 3 we can see that MTP has a very satisfactory convergent property. Normally, it will converge in as few as 4 iterations.

Fig. 3. Convergence of MTP: (a) the recognition rates of experiments A, B, H and J with increasing iteration numbers from 0 to 10; (b) and (c) the differences of the first and second projection matrices between consecutive iterations

4 Conclusions and Future Work

In this paper, we proposed a parameter-free tuning algorithm, MTP, to learn an optimal low-dimensional subspace for high-dimensional gait data. In contrast to DNE, MTP can automatically determine the optimal dimensions of the projection matrices and can obtain the optimal solution of the objective functions at the same time, which improves on other tensor based learning algorithms. As a result, it has the potential for real applications such as surveillance systems because it needs little human help to obtain an optimal solution. Moreover, we prove convergence of MTP. However, its performance for more challenging classifications such as elapse time gait recognition still needs improvement. We should also try to further improve MTP's discriminant ability and extend it with kernel tricks.

Acknowledgements

This paper is sponsored by 863 Project (2007AA01Z176) and NSFC (60635030, 60505002, 60573025).

References

1. He, X., Niyogi, P.: Locality preserving projections. In: *Advances in Neural Information Processing Systems* (2004)
2. Zhang, W., Xue, X., Sun, Z., Guo, Y., Lu, H.: Optimal Dimensionality of Metric Space for Classification. In: *International Conference on Machine Learning* (2007)
3. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph Embedding and Extensions: A General Framework For Dimensionality Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(1), 40–51 (2007)
4. Tao, D.C., Li, X.L., Hu, W.M., Maybank, S., Wu, X.D.: Supervised Tensor Learning. *Knowledge and Information Systems* 13(1), 1–42 (2007)
5. He, X.F., Cai, D., Niyogi, P.: Tensor Subspace Analysis. In: *Advances in Neural Information Processing Systems* (2005)
6. Dai, G., Yeung, D.Y.: Tensor Embedding Methods. In: *Proceedings of the National Conference on Artificial Intelligence* (2006)

7. Yan, S., Xu, D., Lin, S., Huang, T., Chang, S.: Element Rearrangement for Tensor-Based Subspace Learning. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
8. Xu, D., Yan, S., Zhang, L., Lin, S., Zhang, H.J., Huang, T.: Reconstruction and Recognition of Tensor-Based Objects With Concurrent Subspaces Analysis. *IEEE Transactions on Circuits and Systems for Video Technology* 18(1), 36–47 (2008)
9. Wang, L., Tan, T.N., Ning, H.Z., Hu, W.M.: Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(12), 1505–1518 (2003)
10. Han, J., Bhanu, B.: Individual recognition using gait energy image. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 28(2), 316–322 (2006)
11. Boulgouris, N.V., Chi, Z.X.: Gait recognition using radon transform and linear discriminant analysis. *IEEE Transactions on Image Processing* 16(3), 731–740 (2007)
12. Li, X., Lin, S., Yan, S., Xu, D.: Discriminant Locally Linear Embedding With High-Order Tensor Data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 38(2), 342–352 (2008)
13. Xu, D., Yan, S., Tao, D., Lin, S., Zhang, H.: Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content Based Image Retrieval. *IEEE Transactions on Image Processing* 16(11), 2811–2821 (2007)
14. Tao, D.C., Li, X.L., Wu, X.D., Maybank, S.J.: General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(10), 1700–1715 (2007)
15. Lathauwer, L.D.: Signal Processing Based on Multilinear Algebra, PhD thesis, Katholieke Universiteit Leuven (1997)
16. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The Humanid Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(2), 162–177 (2005)
17. Liu, Z., Sarkar, S.: Simplest Representation yet for Gait Recognition: Averaged Silhouett. In: *IEEE International Conference on Pattern Recognition* (2004)
18. Ye, J., Janardan, R., Li, Q.: Two-Dimensional Linear Discriminant Analysis. In: *Advances in Neural Information Processing Systems* (2005)