RANKING 2DLDA FEATURES BASED ON FISHER DISCRIMINANCE

Mohammad Shahin Mahanta, Konstantinos N. Plataniotis

Email: {mahanta, kostas} @ece.utoronto.ca
The Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto

ABSTRACT

In classification of matrix-variate data, two-directional linear discriminant analysis (2DLDA) methods extract discriminant features while preserving and utilizing the matrix structure. These methods provide computational efficiency and improved performance in small sample size problems. Existing 2DLDA solutions produce a feature matrix which is commonly vectorized for processing by conventional vector-based classifiers. However, the vectorization step requires a one-dimensional ranking of features according to their discriminance power. We first demonstrate that independent column-wise and row-wise ranking provided by 2DLDA is not sufficient for uniquely sorting the resulting features, and does not guarantee the selection of the most discriminant features. Then, we theoretically derive the desired global ranking score based on Fisher's criterion. The current results focus on non-iterative solutions, but future extensions to iterative 2DLDA variants are possible. Face recognition experiments using images from the PIE data set are used to demonstrate the theoretically proved improvements over the existing solutions.

***Index Terms*—** Fisher's criterion, linear discriminant analysis, 2DLDA, discriminance score, separable covariance.

1. INTRODUCTION

In many applications such as face recognition, the dimensionality of the data is larger than the number of available training samples. Therefore, a feature extractor is commonly used to eliminate the features that do not significantly contribute toward discrimination of the classes. This step helps to improve the accuracy and efficiency of the classification stage [1].

In face recognition and many other applications [2, 3] the data are inherently matrix-variate. In these applications, a bidirectional feature extractor such as the commonly used two-directional linear discriminant analysis (2DLDA) solutions [4, 5, 6, 7, 8, 9, 10, 11] can be used. These methods require lower computational complexity for training and utilize the inherent matrix-variate structure of the data [11]. For facial images, 2DLDA exploits the correlations between neighboring pixels in both column and row directions.

Almost all the 2DLDA feature extractors have been used along with an explicitly or implicitly vectorial classifier. Notably, classifiers based on vector-induced norms such as Frobenius norm are also effectively implying vectorial data. As a result, the 2DLDA feature matrix is effectively concatenated into a vectorial classifier input. Since the 2DLDA features are not correlated anymore, no significant matrix-variate structure will be lost during the vectorization. However, it is required to rank the components of the feature matrix into a vector so that the most discriminant components are passed to the classifier. Although the 2DLDA features are generally sorted along rows and along columns independently, but we will demonstrate that an overall joint sorting is required in order to avoid elimination of discriminant features.

The sorting process is a critical part of feature selection. Feature extractor transforms the data into the feature domain where the most discriminant features are separated. But if the extracted features are not ranked properly according to their discriminance power, the more discriminant features might be lost during dimensionality reduction [12]. Furthermore, as it will be demonstrated in Section 3.1, reducing the dimensionality independently across row and column directions can lead to loss of highly discriminant features.

The currently existing 2DLDA sorting scores follow a mainly heuristic formulation. Therefore, they have been used only as a postprocessing step after the 2DLDA features are already selected [5]. A discriminance score based on Fisher criterion has been formulated in [12], but it is limited to separate reordering of the row and column directions. Similar to the 2DLDA algorithm, [12] does not provide a unique sorting of the features as needed for proper vectorization.

In this paper, using the Fisher's criterion as the discriminance measure, we will theoretically formulate the discriminance score of different 2DLDA feature components. The proposed discriminance score can sort the 2DLDA features simultaneously along both rows and columns. The corresponding formulation is based on parameters that are already calculated by 2DLDA, hence the additional computational complexity is negligible.

For simplicity, this paper focuses on non-iterative 2DLDA [11, 7, 10], although the results can be extended to iterative 2DLDA solutions [4, 8] in the future.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

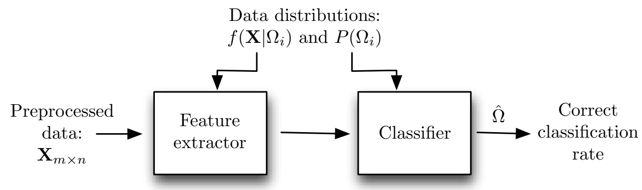

Fig. 1: Overview of the classification system.

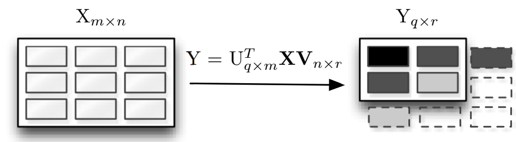

Fig. 2: A practical scenario where the independent row-wise and column-wise sorting of 2DLDA features fails to capture the most discriminant features. Darkness represents discriminance of each feature.

This paper follows with the definition of the classification problem in Section 2, review of related literature in Section 3, derivation of the proposed method in Section 4, experimental analysis in Section 5, and final comments and future directions in Section 6.

2. PROBLEM DEFINITION

This paper considers the problem of classification of matrix-variate data, e.g. preprocessed gray-scale facial images, $\mathbf{X} \in \mathbb{R}^{m\times n}$, into the underlying classes $\Omega_i, 1 \leq i \leq C$. [1] We assume disjoint classes $\Omega_i \cap \Omega_j = \emptyset$, $1 \leq i < j \leq C$, in consistence with many applications [13]. There are N_i training samples available, $\mathbf{X}_{ij}, 1 \leq j \leq N_i$, which are known to belong to each class Ω_i. The available $N = \sum_{i=1}^{C} N_i$ training samples can be used to estimate the likelihoods $f(\mathbf{X}|\Omega_i)$ and prior probabilities p_i for each class. As outlined in Fig. 1, the classification system assigns each input $\mathbf{X}$ to a class $\hat{\Omega}$ such that the correct classification rate is maximized. Using a feature extractor, the most discriminant features are extracted to enhance both estimation accuracy and computational efficiency of the classifier.

This paper focuses on the feature extractor design; hence a simple classifier, i.e., minimum-mean-distance is utilized to facilitate the comparison of different feature extractors. We adopt the commonly used non-iterative 2DLDA (N2DLDA) approach [11, 7, 10] to transform the data into an uncorrelated feature space. The objective of this work is to theoretically derive a discriminance score that can be used for global ranking and hence proper vectorization of the feature matrix components.

3. PRIOR WORK

In this section, the shortcomings of N2DLDA are analyzed, and also Yang's 2DLDA [5] is reviewed as an existing 2DLDA solution with feature ranking.

[1] In this paper, scalars, vectors, and matrices are respectively shown in regular lowercase/uppercase (e.g. a or A), boldface lowercase (e.g. $\mathbf{a}$), and boldface uppercase (e.g. $\mathbf{A}$). The transpose of $\mathbf{A}$, trace of $\mathbf{A}$, null (kernel) space of $\mathbf{A}$, and Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ are respectively denoted by $\mathbf{A}^T$, $\mathrm{tr}(\mathbf{A})$, $\mathrm{Null}(\mathbf{A})$, and $\mathbf{A}\otimes\mathbf{B}$. The vectorized representation of a matrix $\mathbf{A}$ through concatenation of its columns is shown as $\mathrm{vec}(\mathbf{A})$. The symbol for the identity matrix is $\mathbf{I}$.

3.1. Non-iterative 2DLDA Method (N2DLDA)

In the N2DLDA method [11, 7, 10], given the matrix-variate data $\mathbf{X}_{m\times n}$, the mean of the data in each class and the overall mean are found as $\mathbf{M}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{X}_{ij}$ and $\mathbf{M} = \sum_{i=1}^{C}\frac{N_i}{N}\mathbf{M}_i$. The left and right within-class scatters are defined as

$$\begin{aligned}\mathbf{\Phi} &= \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\mathbf{M}_i)(\mathbf{X}_{ij}-\mathbf{M}_i)^T,\\ \mathbf{\Psi} &= \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\mathbf{M}_i)^T(\mathbf{X}_{ij}-\mathbf{M}_i). \end{aligned} \tag{1}$$

Similarly, the left and right between-class scatter matrices are defined as follows:

$$\begin{aligned}\mathbf{S}_{BL} &= \sum_{i=1}^{C}\frac{N_i}{N}(\mathbf{M}_i-\mathbf{M})(\mathbf{M}_i-\mathbf{M})^T,\\ \mathbf{S}_{BR} &= \sum_{i=1}^{C}\frac{N_i}{N}(\mathbf{M}_i-\mathbf{M})^T(\mathbf{M}_i-\mathbf{M}). \end{aligned} \tag{2}$$

The left and right operators, $\mathbf{U}_{m\times q}$ and $\mathbf{V}_{n\times r}$, are constructed with their q columns and r columns respectively as the eigenvectors of $\mathbf{\Phi}^{-1}\mathbf{S}_{BL}$ and $\mathbf{\Psi}^{-1}\mathbf{S}_{BR}$ that correspond to the largest eigenvalues. Furthermore, the uncorrelated eigenvectors are chosen such that $\mathbf{U}^T\mathbf{\Phi}\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{\Psi}\mathbf{V} = \mathbf{I}$.

The matrix-variate N2DLDA features $\mathbf{Y}_{q\times r}$ are calculated through the bilinear operation $\mathbf{Y} = \mathbf{U}^T\mathbf{X}\mathbf{V}$. The row-wise and column-wise directions in $\mathbf{Y}$ are sorted according to Fisher discriminance. However, as demonstrated by Fig. 2, independent selection of the sorted columns and rows does not necessarily result in the selection of the most discriminant features. Furthermore, knowledge of the independent row-wise and column-wise ranking is not sufficient for the vectorized sorting of the features, as required by a vectorial classifier. In Section 4, a global discriminance score is proposed that provides a unique ranking of the N2DLDA features jointly along row and column directions.

3.2. Yang's 2DLDA Method (Y2DLDA)

The Y2DLDA method [5] provides a unique sorting for its features. The feature matrix is calculated as $\mathbf{Y}_{q\times r} = \mathbf{U}'^T\mathbf{X}\mathbf{V}$. The operator $\mathbf{V}_{n\times r}$ is found similar to N2DLDA's right operator. But $\mathbf{U}'_{m\times q}$ is found after $\mathbf{V}$ is applied:

$$\mathbf{\Phi}' = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\mathbf{M}_i)\mathbf{V}\mathbf{V}^T(\mathbf{X}_{ij}-\mathbf{M}_i)^T,$$

$$\mathbf{S}'_{BL} = \sum_{i=1}^{C}\frac{N_i}{N}(\mathbf{M}_i-\mathbf{M})\mathbf{V}\mathbf{V}^T(\mathbf{M}_i-\mathbf{M})^T, \qquad (3)$$

and columns of $\mathbf{U}'$ consist of the q most significant eigenvectors of $\mathbf{\Phi}'^{-1}\mathbf{S}'_{BL}$. Denote the eigenvalues of $\mathbf{\Phi}'^{-1}\mathbf{S}'_{BL}$ and $\mathbf{\Psi}^{-1}\mathbf{S}_{BR}$ by λ'_i and γ_j respectively, corresponding to the columns $\mathbf{U}'_i$ and $\mathbf{V}_j$ from $\mathbf{U}'$ and $\mathbf{V}$. Then, Y2DLDA's score for the (i,j) element of $\mathbf{Y}$ is calculated as $\lambda'_i\gamma_j$.

It should be noted that although the above heuristic formula resembles equation (9) derived in Section 4, our theoretically derived discriminance score represents the Fisher's discriminance only for our proposed framework. The dependence of $\mathbf{\Phi}'$ and $\mathbf{S}'_{BL}$ and hence λ'_i on $\mathbf{V}$ in Y2DLDA formulation causes it to deviate from our derived Fisher discriminance score. Also, Y2DLDA uses an intermediate matrix-variate reduction of dimension and hence marginally suffers from the limitations outlined in Fig. 2. The practical consequences of these theoretical differences are demonstrated using the experimental comparisons in Section 5.

4. PROPOSED FEATURE RANKING METHOD

We use a bilinear feature transformation operator of the form

$$\mathbf{Y}_{m\times n} = \mathbf{U}^T_{m\times m}\mathbf{X}_{m\times n}\mathbf{V}_{n\times n}. \qquad (4)$$

In this derivation, we adopt the $\mathbf{U}$ and $\mathbf{V}$ operators provided by the N2DLDA algorithm, where the i^{th} column of $\mathbf{U}$ (i.e., $\mathbf{U}_i$) corresponds to the eigenvalue λ_i of $\mathbf{\Phi}^{-1}\mathbf{S}_{BL}$, and the j^{th} column of $\mathbf{V}$ (i.e., $\mathbf{V}_j$) corresponds to the eigenvalue γ_j of $\mathbf{\Psi}^{-1}\mathbf{S}_{BR}$. The notable difference with N2DLDA is that there is no dimensionality reduction at this stage. Therefore, we avoid the N2DLDA's inefficiency in selection of the most discriminant features. Instead, a unique global ranking of the features is designed to select the most discriminant features. Notice that after the above transformation, the components in $\mathbf{Y}$ are almost uncorrelated along both column and row directions; therefore, sorting the components of $\mathbf{Y}$ into a vector does not violate the objectives of the matrix-variate approach.

To design a sorting score for the elements of $\mathbf{Y}$, using (4), the (i,j) element of $\mathbf{Y}$ (i.e., Y_{ij}) can be written as [14]

$$Y_{ij} = (\mathbf{V}_j \otimes \mathbf{U}_i)^T\text{vec}(\mathbf{X}),$$

where $\text{vec}(\mathbf{X})$ denotes the vector obtained by concatenation of the columns of $\mathbf{X}$. The vectorial scatter matrices can be written as

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}\text{vec}(\mathbf{X}_{ij}-\mathbf{M}_i)\text{vec}(\mathbf{X}_{ij}-\mathbf{M}_i)^T,$$

$$\mathbf{S}_B = \sum_{i=1}^{C}\frac{N_i}{N}\text{vec}(\mathbf{M}_i-\mathbf{M})\text{vec}(\boldsymbol{\mu}_i-\boldsymbol{\mu})^T. \qquad (5)$$

In most applications, we can approximate these scatter matrices as a separable structure composed of row-wise and column-wise scatters [11, 15]: $\mathbf{\Sigma} = \alpha_1\mathbf{\Psi}\otimes\mathbf{\Phi}$ and $\mathbf{S}_B = \alpha_2\mathbf{S}_{BR}\otimes\mathbf{S}_{BL}$, where $\alpha_1 = \frac{1}{\text{tr}(\mathbf{\Phi})} = \frac{1}{\text{tr}(\mathbf{\Psi})}$ and $\alpha_2 = \frac{1}{\text{tr}(\mathbf{S}_{BL})} = \frac{1}{\text{tr}(\mathbf{S}_{BR})}$. Therefore,

$$\mathbf{\Sigma}^{-1}\mathbf{S}_B = \frac{\alpha_2}{\alpha_1}(\mathbf{\Psi}\otimes\mathbf{\Phi})^{-1}(\mathbf{S}_{BR}\otimes\mathbf{S}_{BL})$$
$$= \frac{\alpha_2}{\alpha_1}(\mathbf{\Psi}^{-1}\mathbf{S}_{BR})\otimes(\mathbf{\Phi}^{-1}\mathbf{S}_{BL}). \qquad (6)$$

The following lemma is used to identify $\mathbf{V}_j\otimes\mathbf{U}_i$ as the eigenvector of $\mathbf{\Sigma}^{-1}\mathbf{S}_B$ corresponding to the eigenvalue $\frac{\alpha_2}{\alpha_1}\gamma_j\lambda_i$.

Lemma 1 ([16, p. 27]). *Let $\mathbf{A}_{m\times m}$ and $\mathbf{B}_{n\times n}$ be two square matrices. Denote their eigenvalues by g_i and h_j respectively, with corresponding eigenvectors $\mathbf{e}_i$ and $\mathbf{f}_j$, where $1\le i\le m$ and $1\le j\le n$. Then, the eigenvalues of $\mathbf{B}\otimes\mathbf{A}$ and their corresponding eigenvectors consist of g_ih_j and $\mathbf{f}_j\otimes\mathbf{e}_i$.*

Using the above results, we formulate the Fisher discriminance score [17, 18] for Y_{ij} as

$$F_{ij} = \frac{(\mathbf{V}_j\otimes\mathbf{U}_i)^T\mathbf{S}_B(\mathbf{V}_j\otimes\mathbf{U}_i)}{(\mathbf{V}_j\otimes\mathbf{U}_i)^T\mathbf{\Sigma}(\mathbf{V}_j\otimes\mathbf{U}_i)} \qquad (7)$$
$$= \frac{\alpha_2(\mathbf{V}_j^T\otimes\mathbf{U}_i^T)(\mathbf{S}_{BR}\otimes\mathbf{S}_{BL})(\mathbf{V}_j\otimes\mathbf{U}_i)}{\alpha_1(\mathbf{V}_j^T\otimes\mathbf{U}_i^T)(\mathbf{\Psi}\otimes\mathbf{\Phi})(\mathbf{V}_j\otimes\mathbf{U}_i)}$$
$$= \frac{\alpha_2}{\alpha_1}\frac{(\mathbf{V}_j^T\mathbf{S}_{BR}\mathbf{V}_j)}{(\mathbf{V}_j^T\mathbf{\Psi}\mathbf{V}_j)}\frac{(\mathbf{U}_i^T\mathbf{S}_{BL}\mathbf{U}_i)}{(\mathbf{U}_i^T\mathbf{\Phi}\mathbf{U}_i)} \qquad (8)$$
$$= \frac{\alpha_2}{\alpha_1}\gamma_j\lambda_i. \qquad (9)$$

The expression in (8) follows from the mixed-product property [14], while (9) uses the generalized Rayleigh quotient [19]. The '$\text{tr}(.)$' and '$\otimes$' operators needed in (7) and (8) respectively are dropped as the operands are scalar values.

As summarized in Fig. 3, for a given $d\le mn$, we select d components Y_{ij} from $\mathbf{Y}$ in (4) that correspond to the highest scores F_{ij} in (9). Since the selected features are uncorrelated, they are sorted in a vector $\mathbf{y}^*_{d\times 1}$. As outlined in Fig. 3, the vectorized features produced by our proposed matrix-to-vector LDA (MVLDA) method contain the highest Fisher discriminance power and do not suffer from the limitations of N2DLDA method (ref. Fig. 2).

5. EXPERIMENTAL RESULTS

Face recognition is by far the most common application for 2DLDA methods. Therefore, we have compared MVLDA

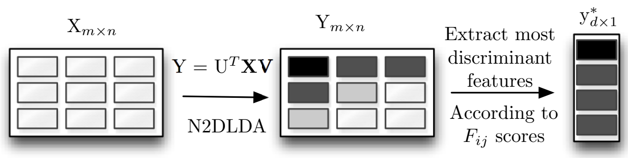

Fig. 3: Schematic view of the proposed MVLDA framework for one-dimensional sorting of the N2DLDA features. Darkness of the extracted features represent the corresponding Fisher score value F_{ij} in (9).

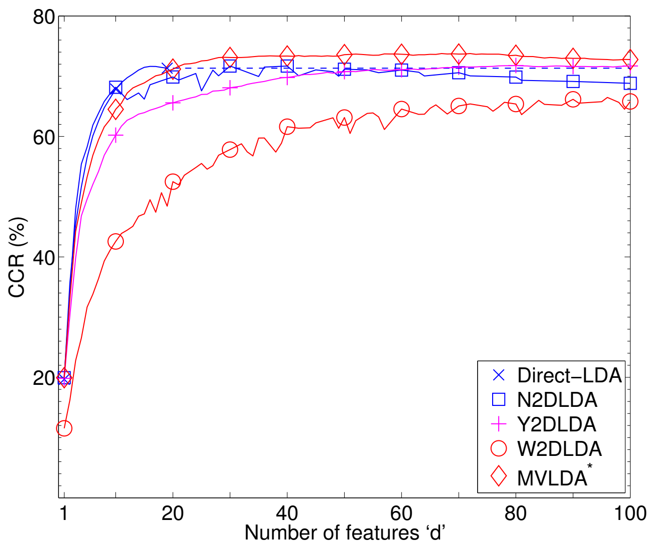

Fig. 4: Correct classification rate for different methods, and the effectiveness of the proposed feature ranking in MVLDA.

with existing methods using face images from the commonly used pose, illumination, and expression (PIE) database [20] which contains 41,368 images from 68 subjects. The facial regions are manually cropped and aligned with a common resolution of 32 × 32 (ref. to [21]), and undergo histogram equalization. A subset of the data set is used that contains 20 randomly selected subjects with at least 147 images per subject. It is known that 2DLDA solutions have the best effect when a small training set is used [22]. Therefore, only 5 images from each of the 20 subjects are randomly selected for training, and all the remaining images are used for testing. This random selection of training samples is repeated 20 times. For each selected training set, the correct classification rate (CCR) using each feature extractor is calculated. The CCR for each method is averaged over the 20 repetitions and the average CCR is plotted in Fig. 4 versus the number of extracted features.

A commonly used one-directional LDA, Direct-LDA [23], has also been included as a benchmark, although its training requires significantly more computations. In addition to N2DLDA and Y2DLDA, the results for Wiriyathammabhum's 2DLDA (W2DLDA) [12] is also included. The latter method provides a reordering of the column-wise and row-wise N2DLDA directions based on Fisher score of the features. It still results in a feature matrix.

Overall, MVLDA achieved the best average CCR of 73.76% with a standard deviation (std.) of 2.66 in Fig. 4. The next standing method, Y2DLDA, could achieve CCR of 71.77% with std. of 2.72. Based on two-sample Student's t-test [24], the improvement provided by MVLDA is statistically significant with p-value of 0.014. This performance gap of Y2DLDA remains significant regardless of the selected values for its intermediately reduced column and row dimensionality. The results in Fig. 4 for Y2DLDA use an intermediate dimensionality of 20 in each direction.

In Fig. 4, the CCR of N2DLDA and W2DLDA is sensitive to the number of features d. Since these methods result in matrix-variate features, for each value of d, the best number of rows and columns (q and r) which result in the highest CCR were selected. The apparent fluctuations can be explained by sudden changes in the selected number of rows and columns, and also the non-feasibility of certain d values that do not factorize into a row and column dimensionality.

The relatively low performance of W2DLDA in this experiment demonstrates that if independent row-wise and column-wise dimension reduction is performed, an independent sorting as suggested by N2DLDA might still outperform a joint row-wise and column-wise Fisher score.

6. CONCLUSIONS AND FUTURE WORK

In conclusion, the proposed ranking based on Fisher score for N2DLDA features helps to select significantly more discriminant features compared to N2DLDA's feature matrix. This improvement is achieved by providing a unique sorting that is missing in N2DLDA's feature matrix. Furthermore, as demonstrated in Fig. 2, a one-dimensional sorting is essential in order to alleviate the shortcomings of a matrix structure in capturing the most discriminant features. Compared to similar heuristic ranking scores, the proposed score is the first one-dimensional sorting score that is theoretically derived based on the Fisher score associated with each N2DLDA feature.

The additional calculations for the proposed Fisher scores are negligible compared to the computational complexity of the N2DLDA algorithm. Furthermore, since the N2DLDA features are uncorrelated, the proposed ranking process in MVLDA does not disregard any major matrix structure.

In this work, we only focused on non-iterative 2DLDA solutions. Consistent Fisher scores for ranking the features of iterative 2DLDA solutions [4, 8] can be derived in the future. Furthermore, depending on the application, the column-wise and row-wise contributions involved in the Fisher score in (9) can be theoretically adjusted so that the column-wise or row-wise separability is emphasized.

7. REFERENCES

[1] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[2] R. Plamondon and S.N. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[3] E.A. Curran and M.J. Stokes, "Learning to control brain activity: A review of the production and control of EEG components for driving braincomputer interface (BCI) systems," *Brain and Cognition*, vol. 51, no. 3, pp. 326 – 336, 2003.

[4] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis.," in *Advances in Neural Information Processing Systems*, 2004.

[5] J. Yang, D. Zhang, X. Yong, and J. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognition*, vol. 38, no. 7, pp. 1125 – 1129, 2005.

[6] K. Inoue, K. Hara, and K. Urahama, "Non-iterative symmetric two-dimensional linear discriminant analysis," *IEICE Transactions on Information and Systems*, vol. 94-D, no. 4, pp. 926–929, 2011.

[7] K. Inoue and K. Urahama, "Non-iterative two-dimensional linear discriminant analysis," in *18th International Conference on Pattern Recognition*, 2006, vol. 2, pp. 540 –543.

[8] D. Luo, C. Ding, and H. Huang, "Symmetric two dimensional linear discriminant analysis (2DLDA)," in *IEEE Conference on Computer Vision and Pattern Recognition*, june 2009, pp. 2820 –2827.

[9] S. Chowdhury, J.K. Sing, D.K. Basu, and M. Nasipuri, "Face recognition by generalized two-dimensional FLD method and multi-class support vector machines," *Applied Soft Computing*, vol. 11, no. 7, pp. 4282 – 4292, 2011.

[10] S. Noushath, G. Hemantha Kumar, and P. Shivakumara, "(2D)2LDA: An efficient approach for face recognition," *Pattern Recognition*, vol. 39, no. 7, pp. 1396 – 1400, 2006.

[11] J. Zhao, P.L.H. Yu, L. Shi, and S. Li, "Separable linear discriminant analysis," *Computational Statistics and Data Analysis*, vol. 56, no. 12, pp. 4290–4300, 2012.

[12] P. Wiriyathammabhum and B. Kijsirikul, "Basis selection for 2DLDA-based face recognition using Fisher score," in *Neural Information Processing*, vol. 5863 of *Lecture Notes in Computer Science*, pp. 708–715. Springer Berlin Heidelberg, 2009.

[13] P.J. Phillips, H. Wechsler, J.S. Huang, and P.J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, pp. 295–306, 1998.

[14] A.K. Gupta and D.K. Nagar, *Matrix variate distributions*, Chapman & Hall/CRC, 1999.

[15] M.S. Mahanta, A.S. Aghaei, and K.N. Plataniotis, "Regularized LDA based on separable scatter matrices for classification of spatio-spectral EEG patterns," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Proceedings*, May 2013, pp. 1237–1241.

[16] A. Graham, *Kronecker products and matrix calculus*, Ellis Horwood, 1981.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press Professional Inc., San Diego, CA, USA, second edition, 1990.

[18] R.A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[19] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley-Interscience, second edition, 2000.

[20] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) database," in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002, pp. 46–51.

[21] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 103–123, Jan. 2009.

[22] W. Zheng, J.H. Lai, and S.Z. Li, "1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based?," *Pattern Recognition*, vol. 41, no. 7, pp. 2156 – 2172, 2008.

[23] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data, with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, Oct. 2001.

[24] N.A.C. Cressie and H.J. Whitford, "How to use the two sample t-test," *Biometrical Journal*, vol. 28, no. 2, pp. 131–148, 1986.