The Islamic University of Gaza
Deanery of Graduate Studies
Faculty of Engineering
Computer Engineering Department

FACE RECOGNITION USING CURVELET AND WAVEATOM TRANSFORM

By
Hana H. Hejazi

Supervisor
Asst. Prof. Mohammed A. Alhanjouri

A Thesis Submitted in partial fulfillment of the Requirements for the degree of Master of Science in Computer Engineering

Gaza, Palestine

(1431, 2010)

ABSTRACT

The field of digital image processing is continually evolving. Nowadays, there is a significant increase in the level of interest in image morphology, neural networks, full-color image processing, image data compression and image recognition. This work deals with image recognition with the application of face recognition.

Some people think that face recognition is an easy task for computer system as for humans, but in reality most of the face recognition systems can't achieve a complete reliable performance because there are many factors affect on the process of recognition like: large variations in facial approach, head size and orientation, and change in environmental conditions, all these factors makes face recognition one of the fundamental problems in pattern analysis, other factors that impact the performance are the accuracy of face location stage and the number of actual face recognition techniques used in each system. So face recognition from still and video images is emerging as an active research area with numerous commercial and law enforcement application.

This research identifies two techniques for face features extraction based on two different multiresolution analysis tools; the first called Curvelet transform while the second is waveatom transform. The resultant features are inputted to train via two famous classifiers; one of them is the artificial neural network (ANN) and the other is hidden Markov model (HMM).

Experiments are carried out on two well-known datasets; AT&T dataset consists of 400 images corresponding to 40 people, and Essex Grimace dataset consists of 360 images corresponding to 18 people. Experimental results show the strength of both curvelets and waveatom features. On one hand, waveatom features obtained the highest accuracy rate of 99% and 100% with HMM classifier, and 98% and 100% with ANN classifier, for AT&T and Essex Grimace datasets, respectively. On the other hand, two levels Curvelet features achieved accuracy rate of 98% and 100% with HMM classifier, and 97% and 100% with ANN classifier, for AT&T and Essex Grimace datasets, respectively.

A comparative study for waveatom with wavelet-based, curvelet-based, and traditional Principal Component Analysis (PCA) techniques is also presented. The proposed techniques supersede all of them. And shows the robustness of feature extraction methods used against included and occluded effects. Also, indicates the potential of HMM over ANN, as they are classifiers.

Acknowledgment

My thanks must first and foremost go to ALLAH who gave me all things to produce this project.

I have the pleasure to have Assistant Professor Mohammed Al Hanjouri as my thesis supervisor. He introduced me to the field of Image Processing with multiresolution transforms, gave me every effort, and guided me patiently throughout this thesis work. I am highly indebted and express my deep sense of gratitude to him for his invaluable guidance, constant inspiration to present this work.

Many thank to any person who has in one way or another contribution to this work. I want to express most sincerely appreciation to Professor Ibrahim Abu Haiba, from IUG computer engineering stuff and Asst. Prof. Hatem Elaydi, from IUG electrical engineering stuff, for their suggestion.

Very big thanks to my university which was my second home since three years and it will still have this value forever.

Finally, I owe my largest debt to my family. I am highly indebted to my parents, brothers, husband, and my loving kids for their love, affection, constant encouragement, and invaluable support throughout this difficult period. I also want to thanks my husband Dr. Hazem who motivated me to the initiate research work. I dedicate this work to all of them.

Table of Contents

List of Figures

List of Abbreviations

AAM	Active Appearance Model.
ANN	Artificial Neural Network.
CS	Compressive Sampling or Compressed Sensing.
CWT	Continuous Wavelet Transform.
DCT	Digital Cosine Transform.
DFT	Discrete Fourier Transform.
DWT	Discrete Wavelet Transform.
EFM	Enhanced Fisher linear discriminant Model.
EM	Expectation-Maximization algorithm.
EMC	Enclosed Model Selection Criterion.
FFT	Fast Fourier Transform.
GFC	Gabor–Fisher Classifier.
GM	The Gaussian mixture modeling.
GWT	Gabor Wavelet Transform.
HMM	Hidden Markov Model.
ICA	Independent Component Analysis.
IGF	Independent Gabor features.
KNN	k-Nearest Neighbor classifier.
LDC	Local Discriminant Coordinates.
LoG	Laplacian-of-Gaussian.
LS-SVM	Least Square Support Vector Machine.
MAP	Maximum a Posteriori logistic model.
ML	Maximum Likelihood.
PCA	Principal Component Analysis.
PDF	Probability Density Functions.
PIN	Personal Identification Number.
RBF	Radial Basis Function.
SVM	Support Vector Machines.
USFFT	Unequally Spaced Fast Fourier Transform.
WFT	Windowed Fourier Transform.

CHAPTER 1
INTRODUCTION

There are many ways that humans can identify each other, and so is for machines. There are many different identification technologies available, many of which have been in commercial use for years. The most common verification and identification methods nowadays are Password/PIN (Known as Personal Identification Number) systems. The problem with that or other similar techniques is that they are not unique and is possible for someone to forget, loose or even have it stolen. In order to overcome these problems they have developed considerable interest in “biometrics” identification systems, which use pattern recognition techniques to identify people using their characteristics. Some of these methods are fingerprints and face recognition.

A face recognition system can be used in buildings or specific area security, a face recognizer could be used at the frontal entrance for automatic access control, and they could be used to enhance the security of user authentication in ATMs by recognizing faces as well as requiring passwords. Also, these systems can be used in the human or computer interface arena workstations with cameras would be able to recognize users, perhaps automatically loading the user environment when he/she sits in the front of the machine.

A face recognition system must operate under a variety of conditions, such as varying illuminations and facial expressions; it must be able to handle non-frontal facial images of both males and females of different ages and races.

Before face recognition is performed, face detection must take place, so the system should determine whether or not there is a face in a given image, once complete this process, face region should be isolated from the scene for the face recognition. The face detection and face extraction are often performed simultaneously. Finally, the classification of the face will take place.

Face recognition can be done in both a still images and video images. Different approaches of face recognition for still images can be categorized into two main groups such as Geometrical approach and Pictorial approach.

In recent years, multiresolution analysis tools, especially wavelets, had been found useful for analyzing the content of images; this leads to use these tools in areas like image processing, pattern recognition and computer vision. Following wavelets, other multiresolution tools were developed like contourlets, ridgelets etc. Curvelet transform is a recent addition to this list of multiscale transforms while the most modern one is called Waveatom Transform. Waveatom transform used in image processing in the field of image denoising, and the results obtained are the best one when compared to the state of art [1].

The whole system of face recognition consist of three main phases, these are: preprocessing, feature extraction, and the classification phases.

1.1 The Problem Statement

The goal of this work is to find the best feature extraction, which performs the smallest feature vector length and gives the highest performance. The features set is obtained using Curvelet transform and Waveatom transform. Artificial Neural Network (ANN) and Hidden Markov Model (HMM) are trained using these features. AT&T as an example of still images, and Essex Grimace as video images are used for training and testing. The aim is to answer the following questions:

1. Can Curvelet transform stand alone as feature extraction or not?
2. How long feature vector length obtained using Curvelet transform?
3. Can Waveatom transform used in face recognition?
4. How long feature vector length obtained using Waveatom transform?
5. Which is the most suitable method to extract feature from a face image; Curvelet or Waveatom?
6. Does ANN give good accuracy with feature vectors obtained using Curvelet and Waveatom Transforms?
7. Does HMM give good accuracy with feature vectors?
8. Which work better as classifier ANN or HMM?

1.2 Literature Review

The subject of face recognition is as old as computer vision, both because of its practical importance and the theoretical interest from scientists, in spite of the fact that other methods of identification (such as fingerprints, or iris scans) can be more accurate.

Face recognition always remains a major focus of research because of its non-invasive nature and because people's primary method of person identification. Since the start of that field of technology there were two main approaches:

- Geometrical approach.
- Pictorial approach.

The geometrical approach uses the spatial configuration of facial features. That means that the main geometrical features of the face such as the eyes, nose, and mouth are first located and then faces are classified on the basis of various geometrical distances and angles between features. On the other hand, the pictorial approach is using the templates of the major facial features and entire face to perform recognition on frontal views of faces.

Perhaps eigenfaces is the most famous early technique used in face recognition systems. Any human face can be presented by linear combination of eigenface images. Eigenfaces is asset of eigenvectors derived from the covariance matrix of a high dimensional vector that represent possible faces of humans. Eigenfaces have advantages over other techniques available, such as the system's speed. The eigenfaces method presented by Turk and Pentland (1991) [2] have found the principal components of a face image.

Unfortunately, these eigenfaces are sensitive to variety in position and scale. For the system to work well, the faces need to be seen from a frontal view under similar lighting. To overcome this problem they suggest using a multi-resolution method in which faces are compared to eigenfaces of varying sizes to compute the best match. Turk and Pentland's paper was very seminal in the field of face recognition and their method is still quite popular due to its ease of implementation [3].

Researchers have starting using Wavelet coefficients as features for face recognition in the last years [4, 5, 6, 7, 8]. The simplest application of using Wavelet in face recognition can be found in [4]. The face image undergoes 2 levels Wavelet transform,

the approximate Wavelet coefficients are used in HMM classifier for recognition. In [5] solved the complexity of PCA training time by using Wavelet packet decomposition, and then used approximated and details coefficients to calculate PCA. Researchers in [6] the face image was inputted into 2 levels Discrete Wavelet Transform (DWT) system. The approximated coefficients projected using PCA into eigenface space; Support Vector Machines (SVM) classifier trained using these PCA for recognition. In [7] Wavelet frequency subbands were searched to found that is insensitive to expression differences and illumination variants ones on faces. The horizontal Wavelet components were found to be a very good feature in face recognition and yielding the highest performance rates. Researchers in [8] tried to measure the discriminability of Wavelet package decomposition with depth 2. They modified a Local Discriminant Coordinates (LDC) by using dilation invariant entropy and Maximum Postiriori (MAP) logistic model.

Beyond Wavelet, Gabor transform were used to extract features [9, 10, 11, 12, 13, 14]. The face image decomposed using Gabor transform in [9, 10, 11]. In [9] the dimensionality of Gabor feature vectors was reduced using Enhanced Fisher linear discriminant model (EFM). Nearest neighbor classifier were used for recognition. Where in [10] the dimensionalities of Gabor feature vectors was reduced using PCA followed by ICA. Researchers in [11] applied AdaBoosted algorithm to Gabor features and solved the imbalance between the amount of the positive samples and that of negative samples by re-sampling scheme. In [12] facial feature points were localized by Active Appearance Model (AAM) and refined the localization by Gabor jet similarity, next Gabor feature vectors extracted at all facial points. In [13] researchers not only applied Gabor transform but also Discrete Cosine Transform (DCT) to face images. A Radial Basis Function (RBF) based neural network was trained using Gabor coefficients and DCT coefficients. As next step to improve the processing speed they enhanced the edge and used non-uniform down sampling to reduce the dimensionality of Gabor coefficients. In [14] multi-scale Harris-Laplace detector was used to evaluate the interest points. Gabor transform was used to extract feature vectors at interest points.

Curvelet transform becomes a very popular multi-resolution transform after implementing its second generation. In face recognition, Curvelet transform seems to be promising [15, 16, 17, 18, 19, 20, 21]. The beginning was in [15]. The face images were

quantized from 256 to 16 and 4 gray scale resolutions, the quantized images were decomposed using Curvelet transform. Three SVMs were trained using Curvelet coefficients and the decision was made by simple majority voting. In [16] the face image undergoes Curvelet transform. PCA was performed on the approximated coefficients. K-Nearest Neighbor classifier was employed to perform the classification task. In [17] as preprocessing step researchers converted face images from 8 bit into 4 bit and 2 bit representations. Curvelet transform was performed to extract feature vectors from these representations, and then the approximated components were used to train different SVMs. Researchers in [18] addressed the problem of identifying faces when the training face database contains one face image of each person. The Curvelet approximated coefficients was framed as a minimization problem. The original image and the reconstructed images of the non-linear approximations were used to generate the training set. A comparative study amongst Wavelet and Curvelet was found in [19]. In [20] the Curvelet sub-bands were divided into small sub-blocks. Means, variance and entropy were calculated from these sub-blocks as statistical measures. Feature vector was constructed by concatenated each block measure. Local discriminant analyses (LDA) was carried out on feature vectors and the city-block distance was used for classification. Researcher in [21] decomposed a face image using Curvelet transform at scale 4. Next Least Square Support Vector Machine (LS-SVM) was trained using Curvelet features.

The results in [17, 18, 19, 20, 21] have showed Curvelet based schemes were better than wavelet based recognition schemes. The results also showed improvement over the previous approach [15].

1.3 Objective and outlines of the Thesis

This thesis introduces new proposed method to extract effective features for face recognition, and it implements many feature extraction techniques to feed several types of classifiers, these procedures will be produced many combined systems for face recognition; therefore, the comparative study is necessary to determine the best one.

The present work was organized as following:

Chapter 2 introduces the proposed Feature Extraction methods based on Curvelet and Waveatom representation of face images. Curvelet theory was presented. Properties of Waveatom as a multiresolution transform were discussed beside its theory.

Chapter 3 overviews the Classification technique used. Hidden Markov model principles were introduced. This chapter gives short summary for the algorithms which HMM depends on such as Expectation Maximization algorithm and Vetribi algorithm.

Chapter 4 talks about the proposed system and its architecture. It explains the how the feature is obtained and used to train both of ANN and HMM. The topology of ANN and HMM was discussed.

Performance of this method is examined on two different standard face databases with different characteristics. Simulation results and their comparisons to well-known face recognition methods are presented in chapter 5.

In chapter 6, concluding remarks are stated. Future works, which may follow this study, are also presented.

CHAPTER 2

FEATURE EXTRACTION TECHNIQUES.

Feature extraction is the most important step for any face recognition system. In reality, using local features is a mature approach to face recognition problem. Feature based methods are based on finding local areas on a face and representing corresponding information in an efficient way. So, choosing feature locations and the corresponding values are extremely critical for the performance of a recognition system. Due to this fact, scientists look for another answer to face recognition problem. The idea of this answer comes from how human vision system perceives both local feature and whole face. Physiological found simple cells, in human visual cortex, that are selectively tuned to orientation as well as to spatial frequency. It appears that Multiresolution transforms could be the response of these cells.

The word ‘transform’ means ‘change’. An ‘image transform’ refers to an alternate way of representing an image. Generally an image is represented in the spatial domain by pixels, but there are alternate representations, the most popular being the frequency domain representation obtained by the Fourier transform. Particularly, the Fourier transform of an image is not very suitable to the field of object recognition. Other transforms like Wavelets, Curvelets and Waveatom provide alternative image representations (other than pixels or frequency). These transforms represent images in such a way that recognition is facilitated.

2.1 Digital Curvelet Transform

Curvelets was proposed by E. Candes and D. Donoho (2000) [22]. The idea of Curvelets is to represent a curve as a superposition of functions of various lengths and widths obeying the scaling parabolic law: $width \cong (length)^2$. Figure 2.1 shows the Curvelet frequency tilling which called Second Dyadic Decomposition (SDD). The length of the localizing windows (colored blue) is doubled at every other dyadic subband. Curvelet Transform in continues domain was defined by using coronae and rotations, see Figure 2.1 (a). The discrete input data was defined on a Cartesian grid, so to able to define Curvelet transform in the discrete domain, concentric squares and shears should be

used instead of concentric circles and rotations see Figure 2.1 (b).The frequency plain is partitioned into radial (circles and squares) and angular (rotations and shears) divisions. Different scales are obtained by radial division; the smallest scale defines the finest resolution while the largest scale defines the coarsest resolution. Angular division divides each scale into different orientation; the maximum number of orientations was found at the finest resolution and the lesser number of orientations was found at coarsest resolution.

Curvelets are designed to represent edges and other singularities along curves much more efficiently than the traditional Wavelet transform which good at representing point singularities. Figure 2.2 shows edge representation by both Wavelets and Curvelet Transforms. It can be noticed, it would take many Wavelet coefficients to accurately represent such a curve while Curvelet needs small number of coefficients; wavelet needs three, six, and twelve coefficients, while Curvelet needs one, two, and four coefficients, in the largest, middle, and smallest scale respectively.

To explain how the Curvelet basis elements align with edges in an image and how this alignment affects the coefficients of the corresponding transform. Figure 2.2 pictorially depicts the alignment. The first box shows the original image. In the second box, the band-passed image (image at a certain resolution) is shown. Finally in the third box the alignment of the Curvelet basis elements with a small section of the edge is shown and will now be briefly explained.

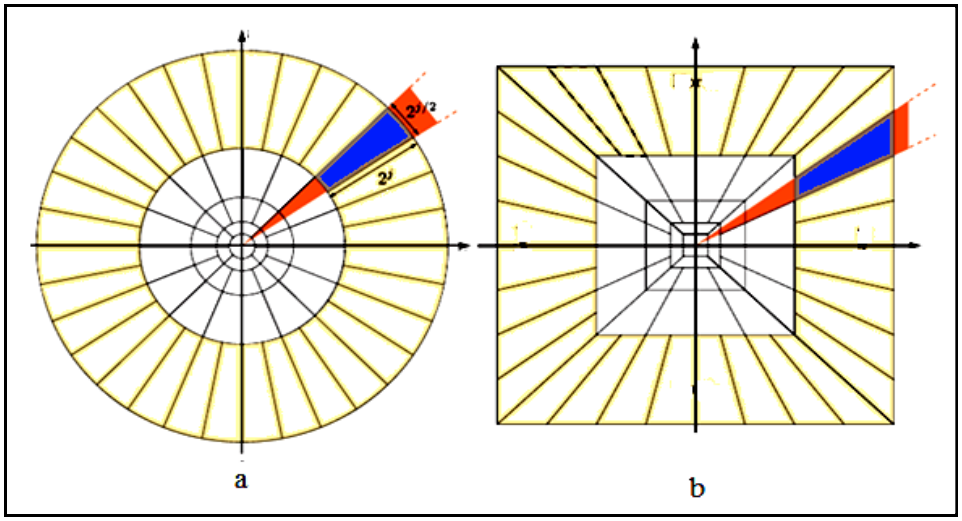


Figure 2.1: Curvelet Frequency Tiling (a): Continuous Domain. (b): Discrete Domain

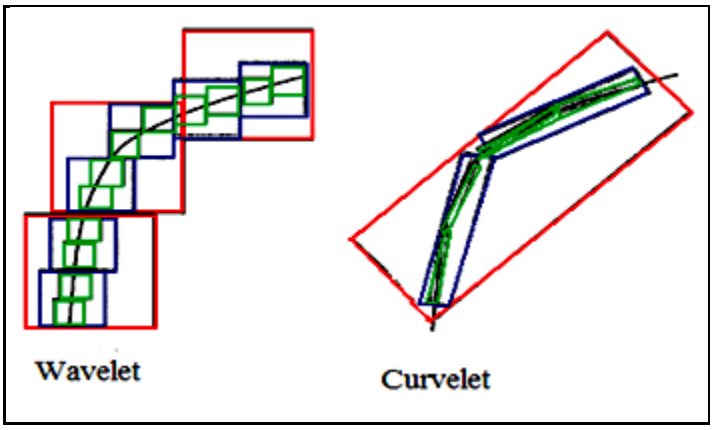


Figure 2.2: Edge Representations

The transform coefficients are inner products of the original image and basis elements. Figure 2.3 show that when a basis element 'c' is aligned with an edge, the corresponding coefficient in the transform domain has a high value (c). Whereas if the basis element such as 'a' or 'b' is misaligned, the corresponding transform coefficient is nearly zero. Thus, the values of the transform coefficients provide an estimate of the 'edginess' of the image at a particular scale and orientation.

Practically, Curvelet Transform is multi-scale geometrical transform in which units are indexed by their position, scale and orientation. To define it in continues domain, suppose we work in the space $\mathbf{R}^2$ with special variable x, frequency variable ω and polar coordinates r, θ in the frequency domain. Define two smooth, non-negative and real valued window functions: the first one is $W(r)$ called radial window, and the second one is $V(t)$ called angular window. The function $W(r)$ takes positive real arguments and is supported on $r \in [1/2, 2]$. The function $V(t)$ takes real arguments and is supported on $t \in [-1, 1]$. These windows should obey the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1 \ , \quad r > 0 \qquad (2.1)$$

$$\sum_{j=-\infty}^{\infty} V^2(t - l) = 1 \ , \quad t \in \mathbf{R} \qquad (2.1)$$

For each j, introduce the frequency window U_j defined in the Fourier domain by:

$$U_j(r,\theta) = 2^{-3j/4}\, W\left(2^{-j} r\right) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right), \lfloor j/2 \rfloor \text{ is the integer part of } \frac{j}{2} \qquad (2.3)$$

Thus the support of U_j is a polar "wedge" defined by the support of W and V at given scale.

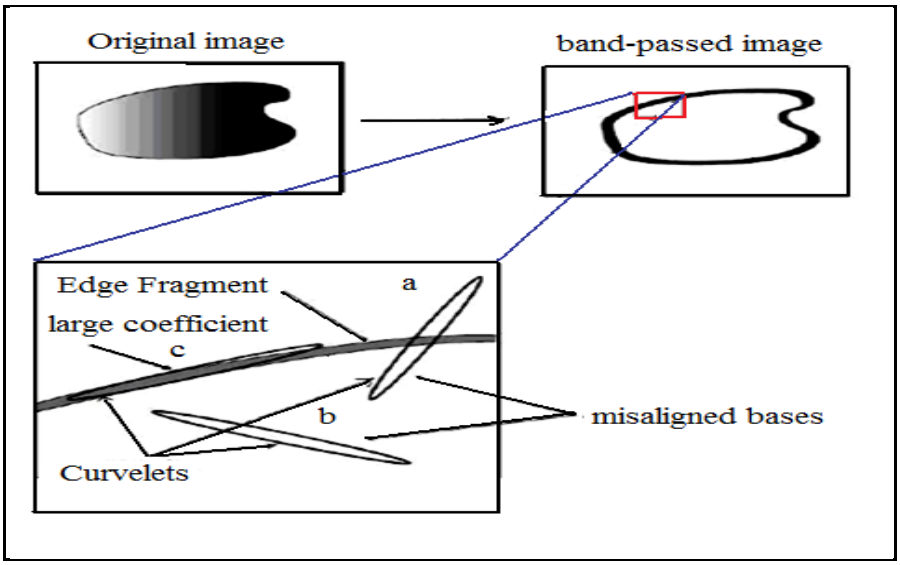


Figure 2.3: Curvelet Alignments

A waveform $\varphi_j(x)$ can defined by letting $\hat{\varphi}_j(\omega) = U_j(\omega)$. The Curvelet transform can be defined as a function of $x = (x_1, x_2)$ at scale 2^{-j} , orientation θ_t where $\theta_t = 2\pi . 2^{\lfloor j/2 \rfloor} . l\,, l = 0,1, \ldots,\ 0 < \theta_t < 2\pi$, and position $x_k^{(j,l)} = \mathrm{R}_{\theta_t}\left(\mathrm{k}_1 . 2^{-\mathrm{j}}, \mathrm{k}_2 . 2^{-\frac{\mathrm{j}}{2}}\right)$ where $k = (\mathrm{k}_1, \mathrm{k}_2)$ is the shift parameter by:

$$\varphi_{j,l,k}(x) = \varphi_j\left(\mathrm{R}_{\theta_t}\left(x - x_k^{(j,l)}\right)\right) \qquad (2.4)$$

So the coefficients of the continous Curvelet are given by:

$$c_{j,l,k} := \langle f, \varphi_{j,l,k} \rangle = \int_{\mathbb{R}^2} f\ (x)\, \overline{\varphi_{j,l,k}(x)}\ \ dx \qquad (2.5)$$

Based on Plancherel's Theory,

$$c_{j,l,k} := \frac{1}{2\pi^2} \int_{\mathbb{R}^2} \hat{f}\ (\omega)\, \overline{\varphi_{j,l,k}(x)}\ \ d\omega \qquad (2.6)$$

$$c_{j,l,k} := \frac{1}{2\pi^2} \int_{\mathbb{R}^2} \hat{f}\ (\omega)\, U_j\left(\mathrm{R}_{\theta_t}\omega\right) e^{j\,\langle x_k^{(j,l)}, \omega \rangle}\ \ d\omega \qquad (2.7)$$

If the input $f\ [x_1, x_2]$ $(0 \leq x_1\ ,\ x_2 < n)$ in the spatial Cartesian is an image, then the discrete form of the continuous Curvelet transform can be expressed as the following:

$$c_{j,l,k}^{D} := \sum_{0 \leq x_1\ ,\ x_2 < n} f\ [x_1, x_2]\ \overline{\varphi_{j,l,k}^{D}[x_1, x_2]} \tag{2.8}$$

There is two generations of Curvelet transform. The first generation defines Curvelet between Wavelet and multiscale Ridgelet. The digital Curvelet transform is obtained by the implementation of three steps [22]:

1. Sub-band decomposition: The object $f[t_1\ , t_2]$, and $t_1 \geq 0\ , t_2 \leq n$, is decomposed into sub-bands with wavelet transform to obtain $\hat{f}[n_1\ , n_2]$.
2. Smooth partitioning: each sub-band is smoothly windowed in to "squares" of an appropriate scale of side-length 2^{-j}. The side-length of these windows is doubled at every other dyadic sub-band.
3. Renormalization: each resulting square is renormalized to unit length and of width 2^{-j}.
4. Ridgelet analysis: Each square is decomposed by Ridgelet transform.

In the second generation, two different implementations of Curvelet were founded: The first digital transformation is based on Unequally Spaced Fast Fourier Transform (USFFT), while the second is based on the wrapping of specially selected Fourier samples. The two implementations essentially differ by the choice of spatial grid used to translate Curvelets at each scale and angle. Where, a tilted grid mostly aligned with the axes of the window which leads to the USFFT. On the other hand, a grid aligned with the input Cartesian grid which leads to the wrapping-based. Both digital transformations having the same output, but the Wrapping Algorithm gives a more intuitive algorithm and faster computation time [23]. Because of this, Curvelet via wrapping will be used for this work.

If we have the object $g[t_1,t_2]$, $t_1 \geq 0$, $t_2 < n$ as Cartesian array and $\hat{g}\ [n_1,n_2]$ to denote its 2D Discrete Fourier Transform, then the architecture of Curvelets via wrapping is as follows:

1. 2D Fast Fourier Transform (FFT) is applied to $g[t_1,t_2]$ to obtain Fourier samples $\hat{g}[n_1,n_2]$.

2. For each scale j and angle l, the product $\tilde{U}_{j,l}[n_1,n_2]\,\hat{g}[n_1,n_2]$ is formed, where $\tilde{U}_{j,l}[n_1,n_2]$ is the discrete localizing window (Figure 2.4 (a)).
3. This product is wrapped around the origin to obtain $\check{g}_{j,l}[n_1,n_2] = W(\tilde{U}_{j,l}\,\hat{g})\,[n_1,n_2]$; where the range for n_1,n_2 is now $0 \leq n_1 < L_{1,j}$ and $0 \leq n_2 < L_{2,j}$; $L_{1,j} \approx 2^j$ and $L_{2,j} \approx 2^{j/2}$ are constants (Figure 2.4 (b)).
4. Inverse 2D FFT is applied to each $\check{g}_{j,l}$, hence creating the discrete Curvelet coefficients.

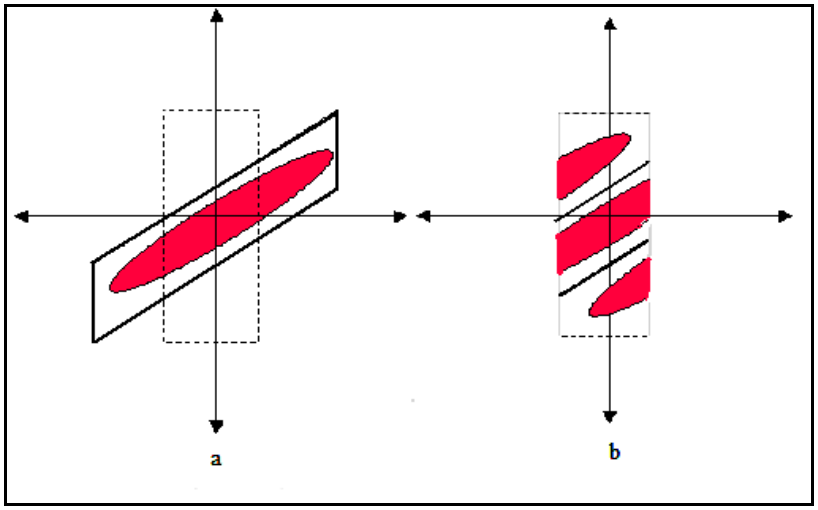


Figure 2.4: Discrete Localizing Window (a): before Wrapping (b): after Wrapping

In spite of good capturing edges and curves, the Curvelet problem is that it is over complete and its redundancy still high (7.2 in 2D and 24 in 3D).

2.2 Waveatom Transform

Waveatom was presented by Demanety and Ying [24], it is a new member in the family of oriented, multiscale transforms for image processing and numerical analysis. Waveatoms come either as an orthonormal basis or a tight frame of directional wave packets, and are particularly well suited for representing oscillatory patterns in images. They also provide a sparse representation of wave equations, hence the name wave atoms.

2.2.1 Waveatom Transform Properties

To be a Multiresolution image transforms, five properties should be satisfied:

1. Multiresolution: The transform should allow images to be successively approximated, from coarse to fine resolutions.
2. Localization: The basis elements of the transforms should be localized in both the spatial and the frequency domains.
3. Critical sampling: For some applications (e.g., compression), the transforms should form a basis, or a frame with small redundancy.
4. Directionality: The transform should contain basis elements oriented at a variety of direction.
5. Anisotropy: When a physical property changes with direction, that property is said anisotropy. For image transforms, anisotropicity means that the basis elements of the transforms should not be circular (similar in all directions) but may be elliptical (more along the major axis and less along the minor axis); the circular basis repeats itself many times and this does not the case of elliptical one.

The core of the Waveatom transform [24] is displayed in Figure 2.5. On the right of Figure 2.5, the frequency plane is divided into wedges. The wedge is formed by partitioning the frequency plane into radial and angular divisions. The radial divisions (concentric circles) are for band-passing the image at different resolution/scales. The angular divisions divide each band-passed image into different angles. To consider each wedge, the band-passed image should be analyzed at scale j and angle θ.

The second property of the wish-list requires the transform to be localized both in the frequency and the spatial domain. However, a signal that is perfectly localized in one domain is spread out in the other. So, one can only expect the transform to be approximately localized in both domains. The image on the right of Figure 2.5 shows how the frequency plane is divided into wedges. If a wedge has an abrupt boundary in the frequency domain, it will be spread in the spatial domain. To avoid that, the boundary of the wedge is tapered as shown in Figure 2.6. The dotted line shows an abrupt wedge boundary, the continuous lines show the actual wedge boundary which is tapered off. This smooth tapering allows for localization in both the frequency and spatial domains.

From Figure 2.5, once the scale and the angle are defined, the wedge is identified. The wedge is inverted to the spatial domain (left side of Figure 2.5) by Inverse Fourier transform. The inverse Fourier transform of the wedge are the Waveatoms corresponding to the wedge at particular scale and angle. The Waveatoms are periodic and repeated infinitely. In Figure 2.5 they are shown as ellipses, the centers of the ellipses are shown as dots.

Figure 2.8 shows a wedge (in the frequency domain) on the right, and a Waveatom (in spatial domain) corresponding to the wedge on the left. As we can see, the Waveatom is not exactly an ellipse as depicted in Figure 2.5. It is elongated in one direction and wave-like in the other but its effective support is elliptical as shown in the left of Figure 2.5 and 2.7. The relationship between the length of major and minor axes of the ellipse follows a parabolic scaling law, i.e. *major axis length*$\approx$ *(minor axis length)* ^{2}. However, the values of the Waveatom coefficients are determined by how much the Waveatom and the actual image are aligned.

What sets them apart from other transform architectures like Wavelets or Curvelets; Waveatoms have a sharp frequency localization that cannot be achieved using wavelet packets and offer a significantly sparser expansion for oscillatory functions than wavelets and Curvelets. Waveatoms capture the coherence of patterns across and along oscillations whereas Curvelets capture coherence along oscillations only. Waveatoms precisely interpolate between Gaboratoms and wavelets means that the period of oscillations of each wave packet is related to the size of essential support by parabolic scaling i.e. wavelength ~ (diameter)^2, which is known as the scaling law [25].

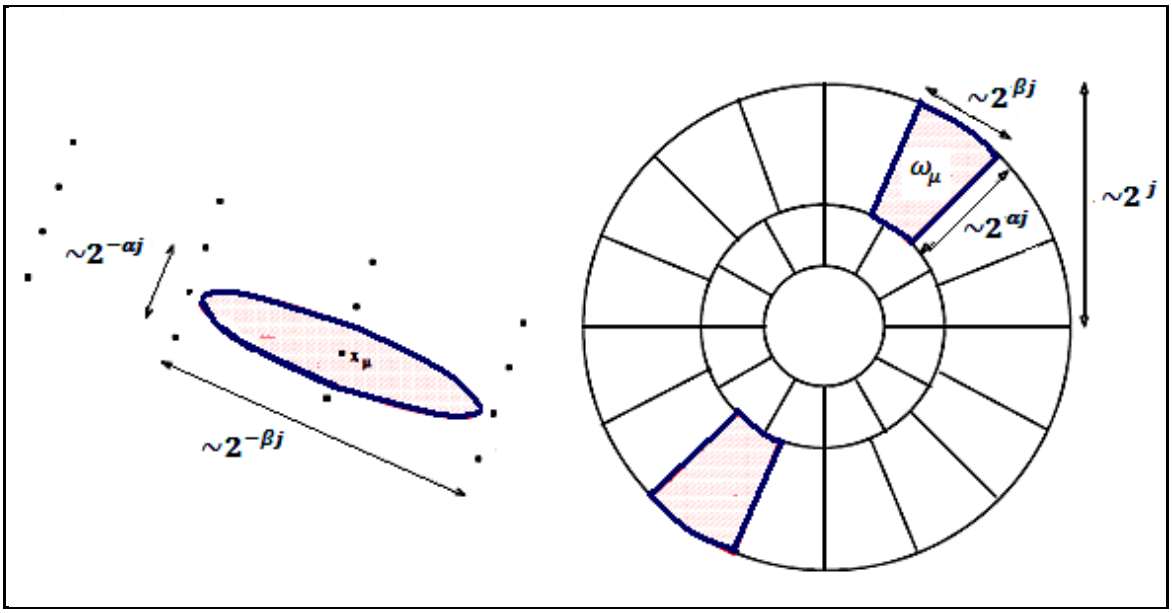


Figure 2.5: Waveatom Transform in Spatial Space (left) and in Frequency (right)

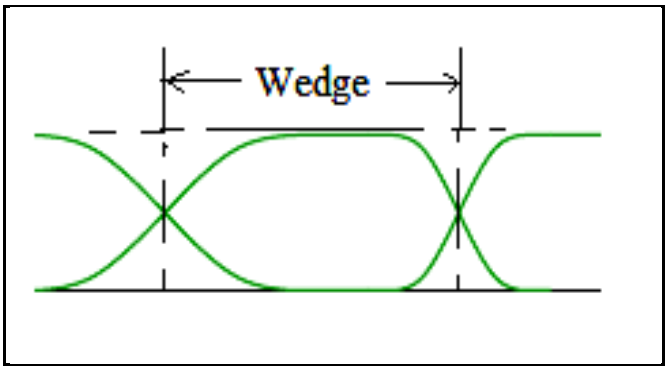


Figure 2.6: Boundary of Wedge

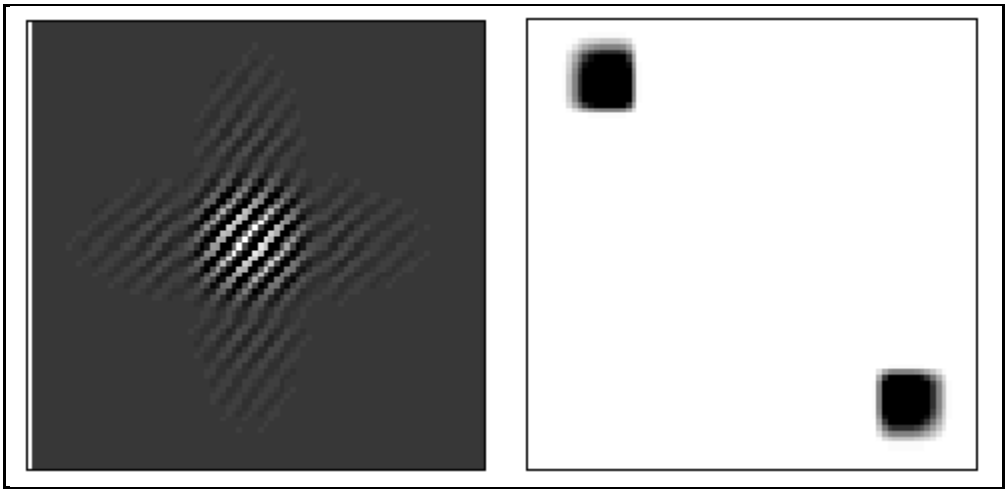


Figure 2.7: Waveatom at Increasing Finer Scale. Waveatom in Spatial Domain (left) and it's Frequency Domain (right)

2.2.2 Waveatom Theory

Suppose j, l, k are integer valued where j is the cutoff in scale, k is the cutoff in space and l labels the different wedges within each scale. Now consider a one-dimensional family of wave packets $\varphi_{l,k}^{j}(x), j \geq 0, l \geq 0, k \in \mathrm{K}$, centered in frequency around $\pm\omega_{j,l} = \pm\pi 2^{j} l$ with $c_1 2^{j} \leq l \leq c_2 2^{j}$ where $c_1 < c_2$ are positive constants, and centered in space around $x_{j,k} = 2^{-j} k$. One-dimensional version of the parabolic scaling states that the support of each bump of $\varphi_{l,k}^{j}(\omega)$ is of length $O(2^{j})$ while $\omega_{j,l} = O(2^{2j})$. Starting with Villemoes wavelets $\hat{\varphi}_{l}^{0}(\omega)$ in frequency plane,

$$\hat{\varphi}_{l}^{0}(\omega) = e^{-j\omega/2}\left[e^{j\alpha_l} g\left(\epsilon_l\left(\omega - \pi(l+1/2)\right)\right) + e^{-j\alpha_l} g\left(\epsilon_{l+1}\left(\omega + \pi(l+1/2)\right)\right)\right] \tag{2.9}$$

Where $\alpha_l = \pi/2\ (l+1/2)$, $\epsilon_l = (-1)^{l}$ and $\{\varphi_l(t-k)\}$ form an orthonormal basis of $L^2(\mathbb{R})$.

Dyadic dilates and translates of $\hat{\varphi}_{l}^{0}$ on the frequency axes are combined and basis functions, written as:

$$\varphi_{l,k}^{j}(x) = \varphi_{l}^{j}\left(x - 2^{-j}k\right) = 2^{j/2}\varphi_{l}^{o}\left(2^{j}x - k\right) \tag{2.10}$$

The transform $WA : L^2(\mathbb{R}) \rightarrow l^2(\mathbb{Z})$ maps a function u onto a sequence of waveatom coefficients, and by Plancherel's theorem:

$$c_{j,l,k} = \int_{-\infty}^{\infty} u(x)\,\varphi_{l,k}^{j}(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i2^{-j}k\omega}\,\overline{\hat{\varphi}_{l}^{j}(\omega)}\,\hat{u}(\omega)\,d\omega \tag{2.11}$$

If the function *u* is discretized at *xn* = *nh*, *h*=1/*K*, *n* =1....*K* , then with a small truncation error (2.9) is modified as:

$$C_{j,l,k}^{D} = \sum_{n=2\pi(-\mathrm{K}/2+1:1:\mathrm{K}/2)} e^{i2^{-j}kn}\,\overline{\hat{\varphi}_{l}^{j}(k)}\,\hat{u}(k) \tag{2.12}$$

A simple wrapping trick is used for the implementation of discrete wavelet packets and the steps involved are:

1. Perform an FFT of size *K* on the samples of (n) .
2. For each pair (j,l) wrap the product $\hat{\varphi}_l^j \hat{u}$ by periodically inside the interval $[-2^j\,\pi\,,\ 2^j\,\pi]$ then perform inverse FFT of size 2^j of the result to obtain $C_{j,l,k}^D$.
3. Repeat step 2 for all pairs(j,l) .

The positive and negative frequency components represented by:

$$\hat{\varphi}_{l,k}^j(\omega) = \hat{\varphi}_{l,k}^+(\omega) + \hat{\varphi}_{l,k}^-(\omega) \tag{2.13}$$

Hilbert transform $H\hat{\varphi}_{l,k}^j(\omega)$ of eq. (2.11) represents an orthonormal basis $L^2(\mathbb{R})$ and is obtained through a linear combination of positive and negative frequency bumps weighted by ***i*** and ***–i*** respectively.

$$H\,\hat{\varphi}_{l,k}^j(\omega) = -i\hat{\varphi}_{l,k}^+(\omega) + i\hat{\varphi}_{l,k}^-(\omega) \tag{2.14}$$

To extend Waveatom to be 2D, let $\mu = (j,\mathbf{l},\mathbf{k})$, where $\mathbf{l} = (l_1,l_2)$ and $\mathbf{k} = (k_1,k_2)$, so from equation 2.8.

$$\varphi_\mu^+(x_1,x_2) = \varphi_{l_1}^j\left(x_1 - 2^{-j}k_1\right)\varphi_{l_2}^j\left(x_2 - 2^{-j}k_2\right) \tag{2.15}$$

Hilbert transform was used to define the dual orthonormal basis,

$$\varphi_\mu^-(x_1,x_2) = H\,\varphi_{l_1}^j\left(x_1 - 2^{-j}k_1\right) H\varphi_{l_2}^j\left(x_2 - 2^{-j}k_2\right) \tag{2.16}$$

Now the bases function problem is that they oscillate in two directions instead of one in x space. To solve this problem we combine the primal and dual (Hilbert-transformed) basis.

$$\varphi_\mu^{(1)} = \frac{\varphi_\mu^+ + \varphi_\mu^-}{2} \quad ,\varphi_\mu^{(2)} = \frac{\varphi_\mu^+ - \varphi_\mu^-}{2} \tag{2.17}$$

By now, basis functions have two bumps in the frequency plane and they are symmetric with respect to the origin, so we get purely directional wave atom. $\varphi_\mu^{(1)}$ and $\varphi_\mu^{(2)}$form the wave atom frame.

The discretization of wave atoms closely follows the strategy of frequency sampling and wrapping used for curvelets [26].

CHAPTER 3
HIDDEN MARKOV MODEL

Although there has been several modeling techniques for face identification, this section focus only on one approach that view a human face as a time series sequence represented by a Markovian process.

3.1 Hidden Markov Model Principles

In spite of studied in the late 1960s, statistical methods of Markov source or hidden Markov modeling have become increasingly popular in the last decade.

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. HMM consist of two interrelated processes:

a. An unobservable Markov chain with limited number of status in the model, the observation symbol probability matrix B, a state transition probability matrix A, and initial state distribution Π.
b. A set of Probability Density Functions (PDF) associated with each state.

Using these notations, a HMM is defined as the triplet λ = (A, B, Π). The states in HMM are hidden and only emitting symbols are observed.

Let N is the number of states, and the state at time t is given by q_t ,$1 < t < T$, where T is the length of the observation sequence. Then the initial state distribution $\Pi = \{\pi_i\}, where\ \pi_i = p\{q_i = i\}, 1 < i < N.$ Now the state transition probability matrix becomes $A = \{a_{ij}\}$, where $a_{ij} = p\{q_{t+1} = j \mid q_t = i\}, 1 < i, j < N, 0 < a_{ij} < 1, and \sum_{j=1}^{N} a_{ij} = 1.$ finally the observation symbol probability matrix $B = \{b_j(o_t)\}$ is approximated by the weighted sum of M Gaussian distributions where $b_j(o_t) = \sum_{m=1}^{M} c_{j,m}\ N(\mu_{j,m}, \sigma_{j,m}, o_t)$ where $N(\mu_{j,m}, \sigma_{j,m}, o_t)$ is a Gaussian pdf mean vector $\mu_{j,m}$ and covariance matrix $\sigma_{j,m}$, $c_{j,m}$ is weighted coefficients for the mth mixture in state j with constraints $c_{j,m} \geq 0,\ 1 < j < N,\ 1 < m < M\ and\ \sum_{m=1}^{M} c_{j,m} = 1.$

HMMs are typically used to address three unique problems [27]:

1. Evaluation: Given a model λ and a sequence of observations *O*, how one could efficiently compute *P (O*| λ*)*.
2. Decoding: Given a model λ and a sequence of observations *O*, what is the hidden state sequence q^* most likely to have produced *O*, i.e., q^* = arg max$_q$ [*P (q*| λ*, O)*].

3. Parameter estimation: Given an observation sequence *O*, what model λ is most likely to have produced *O*.

The first problem is typically used for pattern recognition tasks; a number of distinct HMMs used to generate the probability of an observation sequence, each of which corresponds to a class of pattern. The pattern is classified as belonging to the same class as the HMM which produces the highest probability. The second problem can be used to find the optimal state sequence in the application or to learn about the structure of a model. The last problem is referred to as training, because the model's parameters are adjusted until some convergence criterion is reached. Typically, a number of observation sequences are used to train a model.

In deeply manner, in the first process, each state j has an associated observation probability distribution $b_j(o_t)$ which determines the probability of generating observation o_t at time t and each pair of states i and j has an associated transition probability a_{ij} . Figure 3.3 shows an example of this process where the six state model moves through the state sequence $X = 1; 2; 2; 3; 4; 4; 5; 6$ in order to generate the sequence o_1 to o_6 [28].

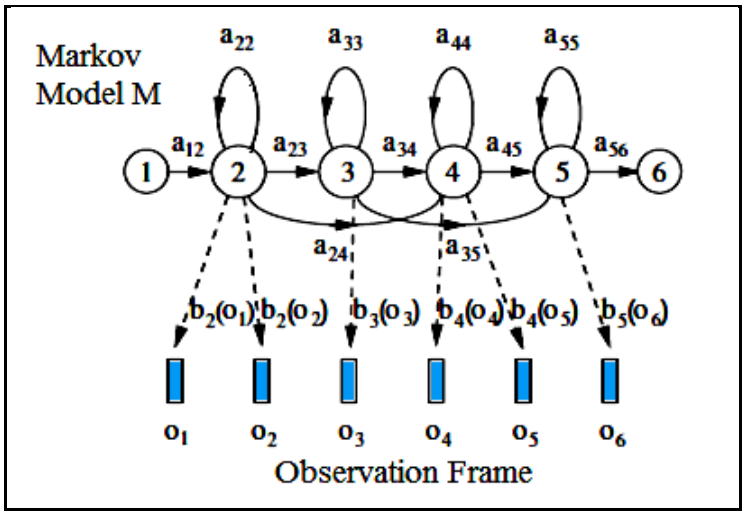


Figure 3.1: A Simple HMM

The observation probability is represented by a mixture Gaussian density. The mathematical form of an m component Gaussian mixture for D dimensional input vectors is,

$$p\langle \mathrm{x}|\mathrm{M}\rangle = \sum_{n=1}^{m} \left(a_n \frac{1}{(2\pi)^{\mathrm{D}/2}|\sum \mathrm{n}|^{1/2}}\right) \mathrm{e}^{\left(-\frac{1}{2}(\mathrm{x}-\mu)^{\mathrm{T}} \sum_{\mathrm{n}}^{-1}(\mathrm{x}-\mu_{\mathrm{n}})\right)} \qquad (3.1)$$

Where $p\langle x|M\rangle$ is the likelihood of $\boldsymbol{x}$ given the mixture model, *M*. The mixture model consists of a weighted sum over *m* unimodal Gaussian densities each parameterized by the mean vectors, μ_n, and covariance matrices, σ_n. The coefficients, a_n, are the mixture weights, which are constrained to be positive and must sum to one. The parameters of a Gaussian mixture model, a_n, μ_n and σ_n for n=1…m may be estimated using maximum likelihood criterion via the iterative Expectation-Maximizations (EM) algorithm. In general, fewer than ten iterations of the EM algorithm will provide sufficient parameter convergence [29].

In the second process, each state acts as PDF with its own parameters. The Gaussian mixture (GM) modeling is commonly used for each state, where each state has its own GM. When a feature vector enters a state, the PDF of that vector is performed according to the GM model of that state.

3.2 Parameters Estimation

HMM initializes mean and variance of each GM by using Baum-Welch formulae, then used EM algorithm to re-estimate more accurate parameters.

To go deep, the training observation vectors o_t were assigned to all states. Using Baum-Welch formulae, calculate mean μ_n and variance σ_n of state n (equations 3.2 and 3.3).

$$\hat{\mu}_n = \frac{\sum_{t=1}^{T} L_n(t)\, o_t}{\sum_{t=1}^{T} L_n(t)} \tag{3.2}$$

$$\hat{\sigma}_n = \frac{\sum_{t=1}^{T} L_n(t)(o_t - \hat{\mu}_j)(o_t - \hat{\mu}_j)'}{\sum_{t=1}^{T} L_n(t)} \tag{3.3}$$

Where $(o_t - \hat{\mu}_j)'$ is the transpose of $(o_t - \hat{\mu}_j)$, and $L_n(t)$ is the probability of being in state n at time t. forward-backward algorithm was used to calculate $L_n(t)$. Suppose $\alpha_n(t)$ is the joint probability of observing the first t vectors and being in state n at time t, for such a model M and N state,

$$\alpha_n(t) = P(o_1, \dots\dots, o_t, x(t) = n|M) \tag{3.4}$$

By induction, we have,

$$P(O|M) = \sum_{t=1}^{T} \alpha_n(t) \tag{3.5}$$

The backward probability $\beta_n(t)$,

$$\beta_n(t) = P(o_{t+1}, \dots\dots, o_T, x(t) = n, M) \quad (3.4)$$

Once again inductively we have,

$$\beta_n(t) = \sum_{j=2}^{N-1} \alpha_{nj}\, b_j(o_{t+1})\beta_j(t+1) \quad (3.5)$$

By now,

$$L_n(t) = {}^{1}/_{P}\ \alpha_n(t)\,\beta_n(t) \quad , \qquad where\ P = P(O|M) \quad (3.6)$$

After parameters initialization, HMM uses viterbi algorithm to find the best path given such a sequence. To do so, the EM algorithm was used to re-estimate mean and variance. The EM algorithm is an iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. Each iteration of the EM algorithm alternates between performing two processes:

1. The expectation step (E-step).
2. The maximization step (M-step).

In E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimates of the missing data from the E-step are used in instead of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood during the iterations. However, the great advantage of the algorithm is that its convergence is smooth.

The Maximum likelihood Gaussian mixture model, to estimate probability density function $p\langle \mathrm{x}|\mathrm{M}\rangle$. It was estimated by the mean log likelihood over the sequence, $X=\{X_1 \dots X_N\}$,

$$S(X) = \log\ \mathrm{P}\langle \mathrm{X}|\mathrm{M}\rangle = \frac{1}{N}\sum_{n=1}^{m} \log\ \mathrm{P}\langle \mathrm{X_n}|\mathrm{M}\rangle \quad (3.7)$$

Suggest that the states are already trained. In Figure 3.3 the observation sequence had 6 frames; states and frames were represented by vertical and horizontal dimensions, respectively. Dots and arcs represent the log observation probability and log transition probability, respectively. The log probability of any path is the summation of all log

probabilities (observation and transitions) were contained on that path, the best time-alignment of the frames is obtained, and this is so-called Viterbi algorithm.

In recognition stage, preferably, the decision depend on the maximum score path in the model, which is obtained by the viterbi algorithm. At each node in the viterbi trellis, only the best path leading to this node is kept and the rest are dropped. This continues until all frames are done. Thus at each node to select one path and one score is recorded.

3.3 HMM Applications

Neither the theory of HMM nor its application in the field of recognition is new. Hidden Markov Models have been successfully used for speech recognition where data is essentially one dimensional. Extension to a fully connected two dimensional HMM has been shown to be computationally very complex.

HMM seems to be a promising method that works well for images with variation in different lighting, facial expression, and orientation. The system being modeled is assumed to be a Markov process with unknown parameters, and the goal is to find hidden parameters from the observable parameters. Each state in HMM has a probability distribution over the possible output whereas each state in a regular Markov model is observable.

CHAPTER 4
SYSTEM DESIGN

The project contains three main stages: preprocessing stage followed by feature extraction stage and the classification stage. MatlabR2008a were used to implement this project. Three datasets used for testing; AT&T and Essex Grimace datasets. Preprocessing was applied to the datasets. In Feature Extraction stage, the images are decomposed by two levels Curvelet and Waveatom Transform.

4.1 Preprocessing Stage

As preprocessing, first Color images of Essex Grimace database are converted to gray scale images using equation 4.1. Let R, G, B be red, green, blue value of colored image,

$$gray_{value} = 0.2989 \times R + 0.587 \times G + 0.114 \times B \quad (4.1)$$

For each dataset, all face images are quantized into 8 gray levels. The intensity image was scaled and rounds produce an equivalent indexed image. Then cropped face region and resized it into 120×120.

4.2 Feature Extraction Stage

The most important property of multiscale transforms, in the recognition task, is the sparsity. Features choice will be no longer critical task, if we use the sparsity of the transform in correct way. The important here is answering two questions: the first question is if the number of features is large or not. The second one is if the sparse representation was calculated in correct manner or not.

Sparsest representation has discriminant nature, from all subsets of features; it selects the most representative subset and most compactly expressed the input image and rejects all other less compact representations.

The underlying texture structure was zoomed on and out by the analysis of multiresolution transforms, therefore, the extracted texture does not affected by the size of neighboring pixels.

The multiscale transforms theory yields to two important points; the role of feature extraction and the difficulty due to occlusion. Feature extraction role of Curvelet and Waveatom transforms was covered on chapter two. The occlusion difficulty will be covered in the next section, when talking about the classifier.

The feature extraction stage was implemented using two modern multiresolution transform; the first is the Curvelet transform and the second is the Waveatom transform. CurveLab-2.1.2 Matlab package [30] was used to yield Curvelet feature vectors. Waveatom feature vectors were outperformed using WaveAtom-1.1.1 matlab package [31].

4.2.1 Curvelet based Feature Extraction

In order to extract feature vectors, the images are decomposed into its approximate and detailed components using two levels of Curvelet transform. These sub-images thus obtained are called *curvefaces*. These *curvefaces* greatly reduces the dimensionality of the original image. Thereafter only the approximate components are selected to perform further computations, as they account for maximum variance. Thus, a representative and efficient feature set is produced. Figure 4.1 shows the Curvelet coefficients of a face from ORL dataset decomposed at scale = 2 and angle = 8. The image in the first row is the original image. The coarse scale (low frequency) is first image in the second row, which so-called approximate coefficients. The other images are finest scale which so-called detail coefficients.

The face images, belongs to ORL and Essex Grimace dataset, are decomposed using Curvelet transform at scale = 3 and angle = 8. Thus 25 components are produced, including 1 approximate and 24 detailed sub-band. The resolution of the approximate subband is reduced to 42 x 37 and 42 x 42 for images of ORL and Essex Grimace respectively. To further reduce the dimensionality, Curvelet transform, at scale = 3 and angle = 8, was applied once again on these approximate components only. The resolution of the approximated subband became 13x15 and 15x15 for images of ORL and Essex Grimace respectively as shown in Figure 4.2. A total of 195 features of Curvelet sub-images are produced.

4.2.2 Waveatom based Feature Extraction

Waveatom decomposition is used for sparse representation of face images since they belong to a category of images that oscillate smoothly in varying directions. Firstly, discrete 2D Waveatom decomposition is applied to the original face image; to efficiently capture coherence patterns along and across the oscillations. Figure 4.3 shows the Waveatom coefficients of a face from ORL dataset decomposed at different scales. The

highest coefficients value is inserted first in the coefficients array within each scale; the array sorted descending. The upside right block shows the coefficients at scale three, actually the number of feature vector is 8 x 8 features. The down side left block shows scale four coefficients, coefficients array length 50 x 50 coefficients. The block in the downside right shows scale five coefficients which is 120 x 120 coefficients.

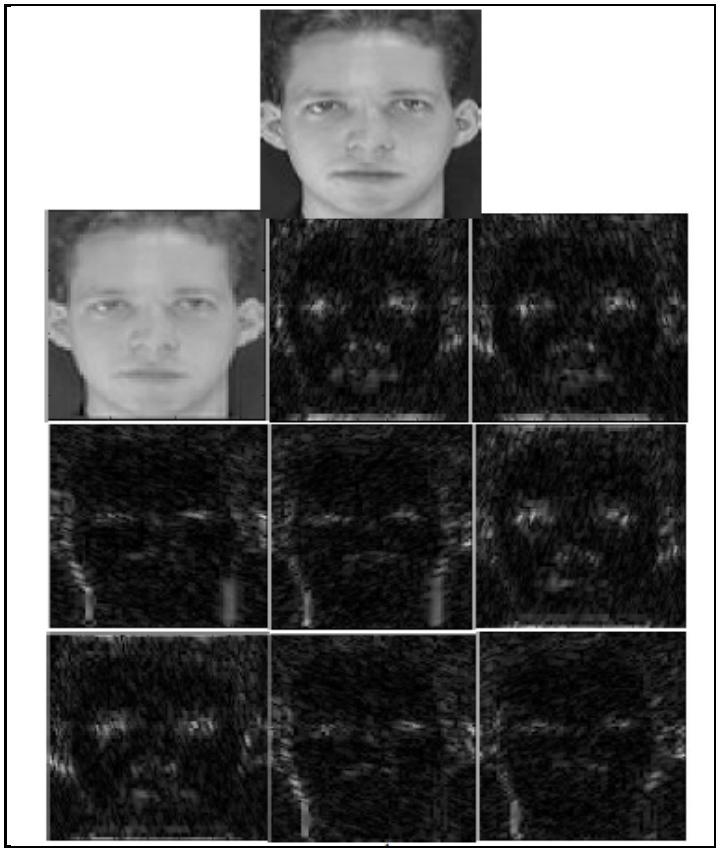


Figure 4.1: Curvelet Coefficients (Approximated and Details)

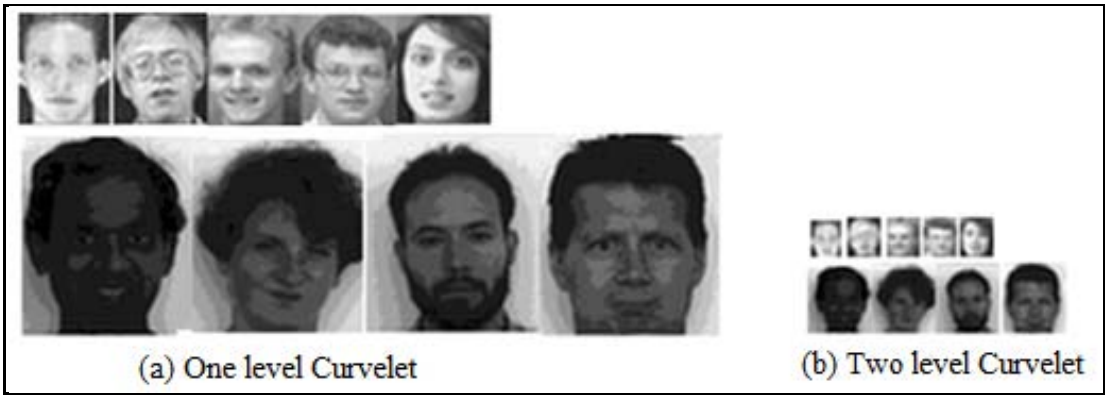


Figure 4.2: Sample Images of Curvelet Transform

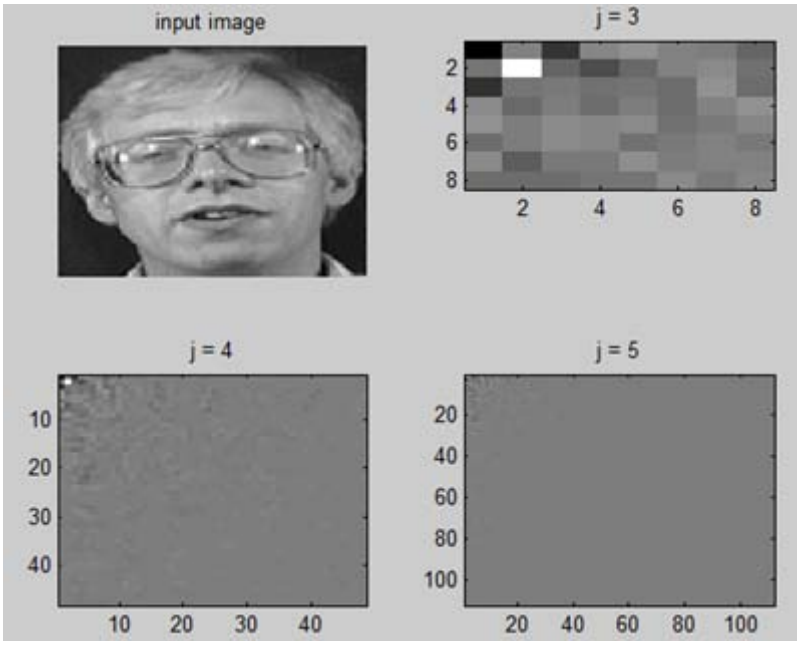


Figure 4.3: Coefficients of Waveatom Transform at Different Scales

To study the marginal statistics of the Waveatom coefficients of images, looking to Figure 4.4 to note that, the histograms of two orientations at two successive scales of a face image (two following scale). The distributions are characterized by a very sharp peak at zero amplitude and extended tails to both sides of the peak (leptokurtic). This leptokurtic behavior is observed on all histograms of all orientations and scales of all images in our test set. This implies that the Waveatom transform is very sparse, as the majority of coefficients have amplitudes close to zero. The kurtoses of these distributions are higher than the Gaussian value of 3. Where the Kurtoses are defined as the fourth central moment divided by the square of the variance as follows:

$$Kurtosis = \frac{\mathrm{E}[(\mathrm{x} - \bar{\mathrm{x}})^4]}{\sigma^4} \qquad (2.18)$$

Thus, the marginal distributions of images in Waveatom domain are highly non-Gaussian.

From all discussions and characteristics that viewed in this chapter, the Curvelet transform and Waveatom transform are used in this thesis as novel techniques for features extraction, as will be discussed later.

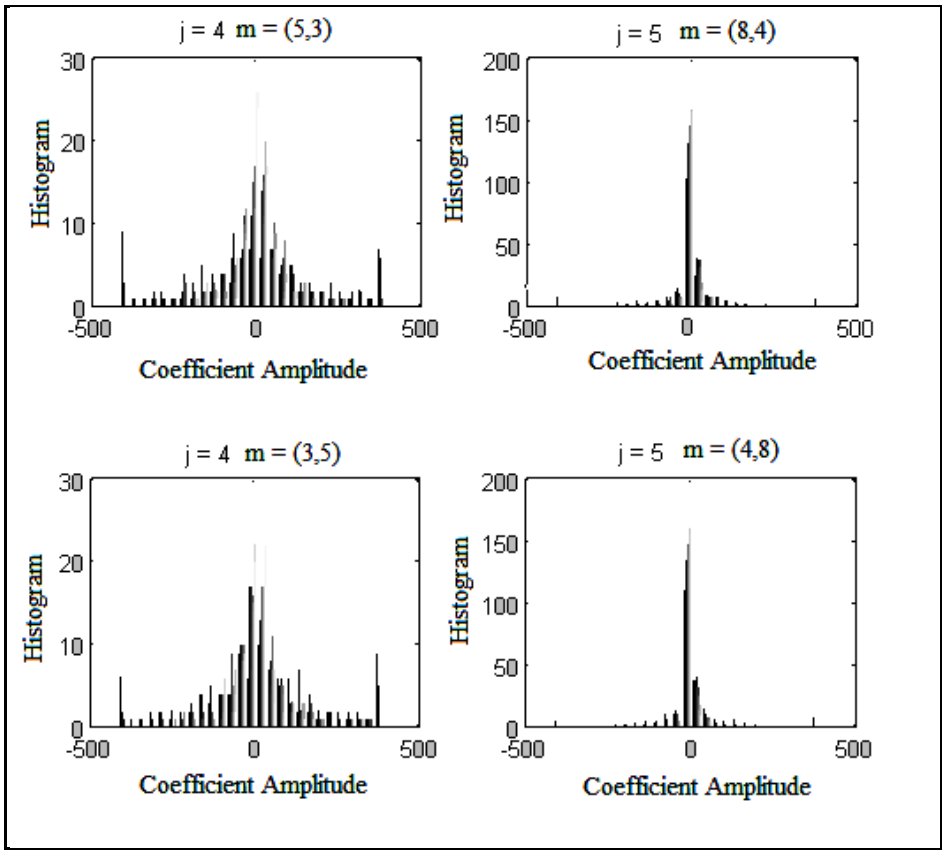


Figure 4.4: Marginal Distributions

4.3 Classification Stage

Before going in to the identification task, In order to be able to compare the feature vectors obtained from Curvelet and Waveatom transforms, they were needed to put into the same scale. Therefore, a normalization of these values using the Gaussian model based Z-score normalization method was performed. The following formula was used for normalization:

$$normalized_{z-score} = \frac{row_{score} - \mu}{\sigma} \qquad (4.2)$$

Where μ and σ are the mean and standard deviation for such a row. This normalization maps the score onto a common scale and removes the dependencies of the scores too.

The classification is performed by HMM. Many efforts spend to utilize the HMM technique to the classification task. In particular, a new application of the well-known HMM, which is implemented in HMMall matlab toolbox [32].

4.3.1 Artificial Neural Network

Neural networks are composed of simple elements operating in parallel. The neuron model shown in Figure 4.5 is the one that widely used in artificial neural networks.

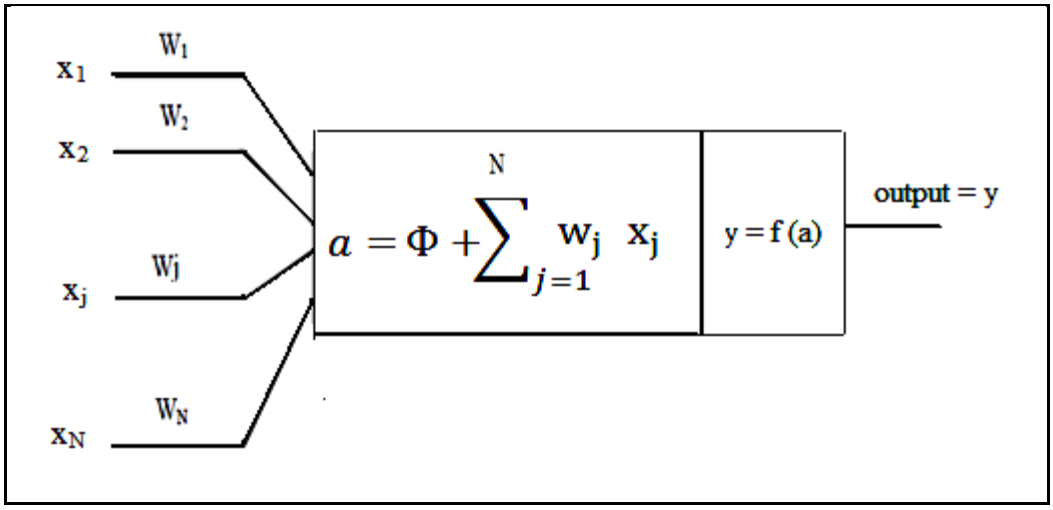


Figure 4.5: Artificial Neurons

The neuron given in this figure has N input, denoted as x_1, x_2…x_N. Each line connecting these inputs to the neuron is assigned a weight, which is denoted as w_1, w_2…. w_N respectively. The threshold in artificial neuron is usually represented by Φ and the activation is given by the formula:

$$a = \Phi + \sum_{j=1}^{n} w_j \; x_j \qquad (4.3)$$

Both of the inputs and weights are real values. If Φ is positive, it is usually referred as bias. Due to its mathematical convenience (+) sign is used in the activation formula. Sometimes, the threshold is combined into the summation part by assuming an imaginary input $x_0 = +1$ and a connection weight $w_0 = \Phi$. Equation 4.4, which is vector notation, used to update the neuron.

$$A = W^T + \Phi \qquad (4.4)$$

Function implementations can be done by adjusting the weights and the threshold of the neuron. Furthermore, by connecting the outputs of some neurons as inputs to the others, neural network will be established, and any function can be implemented by these networks. The last layer of neurons is called the output layer and the layers between the input and output layer are called the hidden layers. The input layer is made up of special input neurons, transmitting only the applied external input to their outputs.

The architecture of artificial neural network used in this thesis is a multi layer perceptron with steepest descent back-propagation training algorithm with adaptive learning rate. Figure 4.6 shows the architecture of a typical Back-propagation neural network. It contains from three layers; input layer, one hidden layer, and output layer. Tansigmoid transfer function was used in the hidden layer. Linear transfer function was used in the output layer.

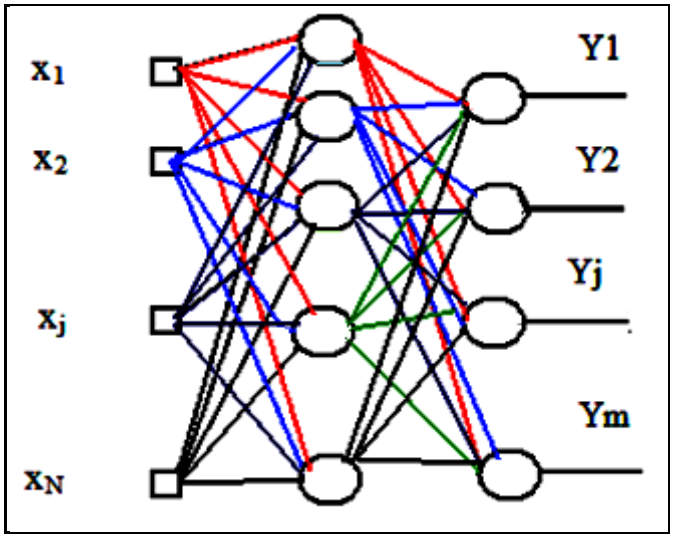


Figure 4.6: Back-propagation Neural Network

Back-propagation algorithm is used as the training method of the designed artificial neural network. The back-propagation algorithm includes the following steps:

1. Initialize weights and biases to small random numbers.
2. Present a training data to neural network and calculate the output by propagating the input forward through the network used.
3. Propagate the sensitivities backward through the network:

$$S^M = -2\,\dot{F}^M(n^M)(t-a) \tag{4.5}$$

$$S^m = \dot{F}^m(n^m)(W^{mt})^T S^{m+1} \quad \text{for m} = \text{M}-1,\ldots.,2,1 \tag{4.6}$$

Where

$$A = \begin{bmatrix} \dot{f}^m(n_1^m) & 0 & \ldots & 0 \\ 0 & \dot{f}^m(n_2^m) & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \dot{f}^m(n_{j^m}^m) \end{bmatrix} \tag{4.7}$$

$$\dot{f}^m\left(n_j^m\right) = \frac{\partial\, f^m(n_j^m)}{\partial\, n_j^m} \tag{4.8}$$

4. Calculate weight and bias updates

$$\Delta W^m(k) = -\alpha\, s^m (a^{m-1})^T \quad (4.9)$$

$$\Delta b^m(k) = -\alpha\, s^m \quad (4.10)$$

5. Update the weights and biases

$$W^m(k+1) = W^m(k) + \Delta W^m(k) \quad (4.11)$$

$$b^m(k+1) = b^m(k) + \Delta b^m(k) \quad (4.12)$$

6. Repeat steps 2-5 until convergence.

4.3.2 HMM for Face Recognition

In this section, the goal is to build HMM model assigned to the different faces contained in the training sets. One HMM is built for each subject; the training process is performed using the Baum–Welch algorithm. Each feature vector is modeled by *continuous left-to-right* HMMs. Each HMM state generates a *mixture of Gaussian densities*. The number of states and the number of densities per state that are appropriate to model each class depend on the amount of training data available for that class.

In the feature extraction phase of HMM model, each individual in the dataset is represented by a HMM face model. First, the HMM is initialized. The image is segmented from top to bottom. Each segment is fixed-size 2D window that belongs to some predefined facial regions as shown in Figure 4.7. In other words, extract the observation vector sequences from the entire face image. Each observation sequences a multidimensional vector. Each state is assigned one and only one facial region. Actually, facial regions are all clusters of vectors that are formed using k-means algorithm. Then the image data within a region is modeled by a multivariate Gaussian distribution. Next, model parameters are re-estimated by a process called E-M procedure to maximize the model probability until convergence. In other words, one state is responsible for characterizing the observation vectors of human foreheads, and another state is responsible for characterizing the observation vectors of human eyes.

In the training phase, HMM builds a model for each person in the database. Each HMM is trained with different face images of the same person. Each parameter of this

model will be trained through two levels Curvelet feature vectors and Waveatom feature vectors.

The HMM testing phase consists of partitioning each test image into blocks and assigning a facial region to each of them. K-means clustering algorithm has the responsibility to perform this phase; k is the number of selected facial region. This sequence of blocks with their facial regions was used to express a face image. Now, the recognition phase will be performed by founding the model within the training sets that maximizes the likelihood of a tested face image.

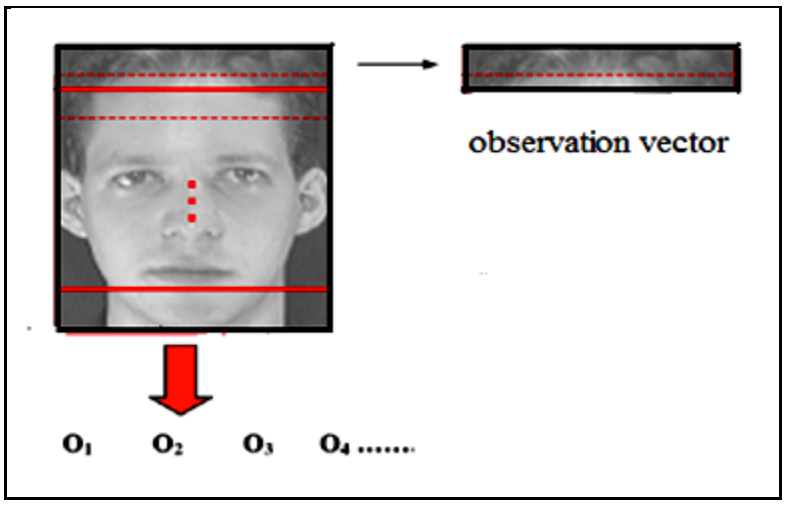


Figure 4.7: HMM observation vector

HMM chooses the Maximum Likelihood model (ML) using the Maximum Likelihood criterion as follows:

$$ML = \frac{1}{N} \log p(X|\theta_{ml}) \tag{4.13}$$

Where X is the dataset and θ_{ml} is maximum likelihood estimate parameters. HMM likelihood increases with the number of parameters in the model. The shortcoming of ML criterion is that there is no penalty to prevent the number of model parameters from increasing. On the other side, maximum likelihood estimation methods two attractive points; the first point is that they have good convergence properties as the number of training samples increases. The second point is that maximum likelihood estimation often can be found simply.

HMM is good to deal with noisy or distorted data but it tends towered outputting high probability not for the proper class data, but for out-class data too. So HMM discriminant is reduced.

To overcome this problem, it's needed to build enclosed models (out-class) for each class using included and occluded data. The procedure as follows: first of all, HMM standard model was trained using in-class data. Second, standard models recognition results are used to defined the confusion sets data. Again, HMM standard model was trained using out-class data. Finally, for each class, the final model (ML_{fin}) was calculated using the following equation:

$$ML_{fin} = \frac{1}{N} \log p(X|\theta_{ml}) - \frac{1}{N'} \log p(X'|\theta'_{ml}) = ML_{in} - ML_{out} \quad (4.14)$$

Where θ_{ml} and θ'_{ml}, X and X' and N and N', are the maximum likelihood estimate of parameters for in-class and out-class, in-class and out-class training data sets, and number for in-class and out-class training data sets respectively. ML_{in} and ML_{out} are in-class model and out-class model respectively.

4.4 Summary

The procedure can be summarized as follows:

1. Preprocessing Stage
    I. Convert colored images into gray scale, each face image, crop the face region, quantize into 8 gray levels, and resize it.
    II. Original data set will be in-class data sets.
    III. Add inclusions and occlusions to original data sets to define out-class sets.
2. Feature Extraction Stage
    I. Two levels Curvelet decomposition applied to all face images, approximated coefficients will be the feature vector.
    II. Decompose all face images using Waveatom transform, coefficients at scale three will be the feature vector.
3. Classification Stage
    I. Part A

    Train ANN using both ORL and Essex Grimace data sets.

    II. Part B

Train HMM using both ORL and Essex Grimace data sets.

III. Part C
Using in-classes data sets, train HMM
Using out-classes data sets, train HMM
Compute ML_{fin} using equation 4.14

End.

CHAPTER 5
SIMULATION RESULTS

This chapter talks about the Experimental results obtained and how it obtained, discusses it in deep, and compares it to show the benefit to use the proposed method.

5.1 Data Collection

Experiments were carried out using three datasets from different sources: ORL (AT&T) database, Essex Grimace database and Yale database, both sets are used to implement different Algorithms to recognize the human face.

ORL (AT&T) database [33] contains distinct face images sets for 40 persons with dimension of 92×112, and each set consists of 10 different images for the same person. For some persons, images were taken at different times varying the lighting, facial expression (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the faces in an upright, frontal position (with tolerance for some side movement). Sample images of this dataset are shown in Figure 5.1.

Figure 5.1: Sample Images from ORL database

Essex Grimace database [34] contains sequence face images for 18 persons each one has 20 images (180×200), all images taken with a fixed camera for male and female. During the sequence, the subject moves his/her head and makes grimaces which get more extreme towards the end of the sequence. Images are taken against a plain background, with very little variation in illumination. Sample images of this database are shown in Figure 5.2. For the purposes of the experiments carried out, the Essex faces were converted to grayscale prior to training.

Experiments were carried out using Matlab R2008a on 2 Due CPU, 2.27 GHz laptop processor, and 2 GB RAM.

First of all, the preprocessing is performed on all face images and for all datasets.

Figure 5.2: Sample Images from Essex Grimace Database

First of all, the preprocessing is performed on all face images and for all datasets. In order to build a confusion datasets; inclusion data sets were built by add gray level to face image in different parts of faces while occlusion data sets were built by cropping different parts of faces. Bellow, Figure 5.3 shows samples of out-class image from ORL and Yale datasets. Indeed, for each dataset, each person has two enclosed data classes; the first class is included data class which contains four included face images and six normal face images, the second class is occluded data class which contains four occluded face images and six normal face images.

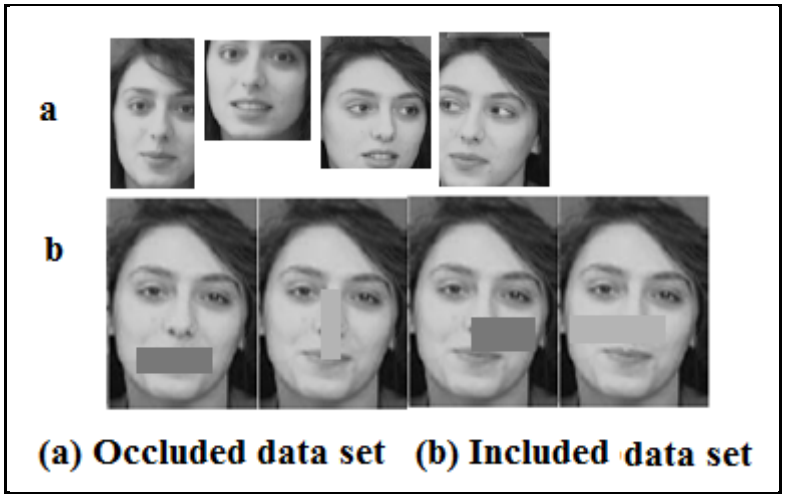


Figure 5.3: Out-class Image Samples

In order to assess the efficiency of the proposed technique described in the previous chapter, a series of experiments were carried out using all databases separately. Therefore, with AT&T database, six images were used for training and four images for testing during each run. When using the Essex 95 database, nine images were used for training and eleven images for testing during each run. One HMM model was trained for each individual in the database.

After feature extraction take place, it's time to perform the identification task. The classification is performed by two different techniques; the first is gradient descent back-propagation neural network with adaptive learning rate, while the second is HMM.

5.2 Neural Network for Face Recognition

Neural networks have been widely used in many pattern recognition problems, such as optical character recognition, and object recognition. Since face detection can be treated as a two class pattern recognition problem, various neural network architectures have been proposed. The advantage of using neural networks for face detection is the feasibility of training a system. This feasibility used to capture the complex class conditional density of face patterns. However, one drawback is that the network architecture has to be extensively tuned (number of layers, number of nodes, learning rates, etc.) to get exceptional performance.

The aim of initial experiments was to investigate the efficiency of using both of two levels Curvelet transform and Waveatom transform for extract features with gradient descent backpropagation ANN based face recognition task.

The gradient descent with momentum training function was used to update weight and bias values. Tan sigmoid transfer function and linear transfer function were used in hidden layers and output layer respectively. The number of hidden layer and the number of nodes in each layer are varied to obtain the optimal back-propagation neural network topology for highest performance with both 2-level Curvelet Transform and Waveatom Transform. As shown in Figures 5.4 and 5.5, the number of hidden layer and the number of nodes in each one are important parameter to improve the recognition rate (success rate).

From Figure 5.4 (a) for both ORL and Essex Grimace data sets, when the ANN topology has one hidden layer; starting with 20 nodes in the first hidden layer and increased them to be 80 nodes. It can be seen that the curves in Figure 5.4 (a) is a linear piecewise curves. The success rate obtained was 87% and 88% for ORL and Essex Grimace databases respectively using 20 nodes in the first hidden layer. With 30 nodes, the success rate increased to become 89% for ORL database and 90% for Essex Grimace database. With 40 nodes, the success rate increased to become 91% for ORL database and 95% for Essex Grimace database. Once again it increased to be 97% and 100% for

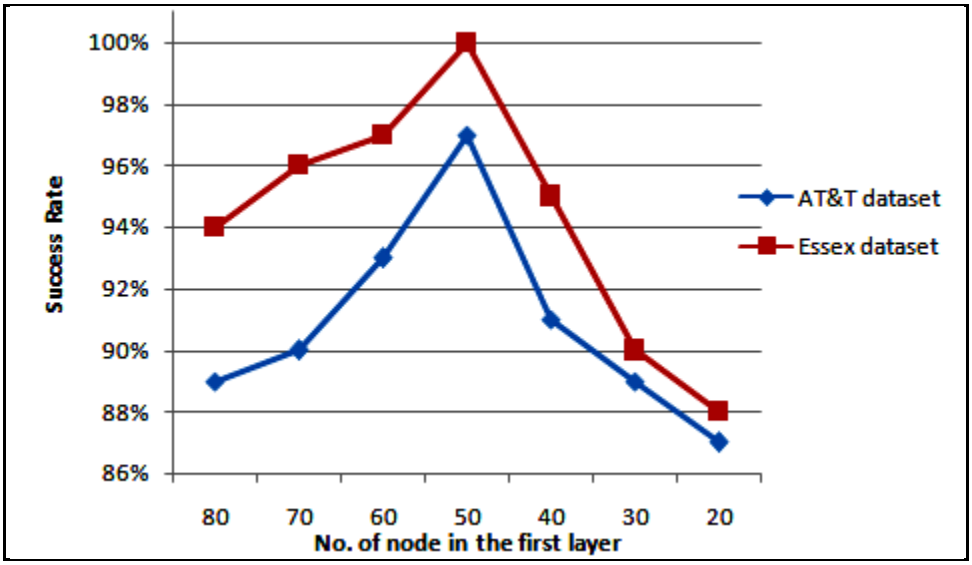
ORL and Essex Grimace databases respectively. The success rate decreased to 93% and 97% for ORL and Essex Grimace databases respectively using 60 nodes in the first hidden layer. With 70 nodes the success rate was 90% for ORL database and 96% for Essex Grimace database. With 80 nodes it decreased to 89% and 94% for ORL and Essex Grimace databases.

From Figure 5.4 (b) for both ORL and Essex Grimace data sets, when the ANN topology has two hidden layers; the number of nodes is 50 nodes in the first hidden layer. The number of nodes will be varied to found the optimum number of nodes for the second hidden layer. Beginning with 20 nodes in the second hidden layer and increased them to be 80 nodes. It can be seen that the curves in Figure 5.3 (b) is a linear piecewise curves. The success rate obtained was 85% for ORL and Essex Grimace databases. using 20 nodes in the first hidden layer. With 30 nodes, the success rate increased to become 87% for ORL database and 88% for Essex Grimace database. With 40 nodes, the success rate increased to become 90% for ORL database and 92% for Essex Grimace database. Once again it increased to be 95% and 98% for ORL and Essex Grimace databases respectively. The success rate decreased to 92% and 95% for ORL and Essex Grimace databases respectively using 60 nodes in the first hidden layer. With 70 nodes the success rate was 90% for ORL database and 93% for Essex Grimace database. With 80 nodes it decreased to 88% and 91% for ORL and Essex Grimace databases.

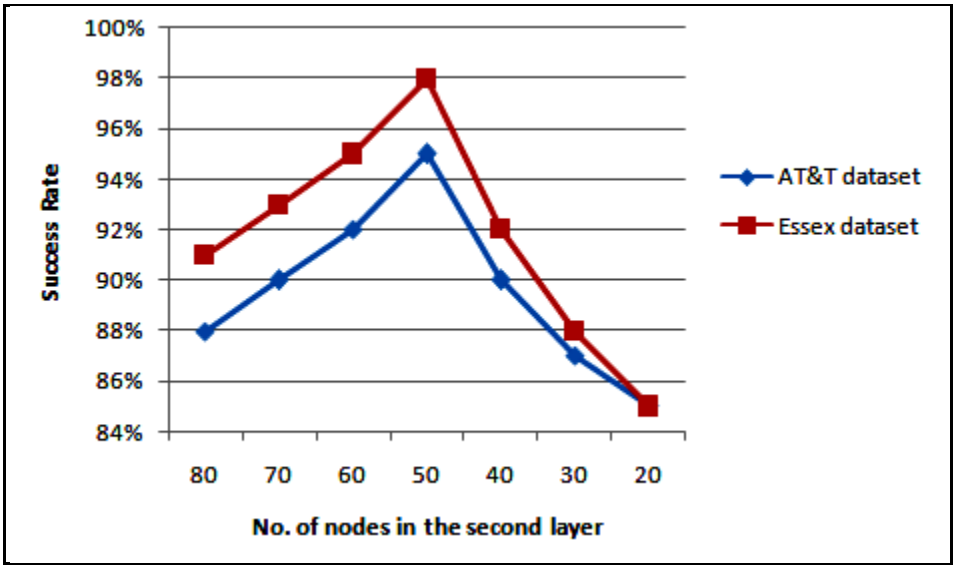
The success classification rate increased with increasing nodes till 50 nodes then decrement again. The best recognition rate with 2-level Curvelet Transform achieved 97% at 50 nodes in the first hidden layer.

On Waveatom side, number of nodes and numbers of hidden layer were varied to obtain the best neural network topology. Beginning with one hidden layer contained 10 nodes; the success rate was 90% and 98% for AT&T and Essex Grimace database. Piecewise linear incremental success rate function was shown in Figure 5.5 (a) till reaches 20 nodes in the first hidden layer. With 15 nodes, the success rate increased to become 94% for ORL database and 99% for Essex Grimace database. Once again it increased with 20 nodes to 98% and 100% for ORL and Essex Grimace databases respectively. The success rate decreased to 98% and still 100% for ORL and Essex Grimace databases respectively using 30 nodes in the first hidden layer. With 40 nodes

the success rate was 94% for ORL database and 98% for Essex Grimace database. With 50 nodes it decreased to 88% and 98% for ORL and Essex Grimace databases.



(a)



(b)

Figure 5.4: Results obtained from Optimizing ANN for 2-Level Curvelet Transform, (a): optimizing 1st layer, (b): optimizing 2nd layer

Now the topology would be consisted from 20 nodes in the first hidden layer and numbers of nodes in the second layer will be varied starting with 10 nodes till 50 nodes. As can be seen from Figure 5.5 (b), with 10 nodes the success rate was 87% and 96% for

AT&T and Essex Grimace database respectively. The success rate was 93% for ORL database and 98% for Essex Grimace database with 15 nodes. Once again it increased with 20 nodes to 91% and 100% for ORL and Essex Grimace databases respectively. The success rate decreased to 87% and 98% for ORL and Essex Grimace databases respectively using 30 nodes in the first hidden layer. With 40 nodes the success rate was 90% for ORL database and 97% for Essex Grimace database. With 50 nodes it decreased to 85% and 94% for ORL and Essex Grimace databases.

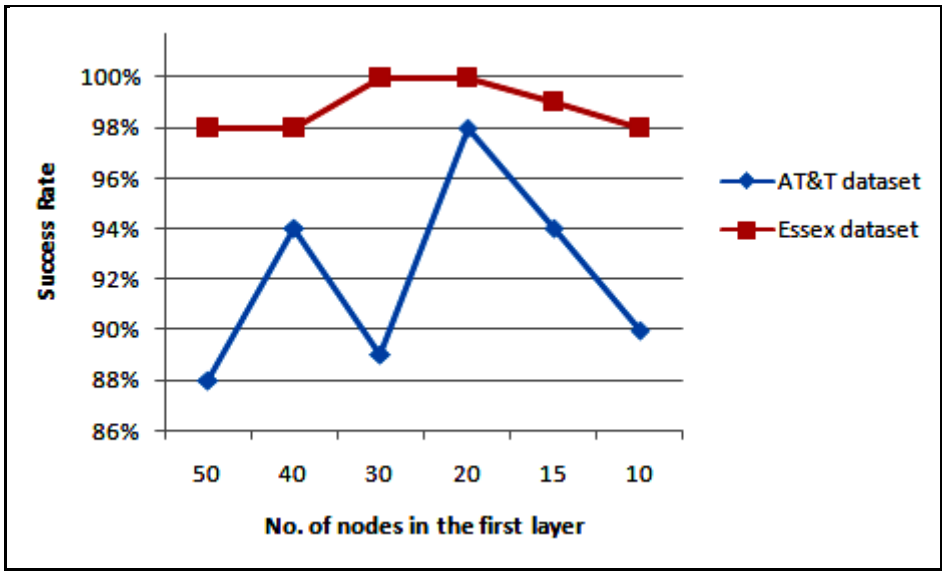
The best recognition rate with Waveatom Transform achieved 98% at 20 nodes in the first layer while the success rate was 93% at 15 nodes in the second layer for ORL database.

Clearly, the ANN topology should contain one hidden layer with 50 and 20 nodes to get the highest performance by using 2-level Curvelet and Waveatom Transforms respectively for ORL dataset. The same results are obtained for Essex Grimace dataset. The comparison between the two databases that used is not fair for many reasons: each database has different size, ORL database are gray images while Essex Grimace data are colored images, and ORL data was captured by photographic camera while Essex Grimace was obtained using video camera.

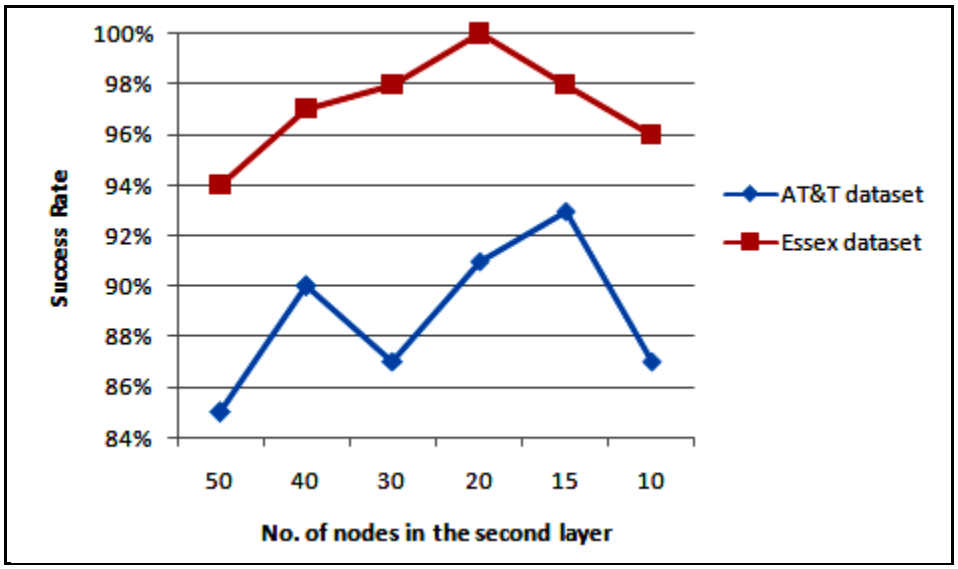
In order to show the robustness of the system, the system was tested using face images from outside training data sets. The reject rate was 1.5% with Waveatom based ANN and 2.75% with two levels Curvelet based ANN.

5.3 Hidden Markov Model

The next set of experiments was designed to investigate the efficiency of using both of two levels Curvelet transform and Waveatom transform for extract features with HMM based face recognition task. The approximated Curvelet coefficients which of size equal to 15×15, were used as feature vector. Scale three Waveatom coefficients which of size equal to 8×8, were used as feature vector. A model which contains five states as discussed in the previous chapter.



(a)



(b)

Figure 5.5: Results obtained from Optimizing ANN for Waveatom Transform, (a): optimizing 1st layer, (b): optimizing 2nd layer

A varied window Waveatom feature vector is used by using five states and optimizing both of feature vector length and the number of Gaussians densities per state. Figure 5.6 shows the classification success rate for varying Gaussian densities per state and several vector lengths for ORL (AT&T) and Essex Grimace databases, respectively. In the case of using HMM with 2 levels Curvelet Transform, one would resize feature

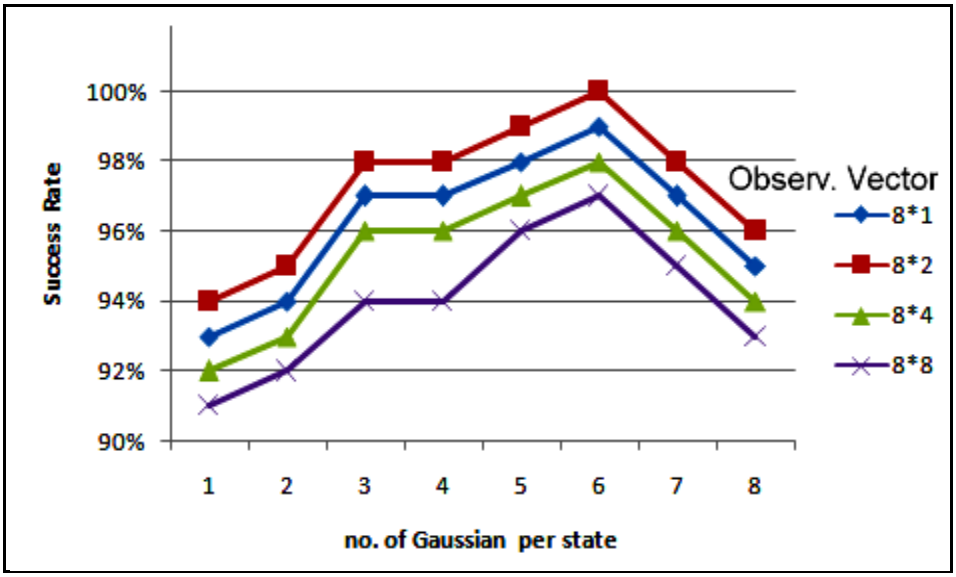
vector to be 12 x 12 and 24 x 24 for AT&T and Essex Grimace respectively; to could divide the observation vector to different lengths. Figure 5.7 shows the classification error rate for varying Gaussian densities per state and several vector lengths for ORL (AT&T) and Essex Grimace databases, respectively.

The curves in Figure 5.6 and Figure 5.7 consist of piecewise continuous lines increasing inside the interval [1,3[of Gaussian densities per state for ORL (AT&T) and Essex Grimace databases of faces, respectively. Whereas it becomes decreasing inside the interval [3,4[of Gaussian densities per state for ORL database. In the other side, it seems constants for Essex Grimace dataset in the last interval. The curves becomes increasing again in the interval [4,5[of Gaussian densities per state for both ORL and Essex Grimace dataset, respectively. Again, the curves become decreasing inside the interval [6,8[of Gaussian densities per state for both datasets.

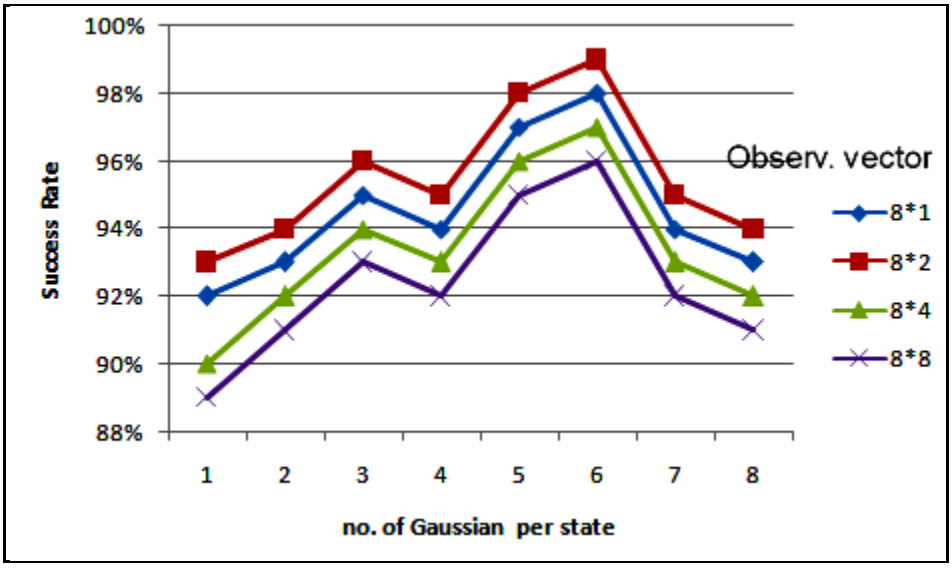
Empirically, Figure 5.6 yields that the best result for success rate by using Waveatom features were obtained using a certain values of several parameters, such as: 5 states in each HMMs, 6 Gaussian densities per each state, and a length of 8 **x** 2 window size as an observation vectors. These HMM parameters was produced a classification rate of 99% and 100% for ORL (AT&T) and Essex Grimace databases of faces, respectively.

In fact, Figure 5.7 summarizes the following: using 5 states in HMM, the best result for success rate when using 2 levels Curvelet features was 98% with AT&T faces and 100% with Essex faces. These results were obtained with a length of 24 x 3 and 12 x 3 window size for feature vectors, for both Essex Grimace and AT&T respectively, (and 6 Gaussian densities per each state).

In order to show the robustness of the system, the system was tested using face images from outside training data sets. The reject rate was 1% with Waveatom based HMM and 2% with two levels Curvelet based HMM.

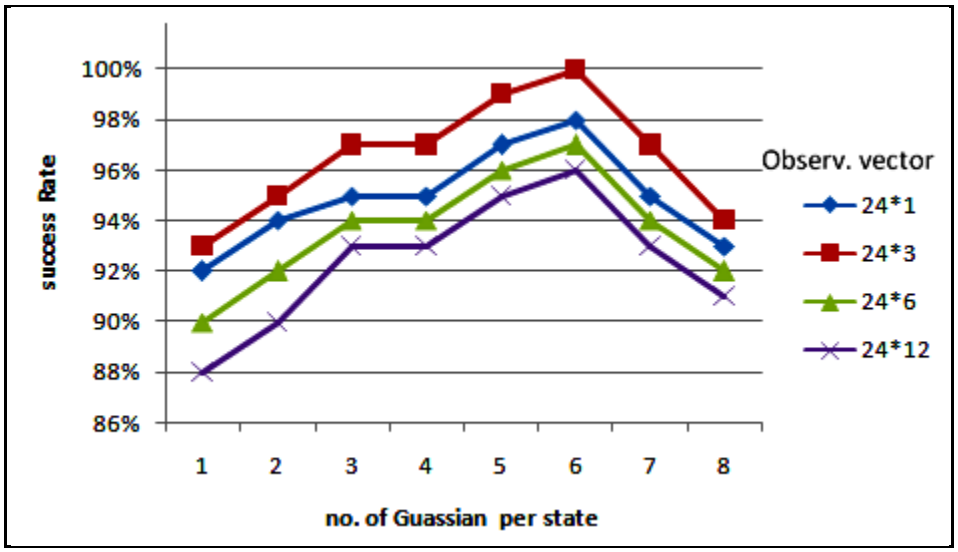


(a)

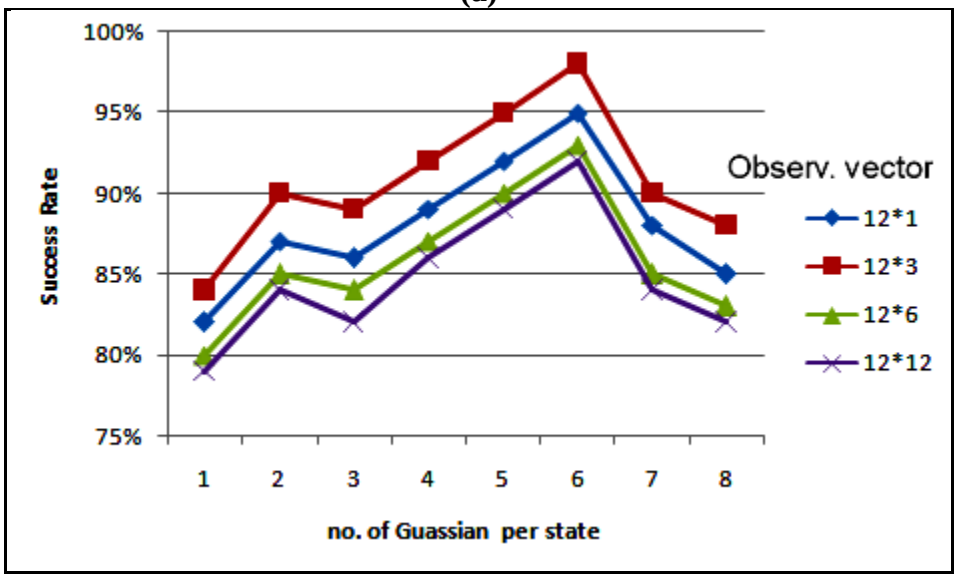


(b)

Figure 5.6: The Classification Success Rate Obtained when using Waveatom Features for Varying Gaussian Densities per State and Several Vector Lengths: (a) For Essex Grimace database of faces (b) For ORL (from AT&T) databases of faces



(a)



(b)

Figure 5.7: The Classification Success Rate Obtained when using 2 Levels Curvelet Features for Varying Gaussian Densities per State and Several Vector Lengths: (a) For Essex Grimace database of faces (b) For ORL (from AT&T) databases of faces

5.4 Comparative Study

In the previous two sections, different results have been presented. In order to show the capability of the proposed method one has compared it against the most popular existing techniques. A 3-level wavelet decomposition using 'Haar' wavelet was performed for wavelet based PCA technique. The best highest 50 eigenvectors were selected for both Curvelet and wavelet based PCA.

Figure 5.8 summarizes the results of the comparative study. It shows the recognition success rate of three different methods, and compares the results with the ANN results obtained in this thesis. These methods are two levels Curvelet Transform and Waveatom Transform. Both of them used as feature extraction techniques. The classification technique was the gradient descent backpropagation ANN. The results show that the proposal methods have the highest Recognition Rate. For two levels Curvelet features, the success rate was 97% and 100% for ORL (AT&T) and Essex Grimace databases of faces, respectively. In the other side, Waveatom features have the highest success rate. The success rate was 98% and 100% for ORL (AT&T) and Essex Grimace databases of faces, respectively. These results was against 93%, 94%, 96% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for ORL (AT&T) database of faces. And it was 70%, 98%, 100% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for Essex Grimace database of faces.

Figure 5.9 compares the recognition rate obtained of three different methods against the rate obtained when use HMM classifier. Feature extraction obtained by both tow level Curvelet Transform and Waveatom Transform. The classification technique was HMM. The results show that Waveatom features have the highest recognition rate once again. For two levels Curvelet features, the success rate was 98% and 100% for ORL (AT&T) and Essex Grimace databases of faces, respectively. In the other side, Waveatom features have the highest success rate. The success rate was 99% and 100% for ORL (AT&T) and Essex Grimace databases of faces, respectively. These results was against 93%, 95%, 96% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for ORL (AT&T) database of faces. And it was 75%, 98%, 100% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for Essex Grimace database of faces.

Backing to Figure 5.8 and 5.9, one note that both of the proposal methods for feature Extraction supersede the state of art in this era, and viewed as promised way for more studying. Actually either use the HMM classifier or the ANN classifier, Waveatom features achieve the highest success rate, behind it the features obtained from two levels Curvelet Transform.

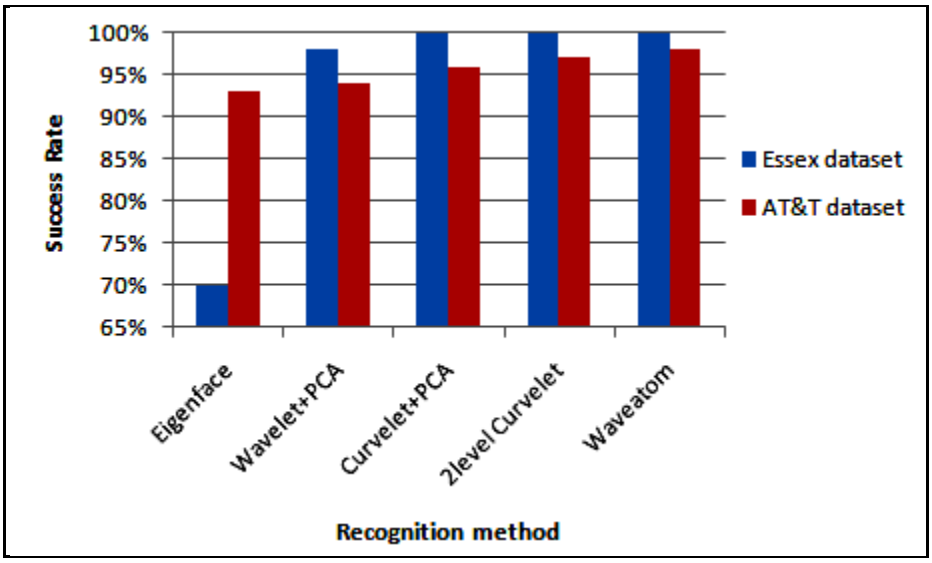


Figure 5.8: Comparative Studies for ANN

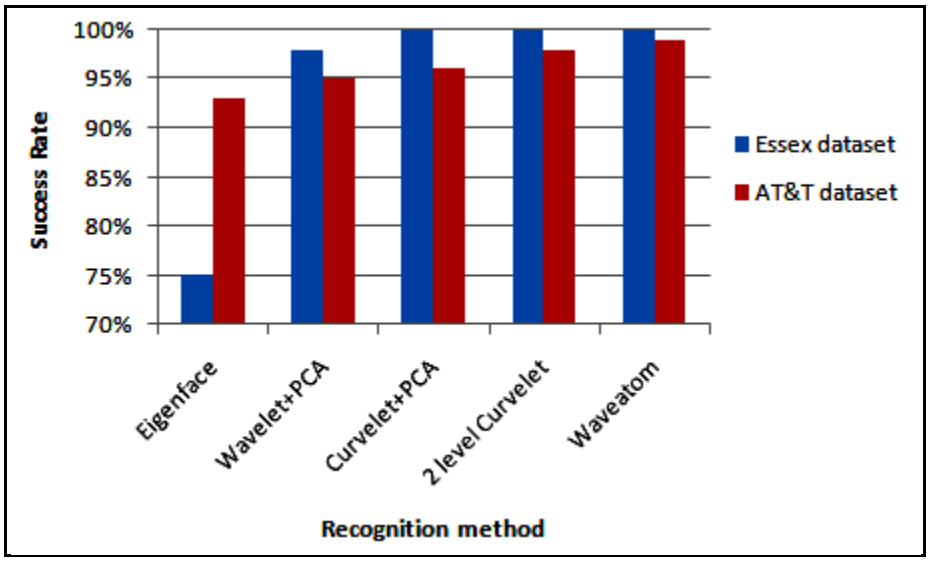


Figure 5.9: Comparative Studies for HMM

5.5 Time Requirements

In addition to identification accuracy, an important factor in face recognition system is time requirements for whole system stages.

For Feature Extraction process, Figure 5.10 shows the time requirements per one face image. Wavelet face + PCA feature extraction time was 0.2698 seconds and 0.177401 seconds for AT&T database and Essex Grimace database respectively. This time is increased when using the first generation Curvelet + PCA to become 0.194903 seconds for AT&T database and 0.311185 seconds for Essex Grimace data base, since the high

redundancy which suffers from. The time application was 0.1887 seconds for AT&T data base and 0.2388 seconds for Essex Grimace with two levels second generation Curvelet transform. Waveatom feature Extraction time was 0.12623 seconds for AT&T database and 0.168477 seconds for Essex Grimace data base.

In particular, two levels Curvelet seems to be faster than all of Wavelet + PCA and Curvelet + PCA. Actually, Waveatom transform is the fastest one of all multiresolution transforms.

For Classification task, two times should be measured; the first is training phase time and the later is testing phase time. In training both ANN and HMM classifier, the time application is measured and summarized in Figure 5.11 and 5.12, respectively.

As can be seen from Figure 5.11, Wavelet face + PCA ANN training time was 221 minutes and 71 minutes for AT&T database and Essex Grimace database respectively. The training time is increased when using the first generation Curvelet + PCA to become 194 minutes for AT&T database and 67 minutes for Essex Grimace data base. The time application was 123 minutes for AT&T data base and 39 minutes for Essex Grimace with two levels second generation Curvelet transform. Waveatom feature Extraction time was 55 minutes for AT&T database and 26 minutes for Essex Grimace data base.

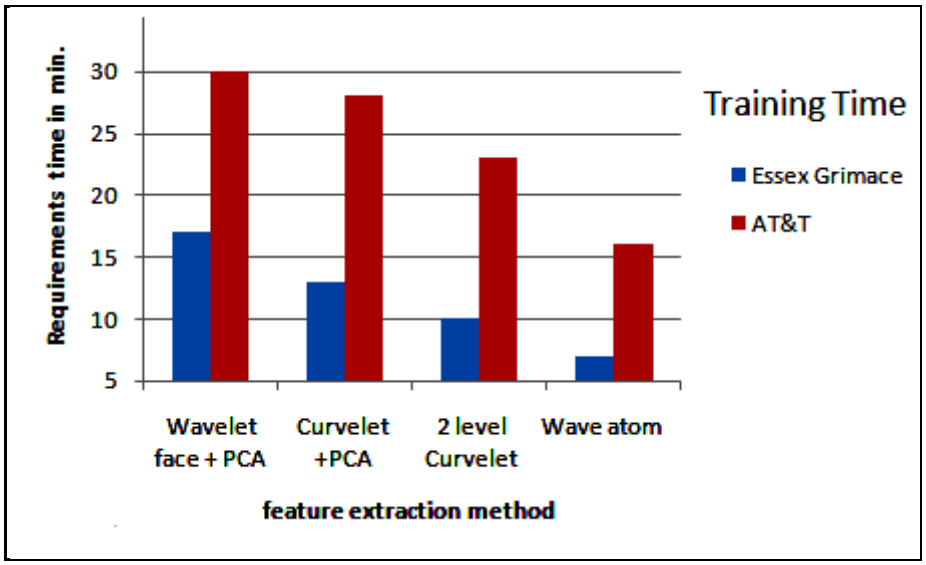


Figure 5.10: Feature Extraction's Time Requirements

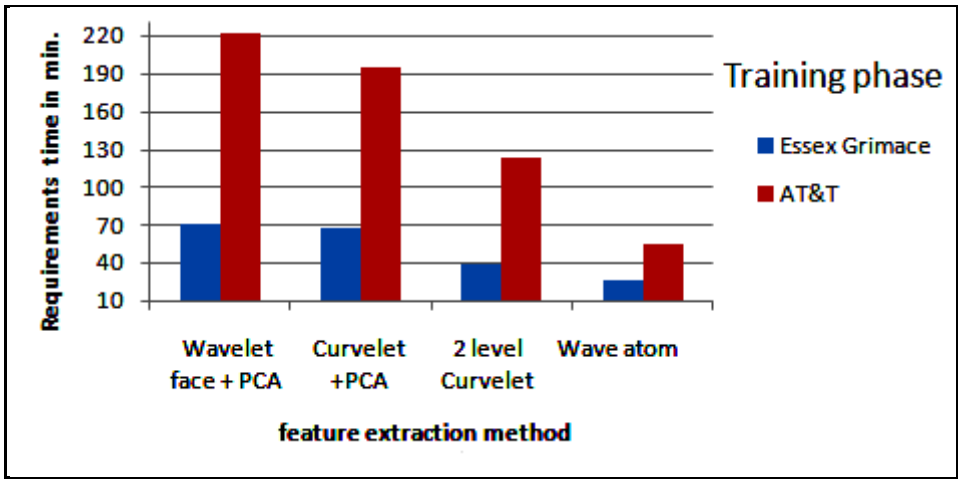


Figure 5.11: Time Requirements in Training ANN

As can be noted from Figure 5.12, Wavelet face + PCA HMM training time was 30 minutes and 17 minutes for AT&T database and Essex Grimace database respectively. The training time is increased when using the first generation Curvelet + PCA to become 28 minutes for AT&T database and 13 minutes for Essex Grimace data base. The time application was 23 minutes for AT&T data base and 10 minutes for Essex Grimace with two levels second generation Curvelet transform. Waveatom feature Extraction time was 16 minutes for AT&T database and 7 minutes for Essex Grimace data base.

In reality, the time needed to train HMM is less that needed to Train ANN. Notably; Waveatom based HMM is the fastest in training stage.

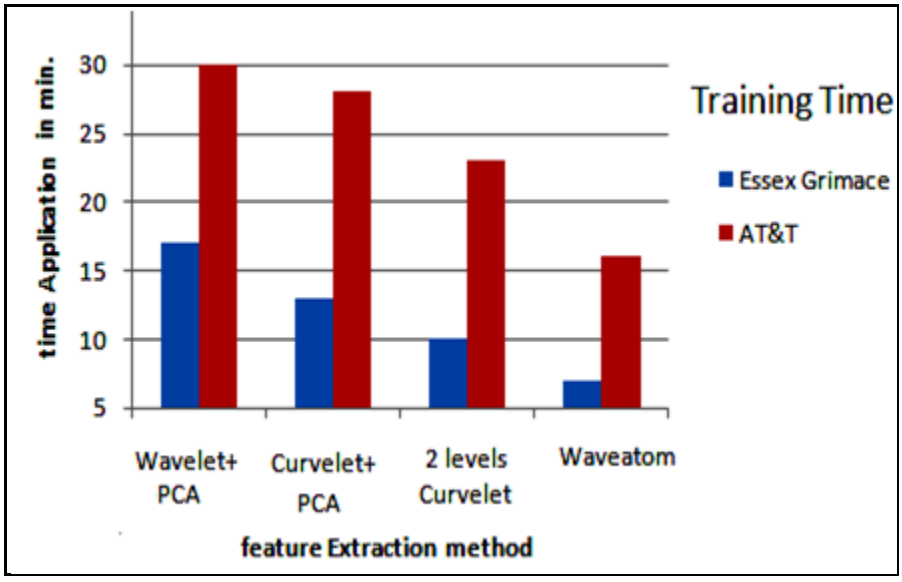


Figure 5.12: Time Requirements in Training HMM

Figure 5.13 and 5.14 summarizes the testing phase required time using ANN and HMM respectively.

As can be seen from Figure 5.13, Wavelet face + PCA ANN testing time was 0.06 seconds and 0.08 seconds for AT&T database and Essex Grimace database respectively. The testing time is increased when using the first generation Curvelet + PCA to become 0.06 seconds for AT&T database and 0.085 seconds for Essex Grimace data base. The time application was 0.054 seconds for AT&T data base and 0.164 seconds for Essex Grimace with two levels second generation Curvelet transform. Waveatom feature Extraction time was 0.053 seconds for AT&T database and 0.06 seconds for Essex Grimace data base.

As can be seen from Figure 5.14, Wavelet face + PCA HMM testing time was 0.018218 seconds and 0.126939 seconds for AT&T database and Essex Grimace database respectively. The testing time is increased when using the first generation Curvelet + PCA to become 0.020292 seconds for AT&T database and 0.028779 seconds for Essex Grimace data base. The time needed was 0.032938 seconds for AT&T data base and 0.02953 seconds for Essex Grimace with two levels second generation Curvelet transform. Waveatom feature Extraction time was 0.019093 seconds for AT&T database and 0.018434 seconds for Essex Grimace data base.

Again, HMM look like to be faster than ANN. For another time, Waveatom based HMM is the fastest in testing stage.

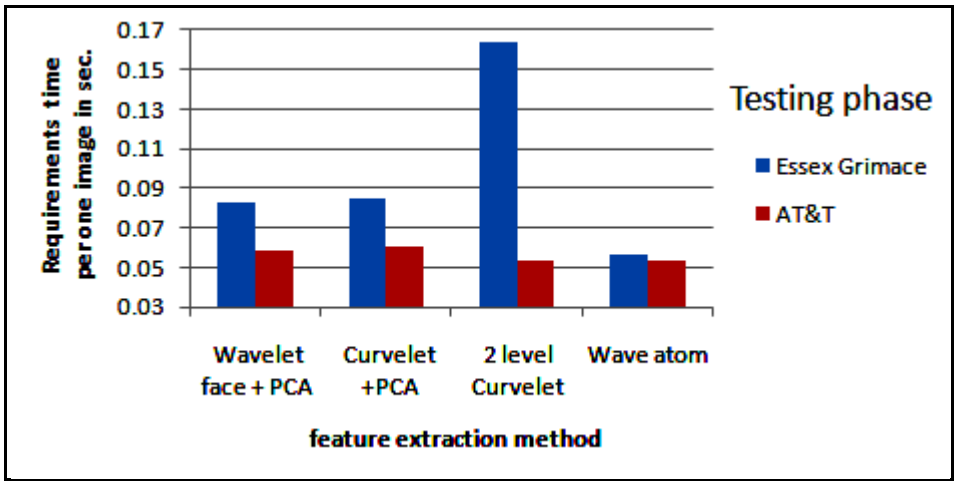


Figure 5.13: Time Requirements in Testing ANN

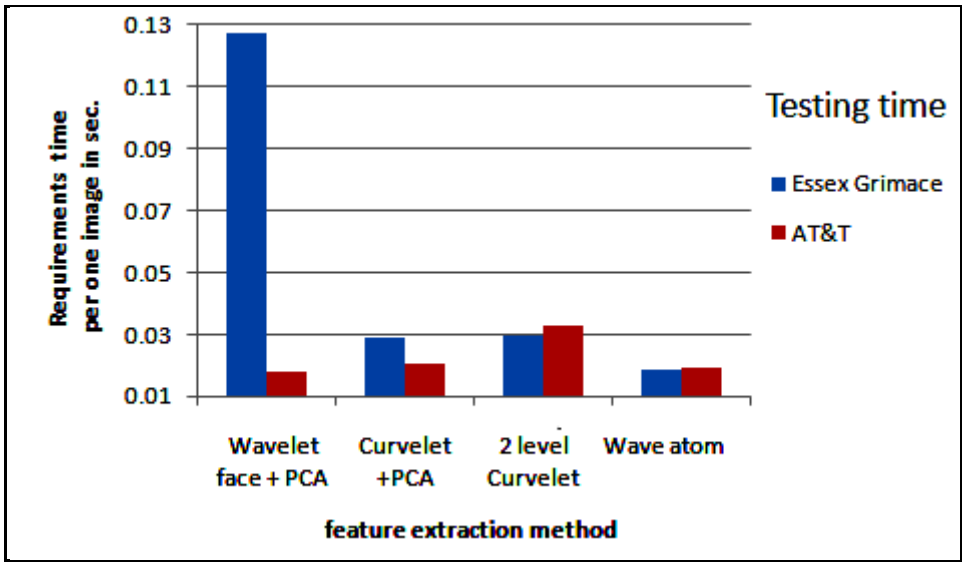


Figure 5.14: Time Requirements in Testing HMM

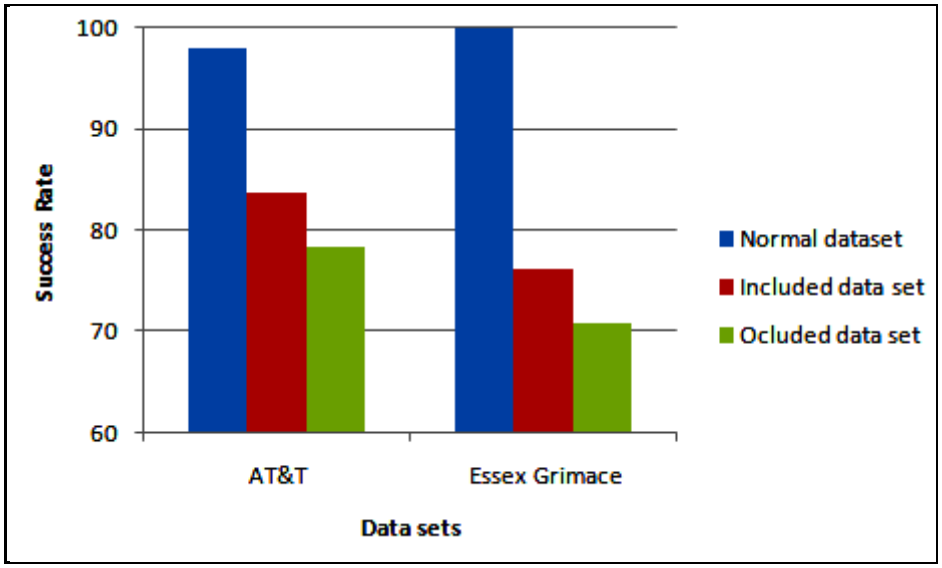
5.6 Enclosed Model Selection Criteria

The purpose of the next set of experiments was to show the effect of enclosed data sets on the accuracy recognition rate. The recognition accuracy of Curvelet and Waveatom features was presented in Figure 5.15. Figure 5.15 (a) listed the results obtained for Curvelet features based HMM. It can be seen the very substantial dropping in performance due to included and occluded data sets. In reality, with AT&T dataset the accuracy decreased of 13% and 19% for included and occluded data sets respectively. Also, for Essex Grimace, the accuracy decreased of 23% and 29% for included and occluded data sets respectively.

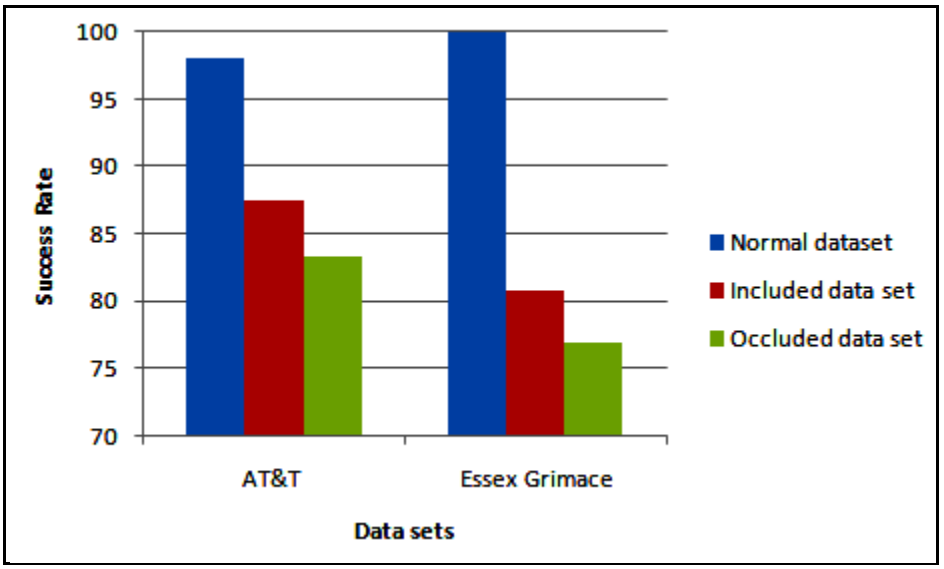
Figure 5.15 (b) viewed the results obtained for Waveatom features based standard SHMM. It can notice that exist of huge decreasing in accurate rate due to inclusion and occlusion data sets. In reality, with AT&T dataset the accuracy rate decreased from 98% to 88% and 83% for included and occluded data sets respectively. Also, for Essex Grimace, the accuracy rate decreased from 100% to 81% and 77% for included and occluded data sets respectively.

The following set of experiments was performed to show the benefit of using enclosed model selection criterion (EMC) HMM for face recognition. Where appropriate, the parameters were still the same as for HMM (such as block size). The experiments were carried out in both the Curvelet domain and Waveatom domain. The recognition accuracy is presented in Figure 5.16 for included data sets. On the side of occluded data sets, the recognition accuracy is presented in Figure 5.17. As can be seen from the tables,

the use of enclosed model selection criterion HMM the increases recognition accuracy in all cases tested.



(a)



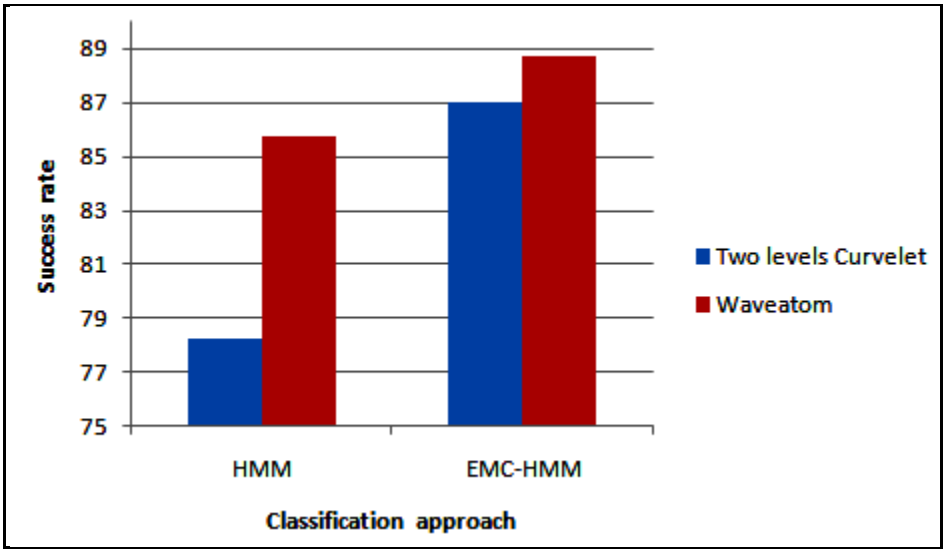
(b)

Figure 5.15: Comparison of HMM identification accuracy (%) using both normal and enclosed datasets (a): Curvelet features (b): Waveatom features

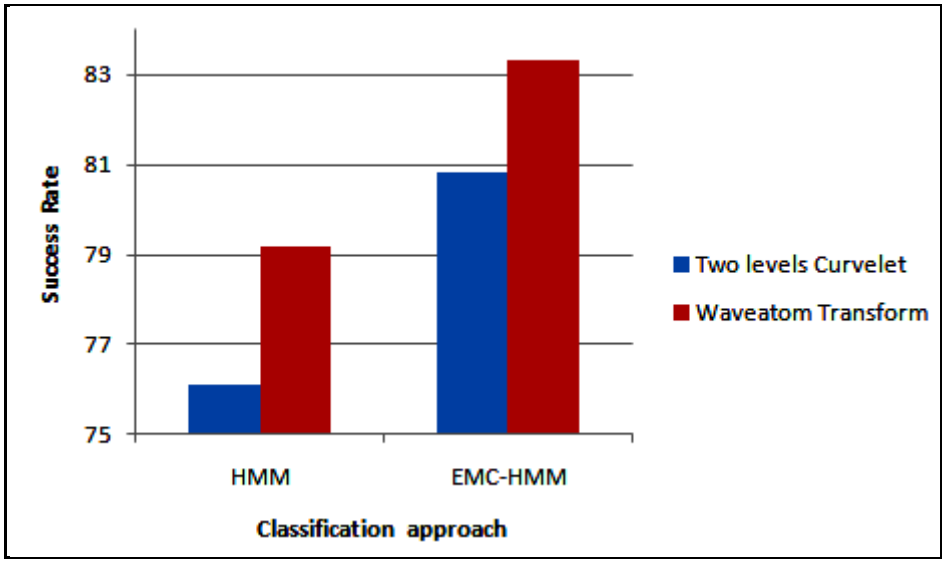
Figure 5.16 (a) reported the accuracy rate when tested using AT&T database. For Curvelet features, the accuracy rate increased from 78% for HMM to 86% for EMC-HMM. The incorrect match rate for EMC is near 13% lower than HMM model. On the

Waveatom side, the performance increased to from 87% for HMM versus 89% for EMC-HMM. There is an evident decreasing of about 12% in the rate of false classification.

Figure 5.16 (b) illustrated the accuracy rate for Essex Grimace database when used for testing. On the Curvelet side, the accuracy rate increased from 76% when HMM was used up to 79% when EMC-HMM used. On Waveatom side, the accuracy rate increased from 81% when using HMM up to 83% when using EMC-HMM model.



(a)

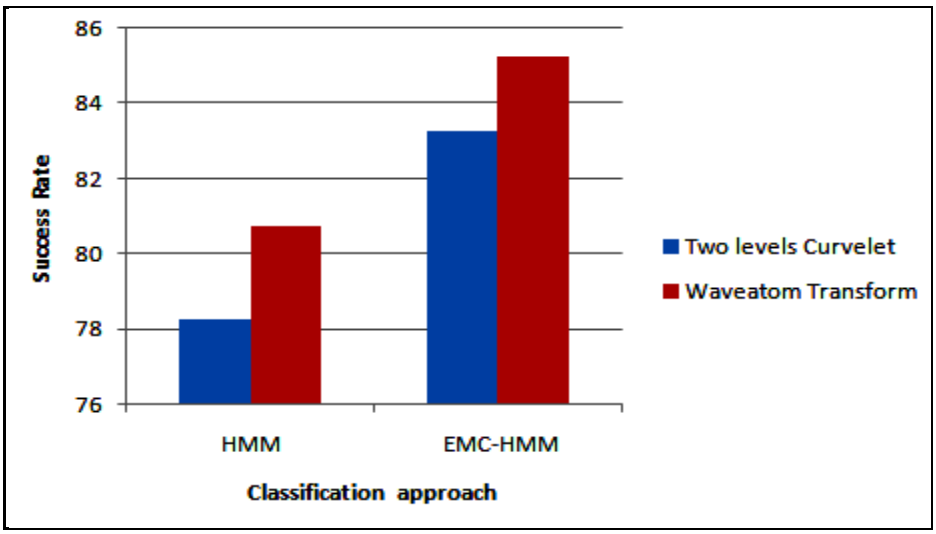


(b)

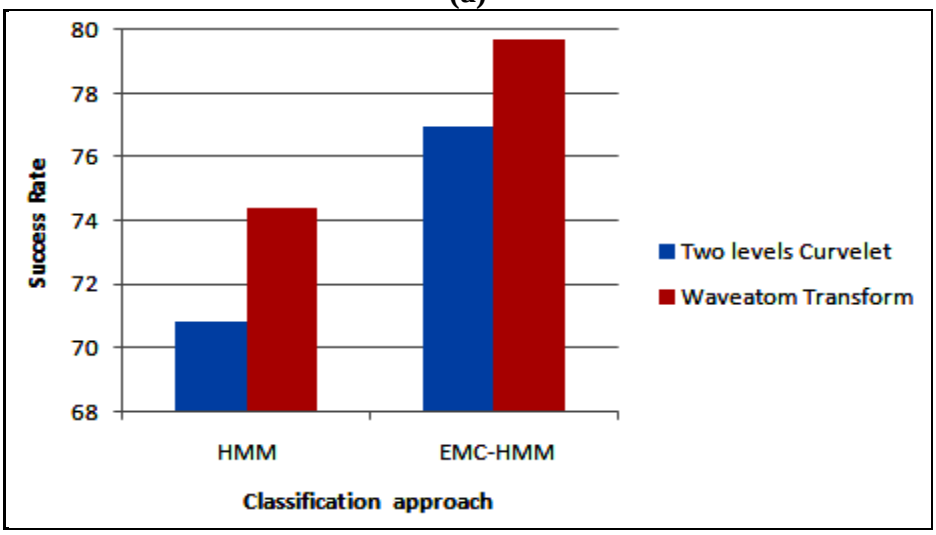
Figure 5.16: Comparison of inclusion identification accuracy (%) using HMM and EMC-HMM (a): On AT&T database (b): On Essex Grimace

Figure 5.17 (a) reported the accuracy rate when tested using AT&T database. For Curvelet features, the accuracy rate increased from 78% for HMM to 81% for EMC-HMM. The incorrect match rate for EMC is near 10% lower than HMM model. On the Waveatom side, the performance increased to from 83% for HMM versus 85% for EMC-HMM. There is an evident decreasing of about 10% in the rate of false classification.

Figure 5.17 (b) illustrated the accuracy rate for Essex Grimace database when used for testing. On the Curvelet side, the accuracy rate increased from 71% when HMM was used up to 74% when EMC-HMM used. On Waveatom side, the accuracy rate increased from 77% when using HMM up to 79% when using EMC-HMM model.



(a)



(b)

Figure 5.17: Comparison of occlusion identification accuracy (%) using HMM and EMC-HMM (a): On AT&T database (b): On Essex Grimace

CHAPTER 6
CONCLUSION

The problem intended here is to solve face recognition. During the past few years, face recognition has received significant attentions. The most cleared two reasons are the wide range of commercial and law enforcement applications, and the availability of feasible technologies after more than three decades of research.

Face recognition systems are affected by many factors like: face position and orientation, lighting conditions, and illumination variations. Face recognition can be applied for both a still images and video images. Many approaches of face recognition can be listed, starting with PCA, Fisherface, and continuing to reach Multiresolution transforms. These transforms found to be useful for analyzing the content of images and extracting feature from it. The most famous one is the Wavelet Transform. Many multiresolution transforms like Contourlets, Ridgelets follow. But actually the most modern one is named Waveatom Transform.

The aim behind this thesis was to find the best feature extraction method, that work well with the best classification technique, to recognize faces and to classify them in correct manner. Feature Extraction obtained by using two different novel techniques; the first using two levels Curvelet Transform, and the second one using Waveatom Transform. Also, two classification techniques used to obtain the performance rate. These are ANN and HMM.

Two well-known databases indicate the potential of these proposed methods; AT&T dataset and Essex Grimace dataset. Both of feature extraction techniques have been found to be robust against extreme expression variation as it works efficiently on Essex database. The subjects in this dataset make grimaces, which form edges in the facial images and both transforms (Curvelet and Waveatom) captures this crucial edge information. The proposed methods also seem to work well for ORL database, which show significant variety in illumination and facial details. These feature extraction techniques were coupled with two different classification techniques. The classification techniques were the gradient descent backpropagation ANN and HMM.

From the comparative study, it was evident that both feature extraction techniques (two levels Curvelet features and Waveatom features) completely outperformed standard

eigenface technique; it also superseded both Wavelet based PCA scheme and Curvelet based PCA scheme. The results indicated that Curvelet transform stood alone as an effective solution to face recognition problem in future. It promises that Waveatom Transform could be new platform in the face recognition field. On the other side, comparative study shows the strongest of HMM over ANN; as they were classifiers. Thus, Waveatom based HMM showed the highest performance rate against other methods when compared to the state of art. Also the enclosed model selection criterion gives acceptable accuracy improvement with included and occluded data sets.

As future work, I propose using another classification technique like SVM with suitable voter. In the side of HMM study the using of Confusion Model Selection Criterion (CMC) instead of using the Maximum Likelihood Criterion (ML). In the side of Waveatom and Curvelet, one could study using them in other field of image processing such as compression, encryption, denoising, and segmentation.

Bibliography

[1] Laurent Demanet and L. Ying, "Waveatom", available on: (www.wafeatom.org), last modification: May, 2008.

[2] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86.

[3] Ilker Atalay, face recognition using eigenfaces, MS. thesis, Istanbul technical university, January, 1996.

[4] Li Bai and Linlin Shen, Combining Wavelets with HMM for Face Recognition, *23rd* International Conference on Innovative Techniques and Applications of Artificial Intelligence (SGAI '03), Cambridge, UK, 13-15 December 2003, pp. 227-234.

[5] Vytautas PERLIBAKAS, Face Recognition Using Principal Component Analysis and Wavelet Packet Decomposition, informatica, vol. 15, no. 2, 2004, pp. 243 – 250.

[6] Majid Safari, Mehrtash T. Harandi, Babak N. Araabi, A SVM-based method for face recognition using a wavelet pca representation of faces, Image Processing, ICIP apos; 04. International Conference, vol.2, 24-27 October. 2004, pp. 853 – 856.

[7] Hazim Kemal Ekenel, Bulent Sankur, Multiresolution face recognition, Image and Vision Computing, vol. 23, 2005, pp. 469–477.

[8] Chao-Chun Liu, Dao-Qing Dai, Hong Yan, Local Discriminant Wavelet Packet Coordinates for Face Recognition, Journal of Machine Learning Research, 2007, pp. 1165-1195.

[9] Chengjun Liu, Harry Wechsler, Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition, IEEE transactions on image processing, vol. 11, no. 4, April, 2002, pp. 467 – 477.

[10] Chengjun Liu and Harry Wechsler, Independent Component Analysis of Gabor Features for Face Recognition , IEEE transactions on neural networks, vol. 14, no. 4, July 2003, pp. 919 – 929.

[11] Peng Yang, Shiguang Shan, and others, Face Recognition Using Ada-Boosted Gabor features, proce. Of the 6th IEEE inter. Conference on Automatic face and gesture recognition, Korea, May, 2004, pp. 356 – 361.

[12] Sanghoon Kim, Sun-Tae Chung, Souhwan Jung, Seoungseon Jeon, Jaemin Kim, and Seongwon Cho, Robust Face Recognition using AAM and Gabor Features, PWASET, vol. 21, January, 2007, pp. 493-497.

[13] Praseeda Lekshmi.V, M.Sasikumar, A Neural Network Based Facial Expression Analysis using Gabor Wavelets, PWASET, vol. 32, August, 2008, pp. 539-597.

[14] Sanjay A. Pardeshi, S.N.Talbar, Automatic Face Recognition Using Local Invariant Features – A Scale Space Approach, Journal of Wavelet theory and applications, vol. 2, no. 1, 2008, pp. 31–39.

[15] T. Mandal, A. Majumdar, Q.M. J. Wu, Face Recognition by Curvelet Based Feature Extraction, Proc of ICIAR, Montreal, Canada, vol. 4633, 22-24 August 2007, pp 806-817.

[16] Tanaya Mandal, Q. M. Jonathan Wu, Face Recognition using Curvelet Based PCA, Pattern Recognition, ICPR, 19th international conference, Tampa, Florida, USA, 8-11 December 2008, pp. 1- 4.

[17] A. Majumdar and A. Bhattacharya, Face Recognition by Multiresolution Curvelet Transform on Bit Quantized Facial Images, International Conference on Computational Intelligence and Multimedia Applications, vol. 2, 13-15 December, 2008, pp. 209-213.

[18] A. Majumdar and R. K. Ward, Single image per person face recognition with images synthesized by non-linear approximation, ICIP, 15th IEEE intrnational conference on digital object identifier, 12-15 October, 2008, pp. 2740-2743.

[19] Angshul Majumdar and Rabab K. Ward, Multiresolution Methods in Face Recognition, Recent Advances in Face Recognition, Book edited by: Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett, ISBN 978-953-7619-34-3, I-Tech, Vienna, Austria, 2008, pp. 79-96.
[20] Mohammed Rziza, Mohamed El Aroussi and other, Local Curvelet Based Classification Using Linear Discriminant Analysis for Face Recognition, International Journal of Computer Science, Volum 4, Number 1, 2009, pp. 72-77.
[21] Jianhong Xie, Face Recognition Based on Curvelet Transform and LS-SVM, Proceedings of the International Symposium on Information Processing (ISIP'09), Huangshan, P. R. China, 21-23 August 2009, pp. 140-143.
[22] E. J. Candes and D. L. Donoho. Curvelets: A surprisingly effective nonadaptive representation for objects with edges, [Online]. Available on http://www.Curvelet.org/papers/Curve99.pdf, 2000, last visit: Nov. 13, 2009.
[23] Emmanuel Candes, Laurent Demanet, and others, Fast Discrete Curvelet Transforms, Technical Report, Cal Tech, March, 2006.
[24] L. Demanety, L. Ying, Wave atoms and sparsely of oscillatory patterns, appear in Appl. Comput. Harm. Anal. , February, 2007.
[25] Laurent Demanety, Lexing Ying, Curvelets and Wave Atoms for Mirror-Extended Images,[Online] available on: www.waveatom.org, july, 2007, last visit: Nov. 13, 2009.
[26] Laurent Demane, Curvelets, Wave Atoms, and Wave Equations, D.hP. thesis, California Institute of Technology, Pasadena, California, May, 2006.
[27] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, IEEE Proc., vol. 77, no. 2, 1989, pp. 257–286.
[28] HTK book for HTK version 3.4, Cambridge University Engineering Department, 2009.
[29] Mohammed A. Alhanjouri, Feature Analysis and Classification of Human Chromosome Images, PhD. Thesis, Mansoura University, 2006.
[30] CurveLab, software package, available on: (http:// www.curvetet.org), last visit: Dec. 12, 2009.
[31] WaveLab, software package, available on :(http:// www.waveatom.org), last visit: Dec. 12, 2009.
[32] HMMall, software package, available on : http://en.pudn.com/downloads162/sourcecode/others/detail739897_en.html
[33] An archive of AT&T Laboratories Cambridge, (http://www.cl.cam.ac.uk/Research/DTG/attarchive), last visit: Dec. 12, 2009.
[34] Libor Spacek [Jun-2008], "Description of the Collection of Facial Images", available at (http://cswww.essex.ac.uk/mv/allfaces/grimace.zip), last visit: Dec. 12, 2009.

الجامعة الاسلامية – غزة
كلية الدراسات العليا
قسم الهندسة
تخصص هندسة الحاسوب

تمييز الوجه بإستخدام تحويل مصغر-المنحنى و تحويل موجة-الذرة

إعداد
هناء حمدان حجازى

إشراف
أستاذ مساعد: محمد أحمد الحنجوري

رسالة مقدمة كجزء من متطلبات نيل درجة الماجستيرفي كلية الدراسات العليا قسم الهندسة
تخصص هندسة الحاسوب

غزة، فلسطين

(1431, 2010)

الملخص

إن مجال معالجة الصور الرقمية يتطور بصورة مستمرة. حيث ظهرت زيادة كبيرة في مستوى الاهتمام بمعالجة وتشكيل الصور، سواءاً كانت هذه الصور ذات التدريج الرمادي أو صور كاملة الألوان، كما طال هذا الإهتمام المتزايد بمعالجة الصور، ضغط بيانات الصور بشتى أنواعها وكذلك مجال التعرف على الصور. و لذلك فإن هذا العمل يتناول مجال تمييز الصورة مع تطبيق التعرف على صور الوجه.

بعض الناس يعتقدون ان التعرف على الوجه هو مهمة سهلة لنظام الكمبيوتر كما للبشر ، ولكن فعلياً معظم نظم التعرف على صورة الوجه لا يمكن أن تحقق أداء كامل موثوق به لأن هناك عوامل كثيرة تؤثر على عملية التمييز مثل : الإختلافات الكبيرة في أشكال الوجوه، وحجم الرأس وإتجاهه، وكذلك التغييرات في الظروف البيئية، كل هذه العوامل تجعل التعرف على الوجه واحدة من المشاكل الأساسية في مجال تحليل الأنماط. وهناك عوامل أخرى قد تؤثر على الأداء والدقة كموقع الوجه وعدد تقنيات التعرف على الوجه الفعلية المستخدمة في كل نظام. إن التعرف على صور الوجه سواءاً الثابتة أوالفيديو يعتبر من مجالات البحث النشطة مع الكثير من التطبيقات المدعومة تجارياً وقانونياً.

يدرس هذا البحث تقنيتين حديثتين من تقنيات استخراج ميزات الوجه إستناداً الى نوعين من أنواع التحليل متعدد النتائج، النوع الأول يدعى تحويل مصغر-المنحنى (Curvelet) والثاني هو تحويل موجة-الذرة (Waveatom). ويتم تدريب واختبار الميزات الناتجة عن طريق اثنين من انواع المصنفات، أحدها هو الشبكة العصبية الاصطناعية (ANN)، والآخر هو نموذج ماركوف المخفي (HMM). بناءاً على ما تم نشره من أبحاث، يعتبر هذا البحث من الخطوات الأولى لاستخدام تحويل موجة الذرة ومقارنتها بتحويل مصغر-المنحنى في مجال التعرف على الأنماط، وخصوصا في مجال التعرف على صور الوجه.

وقد أجريت التجارب في هذا البحث على مجموعتين من البيانات المعروفة عالمياً في هذا المجال، بيانات من شركة (AT&T) وتتكون هذه المجموعة من 400 صورة وجه موزعة على 40 شخصاً بالتساوي، أما المجموعة الثانية من البيانات يطلق عليها بيانات الوجوه المتجهمة (Essex Grimace), وتستخدم في الكثير من أبحاث نظم التعرف على الوجوه، وتتكون من 360 صورة تمثل 18 شخصا. وقد اظهرت النتائج قوة وخصائص كل من تحويل مصغر -المنحنى وتحويل موجة-الذرة، وقد لوحظ أن السمات التي تم الحصول عليها من تحويل موجة-الذرة أعطت أعلى نسب تعرف 99 ٪ و 100 ٪ مع مصنف نموذج ماركوف المخفي، ونسب تعرف 98 ٪ و 100 ٪ مع مصنف الشبكات العصبية الإصطناعية، وذلك لكل من مجموعة بيانات AT&T و مجموعة بيانات الوجوه المتجهمة، على التوالي، ومن ناحية أخرى، أظهر إستخدام مستويين من تحويل المويجة (Wavelet) تحقيق معدل دقة 98 ٪ و 100 ٪ مع مصنف نموذج ماركوف المخفي، و 97 ٪ و 100 ٪ مع مصنف الشبكات العصبية الإصطناعية، وذلك لكل من مجموعة بيانات AT&T و مجموعة بيانات الوجوه المتجهمة، على التوالي.

كما قورنت نتائج دراسة إستخدام تحويل موجة-الذرة مع كلٍ من تقنيات تحويل مصغر-المنحنى وتحويل المويجة ، وتحليل المكونات الرئيسية التقليدية (PCA). وقد وجد أن التقنيات المقترحة في البحث أعطت نتائج أفضل من غيرها. برهنت على قدرتها لإستخراج السمات رغم وجود التشويش وإختلاف تفاصيل الوجه. أيضا أشارت النتائج إلى فعالية مصنف ماركوف المخفي مقابل مصنف الشبكات العصبية الإصطناعية.