CrossMark

On the complex domain deep machine learning for face recognition

B. K. Tripathi[1]

Abstract Biometric based verification and recognition has become the center of attention for many significant applications for security conscious societies, as it is believed that biometrics can provide accurate and reliable identification. The face biometrics are one that possesses the merits of both high accuracy and low intrusiveness. An efficient machine recognition of human faces in big dataset is both important and challenging tasks. This paper addresses an intelligent face recognition system that is pose invariant and can recognize multi-expression, occluded and blurred faces through efficient but compact deep learning. Superior functionality of neural network in a complex domain has been observed in recent researches. My work presents a new approach, which is the fusion of higher-order novel neuron models with multivariate statistical techniques in a complex domain with a sole goal of improving performance of biometric systems. This also aims at reducing the computational cost and providing a faster recognition system. This paper presents the formal algorithms for feature extraction with multivariate statistical techniques in complex domain and compare them their real domain counterpart. This paper also presents a classifier structure (OCON : One-Class-in-One-Neuron) which contains an ensemble of novel higher order neurons, which drastically reduces the complexity of proposed learning machine because only single neuron is sufficient to recognize a subject in the database. This novel fusion in the proposed deep learning machine has thoroughly presented its superiority over a wide spectrum of experiments. Advanced deep learning capabilities, and complex domain implementation in particular, are significantly advancing state-of-art in computer vision and pattern recognition.

1 Introduction

ANN is an attempt to mimic the biological brain, which has evolved to its present form over thousands of years to achieve optimized results. It gives computers the ability to learn things similar to the way we do using software instructions. In fact, this direction has prompted an active development in cognitive science, intelligent system, bioinformatics, computer vision and biometric applications. Biometric based intelligent system identifies a person by biological characteristics, which are difficult to forge [1, 18, 40, 44]. It also increases the user-friendliness in human-computer interaction. The facial features have several advantages over other biometric attributes considered by Hietmeyer [18]. It is natural, passive, and non intrusive. The human face recognition is one of the distinguished and inherent capabilities of the human visual system. Though, humans do it effortlessly in spite of variations in features but it is not easy in computer vision terms. However, the use of deep learning makes this task easy. Deep learning envisaged as a learning system composed of multiple level of non linear operations. Deep machine learning is based on a set of algorithms, which attempt to model higher level abstractions in data by using multiple processing layers with complex structures. This paper presents the

✉ B. K. Tripathi
abkt.iitk@gmail.com

[1] The Nature-Inspired Computational Intelligence Research Group, HBTU Kanpur, 208002, India

model of layer features with an unsupervised mode followed by a supervised mode. Hence, a good trade off between unsupervised and supervised learning is an interesting but challenging task, because it utilizes techniques for feature extraction and classification respectively. The success of machine face recognition techniques depends heavily on the particular choice of feature extractor and classifier [22, 39]. Complex-valued neural networks (CVNN) have presented the second generation of development in neural network [19, 26, 43]. The performance of complex-valued neural networks has scored over real-valued neural networks, even on real-valued problems [4, 15, 31, 42]. Therefore, the strong need for reliable recognition for a large dataset has demanded an important role of a novel synergism of feature extractor and classifier in the complex domain for deep learning.

Over the years, the researchers have contributed a number of methods for feature extraction and classification in the real domain. Few researchers have tried multivariate statistical techniques in complex domain, like complex principal component analysis ($\mathbf{C}PCA$) for 2-D vector field analysis [36] and complex independent component analysis ($\mathbf{C}ICA$) for performing source separation of functional magnetic resonance imaging data [12, 29]. Few attempts have been made to use these techniques for analyzing complex variables or complexified real data. In this paper, I developed formal algorithms for feature extraction using these unsupervised statistical techniques in complex domain and figure out their technical benefits over conventional ones. An efficient solution for many real-valued problems has been achieved with compact network topology of complex-valued neural networks (CVNN) in comparison to real-valued neural networks (RVNN) [19, 26]. The 2D structure of the error propagation method of CVNN reduces the problem of saturation in learning and offers faster convergence [43]. They are more efficient, fault-tolerant, and less sensitive to noise. The new insights available from neuroscience have presented non-linear neuronal activities in a neuron cell body [28]. It motivated us to design higher-order non-linear neurons, which will serve as a basis for the construction of powerful classifiers. Efficient learning and better precision in classification have been yielded by complex-valued higher order neurons. In this paper, I also propose a compact but efficient classifier using higher-order neurons in the complex domain. The success of machine face recognition is limited by variations in facial features imposed by real life situations [35]. These may be due to camera distortion, acquisition in an outdoor environment, different noises, complex background, occlusion, facial expression, illumination and presence of beard, mustache, spectacles, etc. A deep learning system presented in this paper by the combined efforts of considered feature extractors and higher order neurons in the complex domain yields better class distinctiveness, low complexity and high recognition performance even in a natural environment.

Face recognition is one of the most effective applications of image processing and biometric systems. Over the years, researchers have contributed a number of methods for feature extraction and classification. A general review of different literature studies and systems are recently presented in [22, 30] with their strengths and limitations. Most of the work is related to real domain. Apart from them, researchers have tried multivariate statistical techniques in complex domain, like complex-PCA for 2-D vector field analysis [36] and complex-ICA for performing source separation of functional magnetic resonance imaging data [12, 29]. Few attempts have been made to use these techniques for analyzing complex variables or complexified real data. In this paper, I present formal procedures for facial feature extraction using these unsupervised statistical methods in the complex domain and figure out their technical benefits over conventional methods.

Section 2 presents the formal development of a novel intelligent system in the complex domain. This Section is dedicated to formally present the algorithms for complex PCA and complex ICA based feature extraction for large image data set. This section further presents a new recognizer (classifier) based on higher order neuron. The considered single TION1 or TION2 neuron in a complex domain allows the designing of a host (classifier) for each subject. It also presents the classifier structure and its learning rules. Section 3 addresses the performance evaluation through simulation results over three different face data sets. In order to ensure the effective recognition capabilities of considered techniques in a real-world environment, their performance is assessed with variety of blurred and partial occluded images. Finally, this investigation concludes with inferences and discussions in the Section 4.

2 Deep machine learning in complex domain

In this paper, I present a deep learning based intelligent system in the complex domain. The novel fusion of proposed feature extractor and classifier yields a compact recognition system with efficient machine learning. A fundamental problem in digital image processing is to find a suitable representation for image, audio-video or other kind of data. Unsupervised approaches like principal component analysis (PCA) independent component analysis (ICA) find a set of basis images and yields a new representation as a linear combination of these basis images. They reduce the amount of data in the new representation while maintaining sufficient information for meaningful training. PCA aims at the decomposition of a linear mixture of independent source signals into uncorrelated components using only

second order statistics [13], however, higher order relationships still appear in the joint distribution of basis images. In typical pattern classification problems, the significant information may be contained in the higher order relationships among pixels. So, it will be desirable to use methods which are sensitive to these higher order statistics to get better basis images. The assumption of Gaussian sources are also implicit in the PCA, which makes it inadequate, because in the real world, the data often does not follow a Gaussian distribution. ICA is a method for transforming multidimensional random vectors into its components that are both statistically independent and non-Gaussian [20, 25]. This transformation brings out the essential features of image data and makes them more visible or accessible. Therefore, it is expected that these data of reduced dimension will be rich in features.

The real-valued neural network (RVNN) based approaches have been widely applied in [8, 14, 38] for face recognition due to its intrinsic generalization ability and robustness toward ill defined pattern classes. However, one drawback in this supervised approaches is that a huge network architecture has to be extensively tuned (number of layers, number of nodes, weight initialization, learning rates etc.) to get exceptional performance. Thus a reduction in network complexity of neural network is highly demanding. Second generation neural networks like CVNN have thoroughly proved their superiority over conventional ones for classification. As a result, we can not only decrease the computational complexity of an intelligent system, but can also increase the recognition rate using second generation neurocomputing approaches.

For pattern classification, the structure of neural networks can broadly be categorized into two ways viz All-Classes-in-One-Neural Network (ACONN) and One-Class-in-One-Neural Network (OCONN) [8]. In ACONN, one single network (classifier) having as many outputs as the number of classes is designed to classify all the classes (subject) in the databases. In OCONN, an ensemble of ANN is used, where each network is dedicated to recognizing one particular class in the database. It has been observed in literature [3, 16] that OCONN classifier achieves higher performance and is more realistic for large datasets.

The complex domain neural networks are becoming very attractive because of problem of same complexity, they need smaller network topology and lesser training time to yield better accuracy in comparison to equivalent neural networks in a real domain [4, 41–43]. A higher order neuron in complex domain further boost up the capability of the classifier. The face recognizer (classifier) presented in this paper is designed with an ensemble of novel higher-order neurons in a complex domain, where each neuron is dedicated to recognizing one particular class in the database. Thus, proposed classifier is a One-Class-in-One-Neuron (OCON).

2.1 Feature extraction

In most of the applications an image, x, is available in the form of single or multiple views of 2D (p by q) intensity data (ie. pixel values). Thus, inputs to the face recognition system are visual only.

$$x = \{a_k : k \in S\}, \tag{1}$$

where S is a square lattice. Sometimes, it is more convenient to express a face image matrix as a one dimensional vector of concatenated rows of pixels, then an image vector

$$x = \{a_1, a_2, ...a_N\}, \tag{2}$$

where $N = p \times q$ is the total number of pixels in an image. Such a high dimensional image space is usually inefficient and lacks of discriminative power. Therefore, we need to transform 'x' into a feature vector or new representation which greatly reduces the facial feature dimensions, yet maintains reasonable discriminative power by maximizing the spread of different faces within the image subspace. Let $X = \{x_1, x_2, ...x_M\}^T$ be the M by N matrix of image data, M is the number of face images in the training set. Let vector $avg = \frac{1}{M} \sum_{k=1}^{M} x_k$ be the mean of training images. The goal in considering feature extraction techniques is to find a useful representation of a face by minimizing the statistical dependence among the basis vectors.

2.1.1 Feature extraction with complex principal component analysis

Principal component analysis is widely applied for dimensionality reduction and feature extraction on the basis of extracting the preferred number of principal components of multivariate data [2, 27, 45]. If image elements are considered to be random variables and images are seen as a sample of a stochastic process, the PCA basis vectors (eigenfaces) are defined as the eigenvectors of the covariance matrix. They describe a set of axis within the image space and most of the variance of face images are along these axis. The associated eigenvalues define the degree of spread (variance) of the image population along these axis. Eigenvectors corresponding to the higher eigenvalues carry significant information for representation, which best accounts for the distribution of images within the entire image space. The new representation of an image is generated by projecting that image onto eigenfaces.

The established technique of real PCA (**R**PCA) for extracting features of the image data has been extended to proposed complex principal component analysis (**C**PCA). The transformed data in complex domain are obtained from the original data and their Hilbert transform. The amplitude of each spectral component is unchanged, but each

component's phase is advanced by $\pi/2$ [17] in Hilbert transformation. Complex principal components are determined from the complex cross-correlation matrix of image data matrix Z. The $\mathbf{C}PCA$ algorithm for feature extraction from the image data set can be stated as follows:

1. Collect the images in the data matrix X (M by N). Find the mean subtracted data matrix, $A = X - avg$.
2. Determine a complex image data matrix Z, using Hilbert transform.
3. Determine the cross-correlation matrix $\mathcal{C} = Z^{\aleph} Z$, where $Z^{\aleph}$ is the complex conjugate transposition of Z.
4. Compute eigenvectors of $\mathcal{C}$. Even for the moderate size of an image ($N = p \times q$), the dimension of $\mathcal{C}$ will be $pq \times pq$. Hence, calculations will be intractable and computationally extensive.
5. This issue can be circumnavigated by considering the eigenvectors v_i of $Z Z^{\aleph}$, such that $Z Z^{\aleph} v_i = e_i v_i$. Vector v_i is of size M and there are M eigenvectors. Thus computations are significantly condensed from the order of number of pixels (N) in the image to the order of number of images (M) in the training set, $M << N$.
6. Eigenvectors (basis images) from the cross-correlation matrix $\mathcal{C}$ will be $Z^{\aleph} v_i$ (N by M), such that $Z^{\aleph} Z Z^{\aleph} v_i = e_i Z^{\aleph} v_i$.
7. Since, $\mathcal{C}$ is a Hermitian matrix, all its eigenvalues will be real. Thus, greater reduction can be obtained by letting those eigenvectors with large eigenvalues. This subspace capture most of the variation of the training set with a fewer number ($M' < M$) of eigenfaces.
8. Project each input vector onto basis vectors to find a set of M' coefficients that describe the contribution of each vector in the subspace.
9. Therefore, the input vector of size N is condensed to a new representation (feature vector) of size M', that will be used for classification.

2.1.2 Feature extraction with complex independent component analysis

Let face images in X to be a linear mixture of statistically independent basis images S, combined with an unknown mixing matrix A, such that X = A S. ICA tries to find out the separating matrix W such that U = W X. ICA, as an unsupervised learning algorithm learns weight matrix W, which is used to estimate a set of independent basis images in the rows of U. W is roughly the inverse matrix of A. The two preprocessing steps viz Compression and Whitening simplify the estimation of W.

It will be quite useful to reduce the dimension of the data at the same time whitening is done. First, apply PCA on the image matrix X, to reduce the number of data to a tractable number. Pre-applying PCA does not throw away the higher-order relationships, they still exist but not separated. Well, it enhances ICA performance by discarding the eigenvectors corresponding to trailing eigenvalues, which tend to capture noise. This has often the effect of reducing the noise. Let matrix $\mathcal{E}^{\mathcal{T}}$ (M' by N) contains first M' eigenvectors of M face images in its row, then $U = W\mathcal{E}^{\mathcal{T}}$. Another important preprocessing step in ICA is to whiten the observed variables. This step transforms the observed vector linearly so that the obtained new vector is white, i.e. Whitening process transforms the observed vector into a new vector, whose components are uncorrelated and their variances equal unity. The whitening matrix is obtained as $W_w = 2 \times (COV(X))^{-(1/2)}$. Instead of performing ICA on the original images, it should be carried out in a compressed and whitened space, which reduces the number of parameters to be estimated but most of the representative information is preserved. Therefore, a full transformation matrix will be the product of whiten matrix and matrix learned by ICA. The whitening is a simple and standard strategy, much simpler than any ICA algorithms, thus it is better to reduce the complexity of the problem before applying ICA.

The independent component analysis in the complex domain ($\mathbf{C}ICA$) has been used for source separation of complex-valued data, such as fMRI [12], EEG [5] and communication data, yet this concept is not well developed hence demanding more applications. There are various prominent techniques for capturing independent components [7, 23, 25, 33], they have their own strength and weakness [32]. One of the technique, inspired by information theory, for ICA estimation is minimization of mutual information between random variables. I have formulated a complex feature extraction algorithm, $\mathbf{C}ICA$, using Bell and Sejnowski infomax method in real domain [6]. Mutual information is a natural measure of independence between random variables (sources). Infomax algorithm uses it as a criterion for finding the ICA transformation by iteratively optimizing a smooth function (Entropy), whose global optima occurs when the output vectors ($u \in U$) are independent. Finding independent signals by maximum entropy is known as infomax [7]. The CVNN has proved its superiority over RVNN [41, 42].

In this paper, the basic concepts of $\mathbf{R}ICA$ algorithm has been utilized to build the $\mathbf{C}ICA$ algorithm for image processing and vision applications. The CVNN has proved its superiority over RVNN [41, 42]. The $\mathbf{C}ICA$ has been derived from the principle of optimal information transfer through complex-valued neurons based on non-analytic but bounded activation function [26, 31, 42], defined as:

$$f_{\mathbf{C}}(V) = f(\Re(V)) + j \times f(\Im(V)) \qquad (3)$$

The stimulus of taking this function in $\mathbf{C}ICA$ algorithm is that it can lead the fairly accurate joint cdf of the source distribution [10, 11]. The apparent problem in function (3)

comes from the fact that it is real-valued and therefore is not complex differentiable unless it is a constant. The differentiation of $f_\mathbf{C}$ can be conveniently done [9, 37] without separating real and imaginary parts with following complex (partial) differential operator:

$$f'_\mathbf{C} = \frac{\partial f_\mathbf{C}}{\partial z} = \frac{1}{2}\left(\frac{\partial f_\mathbf{C}}{\partial z_\Re} - J\ \frac{\partial f_\mathbf{C}}{\partial z_\Im}\right) \tag{4}$$

In $\mathbf{C}ICA$ based feature extraction, we assume that the image data X is a linear mixture of M statistically independent complex-valued sources S, then Y = $f_\mathbf{C}$ (U = W X), where U, W $\in$ **C**. The basic steps in the $\mathbf{C}ICA$ algorithm for feature extraction in image database can be summarized as follows

1. Let image data matrix X possesses M images in rows and N pixels in a column.
2. Apply $\mathbf{R}PCA$. The PCA basis vectors in $\mathcal{E}^\mathcal{T}$ are analyzed for independent components, where $\mathcal{E}$ (N by M′) be the matrix of M′ eigenvectors.
3. The final transformation matrix will be the product of whitening matrix and optimal unmixing matrix.
4. The sources are reproduced as complex random vectors. Let take sigmoidal complex function $f_\mathbf{C}$, defined in (3), as joint cdf of source signals.
5. Develop a contrast function h from the point of view of CVNN. Perform maximization of joint entropy $h(Y)$. This can be achieved using an extension of the natural gradient in a complex domain.
6. Apply complex infomax on PCA basis X. Thus, the optimal matrix W such that : MAX [h {$f_\mathbf{C}$ (WX)}], this can be done as:

   – Define a surface h {$f_\mathbf{C}$ (WX)}
   – Obtain gradient ∇h with respect to W and ascent it, $\Delta W \propto \nabla h$, then -

   $$\Delta W = \eta\left(\left[W^\mathcal{T}\right]^{-1} + \frac{1}{M}\sum_{i=1}^{M}\frac{f''_\mathbf{C}(u)}{f'_\mathbf{C}(u)}\ X^\mathcal{T}\right),\ u \in U$$

   – When h is maximum or gradient magnitude converges to zero, W is W_{OPT}. Done !

7. The process of maximizing the joint entropy of outputs also minimizes the mutual information between the basis images in U, ie individual outputs. This also presents that how much the extracted signals in U are close to being independent. Thus, $\mathbf{C}ICA$ algorithm produces transformation matrix $W_t = W_{OPT} \times W_w$, such that $U = W_t\ \mathcal{E}^\mathcal{T}$.
8. Let the PC representation of images in X be R = X $\mathcal{E}$, also X = R $\mathcal{E}^\mathcal{T}$. The assumption that W_t is invertible,

we get $\mathcal{E}^\mathcal{T} = W_t^{-1}\ U$. Hence $X = R\ W_t^{-1}\ U$. Estimation of IC representation of images is thus based on the independent basis images in U.

9. The IC representation of a face image is a row in matrix $B = R\ W_t^{-1}$. Thus, statistically independent feature vectors of images have been obtained in B.

2.2 OCON: the classifier

The second layer of deep learning machine involves the supervised training of proposed classifiers for estimation of learning parameters (weights), which are stored for future testing. Hence, it is most desirable to search a structure for classifier which requires minimum weights and yields the best accuracy. The beauty of the proposed classifier structure for image classification system is that it uses an ensemble of proposed single neurons in the complex domain, instead of ensemble of multi layer network. The proposed 'OCON : One-Class-in-One-Neuron' recognizer is designed using two higher order neuron TION1 and TION2 in the complex domain. Every single neuron in recognizer is dedicated to recognizing images (a subject) of its 'own class' assigned to it (OCON). The outputs in the ensemble of neurons are forming an aggregate output of the classifier.

The conventional neuron models in real and complex domain are based on summation and radial basis aggregation functions. The conventional MLP offers a global approximation to input-output mapping, but it may be plagued by slow convergence and has a tendency to get trapped at bad local minima. On the contrast, RBF network offers local approximation to input-output mapping and often provides a faster convergence. However, it will be inefficient in approximating constant-valued functions, as addressed in [24]. The novelty in aggregation function of novel neurons is to take the advantage of the merits of perceptron and radial basis processing [39]. The net potential of novel neuron is a weighted integration of summation and radial basis sub-functions. Thus, input aggregation for this neuron is a functional, which formulates its compensatory structure. The information processed through sub-functions are integrated in desired proportion (λ : γ), linearly ($\bigoplus$) or non-linearly ($\bigotimes$) in the considered models. The compensatory parameters λ and γ specify the contribution of summation and radial basis sub-functions to take into account the vagueness involved. With a view to achieve a robust aggregation function, the parameters λ and γ are itself made adaptable in course of training.

Let $Z = [z_1, z_2....z_L]$ be the vector of input signals and Y be the output of considered neuron. Let $f_\mathbf{C}$ be the complex-valued activation function defined in (3). $Z^\mathcal{T}$ is the transpose of vector Z and $\bar{z}$ is the complex conjugate of z. $W_S = [w_{S1}, w_{S2}....w_{SL}]$ is a vector of weights from input

layer ($l = 1...L$) to summation part of novel neuron and $W_{RB} = [w_{RB1}, w_{RB2}....w_{RBL}]$ is a vector of weights from input layer to radial basis part of novel neuron. All weights, bias and input-output signals are complex numbers.

2.2.1 Learning rules for C-TION1 based classifier

The product is inserted in the formulas for modeling the non-linear operations. Here, it is expressed as $a \otimes b = 1 + a + b + ab$. The C-TION1 neuron is defined as:

$$Y = f_C \left(\lambda \times W_S \, Z^{\mathcal{T}} \bigotimes \gamma \times e^{-\|Z - W_{RB}\|^2} \right) \quad (5)$$

where $\|Z - W_{RB}\|^2 = (Z - W_{RB}) \times (Z - W_{RB})^{\aleph}$. Here superscript $\aleph$ represents the matrix complex conjugate transposition. Let $\eta \in [0, 1]$ be the learning rate and f' be the derivative of the function f. w_0 is a bias and $z_0 = 1 + j$ is the bias input, where j is an imaginary unity. Let error $e = (D - Y)$ be the difference between desired and actual output. The proposed classifier contains only one layer of C-TION1 neurons. The update rules for learning parameters of considered classifier are derived using split-complex error back-propagation algorithm, by minimizing following real-valued cost function:

$$E = \frac{1}{2} |e|^2 \quad (6)$$

Let V_π be the net potential of C-TION1 neuron then from (5) and definition of operation $\otimes$

$$V_\pi = w_0 \, z_0 + \lambda \, W_S \, Z^{\mathcal{T}} + \gamma \, e^{-\|Z - W_{RB}\|^2} + \lambda \, W_S \, Z^{\mathcal{T}} \, \gamma \, e^{-\|Z - W_{RB}\|^2} \quad (7)$$

This net internal potential of TION1 neuron may also be expressed term wise as follows:

$$V_\pi = w_0 \, z_0 + V_{\pi_1} + V_{\pi_2} + V_{\pi_1} \, V_{\pi_2} \quad (8)$$

The update equations for different learning parameters are as follows:

$$\Delta w_{RB} = 2 \, \eta \, e^{-\|Z - W_{RB}\|^2} (z - w_{RB})$$
$$\left\{ \Re(\Gamma_\pi) \left\{ \Re(\gamma) (1 + \Re(V_{\pi_1})) - \Im(\gamma)\Im(V_{\pi_1}) \right\} + \Im(\Gamma_\pi) \left\{ \Im(\gamma) (1 + \Re(V_{\pi_1})) + \Re(\gamma)\Im(V_{\pi_1}) \right\} \right\} \quad (9)$$

$$\Delta w_S = \eta \, \overline{z} \, \overline{\lambda} \, (1 + \overline{V_{\pi_2}}) \, \Gamma_\pi \quad (10)$$

$$\Delta \lambda = \eta \, \overline{(W_S \, Z^{\mathcal{T}})} \, (1 + \overline{V_{\pi_2}}) \, \Gamma_\pi \quad (11)$$

$$\Delta \gamma = \eta \, e^{-\|Z - W_{RB}\|^2} \, (1 + \overline{V_{\pi_1}}) \, \Gamma_\pi \quad (12)$$

$$\Delta w_0 = \eta \, \overline{z_0} \, \Gamma_\pi \quad (13)$$

where

$$\Gamma_\pi = (\Re(e) \, f'(\Re(V_\pi)) + j \, \Im(e) \, f'(\Im(V_\pi))) \quad (14)$$

2.2.2 Learning rules for C-TION2 based classifier

The algebraic sum is inserted in the formulas for modeling the linear operations. Here, it is expressed as $a \oplus b = 1 + a + b$. The C-TION2 neuron is defined as:

$$Y = f_C \left(\lambda \times W_S \, Z^{\mathcal{T}} \bigoplus \gamma \times e^{-\|Z - W_{RB}\|^2} \right) \quad (15)$$

Let V_σ be net potential of C-TION2 neuron in the classifier then from (15)

$$V_\sigma = w_0 z_0 + \lambda \, W_S \, Z^{\mathcal{T}} + \gamma \, e^{-\|Z - W_{RB}\|^2} \quad (16)$$

The update equations for different learning parameters are as follows:

$$\Delta w_{RB} = 2 \, \eta \, e^{-\|Z - W_{RB}\|^2} (z - w_{RB})$$
$$\{\Re(\Gamma_\sigma) \, \Re(\gamma) + \Im(\Gamma_\sigma) \, \Im(\gamma)\}\} \quad (17)$$

$$\Delta w_S = \eta \, \overline{z} \, \overline{\lambda} \, \Gamma_\sigma \quad (18)$$

$$\Delta \lambda = \eta \, \overline{(W_S \, Z^{\mathcal{T}})} \, \Gamma_\sigma \quad (19)$$

$$\Delta \gamma = \eta \, e^{-\|Z - W_{RB}\|^2} \, \Gamma_\sigma \quad (20)$$

$$\Delta w_0 = \eta \, \overline{z_0} \, \Gamma_\sigma \quad (21)$$

where

$$\Gamma_\sigma = (\Re(e) \, f'(\Re(V_\sigma)) + j \, \Im(e) \, f'(\Im(V_\sigma))) \quad (22)$$

3 Performance evaluation

In order to evaluate the performance of the proposed deep learning machine the standard biometric measures, error rate (in another way percentage accuracy), FAR (False acceptance rate) and FRR (False recognition rate), are considered in this section. FAR and FRR are inversely proportional measurements, therefore, variable threshold setting is provided for users to keep balance. It is generally desirable that the recognition system must be much stricter to unauthorized persons and slightly stricter for authorized persons. In all experiments, our selection of threshold is inclined in a direction where we can obtain significantly lower FAR and slightly higher FRR. The number of neurons in ensemble is set to the number of image classes (subject). Each neuron is trained to give output '$1 + j$' for its own class and 0 for another class. The number of nodes in the input layer is equal to the number of elements in the feature vector. The data values in the feature vectors are normalized to lie within the first quadrant of a unit circle. An input face image presented to a neuron associated to its own class is considered as a positive example, while images of other classes to this neuron are considered as a negative example. The classifiers are trained with learning rules presented in Section 2.2. For testing the identity claim of a face image, an M' dimensional

feature vector is extracted from the face image and this vector is given to every neuron or network dedicated to a subject (person). The discriminant function applied in the proposed system calculates the square deviation of the actual output to the ideal output. An unknown feature vector is classified as belonging to a class, when deviation is less than a threshold. In general, the threshold can be determined from the experimental studies.

The three sets of experiments are conducted for relative performance evaluations of proposed deep learning machine in terms of different neuron architectures and feature extraction algorithms presented in this paper. In all experiments, four images of each subject are taken for training and the rest are used for testing. The results presented in this paper are an average outcome of five trainings with different weight settings. For analysis purposes, the number of *learning parameters* in each complex-valued weight counts as two, as in [31].

3.1 Performance with ORL face database

This database is designed by AT & T laboratories for research purpose and available in [34] which contains facial images of 40 different persons (subjects) and 10 different samples of each subject image per person. There are variations in facial expression (open or close eye, smiling or frowning face), facial details (with or without glasses, hairs), scale (up to 10 %), orientation (upto 20°). Each image was digitized and presented by 92×112 pixel array whose gray level ranged between 0 to 255. As an example, few sample images are shown in Fig. 1. A total 160 images (four of each subject) have been used for training and rest 240 for testing.

Eigenvectors associated with higher eigenvalues provide more information on the face variation than those with smaller eigenvalues. Eigenvalues in $\mathbf{R}PCA$ and $\mathbf{C}PCA$ can be used to plot a graph of variance (Fig. 2) captured by each principal component. It is a cumulative distribution for M' components defined in (23). This graph permits us to select the necessary eigenvectors for feature extraction in different techniques. It is found that 48 eigenvectors or 30 % of the maximum possible number of eigenvalues (training set has 160 images) of the $\mathbf{C}PCA$ is enough to account for more than 93 % of the variations among the training set. While, approximately 86 % of the total varience is kept in the same

Fig. 2 Variance captured by each principal component in $\mathbf{R}PCA$ and $\mathbf{C}PCA$ for the ORL face database

number of eigenvectors of $\mathbf{R}PCA$. Thus, with 48 subspace dimensions $\mathbf{C}PCA$ yields better accuracy in recognition as well as better class distinctiveness in comparison to $\mathbf{R}PCA$.

$$\frac{\sum_{i=1}^{M'} e_i}{\sum_{i=1}^{M} e_i} \qquad where \qquad M' < M << N. \quad (23)$$

The results demonstrate the recognition performance in terms of different parameters (training epochs, weights, FRR, FAR, recognition rate) are presented in Table 1. The performance of different ANN architectures is broken down according to the feature extraction methods. The remarkable feature of Table 1 is that classifier based on TION1 and TION2 neurons in the complex domain requires least training epochs and learning parameters with superior recognition rate. The $\mathbf{C}ICA$ based feature extraction always provides the best results. Though, the difference among different feature extraction methods in the context of accuracy is not large, but it is significant because the class distinction is far better in $\mathbf{C}ICA$ as compared to any other method. Thus, in case of $\mathbf{C}ICA$, the deviation of square error from threshold is quite high for class and non-class faces. For roughly comparable FRR, the value of FAR is seen quiet low in case of $\mathbf{C}ICA$ based system; which makes it vigorous to perform more effectively in real life situations. From widespread simulations on the ORL face database, it is vital to make following deductions:

Fig. 1 Example images from the ORL database

Table 1 Assessment of training and testing performance for ORL face data set with different feature extractor and different neuron based classifier

S No	Neuron Type	Network	Parameters	Feature Extraction	Average Epochs	FRR	FAR	Recognition Rate
1	**R**-MLP	48-8-1	40×401	**R**PCA	44000	0.092	0.032	96.6 %
				RICA	6000	0.067	0.014	98.25 %
				RPCA	28000	0.083	0.020	97.8 %
2	**C**-MLP	48-3-1	40×302	**C**PCA	28000	0.092	0.018	98.0 %
				RICA	6000	0.087	0.016	98.2 %
				CICA	6000	0.067	0.0095	98.9 %
				RPCA	6000	0.092	0.020	97.7 %
3	**C**-TION2	48-1	40×198	**C**PCA	6000	0.080	0.014	98.3 %
				RICA	4000	0.083	0.017	98.1 %
				CICA	4000	0.075	0.009	99.0 %
				RPCA	6000	0.086	0.020	97.85 %
4	**C**-TION1	48-1	40×198	**C**PCA	6000	0.075	0.009	98.8 %
				RICA	4000	0.083	0.013	98.4 %
				CICA	4000	0.070	0.006	99.25 %

– Class distinctiveness is very poor in case of $\mathbf{R}PCA$ and $\mathbf{C}PCA$.
– Class distinctiveness with $\mathbf{R}ICA$ is slightly better than PCA but it is best with $\mathbf{C}ICA$.
– The $\mathbf{C}ICA$ based system is much stricter to unauthorized person.

In order to assess the sensitivity of different feature extraction methods on the number of subspace dimension variations, the comparative performance is given in Fig. 3. The results of experiments are typical in part, because the recognition rate does not increase monotonically with the number of subspace dimensions. Figure 3 shows that the
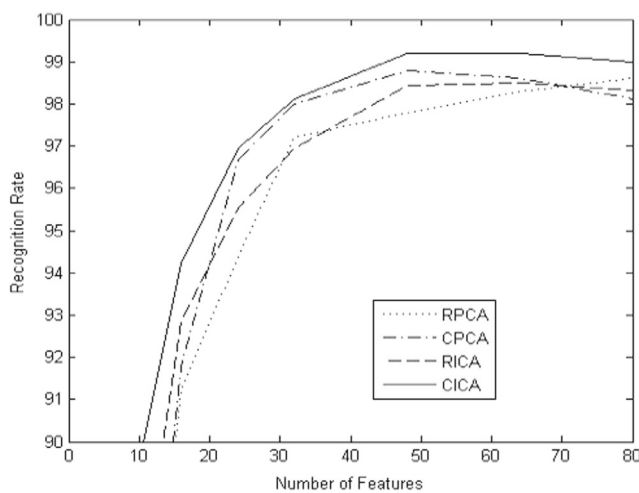
Fig. 3 Recognition rate vs subspace dimension for different feature extraction techniques in the ORL Face Database

feature extraction criterion performs reasonably well, when 48 features of subspace are chosen. The increase in subspace dimension does not improve the overall performance. Though, the recognition rate in $\mathbf{R}PCA$ increases slowly beyond 48 features, but still it is less than the $\mathbf{C}ICA$.

3.1.1 Performance with occluded and blurred images

The occlusion and blurring is introduced (electronically modified) [47] in few images of ORL face data base to examine the performance of the proposed techniques over variations in distortion occurred. Recognition with different feature extraction techniques depends on the degree of occlusion and blurring introduced. In this experiment, we consider the best performing classifier of $\mathbf{C}$-TION1 neurons.

Both PCA based face recognizers have fairly recognized the occluded faces given in Fig. 4. Degree of occlusion in these face is quite less. On increasing the degree of occlusion (Fig. 5), it is observed that both $\mathbf{R}PCA$ and $\mathbf{C}PCA$ based recognizer were not able to identify the set of faces in Fig. 5. But, they are well recognized by ICA based recognizer.

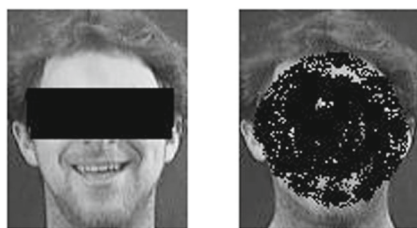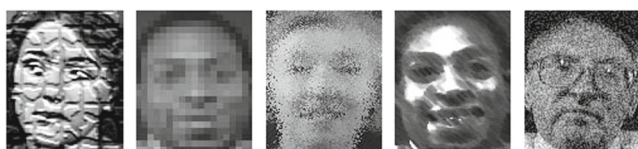**Fig. 4** Occluded images, recognized by both PCA based learning machine

Fig. 5 Occluded images, not recognized by any *PCA* based learning machine but recognized by ICA based learning machine

Fig. 6 Occluded images which are only recognized by **C***ICA* based learning machine, but can not be recognized by **R***ICA* and any *PCA* based learning machine

Fig. 7 The occluded images, not recognized by any feature extractor based learning machine

Fig. 8 The blurred images, recognized by ICA based system, but not recognized by any *PCA* based learning machine

Fig. 9 The blurred images, only recognized by **C***ICA* based learning machine, but can not be recognized by **R***ICA* and any *PCA* based learning machine

Fig. 10 Blurred images, not recognized by any feature extractor based learning machine

The reason was the poor class distinctiveness offered by these methods. In another set of images, shown in Fig. 6, it was observed that only **C***ICA* based recognition system can correctly classify these faces, while other techniques failed to do so. It turns out from these observations that **C***ICA* provides more discriminating power than any other technique. Figure 7 presents faces with very high occlusion. Our experiments found that these images can not be recognized by any considered method.

Similarly, Fig. 8 presents a set of blurred images which were also not correctly identified by both PCA based methods; but recognized by ICA based methods. Another set in Fig. 9 has comparatively high distortions. Only **C***ICA* based methods are able to correctly classify them, while rest methods failed to do so. It is due to robust class distinctiveness yielded by the **C***ICA* based system. Figure 10 presents blurred faces with very high degree of distortion, which were not identified by even **C***ICA*. This limitation is obvious.

3.2 Performance with yale face database

The Yale face database [46] contains 165 grayscale images (11 images per person) of 15 individuals (subjects) from different ethnicity and gender. The images are taken under varying lighting conditions and different facial expressions : center-light, left-light, right-light, with or without glasses, normal, happy, sad, sleepy, surprised and wink. The spatial and gray-level resolution of the images are 320×243 and 256, respectively. As an example, few sample images are shown in Fig. 11.

Figure 12 presents the variation captured by the principal components of the training set in **C***PCA* and **R***PCA*. This information prompted us to select 15 eigenvectors from the set of 60 training images of 15 subjects for further feature extraction. A set of 105 images of 15 subjects is considered for testing. The test results of Yale face data set with different feature extraction techniques and neuron architectures are presented in Table 2. Different measures clearly demonstrate that the recognition system developed in a complex domain outperforms in all respects. Classifiers based

Fig. 11 The example images from Yale Database

on **C**-TION1 and **C**-TION2 neurons give effectively better accuracy with smaller network topology and much less number of training epochs. It is also observed that class discrimination in **C**PCA is better than **R**PCA, because 15 eigenvectors from **C**PCA accounts for more than 93 % of the variations among the training set while, 84 % of the total variance is by **R**PCA. Among different combinations, the superiority **C**ICA and **C**-TION1 is again demonstrated in this experiment. CICA based system is also much stricter to unauthorized people.

The recognition accuracy with respect to dimensionality of subspace among different feature extraction procedure are given in Fig. 13. For every technique, the maximum recognition rate occurs at a different number of subspace dimensions, suggesting that it may not be possible to choose a subspace dimension for all techniques. Figure 13 shows that **C**ICA achieve a higher recognition accuracy with a smaller size of the subspace. It is also observed that **C**PCA performs better at lower dimension, while the performance of **R**ICA and **R**PCA increases smoothly with an increase in subspace dimension. This argument has had correspondence with curve shown in Fig. 12, where **C**PCA has more variance at lower dimension.
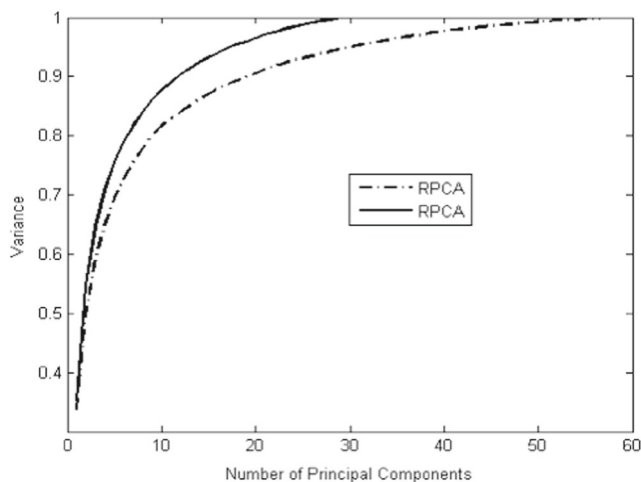
Fig. 12 Graph of variance captured by each principal component of Yale Face Database

3.3 Performance with Indian face database

I consider 500 color images corresponding to 50 subjects of Indian face database [21]. All images (640 × 480 pixels) have a bright homogeneous background and different poses (looking front, looking left, looking right, looking up, looking up toward the left, looking up toward the right, looking down) and emotions (neutral, smile, laughter, sad/disgust). As an example, few sample images are shown in Fig. 14. The number of images considered for training and testing is 200 and 300 respectively. I obviously want to capture as much variations as possible of the training set with fewest number of subspace dimensions. The graph in Fig. 15 allows me to see more clearly how much variation is captured by eigenvectors. This information prompted us to select 60 eigenvectors from the set of 200 training images of 50 subjects for further feature extraction. They correspond for keeping more than 92 % of total variance in the eigenvalues of this data set.

The simulation results of considered database with different feature extractor and neural classifiers are presented in Table 3 in terms of different measures. It again reveals the superiority of the **C**-TION1 based classifier. The feature vectors yielded by ICA based methods provide faster learning in all neural classifiers. It is worthwhile to mention that this face data set have much variations in poses, which decrease its recognition accuracy. But, the performance of **C**ICA in feature extraction is significant in terms of precision in the classification and accuracy. With **C**ICA, classification error is quite low for class faces while it is quite high for non-class faces. From extensive simulations on this face data set, it is imperative to make following inferences:

– Class distinctiveness is very poor in case of **R**PCA and **C**PCA.
– Class distinctiveness with **R**ICA is slightly better than PCA but it is best with **C**ICA.

It allows **C**ICA to perform very well in noisy and blurred images, which are generally captured in real environmental

Table 2 Assessment of training and testing performance for Yale face data set with different feature extractor and different neuron based classifier
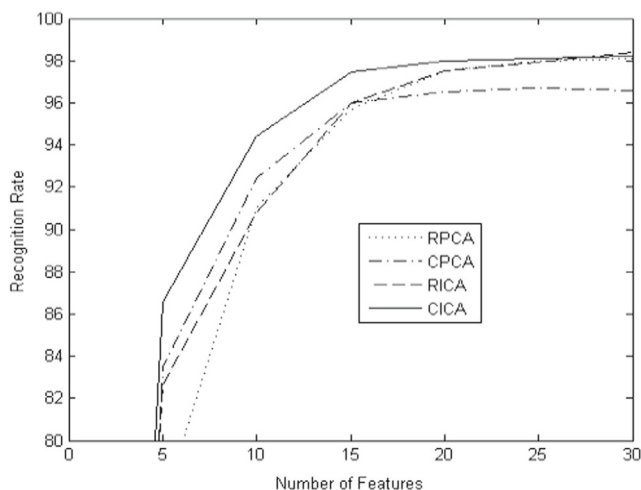
S No	Neuron Type	Network	Parameters	Feature Extraction	Average Epochs	FRR	FAR	Recognition Rate
1	**R**-MLP	15-8-1	15×137	**R**PCA	22500	0.13	0.08	91.75 %
				RICA	10500	0.10	0.05	94.6 %
				RPCA	15000	0.11	0.041	95.5 %
2	**C**-MLP	15-3-1	15×104	**C**PCA	15000	0.10	0.035	96.0 %
				RICA	9000	0.10	0.040	95.60 %
				CICA	9000	0.10	0.026	96.60 %
				RPCA	7500	0.13	0.041	95.10 %
3	**C**-TION2	15-1	15×66	**C**PCA	7500	0.11	0.041	95.35 %
				RICA	5250	0.12	0.036	95.85 %
				CICA	5250	0.10	0.024	97.0 %
				RPCA	7500	0.09	0.040	95.3 %
4	**C**-TION1	15-1	15×66	**C**PCA	7500	0.12	0.034	96.0 %
				RICA	5250	0.10	0.036	96.0 %
				CICA	5250	0.11	0.020	97.4 %

situations. The analysis of different feature extraction methods as a function of the dimension of condensed subspace is presented in Fig. 16. It is obtained from the best performing **C**-TION1 classifier. **C**ICA outperforms over all other methods, it is statistically more significant when we consider a lesser number of subspace dimensions. The relative ordering of the subspace projection techniques depends on the number of subspace dimensions. The performance of the **C**PCA is better in lower dimensions while **R**PCA is better in higher dimensions. It is again observed that the

Fig. 13 Recognition rate vs subspace dimension for different feature extraction techniques in Yale Face Database

recognition rate with **C**ICA is always better along with other parameters.

3.3.1 Performance with occluded and blurred images

The class distinctiveness must also be taken into account in comparing different feature extraction techniques in any classification studies specially where noise is involved. In order to assess the robustness of different feature extraction techniques to noise, distortions and other environmental effects, I carried out a comparative assessment over electronically modified images [47]. Figure 17 presents a set of images in which low distortion is introduced. These images are passed to previously trained **C**-TION1 classifiers with different feature extraction techniques. **C**ICA and **R**ICA based classifiers are able to correctly classify them while any PCA based classifiers are not able to recognize them. Another set of images in Fig. 18 has comparatively more distortions and effects. These images can not be recognized by any PCA and **R**ICA, while **C**ICA based recognition system is able to correctly identify them. It is due to the better class distinctiveness yielded by **C**ICA. Though, there are a few limitations of this technique as there are always in every method. Figure 19 presents some images where **C**ICA also failed to perform proper recognition. Therefore, the basis vectors obtained by **C**ICA are superior to **R**ICA and PCA in the sense that it provides feature vectors, which are more robust to the effect of noise. It is therefore demonstrated in this experiment that **C**ICA outperforms other techniques for recognition especially in real life situations.
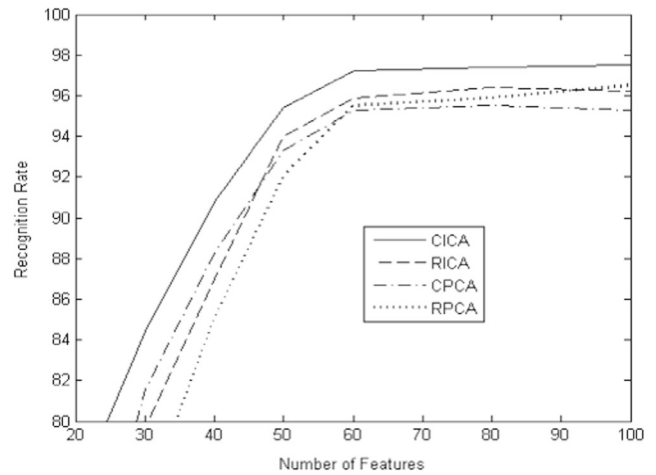
Fig. 14 The example images from Indian face database
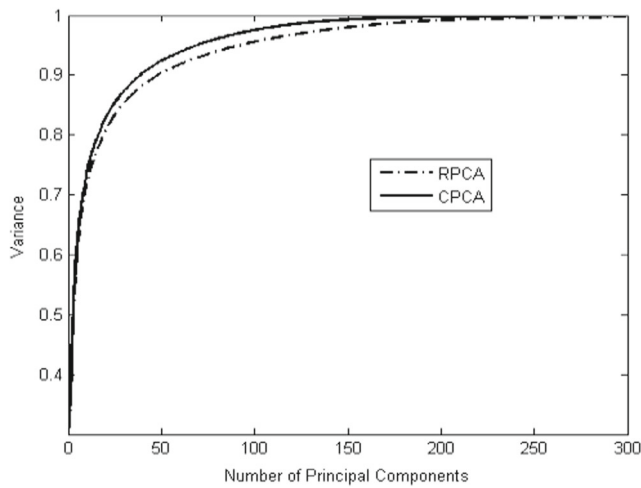
Fig. 15 Graph of variance captured by each principal component of the Indian face database

Fig. 16 Recognition rate vs subspace dimension for different feature extraction techniques in Indian face database

Table 3 Assessment of training and testing performance for Indian face data set with different feature extractor and different neuron based classifier

S No	Neuron Type	Network	Parameters	Feature Extraction	Average Epochs	FRR	FAR	Recognition Rate
1	**R**-MLP	60-8-1	50×497	**R**PCA	45000	0.19	0.055	93.8 %
				RICA	25000	0.17	0.045	95.3 %
				RPCA	30000	0.20	0.044	95.0 %
2	**C**-MLP	60-3-1	50×374	**C**PCA	30000	0.21	0.046	94.8 %
				RICA	10000	0.17	0.040	95.5 %
				CICA	10000	0.17	0.035	96.6 %
				RPCA	20000	0.17	0.042	95.3 %
3	**C**-TION2	60-1	50×246	**C**PCA	20000	0.18	0.046	95.0 %
				RICA	10000	0.20	0.041	95.5 %
				CICA	10000	0.16	0.029	96.8 %
				RPCA	20000	0.15	0.041	95.4 %
4	**C**-TION1	60-1	50×246	**C**PCA	20000	0.16	0.043	95.2 %
				RICA	10000	0.20	0.037	95.9 %
				CICA	10000	0.17	0.025	97.2 %

Fig. 17 Occluded images, not recognized by any PCA based system recognition system; but recognized by **R**ICA and **C**ICA based learning machine

Fig. 18 Occluded images, recognized with only **C**ICA based learning machine

Fig. 19 Occluded images, not recognized with any feature extraction based learning machine

4 Inferences & discussion

This paper presents a state-of-art deep learning machine with different feature extractors and classifiers in the complex domain. A detailed comparative analysis is carried out with real domain feature extractor. I explored the discriminating power in conjunction with thresholding based rejection. Different architectures of ANN ensemble, their suitability for classification and robustness for rejection of unauthorized cases have been throughly examined. The structure for proposed recognizer is a One-Class-in-One-Neuron (OCON) which contains the ensemble of single **C**-TION2 or **C**-TION1 neurons instead of an ensemble of neural networks. The remarkable achievement in proposed recognizer (classifier) is its compact structure, improved learning speed, lesser weights storage and better accuracy in recognition. Performance evaluation presented in this paper by different architecture of recognizer is very strict to unauthorized persons (low FAR) but also strict for authorized person.

Instead, I compared different feature extractors on three face data sets across different recognizers. It is significantly observed that the relative ranking of feature extraction algorithm never changes. When we move from real to complex domain, though it increases the computations in terms of mathematical operations with complex data, but it is not significant in comparison to the improvements we achieve in performance with proposed recognizer. The **C**ICA coefficients have consistently shown a greater class discrimination ability than **R**ICA and **R**PCA coefficients. The **C**ICA based system has also yielded quiet low FAR value for approximately similar FRR value, which is always desirable.

Our extensive experiments have found that **C**ICA is robust enough to recognize images with different distortions and variations. Such distortions generally arise in real life situations. **C**ICA has given significant performance under blurring, partial occlusion and other electronically modified images. Thus, its reduced sensitivity to noise and robust distinctiveness between classes make it work more effectively in real environment. More importantly, I have developed a simple, but efficient recognition system, which also need minimum weight storage for future testing. I conclude that **C**ICA combined with **C**-TION1 outperforms over other techniques in all respects.

References

1. Abate AF, Nappi M, Riccio D, Sabatino G (2007) 2D and 3D face recognition: a survey. Pattern Recognition. Letters 28:1885–1906
2. Ahonen T, Pietikainen M, Hadid A (2004) Face recognition based on the appearance of local regions. In: 17th international conference on pattern recognition, vol 3, pp 153–156
3. Aitkenheada MJ, Mcdonald AJS (2003) A neural network face recognition system. Eng Appl Artif Intell 16(3):167–176
4. Aizenberg I, Moraga C (2007) Multilayer feedforward neural network based on multi-valued neurons (MLMVN) and a back-propagation learning algorithm. Soft Comput 11(2):169–183
5. Anemuller J, Sejnowski T, Makeig S (2003) Complex independent component analysis of frequency-domain electroencephalographic data. Neural Netw 16(9):1311–1323
6. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13(6):1450–1464
7. Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7(6):1129–1159
8. Bhattacharjee D, Basu DK, Nasipuri M, Kundu M (2010) Soft Comput 14(Human face recognition using multilayer perceptron):559–570. Springer
9. Brown JW, Churchill RV (2003) Complex variables and applications, VIIth ed. Mc Graw Hill
10. Calhoun V, Adali T (2006) Complex infomax: convergence and approximation of infomax with complex nonlinearities. VLSI Signal Process, Springer Science 44:173–190
11. Calhoun V, Adali T, Pearlson GD, Pekar JJ (2002) ON complex infomax applied to complex FMRI data. In: Proceedings of ICASSP, Orlando
12. Calhoun VD, Adali T (2006) Unmixing fMRI with independent component analysis. IEEE Eng Med Biol Mag 25(2):79–90
13. Cevikalp H, Neamtu M, Wilkes M, Barkana A (2005) Discriminative common vectors for face recognition. IEEE Trans PAMI 27(1):1–9
14. Er MJ, Chen W, Wu S (2005) High-speed face recognition based on discrete cosine transform and RBF neural networks. IEEE Trans Neural Netw 16(3):679–691
15. Faijul Amin M, Murase K (2008) Single-layered complex-valued neural network for real-valued classification problems. Neurocomputing
16. Haddadnia J, Faez K, Moallem P (2001) Neural network based face recognition with moments invariant. In: IEEE international conference image processing, vol I, Thessaloniki, pp 710
17. Hahn SL (1996) Hilbert transforms in signal processing. Artech House, Boston
18. Hietmeyer R (2000) Biometric identification promises fast and secure processing of airline passengers. The International Civil Aviation Organization Journal 55(9):10–11
19. Hirose A (2006) Complex-valued neural networks. Springer-Verlag New York Inc, New York
20. Huang D, Mi J (2007) A new constrained independent component analysis method. IEEE Trans Neural Netw 18(5):1532–1535
21. Jain V, Mukherjee A (2002) The Indian face database, http://vis-www.cs.umass.edu/vidit/Indianfacedatabase
22. Kasar MM, Bhattacharyya D, Kim TH (2016) Face recognition using neural network: a review. Int J Secur Appl 10(3):81–100
23. Kwak K, Pedrycz W (2007) Face recognition using an enhanced independent component analysis approach. IEEE Trans Neural Netw 18(2):530–541
24. Lee CC, Chung PC, Tsai JR, Chang CI (1999) Robust radial basis function neural network. IEEE Trans Syst, Man Cybern B Cybern 29(6)
25. Lua Wei, Rajapakse JC (2006) ICA with reference. neurocomputing 69:2244–2257