Independent Component Analysis Architectural Feature Evaluation by Correlating Face Images

Independent Component Analysis Architectural Feature Evaluation by Correlating Face Images

Anu Singha, Mrinal Kanti Bhowmik*, **Prasenjit Dhar, and Anjan Kumar Ghosh**

Department of Computer Science and Engineering,
Tripura University(A Central University), Suryamaninagar - 799022, India
{anusingh5012,prasenjitdhar.cse}@gmail.com, mkb_cse@yahoo.co.in,
anjn@ieee.org

Abstract. It is subjective to predict which Independent Component Analysis (ICA) architecture is better for a particular face feature sets. Performance of ICA based feature recognition depends on two architectures ICA-I, ICA-II. Mathematically, it has been found that ICA-I architecture is based on vertically centered Principal Component Analysis (PCA) process and ICA-II architecture is based on horizontally centered PCA process. The paper presents observation and analysis of face feature vectors of two architectures using Pearson correlation coefficient. It has been found that ICA-II has a better independences property through correlation coefficient than ICA-I which may lead better classification performance.

Keywords: Independent component analysis, Architecture, Faces images, Correlation coefficient, Feature vector.

1. Introduction

Face recognition is a very challenging domain in pattern recognition research. Over last few decades face recognition attracts significant attention because of its various applications in biometrics, information security, etc. In face recognition research, some of the techniques (for example PCA, ICA, Linear Discriminant Analysis (LDA)) represent face as a linear combination of some basis images. The feature vectors from these techniques consist of the coefficients that are obtained by simply projecting facial images onto their basis images. Recently, a generalization of PCA, i.e., ICA received wide attention. PCA makes only the data as uncorrelated, but ICA makes data as uncorrelated as well as independent. There are two arguments are needed for ICA based face recognition and face representation. The first argument is that higher order relationships among the image pixels contain information, which is an important task in face recognition. The second argument is that ICA helps find a direction where the projections of data have maximally non-Gaussian distribution.

Bartlett et al. [1] were among the first to apply ICA to face identification task. They used the Infomax algorithm [5] to employ ICA and recommended two ICA based architectures. Both architectures were evaluated on a subset of the FERET database along with PCA, and claims that the two ICA based architectures were equally powerful and both outperformed the PCA. Liu and Wechsler [6] also used FERET database to study the comparative assessment of ICA

*Corresponding author.

.

ISBN: :: doi:

performance through Comon [3] ICA algorithm, and claims that ICA outperform PCA, while other researchers reported differently. Socolinsky et al. [7] has presented a comprehensive analysis of multiple appearance-based face recognition methodologies on visible and long wave infrared (LWIR) imagery. They have observed that LWIR imagery of human faces is not only a valid biometric, but almost surely a superior one to comparably visible imagery. And it is also reported that ICA shows better performance in case of visible images, and PCA shows better in case of infrared images. Draper et al. [4] has compared the PCA and ICA architectures over FERET database in the context of a baseline face recognition system, a comparison is motivated by contradictory claims in the literature. They are able to show that ICA architecture II provides the best results, followed by PCA with L1 or Mahalanobis distance metrics. ICA architecture I and PCA with the L2 or cosine metrics are poorer choices for recognition task. They recommended the FastICA algorithm for ICA architecture II, although the difference between FastICA and Infomax is not large. In recognizing facial actions, the recommendation is reversed: find the best result using Infomax to implement ICA architecture I. Jian Yang et al. [8] also re-evaluated ICA architectures and PCA on the FERET, ORL, AR face databases, and claims as similar to Drapper *et al.*[4]. They construct two PCA baseline algorithms to re-evaluate ICA-based architectures, but observed no significant performance difference between ICA I (II) and PCA I (II).

In the above literature reviews, it has been noted that many of the authors evaluated ICA according to some criterion, for example: image pre-processing, ICA pre-processing steps, effect of different ICA algorithms, distance metrics, different types of images, and so on. As a consequence, it is complicated to predict the best ICA architectures for a particular domain. So, the paper represents a mathematical analysis of feature vectors on two ICA-based architectures and finds out that ICA-II architecture shows potential act than the ICA-I architecture. The paper has analyzed two ICA-based image representation architectures and finds that ICA-I architecture involves a vertically centered data matrix while ICA-II architecture involves a horizontally centered data matrix. After the analysis of the face feature vectors, it has been seen that ICA-II is better compared with ICA-I based on correlation technique applied on feature face vectors.

2. Independent Component Analysis (ICA)

The Independent Component Analysis (ICA) is a computational method for a linear transformation that separates the statistically dependent mixture components into subcomponents such that each subcomponent is statistically independent from each other [2]. To present a rigorous definition of ICA using vector-matrix notation, the mixing model is written as

$$X = AS. \tag{1}$$

The statistical model in Eq. (1) is called ICA model. Assume that mixture vector X as well as independent component vector S are random vectors, where both the mixture vector and the independent component vector have to be centered (zero mean) and unit variance. The mixing matrix A and independent component vector are considered to be unknown, and matrix A should be square pattern. Also assume that the independent component vector S must have a non-Gaussian distribution. In statistical ICA model, the independent components must have a non-Gaussian distribution. This non-gaussianity is an essential and important principle in ICA estimation. In fact, without non-gaussianity the estimation is not probable at all. Till all the observation is about random vector X and task is to estimate both unknown A and S using X. If anyhow estimation of the unknown mixing matrix A is possible, then compute its inverse measured as W, and obtain the independent component vector S simply by Eq. (2).

$$S = A^{-1}X = WX \tag{2}$$

The estimation of unknown matrix W is possible through FastICA algorithm by initializing random vectors. FastICA is a popular algorithm for independent component analysis invented by Aapo Hyvärinen at Helsinki University of Technology [5], and is based on a fixed-point iteration scheme. FastICA for one unit algorithm only is focused on one independent component i.e. using one-unit algorithm of FastICA proceedings of one independent component is only possible. To estimate the multiple number of independent components, the one unit algorithm must be run in iteration of n with different weight vectors as $w_1, ..., w_n$. When the multiple random vectors (say $w_1, ..., w_n$) are estimated, there is a high probability of correlation between the vector values of each other. To prevent different random vectors from converging to the same maxima, it is necessary to de-correlate or orthogonalize the outputs $w_1^T X, ..., w_n^T X$ after every iteration. This de-correlation is done by Gram-Schmidt-Orthogonalization [5].

2.1. *Two Architectures for Performing ICA on Face Recognition*

The goal of ICA two architectures is to find the feature vectors of face images in compressed form. In the experiment each face image is to be organized of as long column vectors with as many dimensions as number of pixels in the image. Regardless of the algorithm that is used, ICA for face recognition will generally operate within one of two different architectures, Architecture I or Architecture II.

2.2. *Architecture I*

In ICA-I architecture, face datasets are to be organized into a matrix where each row vector is a different face images. This approach is illustrated in Fig. 1. In this approach, images are random variables and pixels are trials. In Fig. 1, data matrix X (also called mixture matrix) is the combination of n independent face images. Coefficient matrix W (also called unmixing matrix) is calculated through execution of the FastICA [5] algorithm where coefficient matrix W should be square pattern, and S is the source matrix of n independent basis face images.

In the ICA-I architecture, consider n training samples (image column vectors of length m) to represent a column data matrix $X = (x_1, x_2, \ldots, x_n)$ of dimension $m \times n$. Transposing of data matrix X i.e. $Y = X^T$ is used as a mixture row data matrix of ICA-I architecture. Y is denoted as $Y = (y_1, y_2, \ldots, y_m)$. In centering step, mean vector $\mu_I = \frac{1}{m} \sum_{j=1}^{m} y_j$ is estimated from data matrix Y and this mean vector subtracted from every column vector of data matrix Y i.e. $(y_j - \mu_I) \to \bar{y_j} \to \bar{Y_v}$. $\bar{Y_v}$ denotes the vertically centered row data matrix in the sense that every time the mean vector is subtracting from each column vector and each column vector contains one pixel of each images at a time. Now $\bar{Y_v}$ is a zero-mean image row data matrix where each row is an original image from whose elements the mean of each image has been

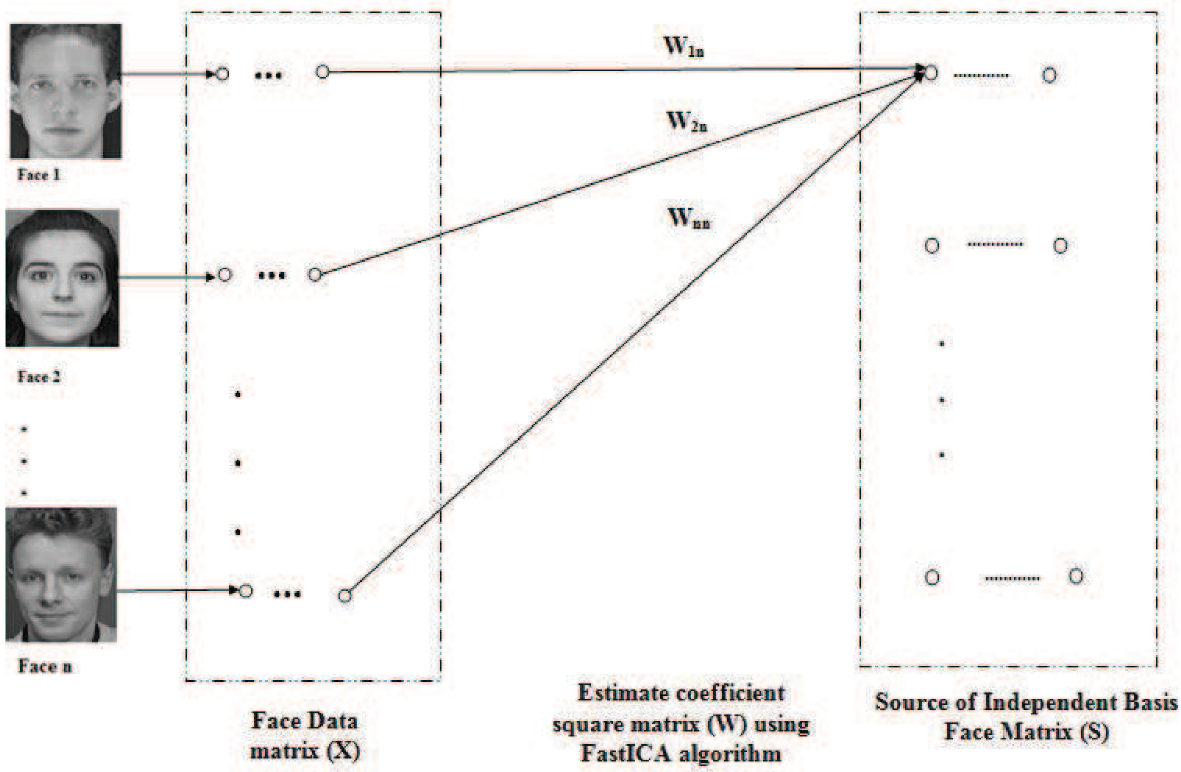


Figure 1: Independent Component Analysis Architecture I.

removed. In next step, whiten the centered data matrix based on the PCA technique. Calculate the orthonormal eigenvectors $V = (v_1, v_2, \ldots, v_n)$ of covariance matrix $\sum_I = \frac{1}{m}\sum_{j=1}^{m} y_j y_j^T$ corresponding to the first p largest positive eigenvalues as $\gamma_1 \geq \gamma_2 \geq ... \geq \gamma_p$. The whitening matrix can be obtained by formulating as $H = VD^{-\frac{1}{2}}$, such that $H^T \sum_I H = I$, where D is the diagonal matrix of p largest positive eigenvalues. After that the data matrix $\bar{Y}_v$ can be whitened using $\tilde{Y} = H^T \bar{Y}_v$ transformation. These preceding steps are the most essential steps of ICA pre-processing. Now ICA will be performed on whitened data matrix H and producing a matrix S_I of p independent basis images in its rows, that is, $S_I = W_I \tilde{Y}$ where W_I is the coefficient matrix of ICA. The row vectors of S_I correspond to the basis images of ICA-I architecture. Finally, the vertically centered data matrix projecting onto these p independent basis images as $Z = \bar{Y}_v {S_I}^T = \bar{Y}_v \tilde{Y}^T W_I$. Each row of Z which represents the feature vectors, is used to represent the compression form of face images for recognition purposes.

2.3. *Architecture II*

In ICA-II architecture, an image data matrix X in which each column vectors consider as different face images is to be organized. So the input face data matrix X is transposed from ICA-I i.e. the pixels are variables and the images are trails, as shown in Fig. 2. Now, each row

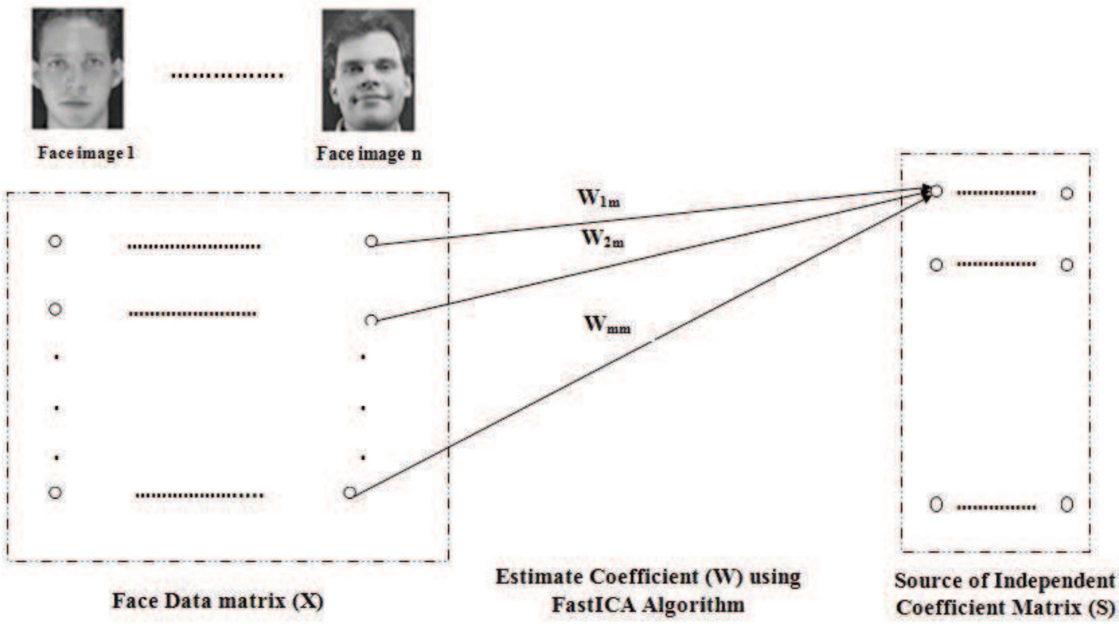


Figure 2: Independent Component Analysis Architecture II.

of the learned weight coefficient matrix W is the basis images, and the statistically independent coefficients that comprise the input images are recovered in the columns of S.

Consider n training samples of length m i.e. each image will take place in the column of data matrix X like $X = (x_1, x_2, \ldots, x_n)$ of dimension $m \times n$. In centering step, mean vector $\mu_{II} = \frac{1}{n}\sum_{j=1}^{n} x_j$ is subtracting from each column vector of data matrix X i.e. $(x_j - \mu_{II}) \to \bar{x}_j \to \bar{X}_h$. $\bar{X}_h$ denoting the horizontally centered column data matrix in the sense that every time the mean vector is subtracting from each column vector and each column vector contains all pixels from one image at a time. Now $\bar{X}_h$ is a zero mean image column data matrix where each column is an original image from whose elements the mean of all image samples has been removed. The illustration of the rest of mathematical basis for ICA-II is roughly analogous to ICA-I. In this architecture, the independent coefficients are recovered in the columns of S_{II} as $S_{II} = W_{II}H^T$. Finally, the horizontally centered data matrix project onto these independent coefficient matrix as $Z = S_{II}\bar{X}_h$. Each column of Z representing the feature vectors, is used to represent the compression form of face images for recognition purposes.

3. Observation and Analysis

The primary task of this paper is to conclude a decision that which one of ICA architectures (ICA-I and ICA-II) is better using correlation technique on feature vectors. When the correlation between two datasets is spread widely, that indicates a weak relationship between the data, and close spread indicates the strong relationship. The correlation value always lies in between +1 to -1. The high relationship value is towards 1 and sign is indication of either

Figure 3: Sample face images from the ORL database.

positively or negatively correlated. This correlation value which can be estimated by Pearson correlation coefficient R^2, is defined as $R^2 = \frac{Cov_{xy}}{\sigma_x \sigma_y}$ where Cov_{xy} indicates the covariance between dataset x and y, and σ_x, σ_y indicates the standard deviations of x and y respectively. The performance of the two architectures depends on the correlation and variance property. These two properties make the sense of independency. The relationship within the datasets i.e. variance should be very strong (nearby one), and the relationship between the datasets should be poor i.e. uncorrelated.

In the experiment and analysis, we have collected 128 face images from ORL face database. These 128 face images have been processed thorough ICA-I and ICA-II, and corresponding 128 feature vectors have been extracted. Total 15 features have been chosen for each feature vector. But in this paper, out of 128 feature vectors only 5 feature vectors have been shown in Table 1. The corresponding face images of these 5 feature vectors have been shown in Fig. 3.

In Table 1, it has been observed that the variation of values within the feature vectors (say, #1 column vector itself) of ICA-I are very large than the variations of ICA-II. So, it is tough to

Table 1: Some (*feature vectors*) from both architectures. (# indicates Feature Vectors.)

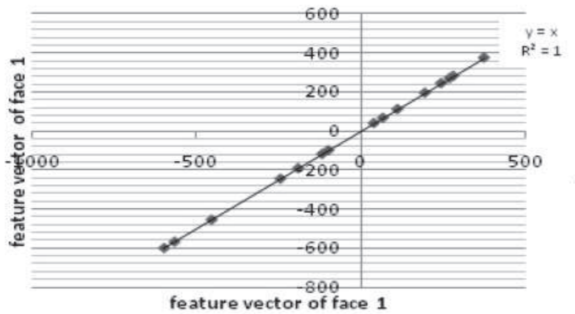
ICA-I					ICA-II				
#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
373.92	313.14	855.53	827.95	604.78	0.29	0.11	0.34	0.12	0.08
280.48	−14.31	253.63	−19.91	483.97	0.17	−0.52	−0.13	0.08	0.33
38.60	55.39	6.90	75.63	−403.29	0.06	−0.14	−0.02	−0.15	−0.32
−246.21	−301.43	250.50	−12.59	42.65	−1.18	−0.91	0.17	−0.88	1.05
−119.28	−50.06	−96.32	−81.72	15.64	−1.22	−0.81	0.71	1.33	0.42
−568.98	−471.67	−332.36	−355.82	−447.34	−0.95	−0.78	1.17	1.21	−1.11
269.22	397.31	147.47	33.02	234.37	0.11	−0.35	0.94	0.05	0.21
−456.10	−440.80	361.34	386.27	217.56	−3.38	−4.04	0.12	−0.50	0.22
−192.06	−43.94	−184.12	47.54	−364.94	0.42	−0.14	0.21	0.55	0.92
194.33	134.58	7.09	128.99	25.68	−0.17	0.06	−0.98	−0.55	−1.42
110.28	51.26	−39.46	−152.43	132.96	0.55	−0.4	−1.16	−2.91	−0.16
−100.33	−130.74	17.49	−59.82	48.21	0.57	−0.59	0.67	0.35	0.82
−600.98	−727.88	−262.24	−245.19	−190.93	0.35	0.37	1.86	2.39	−0.91
243.28	32.92	−405.61	−536.61	−108.31	−0.27	0.65	0.35	−0.03	1.58
66.16	10.32	−14.04	−20.27	167.86	0.23	−0.26	−0.89	−0.30	0.80

measure the correlation property of the feature vector analysis. Although the values within the feature vectors of ICA-I / II are large/ small variations, their scatter plot is shown in Figs. 4(a) and 4(c). Figs. 4(a) and 4(c) are representing the correlation within (i.e. variance) the same feature vector of face image 1 in ICA-I and ICA-II. From both figure, it has been observed that they follows a linear equation with slope of 1, and the points are scattered very closely along this linear equation. This close scattering is an indication of correlation coefficient values of 1 in Figs. 4(a) and 4(c), and is showing a very strong positive correlations in both the architectures.

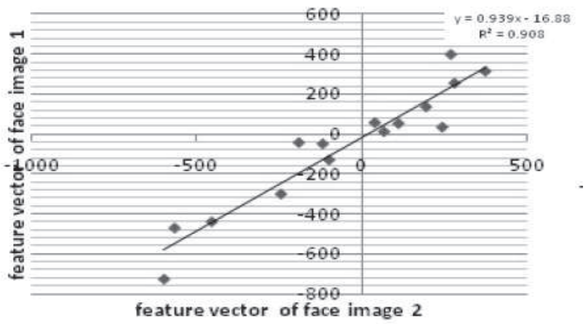
In Table 1, it has also been observed the similar variations between the values of feature vectors from (#1 - #2, #1 - #3, #1 - #4, #1 - #5, #2 - #3, #2 - #4, #2 - #5, #3 - #4, #3 - #5, #4 - #5) both architectures have also been observed. Fig. 4(b) and 4(d) represent the scatter plot correlation between the feature vectors of face image 1 and face image 2 in ICA-I and ICA-II. In Fig. 4(b) the equation means that the intercepting point of y is -16.88 and the slope of the line is 0.939. If the value of x is increased by 1 then the slope of the line increases by 0.939, and if the value of x is increased by 2 then the slope value 0.939 increases twice. The correlation coefficient value of 0.908 in Fig. 4(b) is showing a strong positive correlation between feature vectors of the face image 1 and face image 2 in ICA-I. This strong relationship can also be noticed from the scattering of the data points where the data points are scattered closely along the regression line. But this scatter is not very close in case of ICA-II and not plotted along the regression line in Fig. 4(d). The correlation coefficient value of 0.684 in Fig. 4(d) is showing a comparably weak positive correlations in ICA-II than ICA-I.

Table 2: Comparisons of the One Feature Vector with Other Feature Vectors corresponding to Table 1.

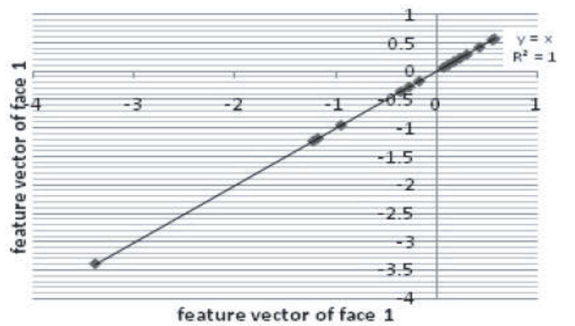
Sl. No.	Feature Vector (#)	Feature Vector (#)	Correlation Coefficient R^2 (ICA-I)	Correlation Coefficient R^2 (ICA-II)
1.	# of image 1	# of image 1	1.000	1.000
2.	# of image 1	# of image 2	0.908	0.684
3.	# of image 1	# of image 3	0.101	0.035
4.	# of image 1	# of image 4	0.074	0.010
5.	# of image 1	# of image 5	0.323	0.002
6.	# of image 2	# of image 2	1.000	1.000
7.	# of image 2	# of image 3	0.113	0.002
8.	# of image 2	# of image 4	0.111	0.031
9.	# of image 2	# of image 5	0.259	0.000
10.	# of image 3	# of image 3	1.000	1.000
11.	# of image 3	# of image 4	0.865	0.673
12.	# of image 3	# of image 5	0.500	0.005
13.	# of image 4	# of image 4	1.000	1.000
14.	# of image 4	# of image 5	0.340	0.026
15.	# of image 5	# of image 5	1.000	1.000



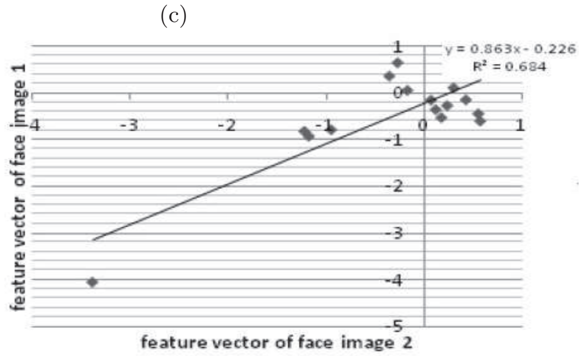
(a)



(b)



(c)



(d)

Figure 4: Correlation of feature vectors from both [(a), (b)] of ICA-I [(c), (d)] of ICA-II.

Other comparative correlation coefficient results of the feature vectors are listed in Table 2. From row 1, 6, 10, 13, 15, the very strong relationship within feature vectors (i.e. variance) has been observed. But it has also been noticed from row 2, 3, 4, 5, 7, 8, 9, 11, 12, 14 that the correlation coefficient values of ICA-II are less as compared to ICA-I. In some cases of ICA-II, coefficient values (row 5, 7, 9, 12) tend to zero (0.002, 0.002, 0.000, and 0.005 respectively). Therefore it has been found that ICA-II has a better independence property through correlation coefficient than ICA-I which may lead to better classification performance.

4. Conclusion and Future Work

Prediction of best ICA architectures is complicated because evaluation of the architectures of many authors are subjective in the sense that evaluation is according to some criterion, for example, image pre-processing, ICA pre-processing steps, effect of different ICA algorithms, distance metrics, and so on. It has been found that ICA-I architecture involves a vertically centered data matrix while ICA-II architecture involves a horizontally centered data matrix. So, the paper represents a mathematical analysis of feature vectors on two ICA-based architectures and finds out that ICA-II Architecture shows potential act than the ICA-I Architecture.

To verify this mathematical assumption, the analyses are being extended using four face databases- the FERET, ORL, CVL, and YALE databases with different expressions, illuminations, and positions based on recognition rate.

Acknowledgments. The work presented here is being conducted in the Biometrics Laboratory and Bio-Medical Infrared Image Processing Laboratory of Department of Computer Science and Engineering of Tripura University (A Central University), Tripura, Suryamaninagar-799022. The research work was supported by the Grant No. 12(2)/2011-ESD, Dated 29/03/2011 from the DeitY, MCIT, Government of India and also supported by the Grant No. BT/533/NE/-
TBP/2013, Dated 03/03/2014 from the Department of Biotechnology (DBT), Government of India. The authors would like to thank Professor Barin Kumar De, Department of Physics, Tripura University (A Central University) and Dr. Debotosh Bhattacharjee, Associate Professor, Department of Computer Science & Engineering, Jadavpur University for their kind support to carry out this research work.

References

1. Bartlett, M.S., Lades, H.M., Sejnowski, T.J.: Independent component representations for face recognition. In: Proceedings of the SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III, San Jose, California (1998)
2. Bell, A., Sejnowski, T.: An information maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)
3. Comon, P.: Independent component analysis: A new concept? Signal Process. 36, 287–314 (1994)
4. Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J. R.: Recognizing faces with PCA and ICA. Comput. Vis. Image Underst. 91, 115–137 (2003)
5. Hyvärinen, A.,Oja, E.: Independent component analysis: Algorithms and applications. Neural Netw. 13, 411–430 (2000)

6. Liu, C., Wechsler, H.: Comparative assessment of independent component analysis (ICA) for face recognition. In: International Conference on Audio and Video Based Biometric Person Authentication, Washington (1999)
7. Socolinsky, D.A., Selinger, A.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. International Conference on Pattern Recognition, Quebec City (2002)
8. Yang, J., Zhang, D., Jing-Yu, Y.: Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance. IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics 37 (2007)