DISCRIMINANT CORRELATION ANALYSIS FOR FEATURE LEVEL FUSION WITH APPLICATION TO MULTIMODAL BIOMETRICS

UNIVERSITY OF MIAMI

Mohammad Haghighat, Mohamed Abdel-Mottaleb, Wadee Alhalabi

Department of Electrical and Computer Engineering, University of Miami

ABSTRACT

In this paper, we present **Discriminant Correlation Analysis (DCA)**, a feature level fusion technique that incorporates the class associations in correlation analysis of the feature sets. DCA performs an effective feature fusion by maximizing the pair-wise correlations across the two feature sets, and at the same time, eliminating the between-class correlations and restricting the correlations to be within classes. Our proposed method can be used in pattern recognition applications for fusing features extracted from multiple modalities or combining different feature vectors extracted from a single modality. It is noteworthy that DCA is the first technique that considers class structure in feature fusion. Moreover, it has a very low computational complexity and it can be employed in real-time applications. Multiple sets of experiments performed on various biometric databases show the effectiveness of our proposed method, which outperforms other state-of-the-art approaches.

INTRODUCTION

Unimodal biometric systems rely on the evidence of a single source of biometrics information, e.g., single fingerprint or face.

Multimodal biometric systems, on the other hand, fuse multiple sources of biometrics information to make a more reliable recognition.

Fusion of the biometrics information can occur at different stages of a recognition system:

- Fusion at the data or feature level
- Fusion at the match score level
- Fusion at the decision level

Feature level fusion is believed to be more effective than the other levels of fusion because the feature set **contains richer information** about the input biometric data than the matching score or the output decision of a classifier [1,2].

FEATURE-LEVEL FUSION USING CCA

Suppose that $X_{(p \times n)}$ and $Y_{(q \times n)}$ are two matrices, containing n training feature vectors. CCA aims to find the linear combinations, $X^* = W_x' X$ and $Y^* = W_y' Y$, that maximize the pair-wise correlations across the two feature sets. The transformation matrices, W_x and W_y, are found by solving the eigenvalue equations [3,4]:

$$S = \begin{pmatrix} cov(x) & cov(x,y) \\ cov(y,x) & cov(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$$

$$\begin{cases} S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx} \hat{W}_x = \Lambda^2 \hat{W}_x \\ S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy} \hat{W}_y = \Lambda^2 \hat{W}_y \end{cases}$$

$$\hat{S} = \begin{pmatrix} 1 & 0 & \dots & 0 & r_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & r_2 & \dots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & r_d \\ r_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \\ 0 & 0 & \dots & r_d & 0 & 0 & \dots & 1 \end{pmatrix}$$

Problems:
- Small Sample Size
  - ➤ PCA + CCA
- Negligence of the class structure
  - ➤ LDA + CCA ?!

DISCRIMINANT CORRELATION ANALYSIS (DCA)

In our method, we incorporate the class structure, into the correlation analysis, which helps in highlighting the differences between classes and at the same time maximizing the pair-wise correlations between features across the two data sets.

Let's assume that the samples in the data matrix are collected from c separate classes. Accordingly, the n columns of the data matrix are divided into c separate groups, where n_i columns belong to the ith class:

$$n = \sum_{i=1}^{c} n_i$$

Let $x_{ij} \in X$ denote the feature vector corresponding to the jth sample in the ith class.

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^{c} n_i \bar{x}_i$$

The between-class scatter matrix is defined as

$$S_{bx(p \times p)} = \sum_{i=1}^{c} n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T = \Phi_{bx} \Phi_{bx}^T$$

$$\Phi_{bx(p \times c)} = [\sqrt{n_1}(\bar{x}_1 - \bar{x}), \sqrt{n_2}(\bar{x}_2 - \bar{x}), \dots, \sqrt{n_c}(\bar{x}_c - \bar{x})]$$

If the classes were well-separated, $\Phi^T \Phi$ would be a diagonal matrix. Since $\Phi^T \Phi$ is symmetric positive semi-definite, we can find transformations that diagonalize it:

$$P^T (\Phi_{bx}^T \Phi_{bx}) P = \hat{\Lambda}$$

where P is the matrix of orthogonal eigenvectors. Let $Q_{(c \times r)}$ consist of the first r eigenvectors, which correspond to the r largest non-zero eigenvalues, from matrix P. We have:

$$Q^T (\Phi_{bx}^T \Phi_{bx}) Q = \Lambda_{(r \times r)}$$

The r most significant eigenvectors of S_{bx} can be obtained with the mapping $Q \to \Phi_{bx} Q$:

$$(\Phi_{bx} Q)^T S_{bx} (\Phi_{bx} Q) = \Lambda_{(r \times r)}$$

If $W_{bx} = \Phi_{bx} Q \Lambda^{-1/2}$, we have:

$$W_{bx}^T S_{bx} W_{bx} = I \qquad X'_{(r \times n)} = W_{bx(r \times p)}^T X_{(p \times n)}$$

X' is the projection of X in a space, where the between-class scatter matrix is I and the classes are separated.

Similarly:

$$W_{by}^T S_{by} W_{by} = I \qquad Y'_{(r \times n)} = W_{by(r \times q)}^T Y_{(q \times n)}$$

Although $S'_{bx} = S'_{by} = I$, the matrices $\Phi'_{bx}{}^T \Phi'_{bx}$ and $\Phi'_{by}{}^T \Phi'_{by}$ are strict diagonally dominant matrices. This makes the centroids of the classes have minimal correlation with each other, and thus, the classes are separated.

Now that the between-class scatter matrices are unitized, we need to make the features in one set have nonzero correlation only with their corresponding features in the other set. Therefore, we diagonalize the between-set covariance matrix $S'_{xy} = X' Y'^T$

$$S'_{xy(r \times r)} = U \Sigma V^T \quad \Rightarrow \quad U^T S'_{xy} V = \Sigma$$

$$(U\Sigma^{-1/2})^T S'_{xy} (V\Sigma^{-1/2}) = I$$
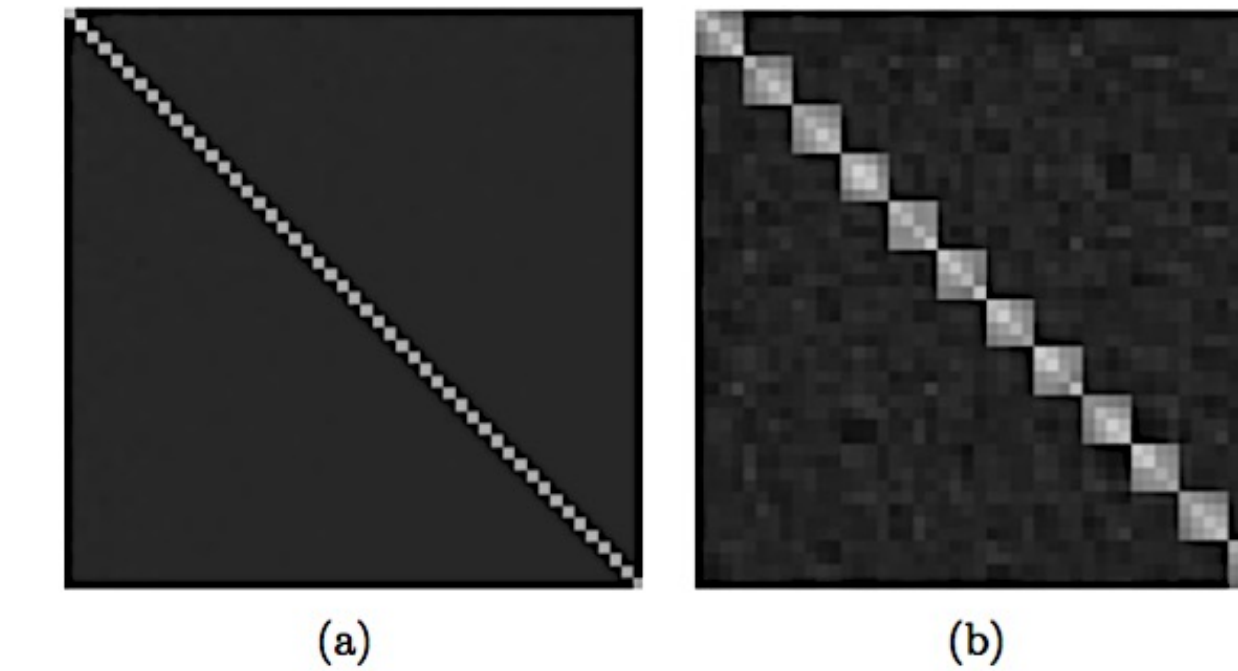$$\underbrace{\quad}_{W_{cx}} \qquad \underbrace{\quad}_{W_{cy}}$$

Transformation:

$$X^* = W_{cx}^T X' = \underbrace{W_{cx}^T W_{bx}^T}_{} X = W_x X$$

$$Y^* = W_{cy}^T Y' = \underbrace{W_{cy}^T W_{by}^T}_{} Y = W_y Y$$

Feature fusion:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} \qquad Z_2 = X^* + Y^* = W_x^T X + W_y^T Y$$


Visualization of covariance matrices. (a) Covariance between features $(X^* X^{*T})$. (b) Covariance between samples $(X^{*T} X^*)$.

EXPERIMENTAL RESULTS

We present two sets of experiments to demonstrate the performance of our proposed feature level fusion technique. First experiment is about the fusion of fingerprint and iris modalities from Multimodal Biometric Dataset Collection, BIOMDATA [5]; and the second experiment is on fusing information from weak biometric modalities extracted from face images in AR face database [6]. The performance of the proposed technique is compared with that of several state-of-the-art methods including the serial feature fusion [7], the parallel feature fusion [8], the CCA-based feature fusion [4,9], and the most recently published JSRC [10] methods.

BIOMDATA Multimodal Biometric Dataset

- 219 subjects having iris and fingerprint modalities
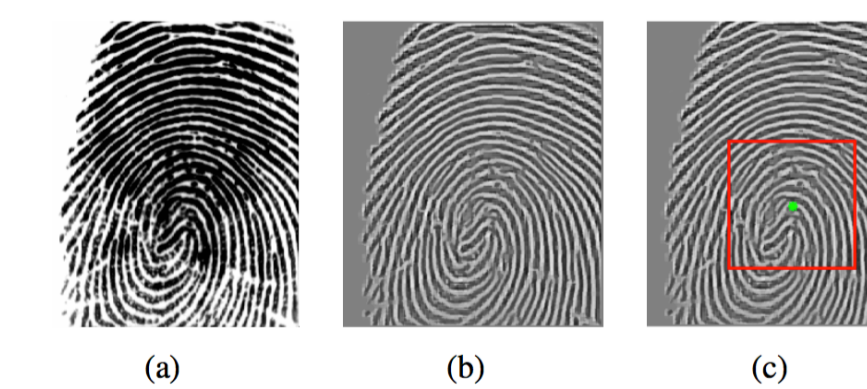- two iris modalities
- four fingerprint modalities
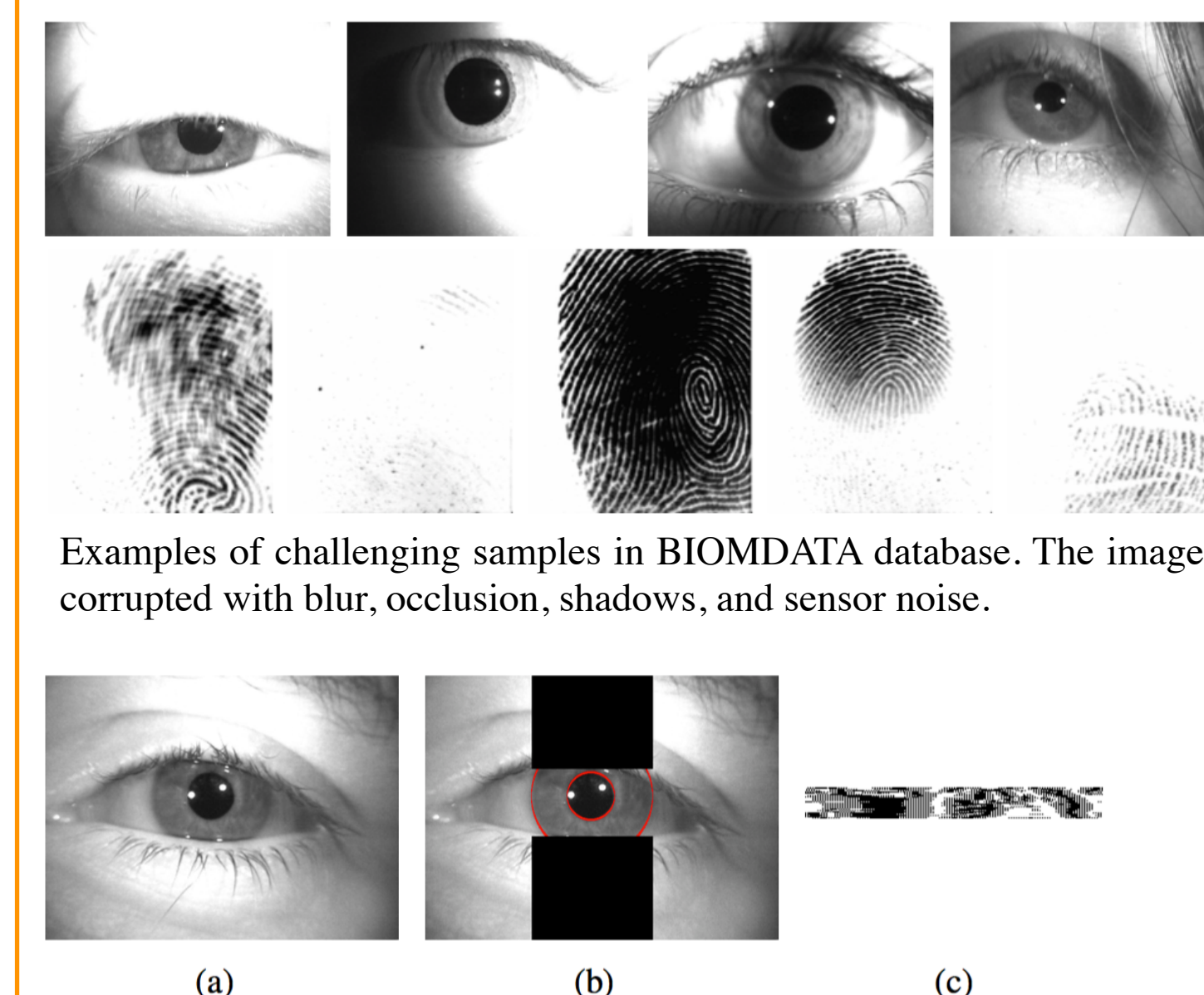
Examples of challenging samples in BIOMDATA database. The images are corrupted with blur, occlusion, shadows, and sensor noise.

(a) (b) (c)
Preprocessing: (a) Original iris image. (b) Segmented iris area. (c) 25×240 binary iris template [11].

Unimodal and Multimodal accuracies in BIOMDATA database obtained by a minimum distance classifier.

Modality	Recognition Rate
Iris (Left)	51.29
Iris (Right)	57.33
Fingerprint (Left thumb)	78.22
Fingerprint (Left index)	90.10
Fingerprint (Right thumb)	79.60
Fingerprint (Right index)	91.29

| Method \ Modality	2 Irises	4 Fingerprints	All 6
Serial + PCA	62.48	94.46	94.85
Serial + LDA	70.31	96.22	96.22
Parallel + PCA	68.22	-	-
Parallel + LDA	72.25	-	-
PCA + CCA	78.51	96.32	97.20
LDA + CCA	78.90	96.40	97.51
JSRC	78.20	97.60	98.60
DCA	84.16	98.71	99.60

(a) (b) (c)
Preprocessing: (a) Original fingerprint image. (b) Enhanced image using the method in [12]. (c) Core point of the fingerprint and the ROI.

AR Face Database

- Fusing weak biometric modalities extracted from face images
- Modalities: 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face
- 100 subjects
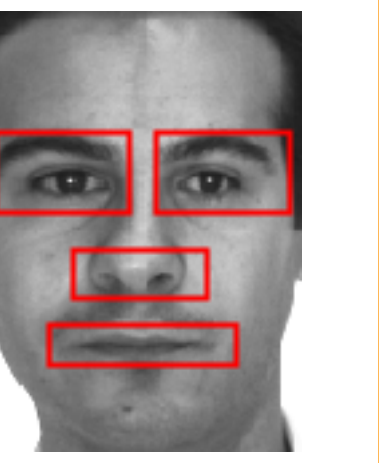
Face mask used to crop out different modalities.

Unimodal and Multimodal accuracies in AR face database.

Modality	1	2	3	4	5
Recognition Rate	84.14	84.29	73.57	74.29	90.57

Method \ Modality	{1,2}	{1,2,3}	{1,2,3,4}	{1,2,3,4,5}
Serial + PCA	85.57	88.71	90.42	90.71
Serial + LDA	89.43	92.14	92.86	93.57
PCA + CCA	90.57	92.86	94.43	96.57
LDA + CCA	91.28	92.57	93.71	97.00
JSRC	92.14	92.86	94.43	98.57
DCA	92.71	93.28	97.43	99.14

CONCLUSIONS

In this paper, we presented a feature fusion technique based on correlation analysis of the feature sets. Our proposed method, called Discriminant Correlation Analysis, uses the class associations of the samples in its analysis. It aims to find transformations that maximize the pair-wise correlations across the two feature sets and at the same time, separate the classes within each set. These characteristics make DCA an effective feature fusion tool for pattern recognition applications. Moreover, DCA is computationally efficient and can be employed in real-time applications. Extensive experiments on various multimodal biometric databases demonstrated the efficacy of our proposed approach in the fusion of multimodal feature sets or different feature sets extracted from a single modality.

REFERENCES

[1] A. Ross and A. Jain, "Multimodal biometrics: An overview," in 12th European Signal Processing Conference, pp. 1221-1224, 2004.

[2] X. Xu and Z. Mu, "Feature fusion method based on KCCA for ear and pro- file face based multimodal recognition," in IEEE International Conference on Automation and Logistics (ICAL), 2007, pp. 620–623.

[3] W. J. Krzanowski, Principles of multivariate analysis: a user's perspective. Oxford University Press, Inc., 1988.

[4] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng, and D.S. Xia, "A new method of feature fusion and its application in image recognition," Pattern Recognition, vol. 38, no. 12, pp. 2437–2448, 2005.

[5] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, "A protocol for multibiometric data acquisition, storage and dissemination," Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering, 2007.

[6] A. Martinez and R. Benavente, "The AR face database," CVC Technical Report, vol. 24, 1998.

[7] C. Liu and H. Wechsler, "A shape-and texture-based enhanced fisher classifier for face recognition," IEEE Transactions on Image Processing, vol. 10, no. 4, pp. 598–608, 2001.

[8] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu, "Feature fusion: Parallel strategy vs. serial strategy," Pattern Recognition, vol. 36, no. 6, pp. 1369–1381, 2003.

[9] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," Expert Systems with Applications, vol. 47, pp. 23–34, 2016.

[10] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 1, pp. 113–126, 2014.

[11] L. Masek and P. Kovesi, "Matlab source code for a biometric identifi- cation system based on iris patterns," The School of Computer Science and Software Engineering, The University of Western Australia, vol. 26, 2003.

[12] S. Chikkerur, C. Wu, and V. Govindaraju, "A systematic approach for feature extraction in fingerprint images," in Biometric Authentication. Springer, 2004, pp. 344–350.