2012 International Conference on Computer Technology and Science (ICCTS 2012)
IPCSIT vol. 47 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V47.20

A Dynamic Data Structure for Real Time Face Recognition

Vandana Agarwal[1+] and Surekha Bhanot[2]

[1]Department of Computer Science and Information Systems
[2]Department of Electrical and Electronics/Instrumentation
Birla Institute of Technology and Science, Pilani-333031 INDIA

Abstract. Intelligent Face Recognition systems, capable of recognizing persons' identity from their face images for surveillance and authentication purposes, are required to respond in real time. The recognition of the identity of an unknown face in real time is a major challenge due to heavy computational requirements of training and testing large number of face images. The data structure plays an important role in reducing the access time to fetch the data for execution by the CPU. In this paper, a dynamic data structure is proposed which enables real time face recognition. Haar wavelets' multi-resolution features are used corresponding to all scales that describe each face image. The proposed work enables extraction of the features for all face images in the given face data base and trains the classifier in maximum 5-6 seconds while it takes only 0.02 second on an average, i.e. near real time, to recognize an unknown face image.

Keywords: Classifier, Computer Vision, Dynamic Data structure, Haar Wavelets.

1. Introduction

Face recognition has been a challenging problem for decades, where researchers from various disciplines such as computer vision, artificial intelligence, pattern recognition and image processing have been experimenting with newer algorithms to automate the face recognition process with higher accuracy. The efforts have also been in handling the variations due to pose, illumination, expression, scale etc. The face recognition process consists of different stages- feature extraction[4,5], feature selection, training and testing [Fig. 1].

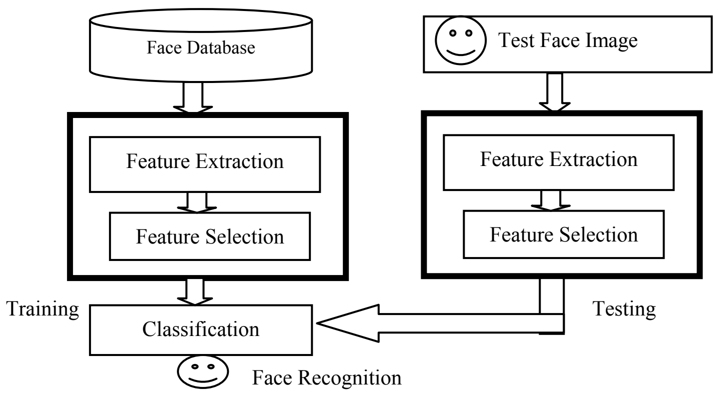

Fig. 1. Face Recognition System

+ Corresponding author.
E-mail address: vandana@bits-pilani.ac.in

The face recognition problem is posed as a classification problem [2, 3]. A classifier is trained to recognize faces of known persons using their face images, called as training samples. An unseen test image is subjected for prediction by the classifier. It has been observed that the wavelets provide sound mathematical basis for efficient feature extraction for face recognition[6,7]. For higher recognition accuracy, the extracted features should have high discriminating power to predict the correct class. Wavelets coefficients capture the information content with respect to the frequency and time parameters[1]. The volume of the wavelet features extracted at each scale and resolution is large. In this paper, a dynamic data structure is proposed which handles the huge memory needs efficiently enabling the real time face recognition.

2. Haar Wavelet Features

The features for face images are extracted using Haar wavelets [1]. The unit height, unit width Haar scaling function φ(x) is described below

$$\varphi(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The Haar wavelet function ψ(x) is described as follows

$$\psi(x) = \begin{cases} 1 & 0 \le x < 0.5 \\ -1 & 0.5 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The face image is represented as the approximation and detail wavelet coefficients. The analysis requires two dimensional scaling function φ(x,y) and three two dimensional wavelets $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$. These wavelets measure the intensity variations along columns (horizontal details), rows (Vertical details) and the variations along the diagonal respectively. A 2-dimensional scaling function is defined as φ(x,y) = φ(x)φ(y). The directionally sensitive 2- dimensional wavelets $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ are defined respectively as products ψ(x)φ(y) , φ(x)ψ(y) and ψ(x)ψ(y). The scaled and translated basis functions are defined as

$$\varphi_{j,m,n}(x,y) = 2^{j/2}\varphi(2^{j/2}\mathrm{X} - \mathrm{m}, 2^{j/2}y - n) \tag{3}$$

$$\psi^{i}_{j,m,n}(x,y) = 2^{j/2}\psi^{i}(2^{j/2}\mathrm{X} - m, 2^{j/2}y - n) \tag{4}$$

Where i = {H,V,D}; m and n represent the translation in x and y directions respectively; H, V and D represent the horizontal, vertical and diagonal directions respectively. The discrete wavelet transform of image f(x,y) of size M×N is given by

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\, \varphi_{j_0 m,n}(x,y) \tag{5}$$

$$W^{i}_{\psi}(j, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\, \psi^{i}_{j,m,n}(x,y) \tag{6}$$

where i = {H,V,D} and j_0 is an arbitrary starting scale. The coefficients $W_\varphi(j_0, m,n)$, for m = 0,1,2,....,M-1 and n = 0,1,2,...,N-1, define the approximation of the original face image f(x,y) at scale j_0. The $W^i_\psi(j, m,n)$ coefficients define the details in the three directions for higher scales j≥j0. A value J is chosen such that J = max($\log_2(M)$, $\log_2(N)$) and j_0 is initialized as 0. The 2D discrete Fast Wavelet Transform (FWT) is implemented using digital filters and down-samplers. This algorithm exploits the relationship between the coefficients of adjacent scales. The highest scale coefficients are assumed to be samples of the function itself. That is $W_\varphi(J, m,n)$ is considered to be the input image $f(x,y)$, where J is the highest scale such that $W_\varphi(J, m,n)$ is initialized as original face image. The rows of $W_\varphi(j,m,n)$ are convolved with the analysis filter functions $h_\varphi(-n)$ and $h_\psi(-n)$, where $h_\varphi(n)$ and $h_\psi(n)$ are the scaling and wavelet vectors. These are defined as follows

$$h_\varphi(n) = \begin{cases} 1/\sqrt{2} & n = 0,1 \\ 0 & otherwise \end{cases} \quad \text{and} \quad h_\psi(n) = \begin{cases} 1/\sqrt{2} & n = 0 \\ -1/\sqrt{2} & n = 1 \\ 0 & otherwise \end{cases} \tag{7}$$

The function $h_\varphi(n)$ represents the low pass filter and $h_\psi(n)$ represents the high pass filter. At each iteration, the rows of the two dimensional approximation coefficients $W_\varphi(j+1,m,n)$ (visualized as a 2- dimensional matrix) are convolved with the functions $h_\varphi(-n)$ and $h_\psi(-n)$ and downsampled along the columns. This generates two subimages; say W_L and W_H respectively, of reduced horizontal dimensions (by a factor of 2). The generated subimage W_L contains low frequency, vertical information, while W_H contains high frequency information with vertical orientation. Each column of both subimages is then convolved with $h_\varphi(-m)$ and $h_\psi(-m)$ and downsampled along the rows to give four subimages of quarter size with respect to the original matrix $W_\varphi(j+1, m,n)$. The newly generated images are $W^D{}_\psi(j, m,n)$, $W^V{}_\psi$ (j, m,n), $W^H{}_\psi$ (j, m,n) and $W_\varphi(j, m,n)$. The images are respectively said to be of resolution HH, HL, LH and LL based on the low(L) and high(H) frequency contents. The sizes of all four matrices thus computed are same and each is one fourth of the size of $W_\varphi(j+1, m,n)$, used for the next iteration. The process starts at scale J and computes intermediate coefficients for every scale $j \geq j0$.

3. Proposed Data Structure

The proposed data structure is a dynamic structure [Fig. 2(a)] with quad lists [Fig. 2(b)] associated with each sample image for each authorized person. Each cell i,j of the two dimensional array, of size P x S contains wavelet coefficients of the j^{th} sample image of the i^{th} person, where P is the total number of persons in the authorized face data base and S is the number of samples used for each person to train the classifier. Each cell contains the address of the record consisting of four fields.

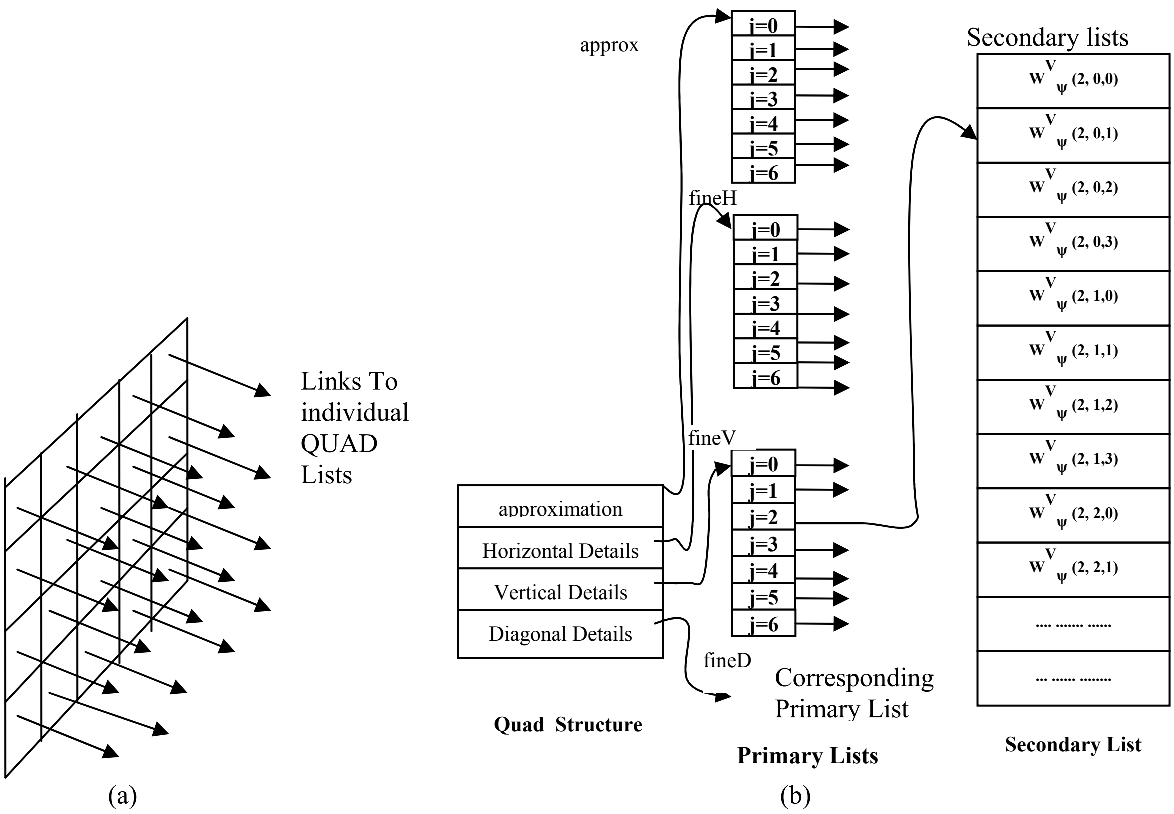

Fig. 2.Proposed Data Structure (a) Links to QUAD Lists (b) QUAD List

Each of these fields contains the starting address of primary lists of fixed sizes equal to J which in turn will point to the second lists . The secondary lists store the corresponding coefficients $W_\varphi(j, m,n)$, $W^H_\psi(j, m,n)$, $W^D_\psi(j, m,n)$ and $W^V_\psi(j, m,n)$ respectively . Primary and secondary lists are dynamically allocated on the heap segment of the main memory (RAM). Each of these four fields points to four primary lists of fixed sizes. Each node [Fig. 2(b)] of this list contains three fields j (scale) , size of the coefficients matrix, and the pointer to the secondary list. Secondary list is a dynamically allocated block of contiguous memory of size, waveMatSize.rows × waveMatSize.cols, where WaveMatSize is a record structure containing two fields, namely 'rows' and 'cols' meant to store the size of the two dimensional structure laid in row major form. This list stores the wavelet coefficients. The size of the secondary list associated with scale j is of order of $2^j \times 2^j$. The two dimensional wavelet coefficients are placed in row major form. Separate secondary lists are constructed dynamically corresponding to each scale j = 0,1,2,....J-1 for placing the coefficients at four different resolutions. The advantage of allocating the memory dynamically over static allocation is that the allocation is as per the size and requirement of the data[8].

The wavelet coefficients $W_\varphi(j,m,n)$, $W^H_\psi(j,m,n)$, $W^D_\psi(j, m,n)$ and $W^V_\psi(j, m,n)$ are accessed in constant time. Constant time access is of order O(1) (O refers to big-oh notation for time complexity) which ensures random access, i.e. the time to access the data from its corresponding location on the memory does not depend on the size of the list and the location of the data on heap. The access is through dereferencing the address of the location where the coefficient is placed. The address of the coefficient , say $W^V_\psi(j,m,n)$, is computed as $Q+((m \times N) +n)$, where Q is the starting address of the secondary list associated with the scale j, preserved in the quad primary list's j^{th} cell. N is computed at compile time with the following computation. Let a variable T denote the 2D mesh [Fig 2(a)] and the variable 'ptr' contain the address of the j^{th} location of the corresponding primary list, computed as (T[i][j].q→fineV + j). Then, N is computed as (ptr→waveMatSize).cols. Each node of the secondary list contains the computed coefficients along with the translation coefficients in both x and y directions.

The proposed data structure proved beneficial in fulfilling the space requirement for all coefficients for 40×10 (400) face images from the AT & T face data base, where each image is of size 112×92. The proposed dynamic data structure outperforms the efficiency of the allocation of 2D arrays for the matrices containing wavelet coefficients as it gets allocated on the call stack segment of memory which has limitations. The call stack is mainly used for allocating space for local variables and the values of the variables is also not accessible after the execution of the function.

4. Results and Discussion

Face database ORL (Olivetti Research Laboratory) was used to test the efficiency of the proposed data structure. The ORL data base contains frontal face images of 40 persons and 10 samples of each person with variations in pose and expressions [9]. The proposed data structure was implemented in C language. The accuracy of recognition was tested at varying percentages of features from different scales 'j' and different resolutions LL, LH, HL and HH. Two sets of features were combined in any pair of resolutions. For example, p% features from resolution LL at scale J-k was combined with (100 - p) percent features from the resolution LH at scale J-k-2 with varying percentage ratio of features at both resolutions, p:(100 - p); p varied from 10 to 100 at a step size of 10. A variation in value of k = 2, 3 and 4 gave a scope of experimenting features from both resolutions at varying scale differences. The results are depicted in Table I.

The accuracy was computed on the basis of correctly classified authorized face images to be accepted by the FRS (say c1) and correctly classified imposter images to be rejected by the FRS (say c2). The accuracy is computed as $[(c1+c2)/T] \times 100$, Where T is the total number of test images. The classifier used for recognition process defines the cluster linear decision boundaries as a parallelepiped. A linear decision boundary is used to demonstrate the potential of the extracted features as we get the accuracy of classification as high as 92.5% (Table I) with a extremely simple linear classifier. The n-dimensional test image feature vector is assigned to class C if the parallelepiped enclosure of the class cluster contains maximum number of features from the test feature vector. An n-dimensional parallelepiped, which is difficult to visualize for n>3, is defined by the minVector and maxVector as vectors containing minimum and maximum respectively, of all the features in a training cluster.

The experiments were conducted to establish the most informative features among all wavelet resolutions i.e. HH, HL, LH and LL. The experiment could be performed due to the capability of the proposed dynamic data structure and the testing time is obtained as 0.02 seconds (Table II).

In this paper, it is also established that the features from the resolutions LL and LH are the most informative and discriminatory (Table I). Also due to the fact that wavelet features have large information packing capability, a very small number of such feature, only 2048 out of 10304 are used for face recognition. The dimensionality is reduced by 80-90% of the total 10304 pixels.

TABLE I: RESOLUTION VS ACCURACY

No. of features used	Dimensionality reduction (%)	Feature set 1	Feature set 2	Accuracy
66	99.35	HL(4) 40%	LH(2) 60%	15.83%
1842	82.12	HH(6) 70%	LH(4) 30%	18.33%
2312	77.56	LH(6) 90%	LL(4) 10%	38.33%
2048	80.12	LL(6) 80%	LH(4) 20%	92.5 %

TABLE II: EXECUTION TIME ANALYSIS FOR FACE RECOGNITION USING ORL FACE DATABASE

No. of samples per person (t)	Total Training Samples (40 × t)	Training time (in seconds)	Testing Time (in seconds)
3	120	2.23	0.0277
5	200	3.75	0.0224
7	280	5.23	0.0221
9	360	6.75	0.0236

It is observed that the features from HH and HL resolutions do not respond well to the recognition result in low accuracy. The features at scales 0 to 3 correspond to coarse details and do not contribute in high accuracy rates. The recognition accuracy is as high as 92.5% using 7 training samples with a combination of the LL and LH resolutions (Table I). It is observed that the combination of features from LL(6) and LH(4) showed poor performance when only 5 samples were used for training and the test samples were the other face images not included for training. The data structure proved beneficial in storing large number of images.

5. Conclusion

The proposed data structure enables real time face recognition. The training time is only 5-7 seconds for as large as 360 images of sizes 112 × 92 each and the testing time for each unknown face image took an average time of 0.02 seconds only (50th of a second) which is in real time.

6. References

[1] R.C. Gonzalez and R.E.Woods, *Digital Image Processing* 3rd Ed. Pearson. Ch 7,pp. 461-510

[2] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, *Face Recognition: A Literature Survey*, ACM Computing Surveys, Vol. 35, No. 4, 2003, pp 399-458.

[3] Rama Chellapa, Charles L. Wilson and Saad Sirohey, *Human and Machine Recognition of Faces,* Proceedings of

the IEEE, Vol. 83, No. 5, May 1995.pp 705-740.

[4] Tao Li, Qi Li, Shenghuo Zhu and Mitsunori Ogihara*, A Survey on Wavelet Applications in Data Mining,* ACM SIGKDD Explorations Newsletter, Volume 4, Issue 2, 2002, pp 49-68.

[5] Randa Atta and Mohammad Ghanbari, *Face Recognition Based on DCT Pyramid Feature Extraction,* 3rd International Congress on Image and Signal Processing (CISP2010), Vol. 2, 2010, pp 934-938.

[6] Cunjian Chen, Jiashu Zhang, *Wavelet Energy Entropy as a new Feature Extractor for Face Recognitio*n, 4th International Conference on Image and Graphics (ICIG2007), 2007, pp 616-619.

[7] Jean Choi, Yun-Su Chung, Ki-Hyun Kim, Jang-Hee Yoo, *Face Recognition Using Energy Probability in DCT Domain,* International Conference on Multimedia and Expo, 2006, pp 1549-1552.

[8] M.H. Gross, O.G. Staadt, R. Gatti**,** *Efficient Triangular Surface Approximations using Wavelets and Quadtree Data Structures*, IEEE Transactions on Visualization and Computer Graphics, Vol.2, Issue 2, 1996, pp 130-143.

[9] Ferdinando Samaria, Andy Harter, *Parameterisation of a Stochastic Model for Human Face Identification,* Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994