Regularized Independent Component Analysis Regularization in Face Verification

Pang Ying Han†, Ooi Shih Yin††, Teo Chuan Chin†††, Low Cheng Yaw††††, Hiew Fu San†††††, Goh Kah Ong††††††

†, ††, †††, ††††,†††††† Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

††††† Infineon Technologies (Malaysia) Sdn. Bhd., Free Trade Zone, Batu Berendam, 75450 Melaka, Malaysia

Summary

In this work, a regularized Independent Component Analysis (coined as RICA) is proposed in face verification. RICA attempts to generate a linearly independent feature representation with minimal within-class variance, leading to better data discrimination. In RICA, information of correlation coefficients between image data is employed to form a Laplacian matrix. This matrix measures the degree of deviation of a data point from its nearby/ adjacent points. In other words, local discriminative features of data could be disclosured through the Laplacian matrix. As the name suggests, independent component analysis (ICA) is adopted as feature descriptor in this approach. Since there are two different architectures of ICA (ICA I and ICA II), RICA is implemented on these two types of feature extractor and the proposed techniques are known as RICA_ICA I and RICA_ICA II, respectively. The efficiency of RICA is assessed based on three face datasets, namely (1) Facial Recognition Technology (FERET), (2) CMU Pose, Illumination, and Expression (CMU PIE) and (3) Face Recognition Grand Challenge (FRGC).

Key words:

Face verification; Correlation Coefficient; Laplacian Matrix; Regularization; Independent Component Analysis

1. Introduction

Face images always represented in a very high dimensional form. Hence, a vast of facial feature extraction techniques has been researched and introduced [1][2][3][4][5][6]. These techniques share a common objective, which is to produce a lower dimensional informative facial description that could well signify the identity of the face.

Principal Component Analysis (PCA) is a well-known unsupervised technique in face recognition [1]. PCA attempts to seek a set of uncorrelated coefficients (known as principal components) for image representation through optimizing maximal data variance. This technique is then further enhanced by incorporating supervised learning criterion for better performance. This supervised technique is known as Linear Discriminant Analysis (LDA) [2]. With supervised learning capability, LDA shows superior performance to PCA.

However, both techniques are sensitive only to second-order dependencies between pixels. So, higher-order dependencies may not be extracted out. In face images, it is believed that these higher-order dependencies may carry significant information [3]. Therefore, Independent Component Analysis (ICA) is proposed [3]. There are two different architectures of ICA implementation in face verification. ICA architecture I treats images as random variables and pixels as outcomes; and, ICA architecture II treats image pixels as random variables and images as outcomes [3].

Yuen and Lai [7] as well as Steward [3] demonstrated the superiority of ICA over PCA in face recognition. However, the ICA performance is yet to be optimal because class specific information is not taken into account during independent component analysis. With no class label information employment, ICA could not well depict the discriminative information.

2. Motivation and Contribution

If class label information is available, an unsupervised technique could be enhanced by incorporating discriminant criterion for discriminative data learning. The most well-known criteria are Fisher discriminant criterion and maximum margin criterion. The representative instances which adopting Fisher discriminant criterion are including LDA, Fisher ICA [8], Marginal Fisher Analysis (MFA) [4], NPDE [9] and etc.; while those adopting maximum margin criterion are including Maximum Margin criterion (MMC) [10], Regularized Local Discriminant Embedding (RLDE) [11], Locally Linear Discriminant Embedding (LLDE) [12], MNMC [13] and

Manuscript received October 5, 2013
Manuscript revised October 20, 2013

etc.

Besides that, another research path for discriminative data learning is based on the improvement of data population estimation and data locality preservation [14][15][16]. In feature analysis, training data is important for feature extraction techniques to understand/ estimate the basic nature of the data. If the estimation is heavily biased, the representation of the data might not be accurate. Theoretically, a large and accurate training set is preferable to ensure the performance of the techniques. However, practically, limited training samples are always resulting in biased estimates. Hence, regularization is introduced to resolve the biasness. In literature, Jiang et al. introduce Eigenfeature Regularization and Extraction (ERE) for the purpose [14]. Besides, Lu et al. also propose a Locality Preserving Projection (LPP) based regularization technique to regulate LPP features [15]. Recently, a discriminant graph embedding technique that regulates sampling data locality is proposed [16].

Inspired by these works, a regularization model is proposed. Information of correlation coefficients between image data is employed to form a Laplacian matrix $\mathbf{L}$. Through this matrix, the deviation between the data point and its nearby/ adjacent points could be retrieved. This is to discover local discriminative features of the data. These features will then be used to regularize the data input before feature extraction. Unlike the works of Jiang, Lu and Pang which require eigenspace decomposition for weighting function formulation, this proposed regularization model directly utilizes eigenspectrum of $\mathbf{XLX}^{\mathrm{T}}$, where $\mathbf{X}$ is the data matrix, to form the weighting function without additional step for eigenspace decomposition. Besides that, their methods require extra free parameters for eigenspace decomposition or/and weighting function formulation. Validation has to be performed for determining a suitable value for the parameter. But, our proposed regularization model is parameter-free model.

In RICA, after data regularization, the regularized data is processed via independent component analysis (ICA) for feature extraction. Since there are two different architectures of ICA (ICA I and ICA II), RICA is implemented on these two types of feature extractor and the proposed techniques are known as RICA_ICA I and RICA_ICA II, respectively. The efficiency of RICA is assessed by using three face datasets: Facial Recognition Technology (FERET) [17], CMU Pose, Illumination, and Expression (CMU PIE) [18] and Face Recognition Grand Challenge (FRGC) [19].

3. Regularized Independent Component Analysis

In pattern recognition, each image data could be represented as a vertex of a graph G [4]. In graph theory, Laplacian matrix could be used to attain some properties of the graph. Given a graph G with n vertices, its Laplacian matrix is defined as,

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{1}$$

where $\mathbf{L}$ is an $n \times n$ matrix, $\mathbf{W} = \{W_{ij}\}_{n\times n}$ is an adjacency/ similarity matrix and $\mathbf{D} = \{D_{ii} = \sum_j W_{ij}\},\ \forall i \neq j$ is a diagonal degree matrix. The Laplacian matrix measures the extent of dissimilarity of a data point from its adjacent points.

In RICA, the similarity matrix $\mathbf{W}$ is defined based on correlation coefficients between adjacent data pairs $\boldsymbol{x}_{ij}$. These correlation coefficients are easy to interpret that showing the strength of the relationship between $\boldsymbol{x}_{ij}$. In this case, adjacency of data pairs is formed if and only if both data are from the same class. W_{ij} is defined as,

$$W_{ij} = \begin{cases} \frac{Cov_{ij}}{\sqrt{Cov_{ii}Cov_{jj}}} & \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ are from the same class} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where Cov is a covariance matrix.

By adopting $\mathbf{L}$, a discriminative structure of data is modelled by computing the eigenvector $\boldsymbol{v}_i$ of local-Laplacian-scatter matrix $\mathbf{XLX}^{\mathrm{T}}$,

$$\mathbf{XLX}^{\mathrm{T}}\boldsymbol{v}_i = \varphi_i \boldsymbol{v}_i \tag{3}$$

where $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n]$ with $\{\boldsymbol{x}_i \in \boldsymbol{R}^d | i = 1,2,\dots,n\}$. Constitution of $\boldsymbol{v}_i$ forms an eigenspace $\mathbf{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_d]$ and φ_i is the eigenvalue that corresponds to $\boldsymbol{v}_i$ where $\varphi_1 > \varphi_2 > \cdots > \varphi_d$ (Figure 1 illustrates the plot of eigenvalues (blue dotted line) versus the dimension).

Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be two data samples that belonged to a same class. These two data are expected to fall into a region (known as local scatter region) in the feature space, i.e. L^2 distances between $\boldsymbol{w}^T\boldsymbol{x}_i$ and $\boldsymbol{w}^T\boldsymbol{x}_j$ provided that $\|\boldsymbol{x}_i - \boldsymbol{x}_j = \boldsymbol{z}\| < \varepsilon$,

$$\begin{aligned} &E\left(\left|\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{w}^T\boldsymbol{x}_j\right|^2 | \|\boldsymbol{x}_i - \boldsymbol{x}_j\| < \varepsilon\right) \\ &= E(|\boldsymbol{w}^T\boldsymbol{z}|^2 | \|\boldsymbol{z}\| < \varepsilon) \\ &= E(|\boldsymbol{w}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{w}| | \|\boldsymbol{z}\| < \varepsilon) \\ &= \boldsymbol{w}^T E(\boldsymbol{z}\boldsymbol{z}^T | \|\boldsymbol{z}\| < \varepsilon)\boldsymbol{w} \end{aligned} \tag{4}$$

The similarity index W_{ij} could be defined in a simple-minded way,

$$W_{ij} = \begin{cases} 1 & \|\boldsymbol{x}_i - \boldsymbol{x}_j\| < \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Let d be the number of non-zero W_{ij} and $\mathbf{D} = \{D_{ii} = \sum_j W_{ij}\}$, $\forall i \neq j$ is a diagonal degree matrix. According to the Law of Large Number, the average of the results obtained from a large number of trials should be close to the expected value. Hence,

$$
\begin{aligned}
&E(\boldsymbol{z}\boldsymbol{z}^T \,|\, \|\boldsymbol{z}\| < \varepsilon) = \frac{1}{d}\sum_{\|\boldsymbol{z}\|<\varepsilon} \boldsymbol{z}\boldsymbol{z}^T \\
&= \frac{1}{d}\sum_{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|<\varepsilon} (\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T \\
&= \frac{1}{d}\sum_{i,j} (\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T W_{ij} \\
&= \frac{1}{d}\left(\sum_{i,j} \boldsymbol{x}_i\boldsymbol{x}_i^T W_{ij} + \sum_{i,j} \boldsymbol{x}_j\boldsymbol{x}_j^T W_{ij} - \sum_{i,j} \boldsymbol{x}_i\boldsymbol{x}_j^T W_{ij} - \sum_{i,j} \boldsymbol{x}_j\boldsymbol{x}_i^T W_{ij}\right) \\
&= \frac{2}{d}\left(\sum_{i} \boldsymbol{x}_i\boldsymbol{x}_i^T D_{ii} - \sum_{i,j} \boldsymbol{x}_i\boldsymbol{x}_j^T W_{ij}\right) \\
&= \frac{2}{d}(\mathbf{XDX}^T - \mathbf{XWX}^T) \\
&= \frac{2}{d}\mathbf{XLX}^T
\end{aligned}
\tag{6}
$$

From here, we could see that local-Laplacian-scatter matrix $\mathbf{XLX}^{\mathrm{T}}$ indeed describes variance of same-class data. Hence, φ_i, which is the eigenvalue of the local-Laplacian-scatter matrix $\mathbf{XLX}^{\mathrm{T}}$, is signifying the local variance of the corresponding eigenvector $\boldsymbol{v}_i$. Larger value of φ_i that corresponding to $\boldsymbol{v}_i$ indicates higher local/ same-class variation in $\boldsymbol{v}_i$. Hence, $\boldsymbol{v}_i$ should be regularized with smaller weights to reduce the same-class data discrepancy. On the other hand, zero value of φ_i implies zero local variation embedded in $\boldsymbol{v}_i$. This zero variance, computed using the training set, is data specified and it might not apply same meaning on other data set. However, the smallest (either nonzero or zero) eigenvalues imply the subspace possessing minimal same-class variation. So, the subspace should be greatly weighted. In other words, increasing-valued weights should be imposed to the increasing-ordered eigenvectors until zero-eigenvalued eigenvectors are reached. In this null subspace, a constant maximal weight is granted. Based on this principle, a weighting model is proposed (Figure 1 illustrates the model (red dashed line)),

$$w_i = \frac{1}{exp(\varphi_i)} \tag{7}$$

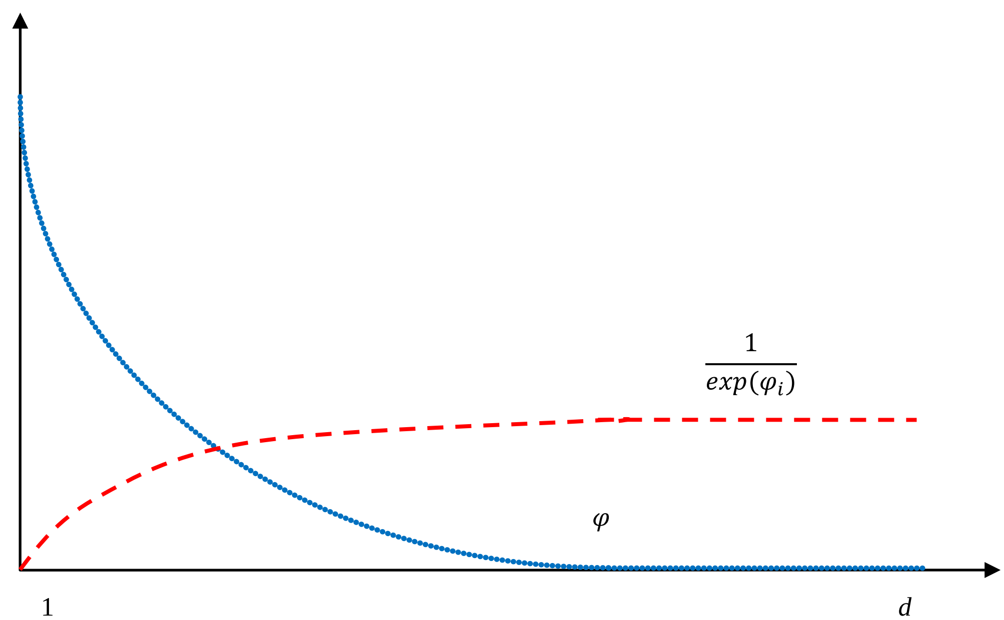

Fig. 1. A plot of φ (blue dotted line) and $\frac{1}{exp(\varphi_i)}$ (red dashed line).

Next, $\boldsymbol{v}_i$ is regularized by the weighting model as $\widetilde{\boldsymbol{v}}_i = \frac{1}{exp(\varphi_i)}\boldsymbol{v}_i$. Before feature extraction, the training data is preprocessed for minimal within-class variance through projecting onto $\widetilde{\mathbf{V}} = [\widetilde{\boldsymbol{v}}_1, \widetilde{\boldsymbol{v}}_2, \ldots, \widetilde{\boldsymbol{v}}_d]$, i.e.

$$\widetilde{\boldsymbol{x}} = \widetilde{\mathbf{V}}^{\mathrm{T}}\boldsymbol{x} \quad (8)$$

In this proposed approach, ICA is employed for feature extraction. ICA attempts to construct a set of linearly independent basis signals from an observed signal [3]. In face analysis, face images are treated as a mixture of unknown statistically independent source signals by an unknown mixing matrix (Figure 2). In this case, the regularized face input is the observed mixture, i.e. $\widetilde{\boldsymbol{x}} = \mathbf{A}\boldsymbol{s}$. ICA seeks the separating matrix $\mathbf{W}$ in such a way that

$$\boldsymbol{u} = \mathbf{W}\widetilde{\boldsymbol{x}} = \mathbf{W}\mathbf{A}\boldsymbol{s} \quad (9)$$

is an estimate of the true source signals.

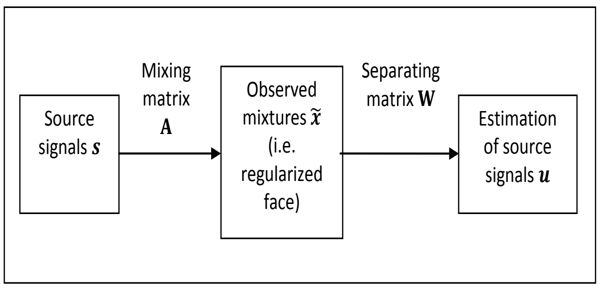

Fig. 2. ICA implementation on face.

In this work, the regularized data input $\widetilde{\boldsymbol{x}}$ will be processed via two different architectures of ICA: ICA I and ICA II for feature extraction. Hence, these proposed techniques are known as RICA_ICA I and RICA_ICA II, respectively.

3.1 RICA_ICA I

Similar to ICA I [3], RICA_ICA I attempts to seek basis signals that are statistically independent. Hence, $\widetilde{\boldsymbol{x}}$ are variables and pixels are observations for the variables. Before ICA analysis, PCA is performed onto the data to reduce data dimension and discard small trailing eigenvalues [3].

Let $\mathbf{R}$ be a $d \times p$ matrix where d is the number of data pixels and p is the first p eigenvectors of a set of n face images. ICA I is implemented on R^T where the data input in the row are treated as variables and the pixels in the column are observations. Independent basis vector $\boldsymbol{u}$ is computed,

$$\boldsymbol{u} = \mathbf{W}\mathbf{R}^{\mathrm{T}} \quad (10)$$

and $\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d]$.

Based on the computed PCA coefficients, $\mathbf{C} = \widetilde{\mathbf{X}}^{\mathrm{T}}\mathbf{R} = \mathbf{X}^{\mathrm{T}}\widetilde{\mathbf{V}}\boldsymbol{R}$, ICA coefficients matrix is calculated,

$$\mathbf{B} = \mathbf{C}\mathbf{W}^{-1} = \mathbf{X}^{\mathrm{T}}\widetilde{\mathbf{V}}\boldsymbol{R}\mathbf{W}^{-1} \quad (11)$$

3.2 RICA_ICA II

RICA_ICA II attempts to seek statistically independent coefficients for the input data. Hence, opposite to RICA_ICA I, $\widetilde{\boldsymbol{x}}$ are observations and pixels are variables. In order to reduce data dimension and discard small trailing eigenvalues, PCA is performed before feature extraction [3]. The statistically independent coefficients are calculated,

$$\mathbf{U} = \mathbf{W}\mathbf{C}^{\mathrm{T}} \quad (12)$$

4. Experimental Results and Discussions

The performance of RICA is evaluated using three different face databases whose images are with significant illumination as well as facial expression variations:

(1) Facial Recognition Technology (FERET) [17]: there are 100 subjects with 10 images per subject. Five images per subject are used as training set and another remaining five images are used for testing.

(2) CMU Pose, Illumination, and Expression (CMU PIE) [18]: there are 67 subjects with 20 images per subject. Half of the images per subject, i.e. 10, are used for training and another half, that is 10, images are used for testing.

(3) Face Recognition Grand Challenge (FRGC) [19]. These are 150 subjects with 10 images per subject. Five images per subject are used as training set and another remaining five images are used for testing.

It is noted that there is no overlap in image between training and testing sets.

In this paper, we also address a performance comparison between RICA and other unsupervised techniques, i.e. PCA, ICA I, ICA II and Locality Preserving Projection (LPP) [20], as well as supervised techniques, i.e. LDA, Supervised LPP [20], MFA, MMC, ERE and RLPDE. In order to have a fair comparison, these techniques are conducted under a same testing strategy and evaluated using a same classifier, i.e. Euclidean metric based nearest neighbourhood classifier.

4.1 Verification Performance

LDA and MMC project data samples onto a subspace constituted by the c-1 largest eigenvectors, where c is the number of class. So, LDA and MMC feature lengths are 99 in FERET database, 66 in CUM PIE database and 149 in FRGC database. Table 1 records the best results, corresponding to the optimal feature dimension *t*, of various techniques on the three different databases.

Table 1: Verification performance of various techniques on FERET database

Technique	Error rate (%) [dimension, t]		
	FERET	CMU PIE	FRGC
Unsupervised technique			
PCA	40.04 [80]	66 [90]	49.44 [110]
LPP	40.37 [60]	45.23 [80]	42 [150]
ICA I	39.26 [50]	60.1 [190]	37.77 [130]
ICA II	38.27 [30]	38.8 [10]	34.64 [30]
Supervised technique			
LDA	39.06 [99]	28.67 [66]	40.38 [149]
SLPP	30.3 [10]	33.73 [70]	37.96 [60]
MFA	35.8 [10]	27.64 [100]	42.2 [100]
MMC	39.17 [99]	35.96 [66]	40.23 [179]
ERE	27.28 [10]	19.85 [10]	22.4 [10]
RLPDE	25.6 [10]	20.4 [10]	20.22 [10]
RICA_ICA I	27.7 [10]	27.5 [10]	24.7 [9]
RICA_ICA II	28.41 [9]	27.7 [9]	26.61 [9]

4.2 The Proposed Regularization Model on Other Feature Extractor, i.e. PCA

Besides implementing the eigenspectrum based regularization model on ICA, we extend the proposed regularization model to other feature extractor that is PCA, in face verification. This integration is namely RICA_PCA. Table 2 records mean error rates of PCA and ICA with their regularized versions on three different face databases.

Table 2: Verification performance of Regularization Model on PCA Feature Extractor

Method	Error rate (%)		
	FERET	CMU PIE	FRGC
Unsupervised technique			
PCA	40.04	66	49.44
ICA I	39.26	60.1	37.77
ICA II	38.27	38.8	34.64
Supervised technique			
RICA_ICA I	27.7	27.5	24.7
RICA_ICA II	28.41	27.7	26.61
RICA_PCA	28.3	27.4	24.9

4.3 Discussions

From the above experimental results, we notice that:

1. RICA obtains its good score in face verification with small number of features. These results indicate that RICA is able to disclosure discriminating features in the lower ordered projection directions/ eigenvectors.
2. RICA_ICA I and RICA_ICA II are the enhanced regularized version of ICA I and ICA II, respectively. From the results, RICA consistently outperforms ordinary ICA on all the tested databases. The regularization in RICA successfully improves the performance of ICA with at least 20%. This validates the effectiveness of the regularization model in RICA to process the data for minimal intra-class variation, leading to better data discrimination.
3. Supervised techniques employ class membership information for data learning during training phase. LDA, MFA and MMC analyse both same-class and different-class information explicitly through discriminant criterion, i.e. Fisher criterion or Maximum Margin criterion. SLLP utilizes class specific information to identity the true neighbourhood of a data (same class data) for disclosing intrinsic data manifold. On the other hand, RICA, which comprising RICA_ICA I and RICA_ICA II, utilizes class specific information to model the discriminative intrinsic data structure based on the labelled training samples. After then, the data eigenspace is regularized for minimal within-class variation before inputting for feature extraction. Experimental results show that the feature regularization of RICA is more effective than the other discriminant function, i.e. Fisher or Maximum Margin criterion, in class discrimination application.
4. From the experimental results, we discover that the proposed RICA shows inferior performance to other regularized techniques, i.e. ERE and RLPDE. Accurate data population estimation from a sample (set of training data) could ensure efficient data learning of feature extraction techniques for data representation. This means that the proposed regularization model of RICA is not as effective as that of ERE and RLPDE for data population estimation. Nevertheless, direct adoption of eigenspectrum, which without subspace decomposition and parameter-free regularization function, for regularization model formulation is a good initiative. Furthermore, RICA has shown its superiority to other supervised techniques that incorporating explicit discriminant criterion, i.e. LDA, MFA and MMC.

5. ICA could be viewed as an extended version of PCA since it not only takes into account second-order dependencies but also high-order dependencies between pixels. Yuen and Lai [7] as well as Steward [3] claim the superiority of ICA over PCA in face recognition. However, different result is obtained in section 4.2. With data regularization prior feature extraction, there is no significant difference in performance between the two techniques. Besides, the data regularization successfully enhances the discriminating capability of PCA in face recognition by showing at least 50% performance improvement.

5. Conclusion

In this paper, a regularized Independent Component Analysis (coined as RICA) on face verification is presented. RICA attempts to construct a discriminative feature representation with minimal within-class variance from facial images. RICA adopts information of correlation coefficients between image data to form a local-Laplacian-scatter matrix. This matrix will then be used to discover local discriminative features. By utilizing this local variance information, the input data is regularized before feature extraction. In this proposed approach, independent component analysis (ICA) is adopted as feature descriptor. RICA is implemented on the two different architectures of ICA (type I and type II) and namely as RICA_ICA I and RICA_ICA II, respectively. From the experimental results, both RICA_ICA I and RICA_ICA II demonstrate superior performance to some other feature extraction techniques, including other discriminant supervised techniques i.e. LDA, MFA and MMC. However, the proposed technique shows inferior performance to ERE and RLPDE. This is because the regularization model adopted in RICA is not as efficient as that in ERE and RLPDE. Thus, seeking a better regularization modelling that could well estimate data population from a limited number of training samples will be the future research direction. Similar to RICA, in formulating this potential regularization model, there is no need of subspace decomposition, but a direct manipulation of eigenspectrum.

6. Acknowledgement

The authors acknowledge the financial support of Telekom Research and Development Sdn. Bhd. of Malaysia

References

[1] M. Turk, A. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, 3(1), pp. 71-86, 1991.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear,” IEEE Transactions on Pattern Analysis and Machine Intelligence 19, pp. 711-720, 1997.
[3] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, “Face Recognition by Independent Component analysis,” IEEE Transactions on Neural Networks 13(6), pp.1450-1464, 2002.
[4] S.C. Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, S. Lin, “Graph Embedding and Extensions: A General Framework for Dimensionality Reduction,” IEEE Trans. On Pattern Analysis and Machine Intelligence 29(1), pp. 40-51, 2007.
[5] Y.H. Pang, T.B.J. Andrew, N. David, “Face authentication system using pseudo Zernike moments on wavelet subband.” IEICE Electronic Express 01/2004, pp. 275-280, 2004.
[6] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Yi Ma, “Robust Face Recognition via Sparse Representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), pp. 210-227, 2009.
[7] P.C. Yuen, J.H. Lai, “Independent Component Analysis
[8] of Face Images,” IEEE Workshop on Biologically Motivated Computer Vision, 1811, pp. 545-553, 2000.
[9] Keun-Chang Kwak, Pedrycz, W., “Face Recognition Using an Enhanced Independent Component Analysis Approach,” IEEE Transactions on Neural Networks 18(2), pp. 530-541, 2007.
[10] Y.H. Pang, B. Andrew, Fazly Salleh Abas, “Neighbourhood preserving discriminant embedding in face recognition,” Journal of Visual Communication and Image Representation, 20(8), pp. 532-542, 2009.
[11] H. Li, T. Jiang, “Efficient and robust feature extraction by Maximum Margin Criterion,” IEEE Trans. on Neural Networks, 17(1) , pp. 157-165, 2006.
[12] Y. Pang, N. Yu, “Regularized Local Discriminant Embedding,” IEEE Int. Conf. On Acoustics, Speech, Signal Processing, 3, (p. III), 2006.
[13] B. Li, C. Zheng, D. Huang, “Locally linear discriminant embedding: an efficient method for face recognition,” Pattern Recognition 41 , pp. 3813-3821.
[14] Y.H. Pang, T. Andrew, “Maximum Neighbourhood Margin Criterion in Face Recognition,” SPIE journal of Optical Engineering 48, pp. 047205, 2009.
[15] X. Jiang, M. Bappaditya, A. Kot, “Eigenfeature Regularization and Extraction in Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), pp. 383-394, 2008.
[16] J. Lu, Y. Tan, “Regularized locality preserving projections and its extensions for face recognition,” IEEE Trans. On System Man. And Cybernetics 40(3) : pp. 958-963, 2010
[17] Y.H. Pang, T. Andrew, Fazly Salleh Abas, “Regularized locality preserving discriminant embedding for face recognition,” Neurocomputing 77(1), pp. 156-166, 2012.
[18] P.J. Phillips, H. Moon, P.J. Rauss, S. Rizvi, “The FERET evaluation methodology for face recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), pp. 1090-1104, 2000.

[19] T. Sim, S. Baker, N. Bsat, "The CMU Pose, Illumination, and Expression Database," IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12) , pp. 1615-1618, 2003.
[20] P.J. Phillips, P.J., Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, "Overview of the face recognition grand challenge," IEEE International Conference on Computer Vision and Pattern Recognition, pp. 947-954, 2005.
[21] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, "Face recognition using laplacianfaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (3), pp. 328-340, 2005.

Pang Ying Han received her Bachelor degree in Electronic Engineering in year 2002, Master of Science degree in year 2005 and Ph.D degree in year 2013 from Multimedia University. She is currently a senior lecturer at Multimedia University. Her research interests include face recognition, manifold learning, image processing and pattern recognition

Ooi Shih Yin received her Bachelor of Information Technology and Master of Science from Multimedia University in 2004 and 2006 respectively. Shih-Yin She is currently a lecturer at Multimedia University and is a PhD student. Her research areas are biometrics, image processing, machine intelligence, computer vision, and data mining.

Teo Chuan Chin received his Bachelor of Information Technology in Software Engineering and Master of Science (Information Technology) from Multimedia University in 2003 and 2006. He is currently a lecturer at Multimedia University, Malaysia. His research interest includes image processing, biometrics and pattern recognition.

Low Cheng Yaw received his Bachelor of Information Technology (Hons) in Data Communications and Master of Science (Information Technology) from Multimedia University in 2004 and 2009. He is currently a lecturer at Multimedia University, Malaysia. His research interest includes biometric watermarking, and image and video processing.

Hiew Fu San received his Bachelor degree, majoring in Computer Engineering in year 2002 and Master of Engineering degree in year 2008 respectively from Multimedia University, Malaysia. His research interests include pattern recognition and remote sensing.

Goh Kah Ong received his Bachelor degree in Information Technology in year 2002, Master of Science in year 2004 and Ph.D degree in year 2013 from Multimedia University, Malaysia. His research interests include biometrics, image processing and pattern recognition.