Nonlinear Feature Extraction Approaches for Scalable Face Recognition Applications

Hima Deepthi Vankayalapati
Institute of Smart Systems Technologies
University of Klagenfurt
9020 Klagenfurt, Austria
hvankaya@edu.uni-klu.ac.at

Kyandoghere Kyamakya
Institute of Smart Systems Technologies
University of Klagenfurt
9020 Klagenfurt, Austria
kyandoghere.kyamakya@uni-klu.ac.at

***Abstract*—The human skill of identifying thousands of people even after so many years excited many researchers to focus on face recognition systems. The majority of real world applications demands more robust, scalable and computationally efficient face recognition techniques which can operate under complex viewing and environmental conditions. The appearance based linear subspace techniques are very useful in data classification and dimensionality reduction tasks; however these algorithms only classify the linear data. The scalability of the linear subspace techniques is limited, as the computational load and memory requirements increase dramatically with the large database. This paper evaluates different nonlinear feature extraction approaches for face recognition application, namely wavelet transform, radon transform and cellular neural networks (CNN). In this work, the combination of radon and wavelet transform based approaches is used to extract the multi-resolution features, which are invariant to facial expression and illumination conditions. The efficiency of the stated wavelet and radon based nonlinear approaches over the databases is demonstrated, with the simulation results performed over the FERET database. This paper also presents the use of CNN in extracting the nonlinear facial features in improving the recognition rate, as well as computational speed, compared to other stated nonlinear approaches over the ORL database.**

***Index Terms*—Feature extraction, Face recognition, Linear subspace techniques, Cellular neural network, Wavelet transform, Radon transform.**

I. Introduction

In computer vision, a feature is a set of measurements. Each measurement contains a piece of information, and specifies the property or characteristics of the object present in the image [1]. The linear features are more advantageous, when the given data is Gaussian distributed in terms of mean. However in most real world face recognition applications, facial features of the face image are not purely Gaussian distributed (they vary with complex viewing and environmental conditions).

Researchers have developed various biometric techniques to identify or recognize persons by their physical characteristics like finger, voice, face etc. These biometric techniques have their own advantages and drawbacks as well [2]. Among all the biometric techniques, the face recognition has a distinct advantage of collecting the required data (i.e image) without any cooperation from the person [3]. The face recognition is a complex visual classification task which plays an important role in computer vision, image processing and pattern recognition.

Research concerning the face recognition started nearly in 1960's [4]. Different face recognition techniques have been proposed during last decades namely feature based, model based and appearance based techniques [5], [6]. In feature based techniques, the overall technique describes the position and size of each feature (eye, nose, mouth or face outline) [7]. In this approach, the extracting features in different poses (viewing conditions) and lighting conditions are very complex tasks. For applications with large databases, we have large set of features with different sizes and positions, making it difficult to identify the required feature points [8]. In the model based approach, a 3D model is constructed based on the facial variations in the image or important information related to the image. The difficulties in this approach are, we need a very expensive camera (Stereo vision) to capture the facial variations clearly; further construction of 3D model is difficult, and it takes more time to construct the model for large databases [6]. The availability of large 3D data is also one of the essential complex tasks that makes the model based methods not suitable for real world applications dealing with large databases.

In 1990's, researchers introduced appearance based linear subspace techniques, statistics related techniques, to solve face recognition problems. The introduction of the linear subspace techniques is a milestone in the face recognition concept. The performance of appearance based techniques heavily depends on the quality of the extracted features from the image [9]. The appearance based linear subspace techniques extract the global features, as these techniques use the statistical properties like the mean and variance of the image [6]. The major difficulty in applying these techniques over large databases is that the computational load and memory requirements for calculating features increase dramatically for large databases [3]. In order to increase the performance of the face recognition techniques, the nonlinear feature extraction techniques are introduced.

In order to improve the performance of the face recognition technique, we have to extract both linear and nonlinear features. We have many nonlinear feature extraction techniques, such as radon transform and wavelet transform. The radon transform based nonlinear feature extraction gives the direction of local features. This process extracts the spatial frequency components in the direction of radon projection is computed

[10]. When features are extracted using radon transform, the variations in this facial frequency are also boosted [10]. The wavelet transform gives the spacial and frequency components present in an image [11]. However these nonlinear feature extraction techniques are computationally expensive. In order to improve the computational speed of the nonlinear feature extraction process, the cellular neural network (CNN) concept is being proposed.

The novel scheme will involve, at its heart, CNN based processors, which will be the key component of the analog computing based ultra-fast solver for image processing tasks. CNN based analog computing has the very attractive advantage of easy implementation or emulation on digital platforms. The objective of this paper is to present the use of CNN in extracting nonlinear features using the ORL database.

The paper is organized as follows: in section 2, the importance and methodologies of the linear subspace techniques are explained briefly. In section 3, the basics and importance of the radon transform are explained briefly. In section 4, wavelet transform is briefly described. In section 5, cellular neural network is introduced. Genetic algorithm based template calculation method is also briefly described in section 5. The experimental simulation results using the FERET and the ORL databases are described in section 6. Section 7 deals with some concluding remarks and outlooks.

II. Linear subspace techniques

Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) are related to the appearance based linear subspace technique [6]. These linear subspace techniques use statistics (mean and co-variance). The calculation of the mean and co-variance is performed by using the train data set to form the data matrix X. In data matrix X, each column x_i represents the image in the train data set. The mean image of the train data set is expressed as shown in Eq. 1.

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

The co-variance matrix C of the random vector x is calculated using Eq. 2.

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)(x_i - m)^T (or) C = AA^T \tag{2}$$

Calculating the co-variance matrix by using Eq. 2 takes high memory because of the dimensions of C. The size of A is LMxN. The size of C is LMxLM, which is very large. So the matrix $L = A^T A$ is considered instead of C. The dimension of L is NxN, which is much smaller than the dimensions of C. After the co-variance matrix, each technique (PCA, ICA and LDA) uses a specific approach to calculate the key parameters of the feature space.

In linear subspace technique, all the images in the train data set are represented as points in the feature space as shown in Fig. 1. The given test image is also represented as a point in the same space and the minimum distance train data set image gives the best match.

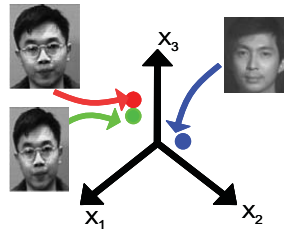


Fig. 1. Image representation in the high dimensional space

A. Principal Component Analysis (PCA)

PCA highlights the similarities and differences between the variables in the data [12], [13]. After calculating the co-variance matrix, we have to calculate the eigenvalues and eigenvectors of the co-variance matrix. Then we arrange all eigenvalues in descending order and we take first few highest eigenvalues and corresponding eigenvectors. This operation is the evaluation of principal components [14]. The eigenvectors e_1, e_2,...e_n are shown in Eq. 3.

$$W_{pca} = [e_1, e_2, ..., e_n] \tag{3}$$

We neglect the remaining less significant eigenvalues and the corresponding eigenvectors. The eigenvalues neglected lead to a very small information loss [15]. The principal component axis passes through the mean values. A new transformation matrix W_{pca} is obtained, by projecting the principal component on to the original data set.

B. Independent Component Analysis (ICA)

ICA uses the higher order statistics of the input data to find the independent components. The independency is distinguished by knowing the uncorrelated data. ICA is a special case of blind source problem [16]. One of the simplest applications of ICA is found in the cocktail party problem. So the ICA technique is a generalization of PCA technique. In this technique we first calculate the PCA transformation matrix W_{pca}, transform the centered matrix $P = [x_1 - m, x_2 - m, ..., x_n - m]$ using W_{pca} and then form a new matrix Z (square matrix with size NXN), which contains the random vector z, whose elements are uncorrelated as shown in Eq. 4.

$$Z = Wpca^T P \tag{4}$$

The next important stage is the rotation stage. In this one, the fixed point algorithm is used to find the W_k [17]. After that, we calculate the overall transformation matrix as shown in Eq. 5.

$$W_{ica} = W_{pca} W_k \tag{5}$$

C. Linear Discriminant Analysis (LDA)

The main objective of the LDA is minimizing the within class variance and maximizing the between class variance in the given data set. In other words it groups the same class images and separates the different class images [18]. A class means the collection of data (images) belonging to the same object or same person. In LDA, we have to calculate the mean image of each class i which is represented as m_i.

$$S_i = \frac{1}{N_i} \sum_{x \in X_i}^{c} (x - m_i)(x - m_i)^T \tag{6}$$

Eq. 6 represents the class dependant scatter matrix and it gives the sum of the co-variance matrix of the centered images in each class. X_i represents the data matrix corresponding to class i. N_i represents the images present in class i. c represents the total number of classes. The within class scatter matrix S_w is calculated from Eq. 7.

$$S_w = \sum_{i=1}^{c} S_i \tag{7}$$

This leads to the evaluation of the amount of variance between the images in each class. S_b represents the between class scatter matrix [3] and it calculates the variance between the classes by using Eq. 8. The co-variance matrix of each class is the difference between the total mean of all classes and the mean of each class. S_b is expressed in Eq. 8.

$$S_b = \sum_{i=1}^{c} (m_i - m)(m_i - m)^T \tag{8}$$

If S_w is non-singular, we should solve the generalized eigen problem of the transformation matrix W by the linear discriminant analysis in Fig. 2. This transformation matrix should maximize the between class scatter matrix and minimize the within class scatter matrix [19]. There are many solutions to solve the generalized eigen problem [20]. One method for solving this eigen problem is to take the inverse of S_w and solve the problem by using $S_w^{-1} S_b W = W\lambda$. This task is derived from Eq. 9.

$$S_b W = S_w W \lambda \tag{9}$$

λ is a diagonal matrix containing the eigen values of the matrix $S_w^{-1} S_b$. The above algorithm is optimal only when the within class scatter matrix is singular. If the within class scatter matrix is non-singular, we should use the direct LDA technique [15]. The direct LDA is performed in the following steps as shown in Fig. 2.

The first step is related to find the eigen vectors of the between class scatter matrix $S_b = P_b^T P_b$, where P_b is calculated by subtracting the mean face images of each class from the mean face image of all images as expressed in Eq. 10.

$$P_b = [m_1 - m, m_2 - m, ..., m_c - m] \tag{10}$$

The second step takes the most significant eigen values and corresponding eigen vectors V. These eigen vectors are used

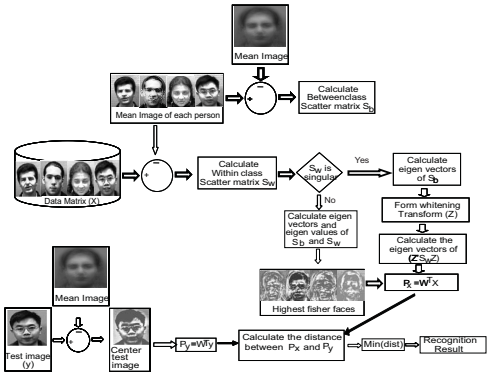


Fig. 2. Linear discriminant analysis technique for face recognition

to calculate $Y = P_b V$ and $D_b = Y^T S_b Y$. This leads to the evaluation of the whitening transform as $Z = Y D_b^{-1/2}$. S_b and S_w are projected onto the new subspace spanned by Z. The small matrix $Z^T S_w Z$ can be diagonalized. The relationship between them is expressed in Eq. 11.

$$U^T Z^T S_w Z U = \lambda_w \tag{11}$$

U and λ_w are the eigen vectors and eigen values of the matrix $Z^T S_w Z$. The corresponding eigen matrix is represented as R. The overall transformation matrix is calculated from $W = ZR$. A new transformation can be performed by using the linear transformation of the original space into a new reduced dimensional feature space $P_x = W^T X$ (i.e project this transformation matrix on to the train data set) [6].

The next operation is concerned with the projection of this transformation matrix on to the test data sets to obtain P_y. The best match is found by calculating the distance between P_x and P_y using the distance measure technique. The overall linear discriminant analysis technique for face recognition is shown in Fig. 2.

Technique	Principal component analysis (PCA)	Independent component analysis (ICA)	Linear discriminant analysis (LDA)
Year	1990	1999	1997
Iterative	No	Yes	No
Class Information usage	No	No	Yes
Order of statistics	Second order	Higher order	Second order
Recognition rate (for 80 persons database)	70%	79%	89%
Speed	medium	very low	high
Scalability	low	low	high

Fig. 3. Comparison of linear subspace techniques (PCA, ICA and LDA)

The performance of different linear subspace techniques like PCA, ICA and LDA is evaluated. Experiments are conducted to understand the performance (recognition rate and speed) of these linear subspace techniques over the FERET database. Among linear subspace techniques, LDA gives both high recognition rate and speed when compared with PCA and

ICA as shown in Fig. 3. But LDA is not scalable, and the recognition rate is also not sufficient for real world applications. In linear subspace techniques, the computational load and memory requirements are dramatically increasing with the size of database.

III. Radon transform

The two dimensional radon transform was introduced by Austrian mathematician Johann Radon in 1917. This transform gives the integral of the set of lines present in a given image [10]. Due to this, it captures the direction of the local features (lines, curves and circles) which are present in the image. This transform is useful in many line, circle and curve detection applications, related to image processing and computer vision [10]. The radon transform of the two dimensional function $f(x,y)$ in (r,θ) plane (Fig. 4(a)) is shown in Eq. 12

$$R(r,\theta)[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos\theta+ysin\theta-r)dxdy \quad (12)$$

Where $\delta(.)$ function is the Dirac function, $r\epsilon[-\infty,\infty]$ is the

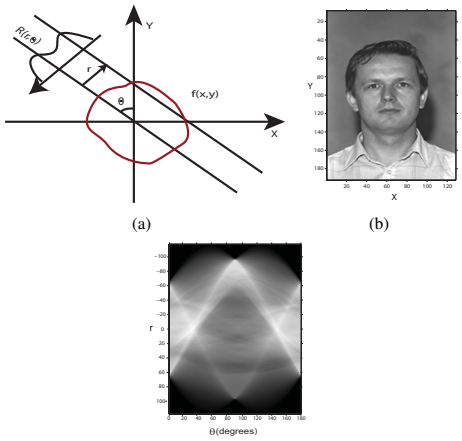


Fig. 4. (a) The radon transform of an image (b) Shows the original image (b) Radon transform of the image with an angle 0 to 180

perpendicular distance of a line from the origin and $\theta\epsilon[0,\pi]$ is the angle formed by the distance vector [10]. The δ function converts the two dimensional integral to a line integral dl along the line $xcos\theta + ysin\theta = r$. The simplified form of $R(r,\theta)[f(x,y)]$ is Rf shown in Eq. 13

$$Rf = \int_{-\infty}^{\infty} f(rcos\theta - lsin\theta - rcos\theta + lsin\theta)dxdy \quad (13)$$

The transformed function (r,θ) is referred to as the sinogram of $f(x,y)$. The δ function transforms the point in f to sinusoidal line δ function in (r,θ) plane. The Rf is defined as a function of straight lines. The radon transform of the two dimensional image shown in Fig. 4(b), extracts the direction of the lines present in that image, as shown in Fig. 4(c). The sinogram (Fig. 4(c)) of the given image has 181 radon projections. Each projection in the image is a feature vector.

IV. Wavelet transform

Morlet introduced the wavelet transform in the early 1980's [21]. Wavelet is named 'ondelette' in French, which means 'small waves' [11]. A wavelet gives both the spatial and frequency information of the images. In the frequency representation, the signal is cut into several parts and each part is analyzed separately. Commonly used discrete wavelets are daubechies wavelets [22]. Wavelets with one level decomposition is performed by using the high pass filter g and the low pass filter h. Convolution with the low pass filter gives the approximation information, while convolution with the high pass filter leads to the detail information [23]. The wavelet decomposition process of two dimensional signal $f(x,y)$ is shown in Fig. 5. The overall process is modeled in Eqs.(14 - 17).

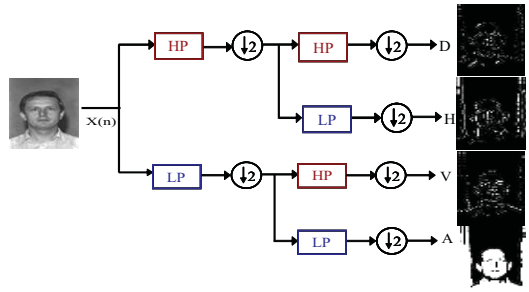


Fig. 5. Wavelet coefficients decomposition in discrete wavelet transform

$$A = [h * [h * f]_x \downarrow 2]_y \downarrow 2 \quad (14)$$

$$H = [g * [h * f]_x \downarrow 2]_y \downarrow 2 \quad (15)$$

$$V = [h * [g * f]_x \downarrow 2]_y \downarrow 2 \quad (16)$$

$$D = [g * [g * f]_x \downarrow 2]_y \downarrow 2 \quad (17)$$

The star $(*)$ represents the convolution operation, and $\downarrow 2$ represents the downsampling by 2 along the direction x or y [11]. To correct this sample rate, the down sampling of the filter by two is performed (by simply throwing away every second coefficient). The daubechies wavelets have many wavelets functions. In this work, $db4$ (because of the symmetry) is used. $db4$ leads to the four wavelet coefficients A, H, V and D and the corresponding images. In this decomposition A gives the approximation information, and the image is a blurred image as shown in Fig. 5. H gives the horizontal features, V gives the vertical features and D gives the diagonal features present in the image. The wavelet coefficient A gives the high performance, when compared to the remaining three wavelet coefficients. Further D gives the less performance. Using the $A+H+V+D$ wavelet coefficients leads to a performance, which is nearly equal to the A's performance.

V. Cellular neural network

The concept of CNN, also called cellular neural networks was introduced in 1988 by Leon O.Chua and Lin Yang. The basic building block in the CNN model is the cell. The CNN model consists of regularly spaced array of cells. It can be identified as the combination of cellular automata [24] and neural networks [25]. The adjacent cells communicate directly through their nearest neighbours and other cells communicate indirectly, because of the propagation effects in the model. The original idea was to use an array of simple, non-linearly coupled dynamic circuits to process, parallely, large amounts of data in real time [25].

Cells are multiple input, single output nonlinear processors. Cells in the CNN processor contain fixed location and fixed topology. Inputs, initial state, and output variables are used to define the CNN processor behavior. Professor Leon O.Chua proposed the diagram of an isolated cell, as shown in Fig. 6. The state variable is not observable from outside the cell itself.

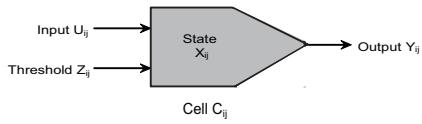


Fig. 6. Representation of an isolated cell

The cell is a lumped circuit, and it contains both linear and nonlinear elements, such as resistors, capacitors and nonlinear controlled sources as shown in Fig. V. The CNN processor is modeled by Eqs.(18 - 19), with x_i, y_i and u_i as state, output and input variables respectively. The schematic model of a CNN cell is shown in Fig.8

$$\dot{x}_{ij} = -x_{ij} + \sum_{c(j)\in N_r(i)} A_{ij}y_{ij} + B_{ij}u_{ij} + I \qquad (18)$$

$$y_{ij} = \frac{1}{2}\left(|x_{ij}+1| - |x_{ij}-1|\right) \qquad (19)$$

The coefficients A_{ij} and B_{ij} values, synaptic weights, completely define the behavior of the network, with given input and initial conditions, as shown in Eq. 18. These values are called the templates. For the ease of representation, they can be represented as a matrix. We have three types of templates: the first one is feedforward or control template, the second is feedback template and the third is bias. All these space invariant templates are called cloning templates. CNNs are particularly interesting, because of their programmable nature i.e. changeable templates.

These templates values and synaptic weights completely define the behavior of the network, with given input and initial conditions. These templates are expressed in the form of a matrix and are repeated in every neighborhood cell. The template set for $r = 1$ CNN contains 19 coefficients (A-template 9, B-template 9 and bias 1).

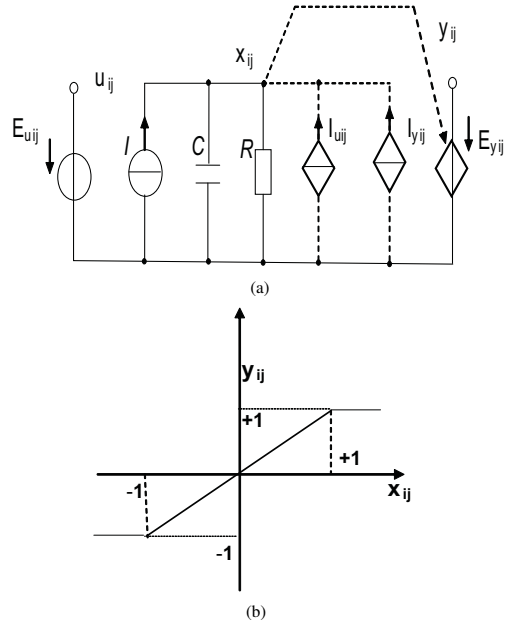


Fig. 7. (a)Electronic circuit model of the isolated cell (b) The classical output nonlinear function for each cell

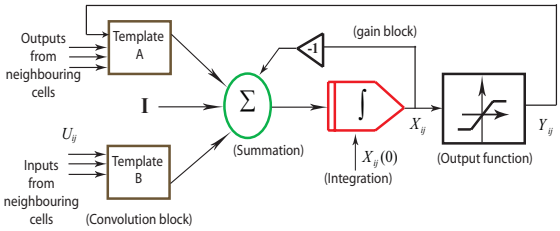


Fig. 8. Schematic representation of the CNN

$$A = \begin{bmatrix} A_{-1,-1} & A_{-1,0} & A_{-1,1} \\ A_{0,-1} & A_{0,0} & A_{0,1} \\ A_{1,-1} & A_{1,0} & A_{1,1} \end{bmatrix};$$

$$B = \begin{bmatrix} B_{-1,-1} & B_{-1,0} & B_{-1,1} \\ B_{0,-1} & B_{0,0} & B_{0,1} \\ B_{1,-1} & B_{1,0} & B_{1,1} \end{bmatrix};$$

The genetic algorithm is used to estimate the A, B and I templates, depending upon the given application. The template set is unique for each application. In this work, we use the genetic algorithm to obtain the template set for the ORL database.

A. Genetic algorithm

In order to extract the facial features from a frontal face image, we assume that the template set values will have symmetrical behavior, as the front view of the face is symmetrical.

Because of this symmetry, instead of 19 template elements, we are calculating the 11 template elements (A-template 5, B-template 5 and bias 1). Each template element is encoded with 32 bit floating point format. Genetic algorithm (GA) uses the papulation of binary strings called chromosomes. In the learning process, initially 72 random chromosomes, with length of $11*32$ bits each, are constructed. Genetic Algorithm is explained in detail in the following steps:

- Construct the random population matrix with size $72X(11*32)$ i.e. each row represents a chromosome (for 11 template elements) of length $11*32 = 352$.
- The IEEE 754 floating point standard is used to calculate the template (A, B and I) elements from each chromosome [24]. In each chromosome first 11 bits represents the first bit of the 11 template elements, and second 11 bits represents the second bit of the 11 template elements so on as given in Eq. 20.

$$S = [A_{11}, A_{12}, A_{13}, A_{21}, A_{22}, B_{11}, B_{12}, B_{13}, B_{21}, B_{22}, I] \quad (20)$$

- After template calculation, these templates are given as input to the CNN. The first CNN works with the template of the first chromosome. After the CNN output appears as stable, cost function is calculated by using this CNN output image P and the target image T. This process is repeated for each chromosome template sets in the population matrix [24]. The cost function is selected as shown in Eq. 21.

$$cost(A, B, I) = \sum_{i}^{m} \sum_{j}^{n} P_{i,j} \oplus T_{i,j} \quad (21)$$

  Here m,n are the number of pixels of the image. $\oplus$ represents the XOR operation.
- After calculating the cost function, the fitness function for each chromosome is evaluated as given in Eq. 22.

$$fitness(A, B, I) = m*n - cost(A, B, I) \quad (22)$$

- The whole process is repeated for each chromosome until the fitness value exceeds the stop criteria. The stop criteria is considered as $stcriteria = 0.99*m*n$. This maximum fitness value of the chromosome in the population matrix is selected.
- The next step is reproduction. In this process, the fitness values corresponding chromosomes are sorted in descending order. All the fitness values are normalized with the sum of the fitness values. The bad fitness value corresponding chromosomes are deleted. The most successful chromosomes will produce the next generation.
- Take the first highest fitness values corresponding chromosomes S_1 and S_2, apply the crossover and mutation operations to generate the children [24]. Crossover operation exchanges the substrings between the two chromosomes S_1 and S_2. In this work, one-point crossover is used and its first cross site is selected with chromosome length of the uniform probability. If the mutation probability is set to 0.01 then 253 bits are selected randomly and then they are inverted.
- Take these new chromosomes and apply the same steps from template calculation to stop criteria.

This learning process is repeated to find the best chromosome. After satisfying the stop criteria, the template elements are calculated from the best chromosome. The template elements to extract features from the frontal face images for ORL database are obtained as:

$$A = \begin{bmatrix} 2.7612 & 7.3152 & 1.7566 \\ 1.5916 & 8.5273 & 1.5916 \\ 1.7566 & 7.3152 & 2.7612 \end{bmatrix};$$

$$B = \begin{bmatrix} -6.1912 & 2.8350 & -7.9270 \\ 1.3044 & -2.7349 & 1.3044 \\ -7.9270 & 2.8350 & -6.1912 \end{bmatrix};$$

$I = 0.4414$

The corresponding best chromosome is

$S =$ [00000101010110011110100011000010100110000101001100001010011000010100110000101001100001010111010110101110101001111000111111000001100001011001010000000111110010101001100001001101110110001101001010101011001101110111110110101001111010111100110010111001100101100110010011001101111000101101000000000110010100100111001010101011011101111010100110001100100100000]

(a) (b)

Fig. 9. (a) The input image of the genetic cellular neural network (b) The genetic cellular neural network output image

The two dimensional image shown in Fig. 9(a), is given as the input image for CNN to extract the important frontal facial features present in that image and the output image with extracted feature set is shown in Fig. 9(b).

VI. Experimental Result

In this section, we evaluate the performance of the wavelet and radon transform based feature extraction approaches using FERET database. The performance of the CNN based approach is compared to other stated face recognition approaches over the ORL database. The performance is evaluated over the FERET database for frontal images (fa or fb), pose variant with an angle 67.5 half left or right shifted images (hr or hl), and pose variant with an angle 90 profile left or right shifted images (pr or pl) [26]. For the ORL database, the performance is evaluated for facial expressions and varying light conditions.

1) Performance evaluation of radon and wavelet transforms: : The radon transform gives the direction of the local features (lines, circles). Radon transform preserves the variation in pixel intensities. While computing the radon projections, the pixel intensities along a line are added. This process extracts the spatial frequency components in the direction of radon projection. When features are extracted using radon transform, the variations in this facial frequency are also boosted. The wavelet transform gives the spacial and frequency components present in an image.

A. Different wavelet functions versus recognition rate

Daubechies wavelets contain different wavelet functions. The recognition rates of two different wavelet functions $db1$ and $db4$ are compared in Fig. 10. $db1$ stands for the haar wavelet and it encodes the constant component. $db4$ encodes both constant and linear components. The $db4$ performance is high when compared to $db1$.

B. Different wavelet coefficients versus recognition rate

In the $db4$ daubechies wavelets, there are four wavelet coefficients. These coefficients vary in terms of the wavelet functions. The four wavelet coefficients are A, H, V and D. The wavelet coefficient A gives approximate information on the features. H, V, and D gives the information about horizontal, vertical and diagonal features present in the given image respectively.

The wavelet coefficient A gives the high recognition rate, when compared to the remaining three wavelet coefficients. Further D gives the less recognition rate (see Fig. 10). Using the $A+H+V+D$ wavelet coefficients leads to a recognition rate, which is nearly equal to the A's recognition rate.

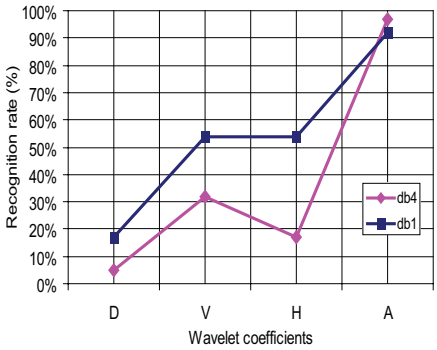
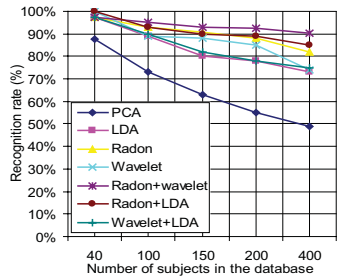
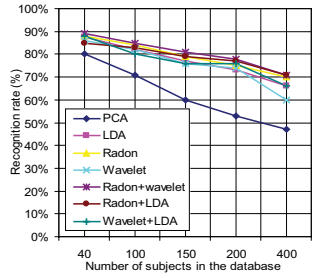


Fig. 10. Performance comparison of different wavelet function $db1$ and $db4$

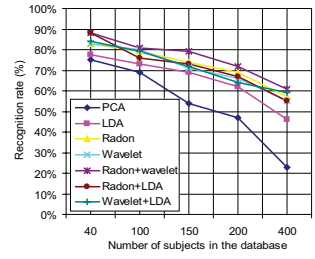
The next experiments are conducted on a FERET database with one frontal image (fb) for each subject as test image, and five images in different poses for each subject in train database. The performance evaluation is shown in Fig. 11(a). The experiments are repeated with pose variant images like hr and pr as test image for each subject, and five images



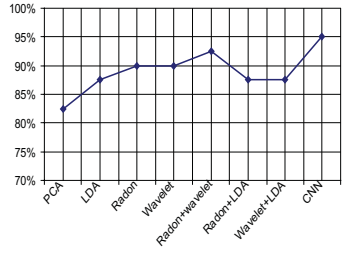
(a)



(b)



(c)



(d)

Fig. 11. (a) Performance comparison of different face recognition approaches with front images (FERET database) (b) Performance comparison of different face recognition approaches with half right images (FERET database) (c) Performance comparison of different face recognition approaches with profile right images (FERET database) (d) Performance comparison of different face recognition approaches with ORL database

excluding the test image for each subject in train database. The results are shown in Fig. 11(b) and Fig. 11(c) respectively. For best matching, the euclidean distance measure is used here.

The recognition rate depends upon the number of subjects in the data set. It is difficult to recognize a subject in the large data set than in the small data set. The experiments are conducted with different sizes of the FERET database, by using linear subspace techniques (principal component analysis (PCA), linear discriminant analysis (LDA)), radon transform and wavelet transform. In applying linear subspace techniques for large databases, computational load and memory requirements increases dramatically with the size of the database. This effects the performance of PCA and LDA on large data sets as shown in Fig. 11.

The radon transform and wavelet transform are mostly independent of size of the database. The combination of radon and wavelet transform gives the multi-resolution features, which are more useful in face recognition. This has been validated with the experimental results shown in Fig. 11. Even though the combination of radon and wavelet transform gives better performance, there is still a need for improvement in pose variant face recognition as shown in Fig. 11(b) and Fig. 11(c).

1) Performance evaluation of cellular neural networks: : The CNN based face recognition approach and other stated approaches are applied on ORL database. The ORL database contains images of 40 subjects. All images are taken in frontal position against a dark homogeneous background. The performance of various algorithms are evaluated using ORL database are shown in Fig. 11(d). CNN with its parallel computing paradigm promises to outperform the other approaches over the ORL database as shown in Fig. 11(d).

VII. Conclusion

The face recognition performance has been systematically evaluated by using different sizes of the database. To improve the performance of the face recognition technique, wavelets, radon and combination of both radon and wavelet transform have been proposed to extract the nonlinear features. The results of the evaluation have shown that the recognition rate is considerably increased with the combination of both radon and wavelet transform compared to PCA and LDA. In addition to these two approaches, this work also shows CNN based feature extraction approach for face recognition outperforms both radon and wavelet transforms for ORL database. However, this should be validated for FERET database, where the images are in different poses. The CNN algorithm should able to detect the pose, and then apply the appropriate template to extract the relevant feature set.

Future work should focus on the recognition algorithm performing over videos, as many applications demand real time recognition. Further, such a system may be integrated in driver assistance system to either recognize the driver of a car, or extract facial expressions that may provide information about his mood or fatigue.

References

[1] I. Guyon and A. Elisseeff, "An Introduction to Feature Extraction," *Zurich Research Laboratory*, (2004).

[2] M. Aleemuddin, "A Pose Invariant Face Recognition system using Subspace Techniques," *Deanship of Graduate studies*, (2004).

[3] G. Shakhnarovich and B. Moghaddam, *Face Recognition in Subspaces*. Springer-verlag, May (2004).

[4] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, (1988).

[5] A. SATO, H. IMAOKA, T. SUZUKI, and T. HOSOI, "Advances in Face Detection and Recognition Technologies," *NEC Journal of Advanced Technology*, vol. 2, no. 1, (2005).

[6] O. Toygar and A. Acan, "Face Recognition using PCA, LDA and ICA approaches on colored images," *Electrical and Electronics engineering*, vol. 3, no. 1, pp. 735–743, (2003).

[7] R. Brunelli, T. Poggio, and I. P. Trento, "Face recognition through geometrical features," in *European Conference on Computer Vision (ECCV)*, pp. 792–800, (1992).

[8] R. Brunelli and T. Poggio, "Face recognition: Features vs. templates.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, (1993).

[9] B. J. Lei, E. A. Hendriks, and M. Reinders, "On Feature Extraction from Images," *Technical Report on MCCWS project*, (1999).

[10] Q. W. Yan CHEN and X. HE, "Human Action Recognition by Radon Transform," *IEEE International Conference on Data Mining Workshops*, May (2008).

[11] N. Shams, I. Hosseini, M. Sadri, and E. Azarnasab, "Low cost fpga-based highly accurate face recognition system using combined wavelets with subspace methods," pp. 2077–2080, (2006).

[12] P. N. Belhumeur, J. a. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, July (1997).

[13] W. S.Yambor, "Analysis of PCA based and Fisher discriminant based image recognition algorithms," *Degree of Master of Science*, (2000).

[14] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with pca and ica," *Comput. Vis. Image Underst.*, vol. 91, no. 1-2, pp. 115–137, (2003).

[15] P. N.Belhumeur, J. P.Hespanha, and D. J.Kriegman, "Eigenfaces vs. Fisherfaces:Recognition Using Class Specific Linear Projection," vol. 19, no. 7, (1997).

[16] J. Kim, M.-J. Choi, M.-J. Yi, and M. Turk, "Effective representation using ica for face recognition robust to local distortion and partial occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1977–1981, (2005).

[17] A. Hyvrinen, "The Fixed-Point Algorithm and Maximum Likelihood estimation for Independent Component Analysis," pp. 1–5, (1999).

[18] W. S. Y. B. A. D. J. R. Beveridge, "Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures," *Department of Computer Science*, (2000).

[19] P. N.Belhumeur, J. P.Hespanha, and D. J.Kriegman, "Eigenfaces vs. Fisherfaces:Recognition Using Class Specific Linear Projection," *European Conference on Computer Vision*, (1996).

[20] W.Zhao, R.Chellappa, and P.J.Phillips, "Subspace Linear Discriminant Analysis for Face Recognition," *Department of Electrical and Electronic Engineering*, (1999).

[21] C. Garcia, G. Zikos, and G. Tziritas, "A wavelet-based framework for face recognition," in *Workshop on Advances in Facial Image Analysis and Recognition Technology, 5 th European Conference on Computer Vision*, pp. 84–92, Publications, (1998).

[22] M. I. M. D. Fatma H. Elfouly, Mohamed I. Mahmoud and S. Deyab, "Comparison between haar and daubechies wavelet transformations on fpga technology," *International Journal of Computer, Information, and Systems Science, and Engineering*, vol. 2, no. 1, pp. 1047–1061, (2006).

[23] C. C. LIU, D. Q. Dai, and H. Yan, "Local Discriminant Wavelet Packet Coordinates for Face Recognition," *Journal of Machine learning Research*, pp. 1165–1195, May (2007).

[24] T. R. Tibor Kozek and L. . Chua, "Genetic Algorithm for CNN Template Learning," *IEEE Transactions on circuits and systems*, vol. 40, no. 6, (1993).

[25] L. Chua and T. Roska, *Cellular neural networks and visual computing: foundations and applications*. Cambridge University Press, (2005).

[26] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," vol. 16, pp. 295–306, April (1998).