Al-Aqsa University Journal (Natural Sciences Series),Vol.15, No.1, Pages 41-66 , Jan. 2011
ISSN 2070-3155

Curvelet and Waveatom Transforms Based Feature Extraction for Face Detection

د. محمد الحنجوري *

الملخص

هذا البحث يدرس تقنيتين حديثتين لاستخراج ميزات الوجه استنادا الى نوعين من أدوات التحليل المتعدد-الحلول، الأول يسمى تحويل المنحنى- المصغر والثاني هو تحويل موجة- الذرة. وقد تم تمرين واختبار الميزات الناتجة باستعمال اثنين من أشهر المصنفات ، الأول هو الشبكة العصبية الاصطناعية (ANN)، أما الثاني فهو نموذج ماركوف الخفي (HMM).

وقد أجريت التجارب على اثنتين من قواعد البيانات المعروفة، الأولى مجموعة البيانات AT&T وتتألف من 400 صورة لأربعين شخص مختلف ، والثانية مجموعة البيانات Essex Grimace للوجوه المتجهمة وتتألف من 360 صورة لحوالي 18 شخص مختلف. وقد أظهرت النتائج قوة وميزات كلٍ من تحويل المنحنى-المصغر وتحويل موجة-الذرة. حيث أعطت المميزات المستخرجة باستخدام وتحويل موجة-الذرة أعلى معدل دقة تمييز للوجوه، حيث بلغ حوالى 94 ٪ و96 ٪ لكلتى مجموعتي البيانات المستخدمة على التوالي وباستعمال نموذج ماركوف الخفي كمصنف، في حين بلغت دقة التمييز حوالى ٪ 90 و ٪ 93 باستخدام الشبكات العصبية الاصطناعية لنفس مجموعتي البيانات على التوالي. في المقابل، حققت المميزات المستخرجة باستخدام مستويين من تحويل المنحنى-المصغر معدل دقة تمييز 92 ٪ و 95 ٪ لصالح مصنف نموذج ماركوف الخفي، و 88 ٪ و 92 ٪ لمصنف الشبكات العصبية الاصطناعية، لمجموعتي البيانات AT&T وEssex Grimace على التوالي.

كما تم عمل دراسة مقارنة لتحويل موجة-الذرة و تحويل المنحنى-المصغر مع كلٍ من تحويل المويجة و طريقة تحليل المكونات الأساسية التقليدية (PCA)، اللذين تم استخدامهما في الكثير من أبحاث التعرف على الوجه. وقد لوحظ ارتفاع معدل دقة التمييز للتقنيات المقترحة في هذه الدراسة. مما يؤكد قوة أساليب استخراج المميزات المقترحة ضد التباين الشديد في التعبير والإضاءة وتفاصيل الوجه المختلفة. أيضا تشير النتائج إلى قدرة نموذج ماركوف الخفي مقابل الشبكات العصبية الاصطناعية كمصنفات.

• قسم الحاسوب – الجامعة الإسلامية - غزة – فلسطين.

ABSTRACT

This work identifies two novel 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 trained and tested 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. In one hand, Waveatom features obtained the highest accuracy rate of 94% and 96% with HMM classifier, and 90% and 93% with ANN classifier, for AT&T and Essex Grimace datasets, respectively. In the other hand, two levels Curvelet features achieved accuracy rate of 92% and 95% with HMM classifier, and 88% and 92% 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 proposal techniques supersede all of them. And proves the robustness of feature extraction methods used against extreme variation on expression and illumination, and different facial details. Also, indicates the potential of HMM over ANN, as they are classifiers.

1. Introduction

Face recognition system can be used in building or specific area security, 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. It is also 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 (Wolde, 2006).

FR is a difficult problem due to the general similar shape of faces combined with the numerous variations between images of the same face. Recognition of faces from an uncontrolled environment is a very complex task: lighting condition may vary tremendously; facial expressions also vary from time to time; face may appear at different orientations and a face can be partially occluded. Further, depending on the application, handling facial features over time (aging) may also be required. Although existing methods perform well under constrained conditions, the problems with the illumination changes, out of plane rotations and occlusions still remain unsolved. Since the techniques used in the best FRS may depend on the application of the system, one can identify at least two broad categories of FRS (Lawrence, et al, 1997):

1. Finding a person within large database of faces (e.g. in a police database). (Often only one image is available per person. It is usually not necessary for recognition to be done in real time.)

2. Identifying particular people in real time (e.g. location tracking system). (Multiple images per person are often available for training and real time recognition is required.)

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 (Ziheng Wang, Xudong Xie, 2010).

In recent years, Multiresolution analysis tools, especially wavelets, have been found useful for analyzing the information content of images; this is lead to be used in areas like image processing, pattern recognition and computer vision. Following wavelets, other multiresolution tools like contourlets, ridgelets etc. were developed. 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 only, exactly with image denoising, and the results obtained are the best one when compared to the state of art (Demanet and ying, 2008).

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 (Starovoitov, et al, 2002):

- 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 (Anbarjafari, 2008).

This work 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.

2. Literature Review

The most famous early example of face recognition system is due to (Kohonen T., 1989), who demonstrated that a simple neural net could perform face recognition for aligned and normalized face images. The type of network that he employed computed a face description by approximating the eigenvectors of the face image's autocorrelation matrix; these eigenvectors are known as "eigenfaces". But the system was not a practical success, however, because of the need for precise alignment and normalization. also these eigenfaces are sensitive to variation in position and scale.

The Gabor–Fisher Classifier (GFC) for face recognition is introduced by (Chengjun L, Harry W, 2002). The GFC method is robust to changes in illumination and facial expression. Actually, they applied the Enhanced Fisher linear discriminant Model (EFM) to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. The importance of this technique comes from three reasons. The first one is that the derivation of an augmented Gabor feature vector, whose dimensionality is further, reduced using the EFM. The second one is that the development of a GFC for multi-class problems. The last reason is the extensive performance evaluation studies. The GFC method achieves 100% recognition accuracy using only 62 features.

Combining wavelet multiresolution analysis and Hidden Markov Model (HMM) for face recognition is introduced by (Li B, and Linlin S, 2003). In this approach a face image is divided into a number of overlapping subimages. Then, wavelet decomposition is performed on each of the resulted subimages. The performance was better than the original Digital Cosine Transform (DCT) based HMM.

Peng presented a face recognition method using AdaBoosted Gabor features (Peng Y., et al, 2004). The main contribution of this method lies in two points: first, AdaBoost is successfully applied to face recognition by introducing the intra-face and extra-face difference space in the Gabor feature space. Second, an appropriate re-sampling scheme is adopted to deal with the imbalance between the amount of the positive samples and that of the negative samples.

(Vytautas P, 2004) used Principal Component Analysis (PCA) and Wavelet Packet Decomposition to deal with the problem of face recognition. The proposed modification of the PCA-based face recognition method could be used in practical applications when the number of training images is too large (thousands or more). For traditional PCA-based, the training becomes too slow with these large databases. With this modification, the training time becomes independent from the number of training images.

(Ekenel H., Sankur B., 2005) searched for the frequency subbands that being insensitive to expression differences and illumination variations on faces. The search for alternative wavelet channels for faces in the presence of illumination variations proves much more effective. this scheme had succeeded with expression differences and illumination variations on faces. But in the other hand, similar studies can be conducted for other facial factors, of aging, accessories, or pose.

(Sanghoon K, et al, 2007) presented a robust face recognition algorithm using Active Appearance Model (AAM) and Gabor features. In localizing facial feature points, AAM is known to be relatively stable, but is affected by illuminations, poses, and expressions. On the other hand, Gabor features are known to be more robust to small variations in scaling, rotation, distortion, illumination, poses, and expressions so that they are popularly employed as features for face recognition. Their results were poor, due to using just a domestic face database. So they should test their algorithm for various face databases and real-time face recognition processing.

(Chao-Chun L, Dao-Qing D, and Hong Y, 2007) proposed face recognition scheme, which used a Local Discriminant Coordinates (LDC) method based on wavelet packet. LDC algorithm used to select uniformly the most discriminant independent coordinates in all spatial frequency subbands, in order to overcome the limitation of the methods based on whole subband. They used the dilation invariant entropy and the Maximum a Posteriori (MAP) logistic model to measure the discriminability. Their results show that the LDC based feature extraction is more effective than WaveletFace and PCA for feature extraction. The tested images must be include more complex expressions, and extend the work to identify the face and its expressions from 3D images.

(Majumdar A, Bhattacharya A, 2008) proposed a multiresolution Curvelet based method for frontal face recognition. Recognition took place in three steps. In the first step the original 8 bit image is quantized to 4 bit and 2 bit representations. In the second step, each of the 3 different resolved versions of the image are subjected to Curvelet transform at five different resolutions. The approximate Curvelet coefficients at each resolution represent a different feature set. In the last step these fifteen (three bit resolved version x five resolutions) sets of coefficients are then used to train separate SVMs. During testing, the results of the fifteen SVMs are fused to determine the final result. This technique produced poor results. They should try to improve the recognition accuracy for these images by introducing some pre-processing steps like head tilt corrections and cropping for removing background variation.

(Rziza M, et al, 2009) presented a face recognition approach based on Curvelet decomposition. Therefore, each face image is described by a subset of band filtered images containing Curvelet coefficients which characterize the face textures. Next they divided the Curvelet sub-bands into small sub-blocks, from which they extracted compact and meaningful feature vectors using simple statistical measures. These feature sets are used in order to create templates with different information content for face recognition (Curvelet database). After that, Fisherface algorithm is carried out on the Curvelet database in which faces with similar statistics will be grouped together by LDA rules. For the purpose of classification they used the city-block distance, they designed experiments specifically to investigate the improvement in robustness against illumination and facial expression changes. The efficiency of this approach is analyzed by comparing the results with those obtained using the well-known subspace reduction based methods PCA, LDA and ICA.

(Jianhong X, 2009) developed a face recognition system based on Curvelet transform and Least Square Support Vector Machine (LS-SVM), they used Curvelet transform to extract features from facial images first, and then applied LS-SVM to classify facial images based on these features. Their results show that the correct recognition rate is up to 96%, and the computational speed is faster. They used ORL dataset only.

The results in all previous techniques showed Curvelet based schemes were better than wavelet based recognition schemes.

3. Digital Curvelet Transform

A new member of the family of multiscale geometric transforms is the Curvelet Transform, The Curvelet Transform is a higher dimensional generalization of the Wavelet Transform designed to represent images at different scales and different angles (Candes and Demanty, 2006).

Curvelets was proposed by E. Candes and D. Donoho (Candes and Donoho, 2000). The idea of Curvelets is to represent a curve as a superposition of functions of various lengths and widths obeying the scaling parabolic law $width \cong length^2$. With this ratio, the anisotropy increases with decreasing scale like a power law. With a dyadic decomposition of the frequency domain, the length of the localizing windows is doubled at every other dyadic subband. Curvelets are designed to represent edges and other singularities along curves much more efficiently than the traditional Wavelet transforms. Figure 1 shows edge representation by both Wavelet and Curvelet Transforms. It can be noticed, it would take many Wavelet coefficients to accurately represent such a curve. In the other side, Curvelets can represent a smooth contour with much fewer coefficients for the same precision.

Practically, Curvelet Transform is multi-scale geometrical transform in which units are indexed by their position, scale and orientation. The directionality concept is integrated and allows an optimal and compact representation of images that contain objects with clear edges. One essential advantage of the Curvelets is their ability to represent smooth edges. Such a gray scale image has a lot of edges; Curvelet Transform will capture this edge information.

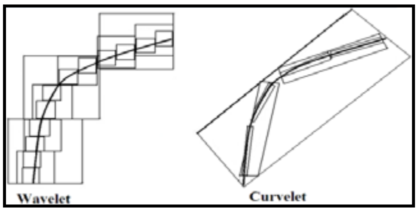


Figure 1: Edge Representations.

To define Curvelet Transform, suppose we work in the space $\mathbf{R}^2$ with special variable x, frequency variable ξ 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:

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

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

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:

$$c^D_{j,l,k} := \sum_{0 \leq x_1, x_2 < n} f[x_1, x_2] \, \overline{\varphi^D_{j,l,k}[x_1, x_2]} \tag{3}$$

There are two implementation of Curvelet: 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. In the other hand, a grid aligned with the input Cartesian grid which leads to the wrapping-based. Both digital transformations return a table of Digital Curvelet coefficients indexed by a scale parameter, an orientation parameter, and a spatial location parameter (Candes and Demanenty, 2006).

In order to compute the Curvelet coefficients, the discrete transform takes as input data defined on a Cartesian grid. The continuous-space definition of the Curvelet transform uses coronae and rotations that are not especially adapted to Cartesian arrays. It is then convenient to replace these concepts by their Cartesian counterparts. That is concentric squares (instead of concentric circles) and shears (instead of rotations), Figure 2 shows these grids.

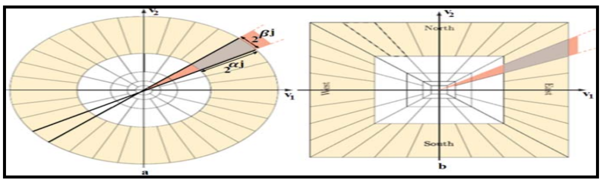


Figure 2: Curvelet Frequency Tiling (a): Continuous Domain. (b): Discrete Domain.

3.1. Digital Curvelet Transform via Wrapping

USFFT Transform finds the Curvelet coefficients by irregularly sampling the Fourier coefficients of an image. Whereas the Wrapping transform, uses a series of translations and a wraparound technique. Both algorithms having the same output, but the Wrapping Algorithm gives a more intuitive algorithm and faster computation time. Because of this, Curvelet via wrapping will be used for this work.

If we have $g[t_1,t_2]$, $t_1 \geq 0$, $t_2 < n$ 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 (Demanet and ying, 2007):

1. 2D Fast Fourier Transform (FFT) is applied to obtain Fourier samples $\hat{g}[n_1,n_2]$ (Figure 2 (b)).
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 3 (a)).
3. This product is wrapped around the origin to obtain $\breve{g}_{j,l}[n_1,n_2] = W(\tilde{U}_{j,l}\,\hat{g})\,[n_1,n_2]$; 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 3 (b)).
4. Inverse 2D FFT is applied to each $\breve{g}_{j,l}$, hence creating the discrete Curvelet coefficients.

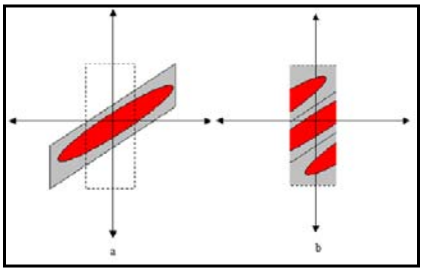


Figure 3: Support of Wedge (a): before Wrapping (b): after Wrapping.

4. Waveatom Transform

Waveatom transform is the newest member in the family of oriented multi-scale transforms for image processing and numerical analysis (Demanety and ying, 2007).

Waveatoms are a recent addition of mathematical transforms of computational harmonic analysis. They 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.

4.1. Waveatom Transform Properties

The 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.

To be a Multiresolution image transforms (Majumdar and Ward, 2008), 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).

All the five properties of the wish-list are covered by the Waveatom Transform. Property 1, i.e. the multiresolution property of the Waveatom Transform, as in the first step of the transform, the entire frequency plane is divided into concentric rings in order to facilitate bandpassing at multiple resolutions. Property 2 shows the actual localization related to the Waveatom in both spatial and frequency domain. Waveatom do not have a compact support in the spatial domain, but the effective support is elliptical. As for property 3 of the wish-list (related to critical sampling) Waveatom do not form a basis, but form tight frames with redundancy two. The directionality of the Waveatom transform is shown in property 4, the entire frequency plane is divided into angular wedges. This allows for analyzing the image at various orientations. As for the last property of the wish-list, from the perspective of image processing, a transform is said to be anisotropic if its basis elements are not circular. Waveatoms are elliptical in shape (the major and the minor axes related by a parabolic scaling law), thus they are anisotropic.

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 (wavelength ~ constant) and wavelets (wavelength ~ diameter) means that the period of oscillations of each wave packet (wavelength) is related to the size of essential support by parabolic scaling i.e. wavelength ~ (diameter)^2, which is known as the scaling law (Demanety and ying, 2007).

Backing to Figure 2, the two distinct parameters α, β represent decomposition and directional ability. They are sufficient for indexing all known forms of wave packet architectures; namely Wavelets, Gabor, ridgelets, Curvelets and Waveatoms. Waveatoms are defined for $\alpha = \beta = 1/2$. α indexes the multiscale nature of the transform, from $\alpha = 0$ (uniform) to $\alpha = 1$ (dyadic). β measures the wave packet's directional selectivity (0 and 1 indicate best and poor selectivity respectively). Waveatoms represent a class of wavelet packets where directionality is sacrificed at the expense of preserving sparsity of oscillatory patterns under smooth diffeomorphisms as shown in Figure 4.

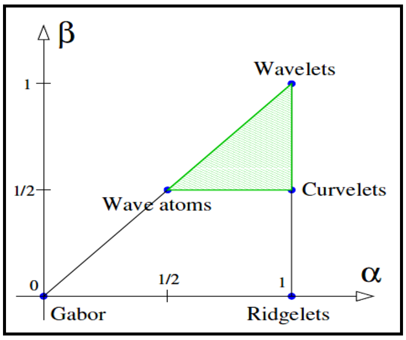


Figure 4: Identification of Wave Packets as α, β.

Corresponding to Waveatoms, the good compromise the point is $\alpha = \beta = 1/2$. This means, $O(N)$ Waveatom coefficients suffice to represent f to some given accuracy. In contrast, we need $O(N^{3/2})$ Curvelets coefficients; or $O(N^2)$ Wavelet coefficients to represent f up to the same accuracy. Actually, Waveatoms are a fixed transform and that the knowledge of N is not needed to define the transform.

5. Back-propagation Neural Network

The architecture of artificial neural network used in this work is a multi layer perceptron with steepest descent back-propagation training algorithm with adaptive learning rate. Figure 5 shows the topology of the used Back-propagation neural network.

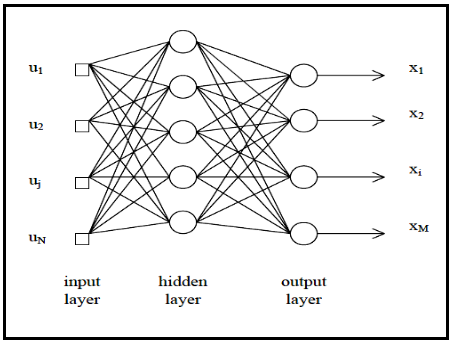


Figure 5: Back-propagation Neural Network.

6. Hidden Markov Model.

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.

HMMs are typically used to address three unique problems (Rabiner, 1989):

1. Evaluation: Given a model λ and a sequence of observations O, how one could efficiently compute $P\ (O|\ \lambda)$.

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., $q^* = \arg\max_q [P(q|\lambda, O)]$.
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. In other word it can used 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 (Young and et al, 2009).

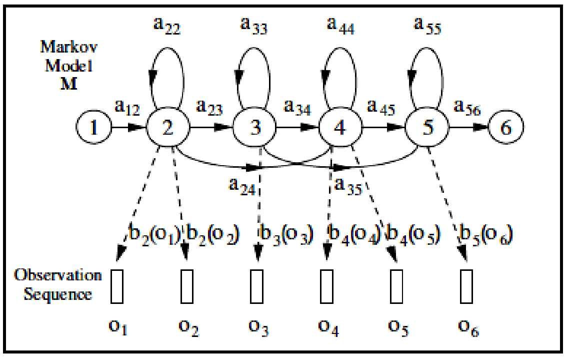


Figure 6: A Simple HMM.

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,

$$p(x|M) = \sum_{n=1}^{m} \left(a_n \frac{1}{(2\pi)^{D/2} |\sum n|^{1/2}} \right) e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma_n^{-1} (x-\mu_n)\right)} \qquad (4)$$

Where $p(x|M)$ is the likelihood of **x** given the mixture model, *M*.

The Maximum likelihood criteria used by EM algorithm to train the Gaussian mixture model, to estimate probability density function, $p(x|M)$. It is estimated by the mean log likelihood over the sequence, $X=\{X_1 ... X_N\}$,

$$S(X) = \log P(X|M) = \frac{1}{N} \sum_{n=1}^{m} \log P(X_n|M) \qquad (5)$$

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.

Suggest that the states are already trained. In Figure 6 the observation sequence had 6 frames. By Viterbi algorithm, the best time-alignment of the frames is obtained. For these 6 states (HMM), one class gives the highest (pdf) score throughout HMM. Thus having one HMM per class (person), for each utterance, the viterbi find out the best alignment of the utterance frames to all HMMs. Each HMM will have a score by its pdfs and transition probabilities. The Biayes decision was used to choose the model with highest score.

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. Figure 7 shows a sample of viterbi trellis, for 6-states, 6 frames (HMM in Figure 6) (Bai and Shen, 2003.

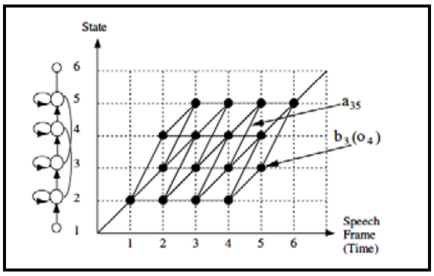


Figure 7: Viterbi Trellis for HMM in Figure 6.

7. Experimental Results

This work uses two types of dataset from different sources: ORL (AT&T) database and Essex Grimace database, both sets are used to implement different Algorithms to recognize the human face.

ORL (AT&T) database (AT&T Lab, 2009) 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 8

Figure 8: Sample Images from ORL database

Essex Grimace database (Spacek, 2008) 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 9.

Figure 9: Sample Images from Essex Grimace Database.

In the first part of the experimental study, 6 images per each subject for ORL and 9 images per each subject for Essex Grimace database are randomly selected as training set; the rest construct the test set. Color images of Essex Grimace database are converted to gray scale images as preprocessing stage.

7.1. Curvelet based Feature Extraction

Curvelet transform has been developed especially to represent objects with *'curve-punctuated smoothness'* i.e. objects which display smoothness except for discontinuity along a general curve; images with edges are good examples of this kind of objects. In a two dimensional image, two adjacent regions can often have differing pixel values. Such a gray scale image will have a lot of "edges" i.e. discontinuity along a general curve and consequently Curvelet transform will capture this edge information. To form an efficient feature set it is crucial to collect these interesting edge information which in turn increases the discriminatory power of a recognition system (Candes and Donoho, 2000).

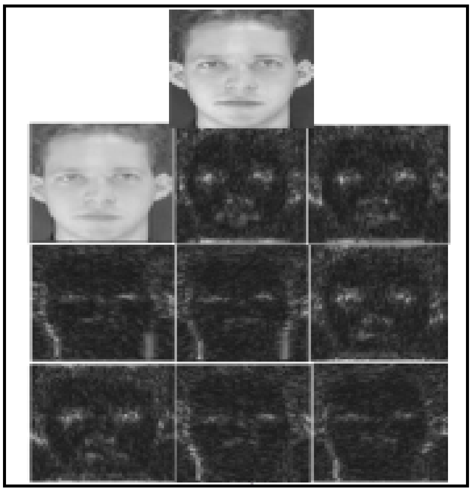


Figure 10: Curvelet Coefficients (Approximated and Details).

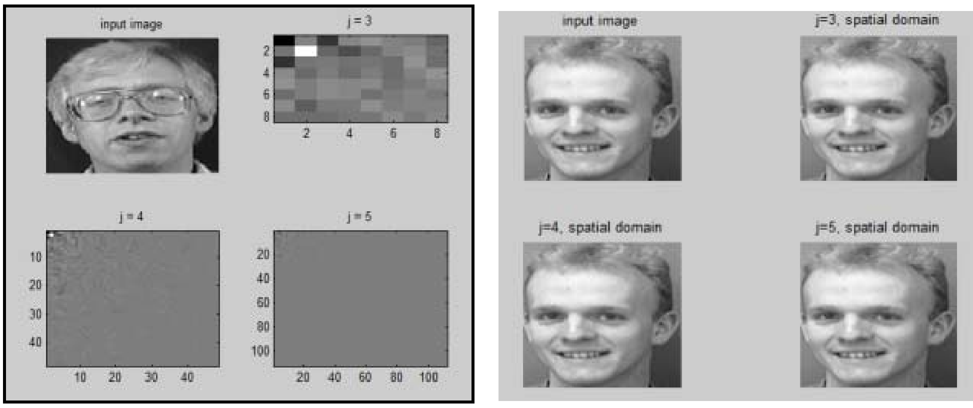
In Feature Extraction stage, the images are decomposed into its approximate and detailed components using two levels of Curvelet transform. These sub-images thus obtained are called *curvefaces*. These *curvefaces* greatly reduces the dimensionality of the original image. Thereafter only the approximate components are selected to perform further computations, as they account for maximum variance. Thus, a representative and efficient feature set is produced. Figure 10 shows the Curvelet coefficients of a face from ORL dataset decomposed at scale = 2 and angle = 8, we can note that the first image is the original image, the first image in the second row is the approximate coefficients and the others are detailed coefficients at eight angles.

In this work, the original images are decomposed using Curvelet transform at scale = 3 and angle = 8. Thus 25 components are produced, including one approximate and 24 detailed sub-band. The resolution of the approximate subband is reduced to 31x37 and 61x67 for images of ORL and Essex Grimace database respectively. To further reduce the dimensionality, Curvelet transform, at scale = 3 and angle = 8, is applied once again on these approximate components only. The resolution of the approximated subband became 11x13 and 21x23 for images of ORL and Essex Grimace database respectively as shown in Figure 11. A total of 143 features of Curvelet sub-images are produced (Hejazi and Alhanjouri, 2010).

Figure 11: Sample Images of Curvelet Transform.

7.2. 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 on the original face image; to efficiently capture coherence patterns along and across the oscillations. Face images are digitized using 256 gray levels therefore a transformation in color space is not required. No further pre-processing is performed to preserve the fine detail in an image. Figure 12 (a) shows the Waveatom coefficients of a face from ORL dataset decomposed at different scales. Figure 12 (b) shows the spatial domain reconstruction of an image when Waveatom coefficient used at different scales. It is notable, the coefficient at scale = 3 gives good reconstruction image. The coefficient size became 8×8 for all face-image sets.



(a) (b)

Figure 12: (a) Coefficients of Waveatom Transform at Different Scales.
(b) Sample Image and its Spatial Domain at Different Scales.

7.3. 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 number of hidden layer and the number of nodes in each one are important parameter to improve the recognition rate (success rate). Empirically, 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 a resultant performance for each number of hidden layers, the ANN topology has one hidden layer produces the best recognition rate with 2-level Curvelet Transform is achieved of 88% at 50 nodes in this layer. While when a second hidden layer is used, we obtained 84% recognition rate at also 50 nodes in the second hidden layer. In the other side, the best recognition rate with Waveatom Transform is achieved of 90% at 20 nodes in the first layer while the success rate was 87% at 15 nodes in the second layer.

Clearly, the ANN topology should be contains 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. Take into consideration, 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.

7.4. HMM for Face Recognition.

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. However, Face Recognition is not a direct application of HMMall since the data need preprocessing, feature extraction and HMM topology design.

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. Due to this, these numbers need some empirical tuning.

In the training phase of face recognition system, each individual in the dataset is represented by a HMM face model. First, the HMM is initialized. The features are modeled by a multivariate Gaussian distribution. An observation sequence consists of all intensity values from each block. Next, model parameters are re-estimated by a process called E-M procedure to maximize the model probability until convergence. For face localization, an HMM is trained for a generic model of human faces from a large collection of face images.

To classify a tested face image, we have to found the HMM with the highest probability in the Viterbi decoding. So each face image would classify with independence of each other. A varied window feature vector is used by using five states and optimizing both of feature vector and the number of Gaussians densities per state. Figure 13 shows the classification success rate for varying Gaussian densities per state and several vector lengths for ORL (AT&T) and Essex Grimace databases, respectively.

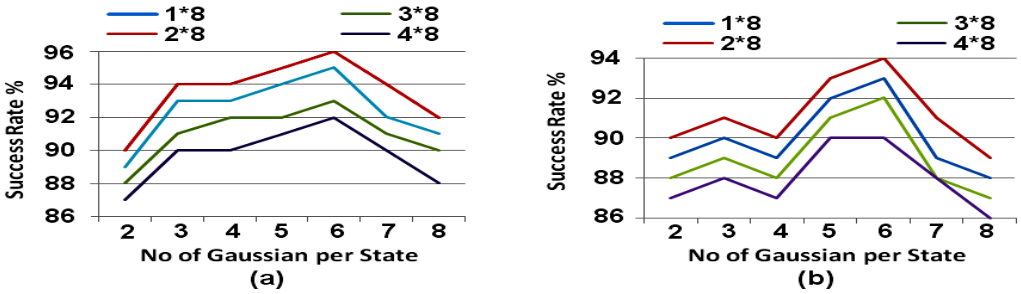


Figure 13: The Classification Success Rate Obtained when using Waveatom Features for Varying Gaussian Densities per State and Several Vector Lengths: (a) For Essex Grimace database of faces (b) For ORL (from AT&T) databases of faces.

Empirically, 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 × 2 window size as an observation vectors. These HMM parameters was produced a classification rate of 94% and 96% for ORL (AT&T) and Essex Grimace databases of faces, respectively.

In the case of using HMM with 2 levels Curvelet Transform, the feature vector resized to be 12 × 12 and 24 x 24 for AT&T and Essex Grimace respectively; to divide the observation (feature) vector to similar lengths. Figure 14 shows the classification error rate for varying Gaussian densities per state and several vector lengths.

In fact, using 5 states in HMM, the best result for success rate when using 2 levels Curvelet features was 92% with AT&T faces and 95% 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).

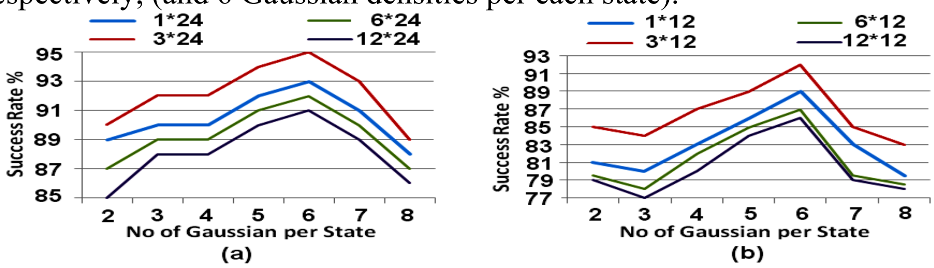


Figure 14: The Classification Success Rate Obtained when using 2 Levels Curvelet Features for Varying Gaussian Densities per State and Several Vector Lengths: (a) For Essex Grimace database of faces (b) For ORL (from AT&T) databases of faces.

7.5. Comparative Study

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 15 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 work. 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 88% and 92% 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 90% and 93% for ORL (AT&T) and Essex Grimace databases of faces, respectively. These results was against 81%, 85%, 87% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for ORL (AT&T) database of faces. And it was 75%, 86%, 90% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for Essex Grimace database of faces.

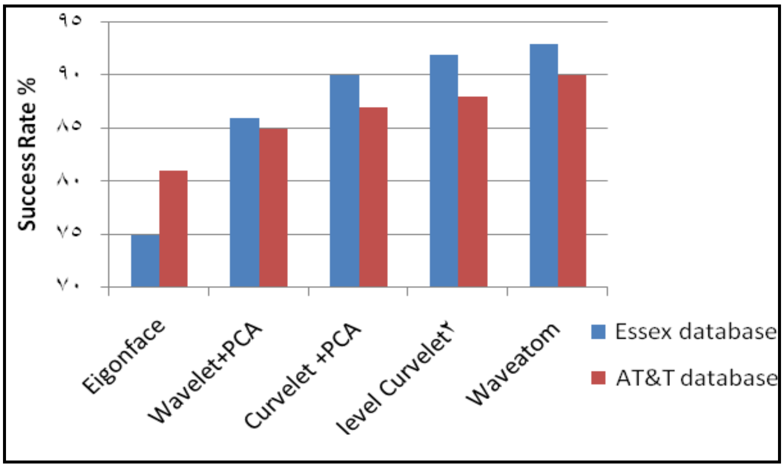


Figure 15: Success Rates for several feature extraction techniques with ANN.

Also, Figure 16 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 92% and 95% 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 94% and 96% for ORL (AT&T) and Essex Grimace databases of faces, respectively. These results was against 82%, 86%, 90% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for ORL (AT&T) database of faces. And it was 80%, 91%, 95% when using eigenfaces features, Wavelet+PCA features and Curvelet+PCA features, respectively, for Essex Grimace database of faces.

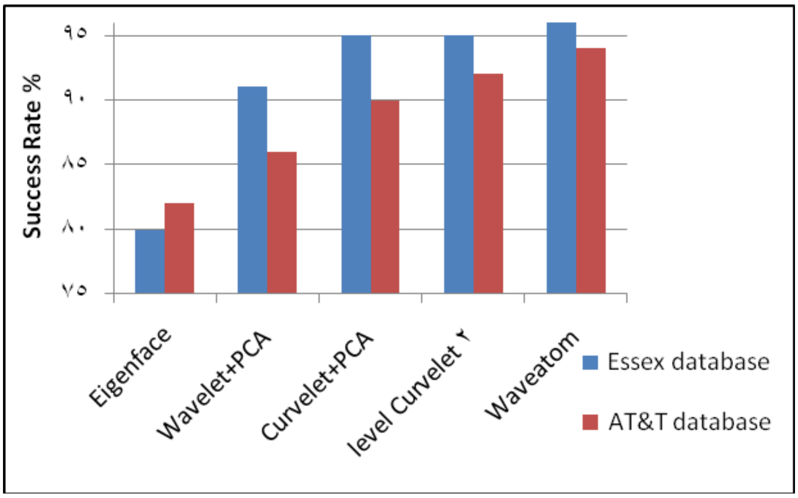


Figure 16: Success Rates for several feature extraction techniques with HMM

From Figure 15 and Figure 16, we can 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.

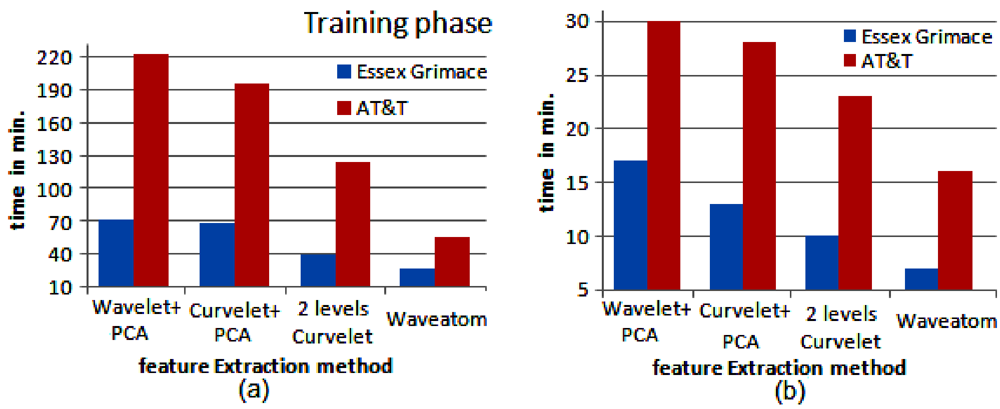


Figure 17: Time-Consuming for Training Phase by (a) ANN, (b) HMM.

As can be noted from Figure 17 (a), the smallest training time by using ANN classifier was achieved by waveatom feature extractor for both AT&T and Essex Grimace databases with 55 minutes and 26 minutes respectively. The same results can be noted from Figure 17 (b) whereas the waveatom features fed HMM classifier to complete the traning phase in 16 minutes and 7 minutes for AT&T and Essex Grimace databases respectively.

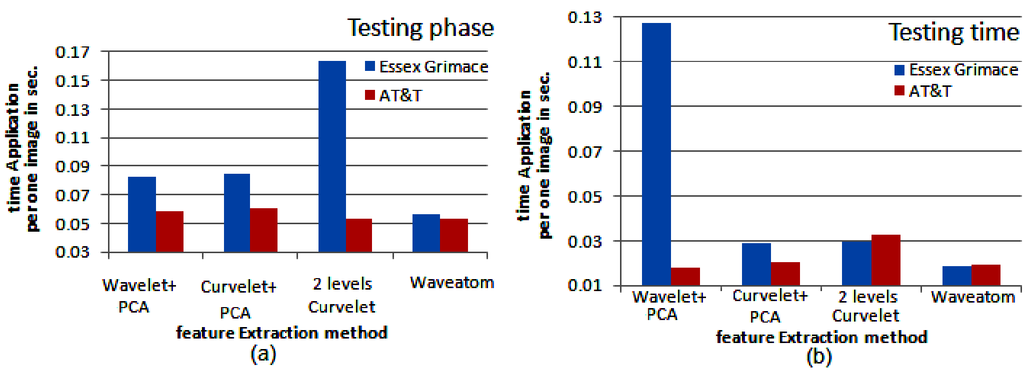


Figure 18: Time-Consuming for Testing Phase by (a) ANN, (b) HMM

.

For testing phase in Figure 18, can be noted that the 2-level Curvelet has the largest testing time-consuming by using ANN, while the worst testing time-consuming is achieved by Wavelet+PCA using HMM. From training and testing phases can be showed that the Waveatom is the best time-consuming by using either ANN or HMM as classifier.

8. Conclusion

The aim behind this work was found 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 proposal 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 shows significant variation in illumination and facial details. These feature extraction techniques are coupled with two different classification techniques. The classification techniques are the gradient descent backpropagation ANN and HMM.

From the comparative study, it is evident that both feature extraction techniques (two levels Curvelet features and Waveatom features) completely outperform standard eigenface technique; it also supersedes both Wavelet based PCA scheme and Curvelet based PCA scheme. The results indicate that Curvelet transform can stand 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. In the other side, comparative study shows the strongest of HMM over ANN; as they are classifiers. Thus, Waveatom based HMM perform the highest performance rate against other methods when compared to the state of art.

REFERENCES

1. Anbarjafari G, 2008: A New Face Recognition System Based on Colour Statistics, M.Sc. Thesis, Eastern Mediterranean University, Gazimağusa.
2. Bai L and Shen L, 2003: Combining Wavelets with HMM for Face Recognition, Inter. Conference on Artificial Intelligence, Cambridge, 2003.

3. Candes E, Demanety L, et al, 2006: Fast Discrete Curvelet Transforms, *Technical Report, Cal Tech*, March, 2006.
4. Candes E and Donoho D, 2000: Curvelets: A surprisingly effective nonadaptive representation for objects with edges, Available on http://www.Curvelet.org/papers/Curve99.pdf ,2000.
5. Chao-Chun L, Dao-Qing D, Hong Y, 2007: Local Discriminant Wavelet Packet Coordinates for Face Recognition, Journal of Machine Learning Research, pp. 1165-1195, 2007.
6. Chengjun L, Harry W, 2002: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition, IEEE transactions on image processing, vol. 11, no. 4, pp. 467-477, April, 2002.
7. Demanety L, Ying L, 2007: Curvelets and Wave Atoms for Mirror-Extended Images, available on: www.waveatom.org, july, 2007.
8. Demanety L and Ying L, 2007: Wave atoms and sparsely of oscillatory patterns, Appl. Comput. Harm. Anal. , February, 2007.
9. Demanety L. and Ying L., 2008: Waveatom, web page available on: www.wafeatom.org
10. Ekenel H., Sankur B., 2005: Multiresolution face recognition, Image and Vision Computing, vol. 23, PP. 469–477, 2005
11. Hejazi H and Alhanjouri M, 2010: Face recognition using Curvelet Transform, IEA/AIE 2010, Springer-Verlag Berlin Heidelber 2010, Spain, 1-4 June, pp. 388-396.
12. Jianhong X., 2009: Face Recognition Based on Curvelet Transform and LS-SVM, Proceedings of the International Symposium on Information Processing, Huangshan, P. R. China, August 21-23, pp. 140-143, 2009
13. Kohonen T., 1989: Self-organization and Associative Memory", Springer-Verlag, Berlin.
14. Lawrence S, Giles C, Tsoi A, and Back A, 1997: Face Recognition: A Convolutional Neural Network Approach, IEEE Trans. on Neural Networks, vol. 8, pp. 98-113, 1997.
15. Li B., Linlin S., 2003: Combining Wavelets with HMM for Face Recognition, International Conference on Artificial Intelligence, Cambridge, 2003.
16. Majumdar A, Bhattacharya A, 2008: Face Recognition by Multiresolution Curvelet Transform on Bit Quantized Facial Images, International Conference on Computational Intelligence and Multimedia Applications, vol. 2, pp. 209-213, 13-15 December, 2008

17. Majumdar A and Ward R, 2008: Multiresolution Methods in Face Recognition, Recent Advances in Face Recognition, Proceeding No ISBN 978-953-7619-34-3, I-Tech, Vienna, Austria, pp. 79-96, 2008.
18. Peng Y., et al, 2004: Face Recognition Using Ada-Boosted Gabor Features, the 6th IEEE int. Conference on Automatic face and gesture recognition, Korea, pp. 356 – 361, May, 2004.
19. Rabiner L, 1989: A tutorial on hidden markov models and selected applications in speech recognition, IEEE Proc., vol. 77, no. 2, pp. 257-286.
20. Rziza M, et al, 2009: Local Curvelet Based Classification Using Linear Discriminant Analysis for Face Recognition, International Journal of Computer Science, Volum 4, Number 1, pp. 72-77, 2009.
21. Sanghoon K, et al, 2007, Robust Face Recognition using AAM and Gabor Features, PWASET journal, vol. 21, pp. 493-497, January, 2007
22. Spacek L, 2008: Description of the Collection of Facial Images, available at (http://cswww.essex.ac.uk/mv/allfaces/grimace.zip), visit: Dec. 12, 2009.
23. Starovoitov V, et al, 2002: Three Approaches for Face Recognition, the 6th International Conference on Pattern Recognition and Image Analysis, October 21-26, 2002, Velikiy Novgorod, Russia, pp. 707-711.
24. Unknown, 2009: An archive of AT&T Laboratories Cambridge, available at (http://www.cl.cam.ac.uk/Research/DTG/attarchive), visit: Dec.12, 2009.
25. Vytautas P, 2004, Face Recognition Using Principal Component Analysis and Wavelet Packet Decomposition , informatica journal, vol. 15, no. 2, pp. 243–250, 2004.
26. Wang Z, Xie X, 2010: An Efficient Face Recognition Algorithm Based on Robust Principal Component Analysis, The 11th International Conference on Control, Automation, Robotics and Vision, 7-10 Dec. 2010, Singapore.
27. Wolde S, 2006: Face Recognition Using Artificial Neural Network, M.Sc. thesis, Faculty of Technology, Addis Ababa University, August 2006.
28. Young S et al., 2009: The HTK Tools and Reference Manuals, version 3.4, Cambridge University Engineering Department, 2009.