Deep Learning HMM Versus Multi-resolution transforms for Soft biometric estimation

Mohammed Ahmed Alhanjouri
Islamic University of Gaza
AlRimal, Gaza, Palestine
00970598923752
mhanjouri@iugaza.edu.ps

Ahmad Fouad El-Samak
Islamic University of Gaza
AlRimal, Gaza, Palestine
00970599347491
af.samak@alaqsa.edu.ps

ABSTRACT
This work is to introduce two different methodologies to estimate the soft biometric traits from face image. The first proposed methodology to extract effective features from facial images using two multi-resolution transforms; waveatom and shearlet, for estimating gender, ethnicity, facial expression and age by Artificial Neural Network (ANN). And the second proposed methodology to use deep learning to extract suitable features by double convolutional and pooling layers to feed Hidden Markov Model (HMM) for classification. To achieve the comparative study, our experiments carried out on a large database collected from three different databases: US Adult Faces, Extended Cohn-Kanade and FG-NET databases. The experimental results show that the multi-resolution waveatom transform was more effective than shearlet transform, but HMM with Deep learning were the best performance and more robust method to classify multi objects together such as in this paper to estimate 13 soft biometrics which clustering in four categories.

Keywords
Waveatom transform; Shearlet transform; Artificial Neural Network; Hidden Markov Model; Deep Learning.

1. INTRODUCTION
Soft biometrics classification using facial images has been in the field of research now days and it is quite interesting. Humans are very good in differentiating the gender, expression, ethnicity and age from facial images. Even if the face of the human is damaged, we can identify them with very high accuracy [1]. Soft biometrics classification such as classifying human face is only challenging for computer.

More recently automated soft biometrics classification from facial images has gained much interest in computer vision, machine language and Image processing community for various reasons, as of the non–intrusiveness, computational efficiency and mostly the need for higher reliability in soft biometric systems (SBSs). Such systems can achieve different goals including database search pruning, human identification, human re–identification and, on a different note, prediction and quantification of facial aesthetics.

Multiresolution analysis tools like wavelets have been found very useful for analyzing the image contents; hence, they widely used in image processing, pattern recognition and computer vision. Over the years, other multi-resolution techniques such as contourlets, ridgelets and curvelets were developed. Shearlet transform is a recent addition. It has been used in the field of image denoising, edge analysis, and image separation, but not much work has been done to solve pattern recognition problems. Another modern multi-resolution technique is Waveatom, which used in the field of image denoising, and the results obtained are the best one when compared to the state of art [2].

The rest of the paper is organized as follows; in Section 2, some related works to classify soft biometrics will be discussed. Section 3 presents the proposed multi-resolution waveatom and shearlet methods with ANN to estimate 13 soft biometric from face image. The deep learning and HMM methodology is illustrated in Section 4. The experimental results are shown in Section 5 and the conclusions are presented in Section 5.

2. LITERATURE REVIEW
There are different methods for the estimation of each trait, e.g., age, gender, facial expression or ethnicity estimation.

In [3], authors have proposed a novel Contourlet Appearance Model (CAM) to localize facial landmarks and extract texture based features from the convex hull bounded by them. Our CAM is shown to be more accurate at reconstructing unseen texture compared to conventional Active Appearance Models (AAM) because of its use of a Modified Active Shape Model for locating the facial landmarks. The results of using our CAM based age estimation technique on databases such as FG-NET and PAL are quite promising and highlight the improved accuracy of our method over several existing age estimation techniques.

A novel gender-specify age estimation model, named Correlated Warped Gaussian Processes (CWGP) was proposed in [4]. This model was shown to be particularly effective in the scenario that age data is very limited for each person.

In [5], multi-scale bandlet and Local Binary Pattern (LBP) based method for gender recognition from faces is proposed. After extracting bandlet coefficients from face images at different scales, LBP is applied to create a histogram, which is used as the feature to a minimum distance classifier. The experiments are performed using FERET grayscale face database, and the highest accuracy of 99.13% is obtained with the proposed method which is better than that of some state-of-the-art methods. However, the proposed method has a high dimension of features, and therefore, is computationally more expensive than PCA or LBP.

In [6], N. S. Lakshmiprabha used four feature extraction methods such as AAM, Gabor wavelets, LBP and Wavelet Decomposition (WD) for gender recognition, age estimation, expression recognition and racial recognition. Experiments carried out on FG-NET, Cohn-Kanade, PAL face databases. Results show that AAM features are better than other features in terms of accuracy and time taken for testing an image. LBP and Gabor give similar performance, whereas LBP is computationally less expensive. In term of time consumption during training and testing, WD is better than other methods. Aging effect in case of gender recognition can be tackled using AAM features. The performance of gender recognition is affected by using different shape landmark points which shows the inconsistence of AAM features.

In [7], authors have proposed a new method to use Local Binary Pattern histogram (LBPH) feature for ordinary binary classification problems. For each LBPH feature, a sample's Chi Square distance to a reference template is used as a measurement of confidence for classification; the positive mean histogram is used as the initialization, and the steep decent method is used to find an optimal reference template. Real AdaBoost is used to train a strong classifier by composing a sequence of LBPH features. The experiments on gender, ethnicity and age classification prove their method's effectiveness. However, a classification of ethnicity, gender, and age groups was executed, with each trait classified independently.

3. PROPOSED MULIT-RESOLUTION METHODS WITH ANN

In this thesis, we present a methodology to extract effective features for soft biometrics estimation from facial images; it can estimate four traits which are: ***age***, ***gender***, ***facial expression*** and ***ethnicity***. The four traits can be classified independently, and there is no relation among them. This approach illustrated by the framework shown in figure 1.

The proposed system consists of four stages which are:

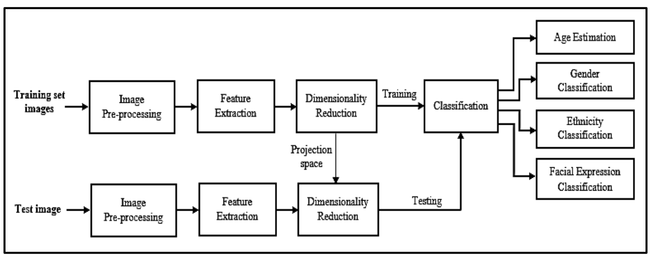
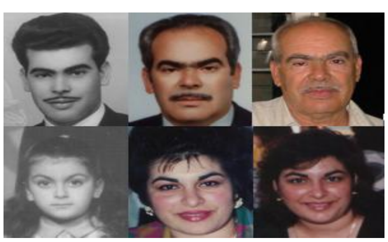


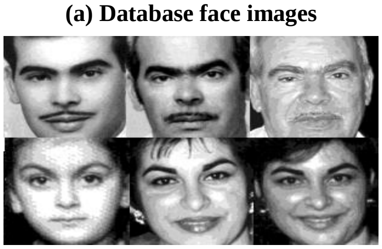
Figure 1. A framework of independent estimation of age, gender, expression and ethnicity, separately.

3.1 Preprocessing Stage

It is necessary to preprocess the face images for biometrics estimation to remove unnecessary features such as background, clothes, and hair. In this stage, we carry out preprocessing to extract only the face area by three steps: (1) convert color images to gray scale images, (2) detect and cropped face region from image, and (3) resized the cropped image into 250×250, as in figure 2.



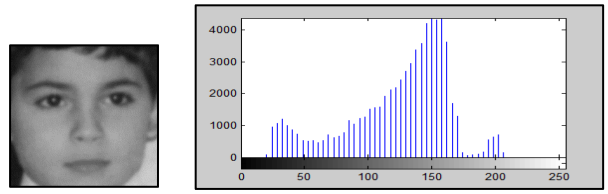
(a) Database face images



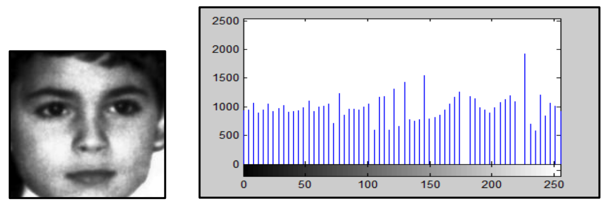
(b) Preprocessed and Normalized face images

Figure 2. Face images from FG-NET database.

Then, we applying histogram equalization on the resized images to improve the contrast and colorfulness of the original images thus can enhance the estimation process. Figure 3 explain histogram equalization technique.



(a) Histogram of face image before normalization



(b) Histogram of face image after normalization

Figure 3. Histogram equalization technique

3.2 Feature Extraction Stage

This paper deals with two multi-resolution transforms as a feature extraction method namely waveatom and shearlet transforms for gender, ethnicity, facial expression and age estimation.

3.2.1 Waveatom Transform based Feature Extraction

Wave atom transform is a new multi-resolution technique with the ability to sparsely represent the face images, since they belong to a category of images that oscillate smoothly in varying directions [8]. Face images are decomposed by directional 2D Wave atoms transform. The cells c{j,d}{m1,m2}(n1,n2) with d = 1, 2 are the waveatoms coefficients at scale j, frequency index (m1,m2) and spatial index (n1, n2). At each scale, the partitioning into cells is indexed by m1 and m2. Each cell contains a matrix of waveatom coefficients. Figure 4 shows the coefficients of waveatom transform at different scales. For example, at scale three and frequency index (5, 3), the coefficients matrix length is 8x8 coefficients. Whilst, the coefficients matrix length, at scale four and frequency index (8, 4); is 16 x 16 coefficients. The last scale has 32x32 coefficients at frequency index (7, 5).

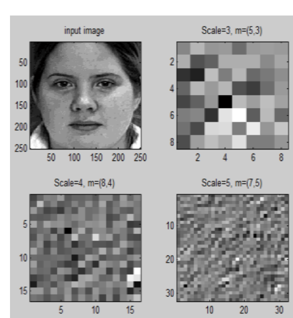


Figure 4. Waveatom transform coefficients at different scales

Face image are decomposed by directional 2D Waveatom transform at scale five, and a feature vector was used to store the extracted features.

3.2.2 Discrete Shearlet Transform based Feature Extraction

Discrete Shearlet Transform is a multi-scale directional representation of an image that provides a highly efficient representation of images with edges [9]. In the feature extraction stage, the preprocessed image was decomposed using shearlet transform into two components namely vertical and horizontal cones at different scales (5 levels). For example, figure 5 shows a face image from CK+ dataset that have been used to be decomposed at different scales such as in figure 6 and 7.

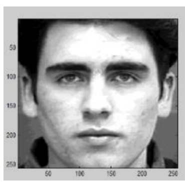


Figure 5: Sample image

Figure 6 shows the shearlet coefficients for vertical cone at five levels. Level 1 is the finest level, while the level 5 is the coarsest level. The first and second levels have 9 directions with 32x128 features and 32x64 features in each direction, respectively. whilst, the number of directions in the third and fourth level are 5 directions with 16x32 features and 16x16 features in each one, respectively. The last level has 5 directions with 8x8 features in each direction.

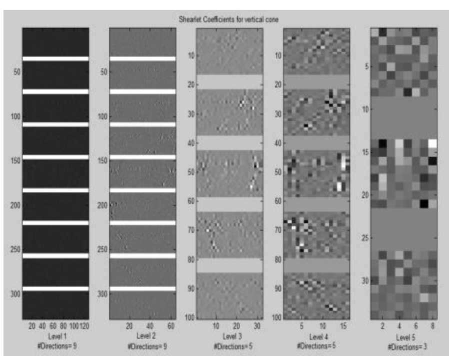


Figure 6. shearlet coefficients for vertical cone at five levels.

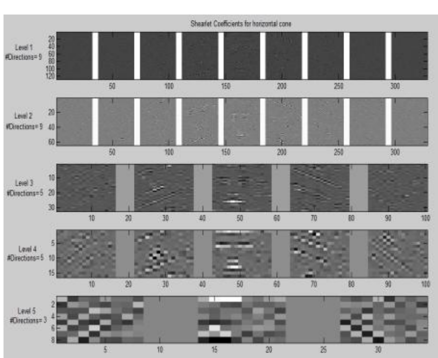


Figure 7: shearlet coefficients for horizontal cone at five levels.

Figure 7 shows the shearlet coefficients for horizontal cone at five levels. Also, level 1 is the finest level, while the level 5 is the coarsest level. The first and second level have 9 directions with 128x32 features and 64x32 features in each direction, respectively. whilst, the number of directions in the third and fourth level are 5 directions with 32x16 features and 16x16 features in each one, respectively. The last level has 5 directions with 8x8 features in each direction.

Finally, feature vector was used to store the features were extracted from the first level (finest level) of horizontal and vertical cones.

3.3 Dimensionality Reduction Stage

The dimensionality reduction stage of features matrix is useful for different reasons. First, it reduces the time and storage space required. Second, removal of multi-collinearity improves the performance of the machine learning model. Third, it becomes easier to visualize the data when reduced to very low dimensions. In this stage, we use the *Principal Component Analysis (PCA)* algorithm as a feature dimension reduction technique [10].

To clarify the work of PCA algorithm in reducing feature dimension, suppose we have *P* images in the training database with the same size of 250x250. After extracting features from each image using feature extraction techniques, we get 1D feature vectors with length *R* and the training set matrix will be *RxP* 2D matrix.

After applying PCA algorithm to the *RxP* 2D matrix, the dimensionality of the training set matrix will be reduced to *PxP*. Note that the feature vectors length that extracted using waveatom transform is 39,936 features, while in shearlet transform is 73,728 features. This is due to the ability of the waveatom transform to selects the most representative features and most compactly expressed the input image and rejects all other less compact representations.

3.4 Classification Stage

Once a matrix of features is obtained for all images, a classifier can be used to perform the final estimation, in this section, artificial neural network is used.

Artificial neural network consists of three layers: input layer, hidden layer and output layer. Initially, we proposed to represent the output layer as 13-bits that represent all biometrics possibilities as shown in figure 8. In which, the hidden nodes are divided into four sets and each set is responsible for one estimation task.

However, the partitioning feature of the hidden nodes is not available with the ANN Matlab tool. So, we use for a separate ANN for each estimating task and testing with the large database, as explain in figure 9.

The training is done using *gradient descent with momentum and adaptive learning rate backpropagation algorithm*. Number of input nodes is equal to the size of the feature vectors.

Tan-sigmoid and *purelin* are the transfer functions that used for the hidden and output layers, respectively. The parameters that have been set for this network are shown in table 1.

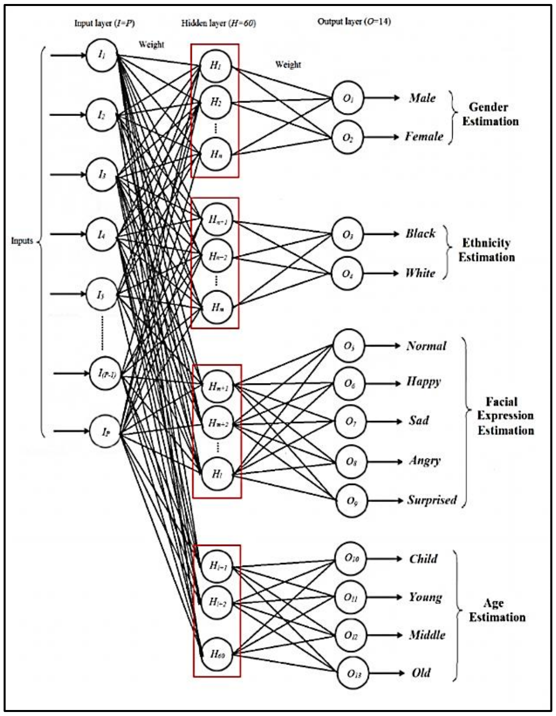


Figure 8. The proposal of partitioning hidden nodes in ANN matlab tool.

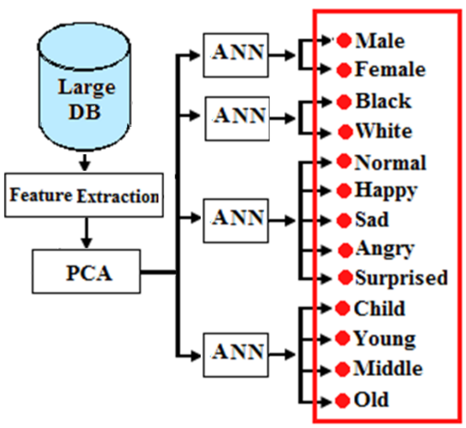


Figure 9. Diagram of the soft biometrics estimation system.

Table 1. ANN parameters

Network Parameter	Value
Hidden nodes	60
Number of iterations	500
learning rate	0.01
Momentum constant	0.93
Achieve goal	1e-5

4. PROPOSED HMM WITH DEEP LEARNING

To achieve our proposed HMM, the pre-processing steps in previous method are needed, such as segmentation that is defined as the process of extracting face from image.

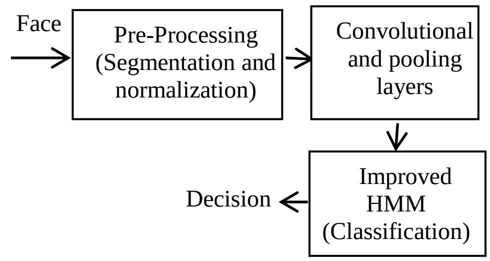


Figure 10. A generic proposed Ear recognition system

The next step is convolutional and pooling layers, which represents the feature extraction step in this models, involves obtaining relevant face features from the data. And the last part is Improved HMM that recognizes the soft biometrics from face images as shown if figure 10. In classification step we implement and train HMM for each soft biometric, which use to evaluate posterior probabilities of unknown test sample to select the model with highest value.

4.1 Convolutional Layer

Figure 11 shows convolutional operation structure with (3×3) filter, the number of output channels corresponds to the number of filters (or kernels) in the filter bank for that particular layer.

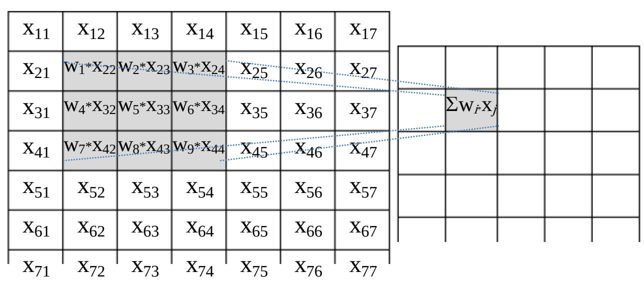


Figure 11. Convolution layer process

To convolve a filter (also known as the kernel) W with an input X we simply place the corresponding filter on the top left corner of our input and slide it across as shown in Figure 7. Every time we re-position our kernel we multiply the filter's weights by the corresponding values in the input tensor and reduce the output by adding up the result of each individual multiplication. If the spatial dimensions of the input are $n \times m$, we use a stride of 1×1 and our kernel's dimensions are $k \times k$ where k is odd and smaller than m and n then the spatial dimensions of the corresponding output are $n - (k - 1) \times m - (k - 1)$ as we don't allow the filter to be placed outside the image partially or otherwise; this mode of operation is referred to as "valid" mode. This mode is particularly helpful while pairing convolution with max-pooling and seeking deeper architectures as it allows us to perform an arbitrary number of convolutions without reducing the spatial dimensions of the current representation [11].

4.2 Max-Pooling

Max-Pooling is traditionally applied in combination with convolution and it is usualy placed in between convolutional layers. We can think of this operation as placing a grid on top of the original input and pooling every element within the corresponding window as illustrated in Figure 8. Notice that every channel is processed independently (i.e. we only apply max-pooling across spatial dimensions) [11].

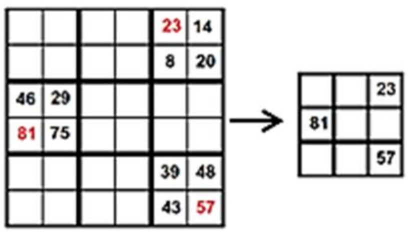


Figure 12. Max Pooling 2

The max-pooling operation reduces the spatial dimensions of the input; reducing the size of the network as we move deeper into the architecture. More specifically, for an m × n input tensor, a $k \times k$ max-pooling operation, where k is smaller than m and n, also k, m, and n are all even, reduces the size of the representation down to $m / k \times n / k$. Figure 12 shows the Max-pooling 2 which is used in this work after each convolutional layer.

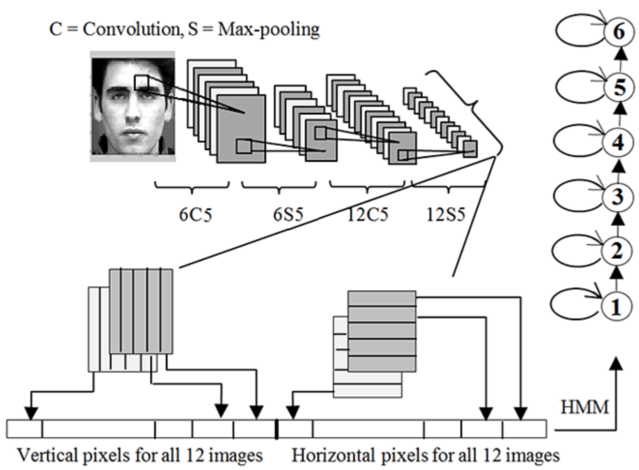


Figure 13. Proposed Deep-HMM System for soft biometric estimation

In our experiment as illustrated in figure 13, we used (6C5) 6 filters (kernel size is 5*5) in convolutional layer to produce 6 images with size of 246*246 that reduced by max pooling layer to 123*123. Second convolutional layer is used to produce 12 images with size of 119*119 and then reduce it to 55*55. Finally we concatenate the resultant pixels horizontally and vertically respectively to feed the HMM.

The HMM consists of 6 states with 4 Gaussian Mixture in each one to model the features of 13 classes represent the soft biometrics that used.

5. EXPERIMENTAL RESULTS

Experiments were carried out using a large database collected from three different databases; US Adult Faces [12], CK+ [13] and FG-NET [14]. This large database contains 800 images for training phase and 265 images for testing phase. The large database contains all the type of biometrics, such as gender (female, male), ethnicity (black, white), expression (normal, happy, sad, angry, surprised) and age (child, young, middle, old).

For the implementation, Matlab R2014a is used to implement the proposed methodology because it contains toolboxes for waveatom and shearlet transforms. ShearLab v1.1 matlab toolbox [17] was used to yield Shearlet feature vectors. Waveatom feature vectors were outperformed using WaveAtom v1.1 matlab toolbox [18]. Experiments are performed on a hp laptop with core i7 CPU (2.5 GHz), 16GB DDR3 RAM and windows 7 operating system.

The performance of proposed multi-resolution methodology using waveatom transform on a large database proved to be more effective than shearlet transform for gender estimation, ethnicity, facial expression, and age estimation, while the deep learning with double convolutional and pooling layers to feed HMM was the best as shown in figure 14.

Deep learning with HMM achieves accuracy rate of 98.5%, 99.25%, 98.5%, and 98.1% for gender estimation, ethnicity, facial expression, and age estimation, respectively.

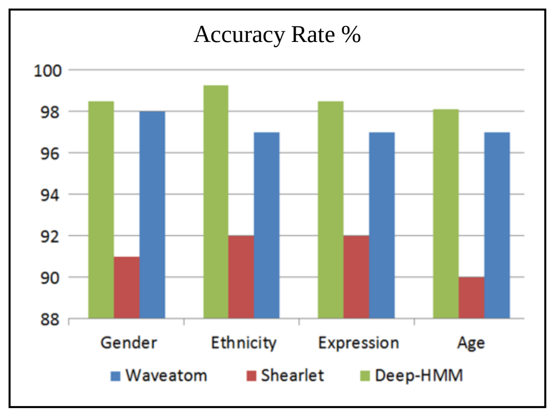


Figure 14: Simulation results of the proposed methods.

6. CONCLUSION

This paper introduced two proposed methodologies to estimate the soft biometrics from face image. The first one depends on using multi-resolution transform to extract features for ANN, in this methodology we used waveatom and shearlet transforms and the results showed that waveatom more accurate than shearlet for face images. The second introduced methodology depends on deep learning to extract suitable features to feed HMM for classification

To train and test our systems; waveatom, shearlet, and deep-HMM, we used big collected database to classify 13 soft biometric traits clustered in four categories; gender, ethnicity, expression, age. The deep-HMM gave the best results with accuracy rate from 98 to 99 % for the four categories.

7. REFERENCES

[1] Michael C. Mangini, Irving Biederman. Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science* 28 (2004) 209–226.

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

[3] Khoa LUUl, Keshav Seshadri2, Marios Savvides, Tien D. Buil, and Ching Y. Suenl, Contourlet Appearance Model for Facial Age Estimation, *IEEE International Joint Conference on Biometrics (IJCB),* 2011.

[4] Difei Gao, Lili Pan, Risheng Liu, Rui Chen, Mei Xie, Correlated warped Gaussian processes for gender-specific age estimation, *IEEE International Conference on Image Processing (ICIP)*, 2015, Pages: 133 – 137.

[5] F. Alomar; G. Muhammad; H. Aboalsamh; M. Hussain; A. Mirza; G. Bebis, Gender recognition from faces using bandlet and local binary patterns, *International Conference on Systems, Signals and Image Processing (IWSSIP),* 2013, Pages: 59 – 62.

[6] N. S. Lakshmiprabha, Face Image Analysis using AAM, Gabor, LBP and WD features for Gender, Age, Expression and Ethnicity Classification, 26 March 2016.

[7] Z. Yang, H. Ai, Demographic classification with local binary patterns, *International Conference on Biometrics (ICB),* 2007, pp. 464–473.

[8] L. Demanet, L. Ying, Wave Atoms and Sparsity of Oscillatory Patterns, to appear *in Appl. Comput. Harmon. Anal*. 2007.

[9] Wang-Q Lim. Discrete Shearlet Transform: New Multiscale Directional Image Representation. Laurent Fesquet and Bruno Torrésani. SAMPTA'09, May 2009, Marseille, France.

[10] Abdi. H., & Williams, L.J. Principal component analysis, Wiley Interdisciplinary Reviews: Computational Statistics, 433–459, 2010.

[11] M. Alhanjouri, Improved HMM by Deep Learning for Ear Classification, International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume-6, Issue-3, May 2018

[12] 10k US Adult Faces Database available on (http://www.wilmabainbridge.com/facememorability2.html)

[13] Extended Cohn-Kanade database available on (http://www.consortium.ri.cmu.edu/ckagree/)

[14] The FG-NET Aging Database available on: (http://wwwprima.inrialpes.fr/FGnet/html/benchmarks.html)

[15] ShearLab, software package, available on (http:// www.shearlab.org)

[16] WaveLab, software package, available on (http:// www.waveatom.org)

Authors' background

Your Name	Title*	Research Field	Personal website
Mohammed Ahmed Alhanjouri	Associate Professor	AI, DSP, Optimization	http://site.iugaza.edu.ps/mhanjouri
Ahmad Fouad El-Samak	master student	AI, Optimization	

***This form helps us to understand your paper better, the form itself will not be published.**

***Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor**