Nonlinear Feature Extraction Approaches with Application to Face Recognition over Large Databases

Hima Deepthi Vankayalapati
Institute of Smart Systems Technologies
University of Klagenfurt
9020 Klagenfurt, Austria

Kyandoghere Kyamakya
Institute of Smart Systems Technologies
University of Klagenfurt
9020 Klagenfurt, Austria

***Abstract*—The extraction of required features from the facial image is an important primitive task for face recognition. This paper evaluates different nonlinear feature extraction approaches, namely wavelet transform, radon transform and cellular neural networks (CNN). The scalability of the linear subspace techniques is limited as the computational load and memory requirements increase dramatically with the large database. In this work, the combination of radon and wavelet transform based approach 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, Cellular neural network, Wavelet transform, Radon transform.**

I. INTRODUCTION

The human skill of identifying thousands of people even after so many years excited many researches to focus on the face recognition system. The majority of the real world person identification and verification applications demand more robust, scalable and computationally efficient face recognition techniques under complex viewing and environmental conditions. Research concerning the face recognition started nearly in 1960's [1]. Different face recognition techniques have been proposed during last decades namely feature based, model based and appearance based techniques [2], [3]. In feature based techniques, the overall technique describes the position and size of each feature (eye, nose, mouth or face outline) [4]. In this approach, extracting features in different poses (viewing condition) and lighting conditions is a very complex task. For applications with large database, we have large set of features with different sizes and positions, making it difficult to identify the required feature points [4]. 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 high 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 database [3]. The availability of large 3D data is also one of the complex task making the model based methods not suitable for real world application dealing with large databases.

The performance of the appearance based techniques heavily depends upon the quality of the extracted features from image. So in these techniques, feature extraction is an important issue. In computer vision, feature is a set of measurements. Each measurement contains piece of information and specifies the property or characteristics of the object present in the image [5]. The appearance based linear subspace techniques extract the global features, as these techniques use the statistical properties like mean, variance of the image [3]. The major difficulty in applying these techniques over large databases is, the computational load and memory requirements for calculating features increase dramatically for large databases [2]. In order to increase the performance of the face recognition techniques, the nonlinear feature extraction techniques are proposed.

The objectives of this paper are threefold: (1) to show the potentials of the nonlinear feature extraction approaches; (2) to evaluate the scalability of nonlinear approaches and linear subspace approaches over the FERET database; and (3) to present the use of CNN in extracting nonlinear features using the ORL database.

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

II. RADON TRANSFORM

The two dimensions and three dimensions radon transform was introduced by Austrian mathematician Johann Radon in 1917. This transform gives the integral of the set of lines present in the given image [6]. Due to this, it captures the direction of the local features (lines, curve and circles) which

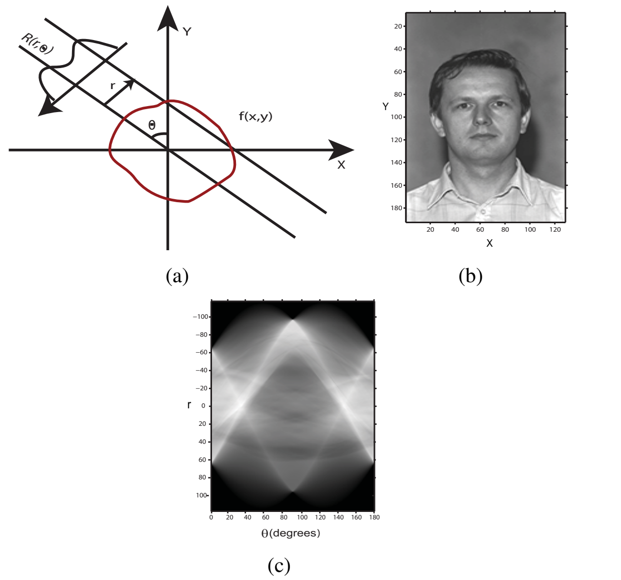


Fig. 1. (a) The radon transform of an image (b) Shows the original image (b) Radon transform of the image with an angle 0 to 180

are present in the image. This transform is useful in many line, circle and curve detection applications related to image processing and computer vision [6]. The radon transform of the two dimensional function $f(x, y)$ in (r, θ) plane (Fig. 1(a)) is shown in Eq. 1

$$R(r,\theta)[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(xcos\theta + ysin\theta - r)dxdy \tag{1}$$

Where $\delta(.)$ function is the Dirac function, $r\epsilon[-\infty, \infty]$ is the perpendicular distance of a line from the origin and $\theta\epsilon[0, \pi]$ is the angle formed by the distance vector [6]. 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. 2

$$Rf = \int_{-\infty}^{\infty} f(rcos\theta - lsin\theta - rcos\theta + lsin\theta)dxdy \tag{2}$$

The transformed function (r, θ) is referred to as 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 function of straight lines. The radon transform of the two dimensional image shown in Fig. 1(b), extracts the direction of the lines present in that image as shown in Fig. 1(c). The sinogram (Fig. 1(c)) of the given image has 181 radon projections. Each projection in the image is a feature vector.

III. Wavelet transform

Morlet introduced the wavelet transform in the early 1980's [7]. Wavelet is named ondelette in french word, which means small waves [8]. 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

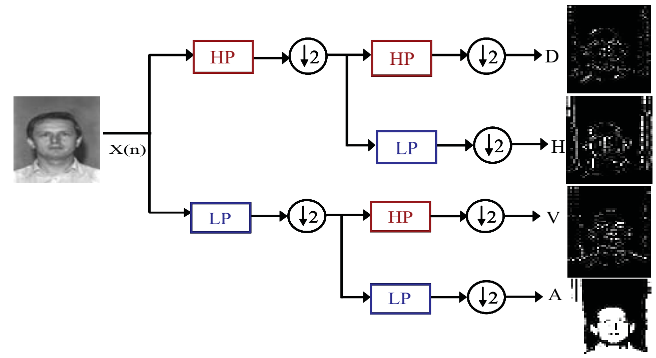


Fig. 2. Wavelet coefficients decomposition in discrete wavelet transform

wavelets [8]. 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 [9]. The wavelet decomposition process of two dimensional signal $f(x, y)$ is shown in Fig. 2. The overall process is modeled in Eqs.(3 - 6).

$$A = [h * [h * f]_x \downarrow 2]_y \downarrow 2 \tag{3}$$

$$H = [g * [h * f]_x \downarrow 2]_y \downarrow 2 \tag{4}$$

$$V = [h * [g * f]_x \downarrow 2]_y \downarrow 2 \tag{5}$$

$$D = [g * [g * f]_x \downarrow 2]_y \downarrow 2 \tag{6}$$

The star $(*)$ represents the convolution operation and $\downarrow 2$ represents the downsampling by 2 along the direction x or y [8]. 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 has 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 blur image as shown in Fig. 2. 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.

IV. Cellular neural network

The concept of CNN, also called cellular neural network was introduced in 1988 by Leon O.Chua and Lin Yang. 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 [10]. It is a large array of interconnected nonlinear dynamic systems called cells. It can be identified as the combination of cellular automata [11] and neural networks [10]. The CNN processor is modeled by Eqs.(7 - 8), with x_i,

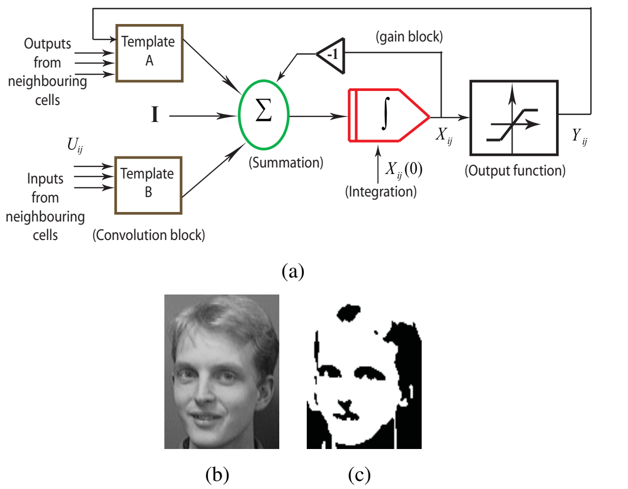


Fig. 3. (a) Schematic representation of the CNN (b) The input image (c) The genetic cellular neural network output image

y_i and u_i as state, output and input variables respectively. The schematic model of a CNN cell is shown in Fig.3(a)

$$\dot{x}_{ij} = -x_{ij} + \sum_{c(j)\in N_r(i)} A_{ij}y_{ij} + B_{ij}u_{ij} + I \tag{7}$$

$$y_{ij} = \frac{1}{2}\left(|x_{ij}+1| - |x_{ij}-1|\right) \tag{8}$$

The coefficients A_{ij} and B_{ij} in Eqs.7 are the feedback template and control or the feedforward template respectively. CNNs are particularly interesting because of the programmable nature i.e. changeable templates.

These templates values, synaptic weights, completely define the behavior of the network with given input and initial conditions. These templates are expressed in the form of 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). 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 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 the template set values will have the 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 [11]. 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. 9.

$$S = [A_{11}, A_{12}, A_{13}, A_{21}, A_{22}, B_{11}, B_{12}, B_{13}, B_{21}, B_{22}, I] \tag{9}$$

- After template calculation, these templates are given as input to CNN. 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 [11]. The cost function is selected as shown in Eq. 10.

$$cost(A, B, I) = \sum_{i}^{m}\sum_{j}^{n} P_{i,j} \oplus T_{i,j} \tag{10}$$

Here m,n are the number of pixels of the image. $\oplus$ represents the XOR operation.
- After calculating the cost function, fitness function for each chromosome is evaluated as given in Eq. 11.

$$fitness(A, B, I) = m * n - cost(A, B, I) \tag{11}$$

- 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 [11].

This learning process is repeated to find the best chromosome. After satisfying the stop criteria, the template elements are calculated from the chromosome by applying IEEE 754 floating point conversion. 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 two dimensional image shown in Fig. 3(b), is given as 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. 3(c).

V. Experimental Result

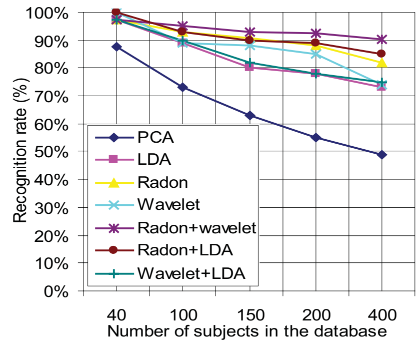
In this section, we evaluate the performance of the wavelet and radon transform based nonlinear feature extraction approaches using FERET database. The performance of CNN based approach is compared to other stated nonlinear approaches over the ORL database. The performance is evaluated over the FERET 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). For the ORL database, the performance is evaluated for facial expression 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 is computed. 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. The low frequency component A in wavelet transform gives higher recognition rate as it gives approximate information present in an image.

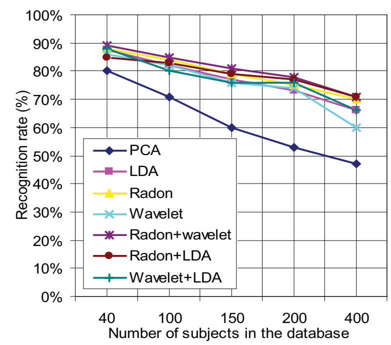
The experiments are conducted on FERET database with one frontal image (fb) for each subject as test image and five images in different poses for each subject in train database. The performance evaluation is shown in Fig. 4(a). The experiments are repeated with pose variant images like hr and pr as test image for each subject and five images excluding the test image for each subject in train database. The results are shown in Fig. 4(b) and Fig. 4(c) respectively. For best matching, the euclidean distance measure is used.

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 size 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. 4.

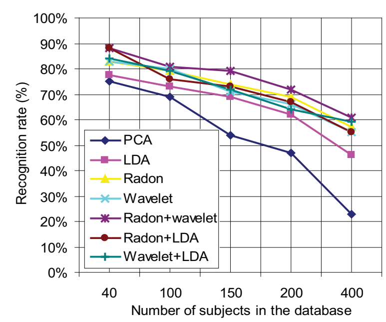
The radon transform and wavelet transform are mostly independent of size of the database. The combination of radon and wavelet transform gives the multi-resolution features which are more useful in face recognition. This has been validated with the experimental results shown in Fig. 4. Even though the combination of radon and wavelet transform gives better



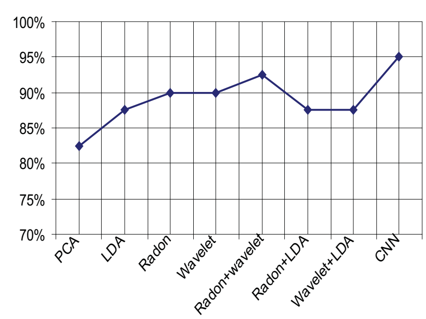
(a)



(b)



(c)



(d)

Fig. 4. (a) Performance comparison of different approaches with front images (FERET database) (b) Performance comparison of different approaches with half right images (FERET database) (c) Performance comparison of different approaches with profile right images (FERET database) (d) Performance comparison of different approaches with ORL database

performance, there is still a need for improvement in pose variant face recognition as shown in Fig. 4(b) and Fig. 4(c).

2) 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. 4(d). CNN with its parallel computing paradigm promises to outperform the other approaches over the ORL database as shown in Fig. 4(d).

VI. 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 for 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 out performs 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 works 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] 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).

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

[3] 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).

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

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

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

[7] 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).

[8] 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).

[9] 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).

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

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