

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/275243450>

# Age-Invariant Face Recognition Using Triangle Geometric Features

ARTICLE *in* INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE · APRIL 2015

Impact Factor: 0.67 · DOI: 10.1142/S0218001415560066

---

READS

61

5 AUTHORS, INCLUDING:



Vijanth Sagayan Asirvadam

Universiti Teknologi PETRONAS

138 PUBLICATIONS 205 CITATIONS

[SEE PROFILE](#)



Mohamed Meselhy Eltoukhy

Universiti Teknologi PETRONAS

17 PUBLICATIONS 195 CITATIONS

[SEE PROFILE](#)

## Age-Invariant Face Recognition Using Triangle Geometric Features

Amal Seralkhatem Osman Ali\*, Vijanth Sagayan Asirvadam,  
Aamir Saeed Malik, Mohamed Meselhy Eltoukhy and Azrina Aziz

*Center of Intelligent Signals and Imaging Research  
Department of Electric and Electronic Engineering  
Universiti Teknologi PETRONAS  
31750 Tronoh, Perak, Malaysia  
\*Amal\_61@hotmail.com*

Received 22 August 2014

Accepted 27 March 2015

Published 26 May 2015

Whilst facial recognition systems are vulnerable to different acquisition conditions, most notably lighting effects and pose variations, their particular level of sensitivity to facial aging effects is yet to be researched. The face recognition vendor test (FRVT) 2012's annual statement estimated deterioration in the performance of face recognition systems due to facial aging. There was about 5% degradation in the accuracies of the face recognition systems for each single year age difference between a test image and a probe image. Consequently, developing an age-invariant platform continues to be a significant requirement for building an effective facial recognition system. The main objective of this work is to address the challenge of facial aging which affects the performance of facial recognition systems. Accordingly, this work presents a geometrical model that is based on extracting a number of triangular facial features. The proposed model comprises a total of six triangular areas connecting and surrounding the main facial features (i.e. eyes, nose and mouth). Furthermore, a set of thirty mathematical relationships are developed and used for building a feature vector for each sample image. The areas and perimeters of the extracted triangular areas are calculated and used as inputs for the developed mathematical relationships. The performance of the system is evaluated over the publicly available face and gesture recognition research network (FG-NET) face aging database. The performance of the system is compared with that of some of the state-of-the-art face recognition methods and state-of-the-art age-invariant face recognition systems. Our proposed system yielded a good performance in term of classification accuracy of more than 94%.

*Keywords:* FRVT; age-invariant; geometrical model; triangular features; FG-NET.

\*Corresponding author.

## List of Abbreviations and Acronyms

- FRVT : face recognition vendor test
- FG-NET : face and gesture recognition research network
- AAM : active appearance model
- PCA : principal component analysis
- TSP : triangle similarity proportion
- SVMs : support vector machines
- LOPO : leave one person out
- CI : confidence intervals
- VR : verification rate
- HTER : half total error rate
- EER : equal error rate
- 1NN : one nearest neighbor
- FP : false positive
- LBP : local binary patterns

## 1. Introduction

Biometric verification technological innovation plays an important role in information privacy and security. Recently, it has gained increasing attention from researchers in the field of computer vision. Biometrics is the technique used to distinctly identify individuals in accordance with some form of unique physiological or potential tendency of characteristics. These include fingerprints, facial recognition, palm geometry, iris identification, vein recognition, signature recognition and voice recognition.<sup>1</sup>

Facial recognition is considered as the most ideal biometric technique.<sup>1</sup> It performs the authentication task based on the personal facial features of human beings. Related studies have involved facial detection, face expression recognition, face tracking and identity recognition. The accuracy of facial recognition is usually affected by substantial intra-class variations due to underlying factors including age, pose, lighting and expression. As a result, the majority of the current works focus on the efforts of minimizing the effects of such variations that could deteriorate the overall performance of the facial recognition. However, the effects of facial aging on the performance of the facial recognition systems have not been thoroughly studied. Thus, there is a need to develop facial recognition algorithms which are generally invariant towards aging variations.<sup>2</sup> These algorithms are called age-invariant facial recognition method.

Facial aging is known as a complex process that varies both shape and texture of the facial area. Shape variations include craniofacial, whereas texture variations include skin coloration, lines and/or wrinkles. Shape and texture are both considered as the common facial aging patterns. Since the aging process occurs throughout different ages, it can be classified into the above-mentioned aging patterns. As a

result, an age-invariant facial recognition method should account for these aging patterns.<sup>3</sup>

The applications of age-invariant facial recognition are lost children investigations, human-computer interaction and passport photo verification. These applications possess two basic attributes: (1) Significant age difference between probe and gallery photos (photos acquired within enrollment and also authentication phases) and (2) failure to acquire the person's facial photo in order to update the template.<sup>2</sup> Furthermore, real-world conditions continue to be challenging, mainly due to the great deal of changes in the process of acquiring faces. For example, criminal offense inspections and society safety organizations must frequently fit a probe photo with registered photos in a database. This may depict significant differences in facial characteristics due to the presence of different age groups. Numerous attempts have already been made to overcome such challenges.

The research on age-invariant face recognition has increased since 2002.<sup>4</sup> One of the first works related to recognizing an aging face from digital images belongs to Kwon and Lobo.<sup>5</sup> In their system, they used certain facial anthropometric measurement to represent shape variations due to aging. Moreover, they addressed texture variations using the thickness of facial lines and wrinkles. Ling *et al.*,<sup>2,6</sup> investigated the effect of age variations on the performance of facial recognition during a passport picture authentication procedure. They presented a non-generative concept, in which a face operator is determined and extracted using image gradient orientations from several image resolutions. Then, support vector machines (SVMs) are used to perform aging facial verification.

Lanitis *et al.*,<sup>7,8</sup> presented a method which generally adopts the active appearance model (AAM) to perform age estimation. They developed a facial aging model that incorporates both texture intensity and shape variations information. Geng *et al.*,<sup>9,10</sup> studied an aging pattern subspace based on the concept that similar faces age in similar ways. Guo *et al.*<sup>11</sup> proposed a kind of age manifold learning pattern for facial age estimation. The facial aging features are extracted and locally altered. They used a robust repressor to predict the ages of human beings. Furthermore, Fu and Huang<sup>12</sup> proposed a technique for manifold learning where a low-dimensional manifold could be identified from a set of facial images that were arranged according to age. Linear and quadratic regression functions were applied on the low-dimensional feature vectors through the particular manifolds to demonstrate facial age estimation. Ramanathan and Chellapa<sup>13</sup> presented a functional two-step strategy to model the aging process of adults. Such model is composed of facial texture and shape variations. In their model, shape variations are modeled by developing physical layouts which usually characterize the abilities of the facial muscular tissues.

Drygajlo *et al.*<sup>14</sup> demonstrated the effective use of a Q-stack classifier to perform facial aging authentication. Mahalingam and Kambhamettu<sup>15</sup> proposed a probabilistic method for facial verification over facial aging. In the proposed method, facial

features descriptors are extracted from hierarchical manifestation of face photos. These descriptors are then applied within a probabilistic model designed for verification applications. Park *et al.*<sup>16</sup> introduced a facial aging simulation approach, which is able to learn shape and texture aging patterns by using principle component analysis (PCA) coefficients.

Suo *et al.*<sup>17</sup> proposed a dynamic model to imitate the facial aging process. The proposed model is a facial growth model which is influenced by age and hair characteristics. The model describes each photo using a multi-layer and-or graph which incorporates adjustments of hairstyle. It also recognizes deformations, shape, and wrinkles variations within the facial components. Singh *et al.*<sup>18</sup> employed a facial aging approach that projects the probe and gallery facial images to polar coordinate space. Moreover, it minimizes the changes in facial characteristics due to facial aging. Tiddeman *et al.*<sup>19</sup> adopted a face age simulation approach. They presented a functional wavelet transformation to model composite facial images. Burt and Perrett<sup>20</sup> proposed a face age simulation technique that uses texture and shape information for forming composite facial images comprising several age ranges. They evaluated and analyzed particular facial cues that are influenced by age.

Ramanathan *et al.*<sup>13</sup> presented an age-difference Bayesian classifier to be used with passport photo verification applications. Most of the previously proposed approaches attempted to eliminate such challenges by imitating the facial aging models. However, these are still far from hands-on utilization. An interesting concept of utilizing the most age-invariant facial region, i.e. Periocular area, was presented in Ref. 21, where the fusion of Walsh–Hadamard transform and LBP (so-called Walsh–Hadamard transform Encoded LBP, WLBP) was applied to this facial segment. Next, the unsupervised discriminant projection (UDP)<sup>22</sup> was used to build subspaces of the images. Fazli *et al.*<sup>23</sup> proposed two methods: discrete wavelet transform and combination discrete wavelet transform with gradient orientation for feature representation, followed by a principal component analysis to reduce the features dimensions, then recognize by Euclidean distance.

Most recently, Bouchaffra *et al.*<sup>24</sup> introduced a novel formalism that performs dimensionality reduction and captures topological features (such as the shape of the observed data) to conduct pattern classification. This mission is achieved by: (1) Reducing the dimension of the observed variables through a kernelized radial basis function technique and expressing the latent variables probability distribution in terms of the observed variables; (2) disclosing the data manifold as a 3D polyhedron via the  $\alpha$ -shape constructor and extracting topological features and (3) classifying a data set using a mixture of multinomial distributions. They have applied their methodology to the problem of age-invariant face recognition.

Our system is designed to deal with the effect of facial aging on the performance of facial recognition systems. We are proposing a simple, fast, yet robust geometric approach. Our work includes developing a mathematical and geometrical model to maintain the degree of similarity between six triangular facial features connecting

and characterizing the main facial features. The system is anticipated to work with real-time applications such as airports identity verification systems where the overall processing time and false positive (FP) rate are critical factors.

The remainder of this paper is organized as follows: Section 2 describes the proposed face recognition geometrical model. Section 3 is dedicated to the feature selection methods, data normalization methods, and classifiers used throughout the experimental part. The results and discussions are presented in Sec. 4. The conclusion and future work are presented in Sec. 5.

## 2. The Proposed Geometric Model

In an endeavor to model the facial variations associated with aging, the proposed model drives a set of triangle ratios. The proposed approach drives a set of triangular regions over the face by connecting some of the facial feature points alternatively. The triangular regions cover areas of the face that are assumed to change in size with aging. A set of mathematical relationships that represent the ratios between the different triangular regions are derived. The outcome of the mathematical relationships is used to build a feature vector for each sample image. Such approach attempts to model the way the main facial features (i.e. eyes, nose and mouth) are correlated with each other as effect of aging. Moreover, the proposed method models the craniofacial growth process, which is usually experienced during childhood as effect of aging, by using a simple yet effective approach.

A block diagram of the main phases of our proposed geometric model is illustrated in Fig. 1. First, the test image will be passed to the system (in the experimental part, all the test images are collected from the face and gesture recognition research network (FG-NET) database). Then, the coordinates for a total of 12 facial feature points are retrieved. The FG-NET<sup>25</sup> database images are available with annotation of a number of facial feature points, where the coordinates of such points are attached to the database in text files. As Fig. 2 shows, there are a total of 68 facial features (focal) points in each annotated sample image in the FG-NET database. However, only 12 facial feature points are considered in our proposed model as illustrated in Fig. 3. The coordinates of the selected 12 facial feature points will be passed to the system. The retrieved coordinates of the facial feature points will be used to form six different triangular regions surrounding the main facial features (i.e. eyes, nose and mouth). Subsequently, the parameters of the triangular regions (i.e. perimeters and areas) are calculated using the standard formulas. Finally, the parameters of the triangles are passed to a set of developed mathematical equations to build the feature vectors for the sample images. The main phases of our proposed system will be discussed in more details in the following section.

The six triangular regions adopted in the proposed model are illustrated in Figs. 4(a)–4(f). The first triangle (T1) covers the distance between the inner corners of the eyes and the distances between the inner corners of the eyes and the center of the nose. The second triangle (T2) mainly covers the size of the nose. The third

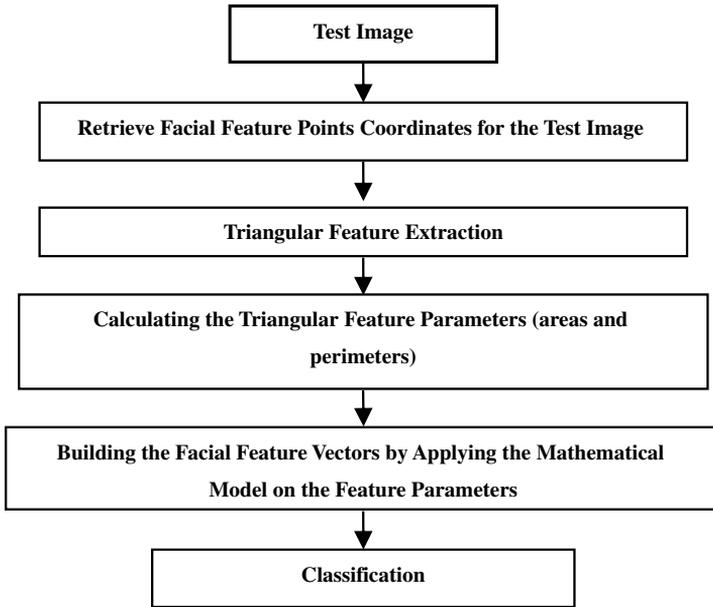


Fig. 1. Block diagram of the proposed system.

triangle (T3) covers the distance between the outer corners of the eyes and the center of the mouth. The fourth triangle (T4) is what is known in anatomy as the danger triangle.<sup>26</sup> Such triangle consists of the area from the corners of the mouth to the bridge of the nose, including the nose and maxilla (the maxilla is the two maxilla bones forming the upper jaw and palate of the mouth<sup>27</sup>). The fifth triangle consists of the area from the outer corners of the eyes and the center of the chin. The sixth triangle consists of the area from the inner corners of the eyes and the center of the mouth.

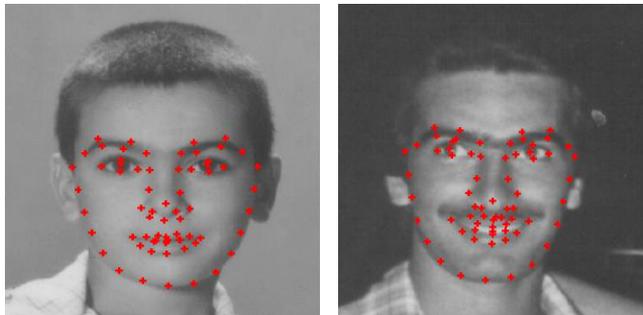


Fig. 2. Examples of the images and landmarks (labeled as red) at different ages of the same subject in the FG-NET database. (a) An image of a subject at age 8; (b) an image of the same subject at age 18.<sup>53</sup>

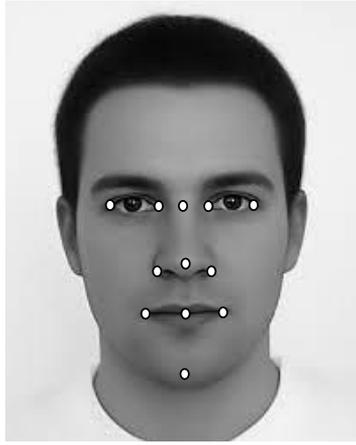


Fig. 3. The main facial feature points used in the proposed model.

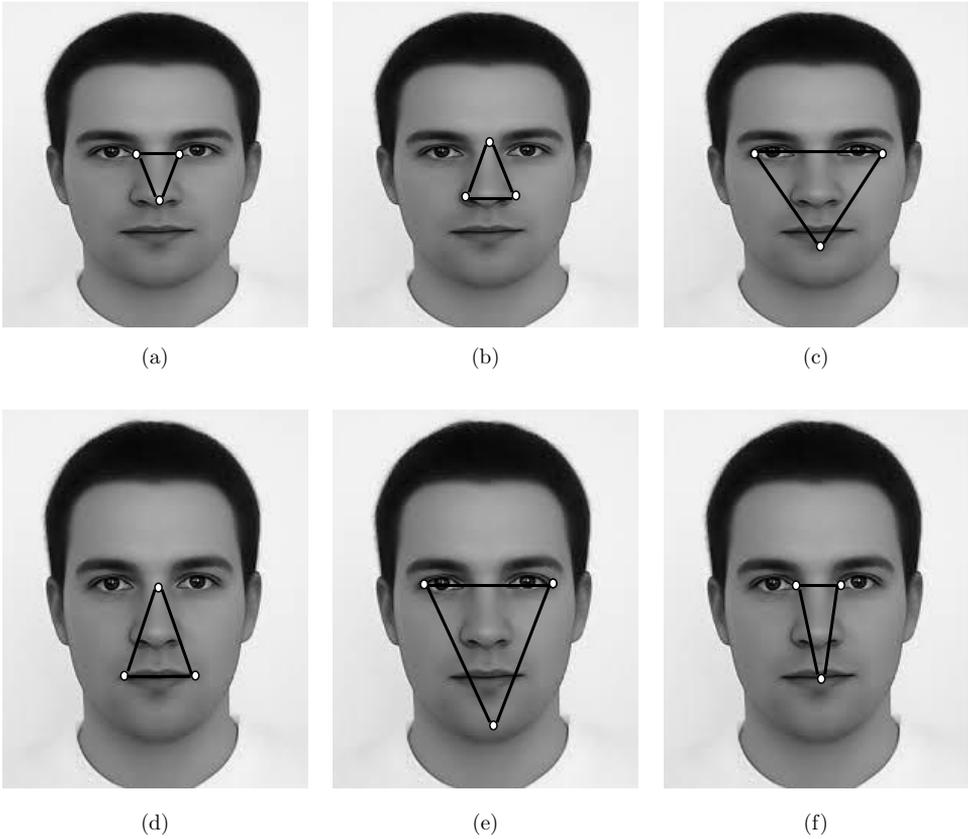


Fig. 4. (a) Triangle 1 (T1), (b) triangle 2 (T2), (c) triangle 3 (T3), (d) triangle 4 (T4), (e) triangle 5 (T5) and (f) triangle 6 (T6).

### 2.1. Calculating the parameters of the triangles

Once the coordinates of the facial feature points are retrieved, the Euclidean distances<sup>28</sup> between the facial feature points are calculated in order to compute the perimeters and areas for each triangle.

As a final stage, the triangles areas and perimeters are passed to a number of mathematical equations to create feature vectors for each sample image. The mathematical equations are discussed in the following section.

### 2.2. Developing the mathematical triangles relationships

In the science of computational geometry, triangles are considered similar when their shapes are similar, but not necessarily having identical dimensions.<sup>29</sup> Such scientific fact motivated us to develop mathematical relations between the six triangular areas. There is a massive human population in our planet estimated recently as about 7 billion. Therefore, it is unrealistic to carry out a direct one-to-one comparison using the triangles parameters to perform face recognition. As a different approach, we propose the use of a set of triangles proportion ratios. As a result, a set of 30 mathematical equations are developed. Those equations are used to estimate the degree of similarity between one triangle and another. Relatively, any two triangles are considered similar if the following geometric relationship in Eq. (1) is fulfilled:

$$\frac{A_i}{A_j} = \frac{p_j^2}{p_i^2}, \tag{1}$$

where  $A$  and  $p$  denote the triangles perimeters and areas successively, whereas  $i$  and  $j$  are the designations of the two triangles. To develop the triangle similarity proportion (TSP), Eq. (2) is used. The TSP is used to quantify the degree of similarity between every two triangles, and it is defined as follows:

$$\text{TSP} = \frac{A_i \times p_j^2}{A_j \times p_i^2}, \tag{2}$$

where TSP denotes the triangle similarity proportion and  $i$  and  $j$  are positive numbers representing the triangles indexes such as,  $i, j = 1, \dots, 6$ .

A statistical analysis was carried out over the collected data from the dataset. The dataset consists of the feature vectors created by using our proposed geometric model. The results of the statistical analysis revealed that there is a considerable distinction in the data related to each different subject. The similarity proportion between every two triangles is calculated using the TSP. A set of thirty TSP relationships are calculated for all of the sample images collected from the FG-NET database and is represented by Eqs. (3)–(32) as illustrated in Table 1.

Table 1. The TSP equations.

|  |      |  |      |  |      |
|--|------|--|------|--|------|
| $(T_1, T_2) = \frac{(A_1 * p_2^2)}{(A_2 * p_1^2)}$ | (3)  | $(T_1, T_3) = \frac{(A_1 * p_3^2)}{(A_3 * p_1^2)}$ | (4)  | $(T_1, T_4) = \frac{(A_1 * p_4^2)}{(A_4 * p_1^2)}$ | (5)  |
| $(T_1, T_5) = \frac{(A_1 * p_5^2)}{(A_5 * p_1^2)}$ | (6)  | $(T_1, T_6) = \frac{(A_1 * p_6^2)}{(A_6 * p_1^2)}$ | (7)  | $(T_2, T_3) = \frac{(A_2 * p_3^2)}{(A_3 * p_2^2)}$ | (8)  |
| $(T_2, T_4) = \frac{(A_2 * p_4^2)}{(A_4 * p_2^2)}$ | (9)  | $(T_3, T_6) = \frac{(A_3 * p_6^2)}{(A_6 * p_3^2)}$ | (10) | $(T_2, T_5) = \frac{(A_2 * p_5^2)}{(A_5 * p_2^2)}$ | (11) |
| $(T_2, T_6) = \frac{(A_2 * p_6^2)}{(A_6 * p_2^2)}$ | (12) | $(T_3, T_4) = \frac{(A_3 * p_4^2)}{(A_4 * p_3^2)}$ | (13) | $(T_3, T_5) = \frac{(A_3 * p_5^2)}{(A_5 * p_3^2)}$ | (14) |
| $(T_4, T_5) = \frac{(A_4 * p_5^2)}{(A_5 * p_4^2)}$ | (15) | $(T_4, T_6) = \frac{(A_4 * p_6^2)}{(A_6 * p_4^2)}$ | (16) | $(T_5, T_6) = \frac{(A_5 * p_6^2)}{(A_6 * p_5^2)}$ | (17) |
| $(T_2, T_1) = \frac{(A_2 * p_1^2)}{(A_1 * p_2^2)}$ | (18) | $(T_3, T_1) = \frac{(A_3 * p_1^2)}{(A_1 * p_3^2)}$ | (19) | $(T_4, T_1) = \frac{(A_4 * p_1^2)}{(A_1 * p_4^2)}$ | (20) |
| $(T_5, T_1) = \frac{(A_5 * p_1^2)}{(A_1 * p_5^2)}$ | (21) | $(T_6, T_1) = \frac{(A_6 * p_1^2)}{(A_1 * p_6^2)}$ | (22) | $(T_3, T_2) = \frac{(A_3 * p_2^2)}{(A_2 * p_3^2)}$ | (23) |
| $(T_4, T_2) = \frac{(A_4 * p_2^2)}{(A_2 * p_4^2)}$ | (24) | $(T_3, T_6) = \frac{(A_3 * p_6^2)}{(A_6 * p_3^2)}$ | (25) | $(T_5, T_2) = \frac{(A_5 * p_2^2)}{(A_2 * p_5^2)}$ | (26) |
| $(T_6, T_2) = \frac{(A_6 * p_2^2)}{(A_2 * p_6^2)}$ | (27) | $(T_4, T_3) = \frac{(A_4 * p_3^2)}{(A_3 * p_4^2)}$ | (28) | $(T_5, T_3) = \frac{(A_5 * p_3^2)}{(A_3 * p_5^2)}$ | (29) |
| $(T_5, T_4) = \frac{(A_5 * p_4^2)}{(A_4 * p_5^2)}$ | (30) | $(T_6, T_4) = \frac{(A_6 * p_4^2)}{(A_4 * p_6^2)}$ | (31) | $(T_6, T_5) = \frac{(A_6 * p_5^2)}{(A_5 * p_6^2)}$ | (32) |

The output of the TSP relationships is used to build a row feature vector for each sample image in the dataset. The set of row features vectors related to the same subject is used to build a class for this particular subject. Since the FG-NET database consists of 82 subjects, a dataset of 82 classes is created.

### 3. Feature Selection, Classification and Data Normalization Techniques

In this section, the first discussion is on the main concept of data pre-processing and its applications. Data pre-processing is a first step for several data mining applications.<sup>30</sup> Then, a number of data pre-processing techniques including normalization, scaling and standardization are discussed. This is followed by a brief description of the feature selection and classification algorithms adopted in the experimental part which is given in terms of operating parameters. Finally, the FG-NET face aging database which is used for validation in the experimental part is described in details.

#### 3.1. Data normalization and standardization

Data normalization usually attempts to give all the features the same weight. Normalization is ideal for classification algorithms, including neural networks as well as distance metrics, for example nearest-neighbor classification and clustering. In distance-based classification methods, normalization assists in avoiding features having massive ranges surpassing features having smaller ranges. Additionally, normalization is beneficial when there is no prior perception of the data.

Typically, data sets are required to be normalized or sometimes standardized to make the entire variables in proportion with each other. The most popular techniques for feature normalization are:

- Basic rescaling.
- Feature standardization (zero-mean and unit variance for every single feature within the dataset).

With basic rescaling, the main objective is to rescale the data along each data dimension to ensure that the final data vectors are within the range  $[0, 1]$  or  $[-1, 1]$  (depending on the dataset). This is often ideal for later processing since several default parameters (e.g. epsilon in PCA-whitening) deal with the data as if it has been scaled to some standard range.<sup>31</sup>

Feature standardization is the process of setting each dimension of the data (independently) to have zero-mean and unit-variance. This is actually the most popular approach for data normalization and is often used by supervised learning algorithms (e.g. with SVMs, feature standardization is usually recommended as an effective preprocessing stage). In practice, feature standardization requires calculating the mean for each dimension (throughout the dataset) and then subtracting the mean from each data element in the same dimension. Subsequently, each dimension is divided by its standard deviation.<sup>32</sup>

### **3.2. Feature selection and classification algorithms optimum parameters setup**

A number of common classification algorithms are adopted in our classification algorithms, namely: K-nearest neighbor (KNN),<sup>33</sup> Naïve Bayes,<sup>34</sup> and SVMs.<sup>35</sup> For feature selection, a type of regularized trees feature selection approach known as random forest feature selection is adopted.<sup>36</sup>

In our experiments, we adopt the one nearest neighbor (1NN) techniques to perform nearest neighbor classification over the dataset. The 1NN classifier that assigns a query image to the class of its closest neighbor in the feature space is the most intuitive nearest neighbor type classifier. As the size of training data set approaches infinity, the 1NN classifier guarantees an error rate of no worse than twice the Bayes error rate (the minimum achievable error rate given the distribution of the data). The SVMs parameters considered in our experiments, the number of categories ( $n$ ), the adopted kernel function, and the accepted error of the SVM ( $\epsilon$ ). The optimum values of the parameters are consistent with those used in Ref. 37 such that,  $n = 82$ , linear kernel and  $\epsilon = 0.1$ . For Naïve Bayes classifier, two attributes were selected. The first attribute enables the use of kernel density estimator rather than normal distribution for numeric attributes and the second attributes enable the use of supervised discretization to process numeric attributes for optimum classification performance.<sup>38</sup>

For feature selection the well-known random forest feature selection algorithm is adopted. Random forest is a classification method which possesses outstanding efficiency. Random Forest was introduced first by Leo Breiman.<sup>36</sup> From a practical point of view, it can be defined as a classifier which combines a collection of classification trees inside an integrated scheme to quantify the significance of every feature.<sup>36</sup> The output of the random forest feature selection algorithm is often visualized by the “Gini importance”<sup>39</sup> which can also be used as an indicator for feature relevance. Such feature importance’s score (Gini importance) offers a relative ranking for the set of features. Random Forest has a tuning parameter,  $mtry$ , which is the number of variables randomly sampled at each node to be considered for splitting. The parameter  $mtry$  is usually chosen to be some function of  $p$ , where  $p$  is the number of features in the input data. In our experiments, cross-validation is used to optimize the choice of the function  $mtry = f(p)$  and the parameter is defined as:  $mtry = p/2$ .

### 3.3. The FG-NET face aging database

The FG-NET database<sup>25</sup> is a publicly available database that has been widely used for evaluating face recognition across aging algorithms. The database has facial images collected at ages in the range from 0 to 69. The study used the FG-NET database in the experiments for validation because it is by far the largest face aging database that covers such a wide age range. Moreover, the FG-NET database provides annotated facial landmarks, as well as age information for each image. In this database, there are 1002 images of 82 subjects.

The age range distribution of the face images in the FG-NET face aging database is illustrated in Table 2. The table shows the percentages occupied by each age span from the total number of images in the database.

About 64% of the images were from children (with ages  $< 18$ ) and around 36% of the images were from the adults (with ages  $\geq 18$ ). For each facial image, there were 68 hand labeled landmarks representing the facial shape. The distribution of the number of images and subjects in different age ranges is shown in Table 3.

Table 2. Age spans distribution in FG-NET database.

| Age Range | FG-NET (%) |
|-----------|------------|
| 0–9       | 37.03      |
| 10–19     | 33.83      |
| 20–29     | 14.37      |
| 30–39     | 7.88       |
| 40–49     | 4.59       |
| 50–59     | 1.5        |
| 60–69     | 0.8        |
| 70–77     | 0          |

Table 3. The distribution of the number of images and subjects in different age ranges.

| Age Range     | 0–5 | 6–10 | 11–15 | 16–20 | 21–30 | 31–40 | 41–50 | 51–60 | 61–70 |
|---------------|-----|------|-------|-------|-------|-------|-------|-------|-------|
| # of Images   | 233 | 178  | 164   | 155   | 143   | 69    | 39    | 14    | 7     |
| # of Subjects | 75  | 70   | 71    | 68    | 84    | 35    | 22    | 8     | 4     |

All the face images in the FG-NET database were properly normalized and pre-processed. The pre-processing stage comprised converting the color input images into 8-bit gray-scale images and normalizing the face images photometrically by eliminating their mean and scaling their pixels to unit variance.<sup>40</sup> Finally, pose correction was performed to the non-frontal face images using the same method adopted by Mahalingam *et al.*<sup>15</sup> Such method uses AAM technique as proposed by Cootes *et al.*<sup>41</sup> As mentioned before, the images from the FG-NET database are annotated with 68 facial feature points, consequently, a generic model is used to fit these points and calculate the pose of the face.

Finally, pose correction is performed by warping the image onto the model. A number of pose corrected sample images from the FG-NET database are illustrated in Fig. 5 where the images in the top row are the original images and the images in the bottom row are the pose corrected images.



Fig. 5. Pose correction using AAM. (a) Original image, (b) posecorrected image.

## 4. Experiments

The experiments are conducted over the publicly available (FG-NET) face aging database.<sup>25</sup> For classification, we used WEKA built in classifiers (WEKA serves as a set of machine learning algorithms pertaining to data mining tasks).<sup>42</sup>

The experiments are conducted in two phases. In the first phase of the experiments, we investigate the significance of the proposed geometric features in different age groups. The main goal of this part of the experiments is to show which features contribute the most in each age group. For this purpose, the FG-NET database is divided into three subsets. The three subsets represent childhood, teenage, and adulthood age groups successively. In the second part of the experiments, a number of classification algorithms namely: KNN, Naïve Bayes and SVMs classifiers are used to estimate and validate the performance of the system. In the next section, the feature selection and classification experiments are discussed.

### 4.1. Features selection experiments

The FG-NET database is divided into three subsets based on the protocol adopted in Ref. 15 for validating face recognition systems over the FG-NET database. The three subsets are as follows:

- FG-NET-8 consists of all the data collected at the ages between 0 and 8. It includes 290 facial images from 74 subjects, in which 580 intra-person pairs and 6000 inter-person pairs are randomly generated for verification.
- FG-NET-18 consists of all the data collected at the ages between 8 and 18. It includes 311 facial images from 79 subjects, in which 577 intra-person pairs and 6000 inter-person pairs are randomly generated for verification.
- FG-NET-adult consists of all the data collected at the ages 18–69 and mainly frontal view. It includes 272 images from 62 subjects, in which 665 interpersonal pairs and about 6000 intra-personal pairs are randomly generated for verification.

This protocol of dividing the FG-NET database is used to determine which features contribute the most in discriminating between the subjects' classes during each age group. To accomplish this, Random Forest algorithm is used to quantify the importance of each feature and to select a subset of the most significant features among the thirty extracted features for each age group.

### 4.2. Classification experiments

In this part of the experiments, classification is conducted using three classifiers. The ability of each of the three classifiers in discriminating between the classes (which are built using the proposed geometric features) is evaluated. In pattern recognition, such procedure is beneficial in deciding which classifier is the most appropriate for classifying particular data patterns.<sup>43</sup> As mentioned above, classification was conducted in the WEKA platform by using three supervised learning classification

algorithms namely: KNN, Naïve Bayes and SVMs. The performance of the three classifiers was evaluated in terms of classification accuracy, FP rate, precision, recall and F-measure. This part of the experiments is conducted in four phases. In each phase, a different data pre-processing method is used.

In the first phase, feature vectors are scaled by using the min–max method. In the second phase, the data is normalized by using Z-score method where the normalization is applied by using the global standard deviation and mean of the whole dataset. Alternatively, the Z-score normalization is performed in the third phase over each column independently where it requires the normalization of each column of the dataset by using their individual standard deviation and mean values. Finally, the feature vectors are scaled by using the unit length rule in the fourth phase. Different pre-processing techniques are considered to evaluate their effect in enhancing the classification accuracy for each of the three classifiers. All of the above preprocessing and classification methods have been tested for performance analysis by using 10-fold cross validation. In 10-fold cross-validation, the data is first partitioned into  $k$  equally (or nearly equally) sized segments or folds. Subsequently, 10 iterations of training and validation are performed in the manners where within each iteration, a different fold of the data is held-out for validation while the remaining nine folds are used for learning.

## 5. Results and Discussion

The results of applying the random forest feature selection algorithm over the three subsets of the FG-NET database are introduced in this section. The importance of the 30-feature vectors extracted from each sample image is quantified by using the Gini index. The higher the numerical value of the Gini index, the greater the significance of the feature. Such factor is significant as it shows how each feature is contributing in discriminating between the dataset classes.

Graphs showing the average value of Gini index for each feature in the subsets are illustrated. The  $Y$  axis represents the average Gini index value related to each feature, whereas the  $X$  axis represents the set of features which are given the designation  $F_i$ . The index of the features is represented by  $i$  which is a positive number such that  $i = 1, \dots, 30$ . For example,  $F_1$  represents the first feature in the data set.

Subsequently, the classification results over the entire FG-NET database using different pre-processing methods are illustrated and discussed. The results of combining each pre-processing method with each classifier are reported.

Finally, the classification results for each FG-NET subset are illustrated in terms of classification accuracy and FP rate. The data is pre-processed by using Z-score normalization since it yielded the best results compared with other pre-processing methods. Moreover, feature selection is performed over the normalized data by using random forest feature selection method.

### 5.1. Feature selection results

This section shows the results of feature selection that used the random forest algorithm. First, the features selection results for the entire FG-NET database are illustrated by using graphs. In the graphs, the most important feature among all other features is highlighted in black to distinguish it from the rest of the features. As mentioned before, the significance of the feature is quantified by using the average value of the Gini index related to each feature.

The results of applying the random forest feature selection algorithm over the FG-NET-8 subset are illustrated in the second part of this section. Subsequently, the feature selection results of the FGN-ET-18 and FGN-ET-Adults subsets are illustrated successively.

#### 5.1.1. FG-NET-8 feature selection results

This section discusses the features ranking results for the FG-NET-8 subset. This subset represents the age group between 0 to 8 years. All the sample images in this subset are for the subjects' childhood stage. Features selection results for the FG-NET-8 subset are illustrated in Fig. 6.

The result of the random forest feature selection algorithm shows the top five features for this subset are respectively as follows: Feature 18, Feature 20, Feature 28, Feature 1 and Feature 16. The features ranking of the FG-NET-8 subset shows that the most important feature is Feature 18 which is produced by using the following Eq. (33):

$$F_{18} = (Tr_4, Tr_1) = \frac{(Ar_4 * pr_1^2)}{(Ar_1 * pr_4^2)} \quad (33)$$

Feature 18 ( $F_{18}$ ) represents the TSP between triangle 4 ( $T_4$ ) and triangle 1 ( $T_1$ ) as illustrated in Fig. 7.

As shown in Fig. 7,  $F_{18}$  incorporates the distances between the inner corners of the eyes and the tip of the nose. Moreover, it includes the distances between the outer

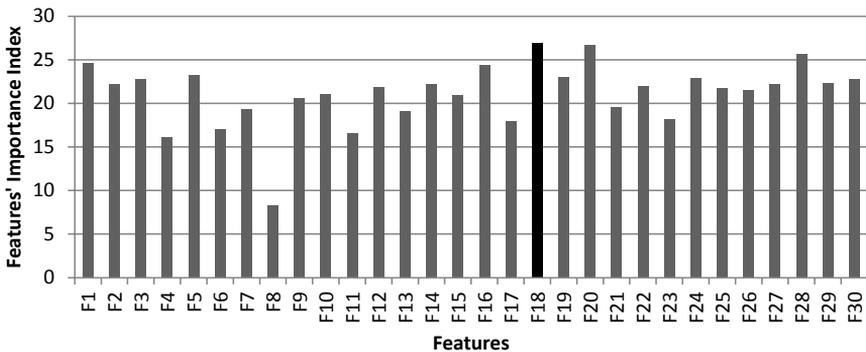


Fig. 6. FG-NET-8 feature selection results using random forest.

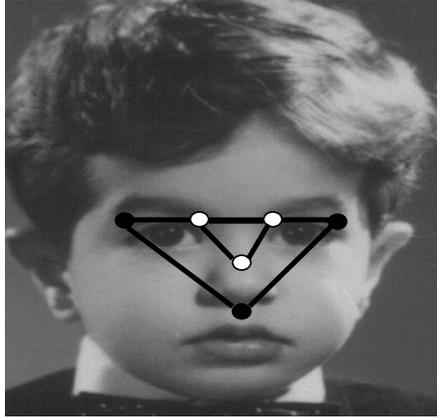


Fig. 7. Feature 18 (F18) triangle combination.

corners of the eyes and the center of the mouth. This indicates that the ratio between those two regions is very important in discriminating between one subject and another during childhood.

#### 5.1.2. FG-NET-18 feature selection results

In Fig. 8, the ranking of the features for FG-NET-18 subset is illustrated. The top five features are successively: Feature 1, Feature 20, Feature 3, Feature 18 and Feature 19. The most important feature is Feature 1 ( $F_1$ ) which is represented by Eq. (34):

$$F_1 = (T_1, T_2) = \frac{(A_1 * p_2^2)}{(A_2 * p_1^2)}. \quad (34)$$

It can be seen from Fig. 9 that the most important feature in the FG-NET-18 subset includes the distances between the inner corners of the eyes and the tip of the nose. It also includes the distances between the center point between the eyes and the

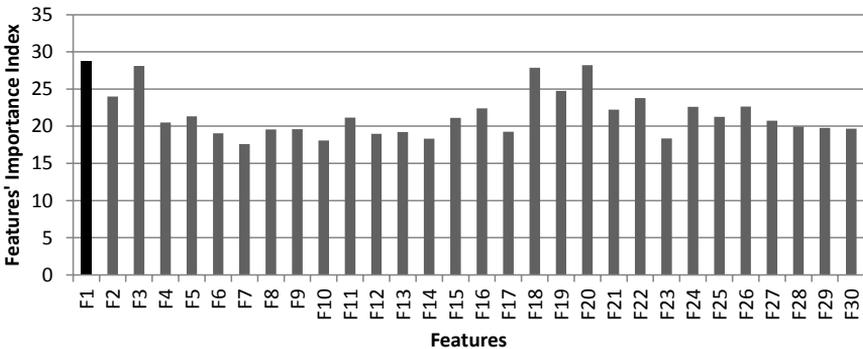


Fig. 8. FG-NET-18 feature selection result using random forest.

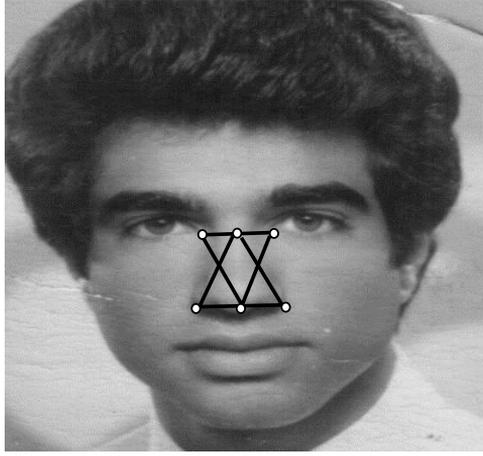


Fig. 9. Feature 1 (F1) triangles combination.

corners of the nose. This indicates that the ratio between those two regions is very important in discriminating between one subject and another during teenage period.

### 5.1.3. FG-NET-adult feature selection results

The result from the random forest feature selection method shows the top five most important features in the FG-NET-adult subset are successively as follows: Feature 1, Feature 11, Feature 2, Feature 13 and Feature 24. This is illustrated in Fig. 10. The most important feature is Feature 1 which is represented by Eq. (34). Although the set of the top five features for the FG-NET-18 and FG-NET-Adult subsets are different, they share the same most important feature ( $F_1$ ). Such result indicates that the similarity proportion between  $T_1$  and  $T_2$  is the most discriminating feature in the age group (18–69).

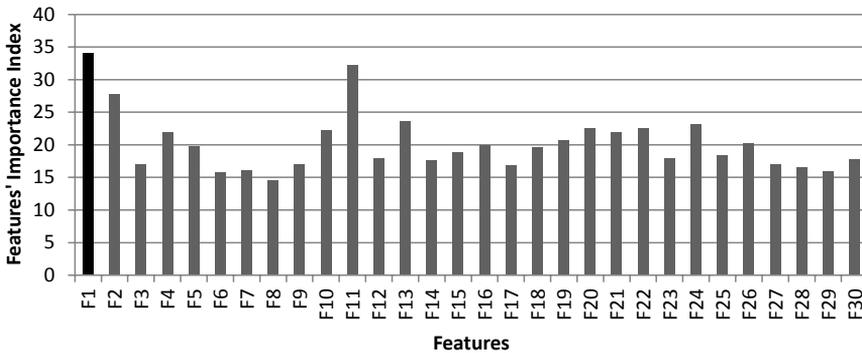


Fig. 10. FG-NET-adult feature selection result using random forest.

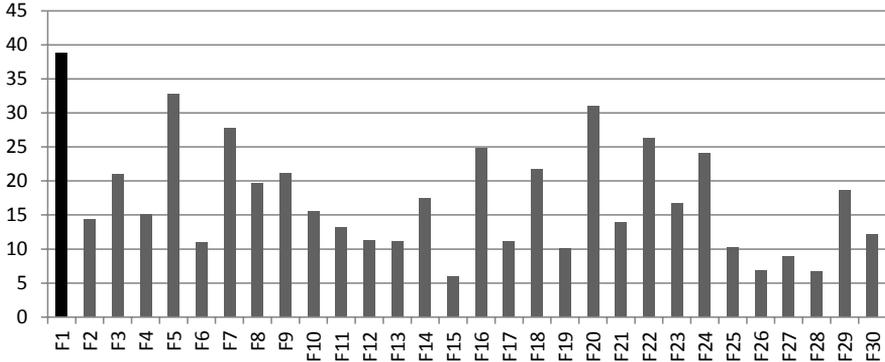


Fig. 11. FG-NET complete set feature selection using random forest.

#### 5.1.4. FG-NET complete set feature selection results

The ranking chart of the thirty developed features for the entire dataset is illustrated in Fig. 11. Features ranking was performed by using the random forest variable importance algorithm. The top five features are successively as follows: Feature 1, Feature 15, Feature 18, Feature 19 and Feature 24.

The most important feature amongst all the age groups was Feature 1 ( $F1$ ) which is represented by Eq. (34). Feature 1 is selected from the total set of features related to the entire database. The database subject of the experiments comprises several age groups. This means that  $F1$  is the most significant feature in our model for different age span.

The values of the most significant features acquired from different sample images related to the same subject vary. But, the rate in which the features' values change due to facial aging is constant for each subject. This is very important in discriminating between the different classes related to different subjects.

The results of applying random forest feature selection methods over the different subsets clearly show that for each age group, there are specific facial features that contribute the most in discriminating between the different subjects. In the subsequent phases of experiments, only a set of the 10 most important features for each FG-NET subset will be considered. The optimized number of the most important features was determined empirically where the best results were achieved using only 10 features for each age group subset.

## 5.2. Classification results

In this section, we discuss the classification performance of the proposed system over the three FG-NET subsets (i.e. FG-NET-8, FG-NET-18 and FG-NET-Adult). As mentioned before, in this set of experiments, three classification algorithms are used in the classification experiments namely, KNN, Naïve Bayes and SVM. Different methods were used for pre-processing the feature vectors. Accordingly, a separate

Table 4. Classification results using min–max scaling for preprocessing.

| Classifier                  | KNN          |               |                  | Naïve Bayes  |               |                  | SVM          |               |                  |
|-----------------------------|--------------|---------------|------------------|--------------|---------------|------------------|--------------|---------------|------------------|
|                             | FG-<br>NET-8 | FG-<br>NET-18 | FG-<br>NET-Adult | FG-<br>NET-8 | FG-<br>NET-18 | FG-NET-<br>Adult | FG-<br>NET-8 | FG-<br>NET-18 | FG-NET-<br>Adult |
| Classification accuracy (%) | 55.50        | 92.28         | 95.34            | 42.90        | 88.33         | 86.95            | 17.40        | 47.00         | 56.17            |
| FP rate                     | 0.006        | 0.002         | 0.002            | 0.011        | 0.002         | 0.002            | 0.017        | 0.012         | 0.009            |
| Precision                   | 0.559        | 0.928         | 0.957            | 0.453        | 0.895         | 0.885            | 0.258        | 0.498         | 0.487            |
| Recall                      | 0.556        | 0.926         | 0.953            | 0.434        | 0.892         | 0.87             | 0.262        | 0.479         | 0.562            |
| F-measure                   | 0.551        | 0.926         | 0.954            | 0.435        | 0.891         | 0.874            | 0.248        | 0.480         | 0.499            |

table is dedicated for each single pre-processing method results. In all the tables, the classification results are introduced in terms of overall classification accuracy, FP rate, precision, recall and F-measure. The classification results of using the min–max scaling method for pre-processing the features vectors are illustrated in Table 4. The features vectors were scaled by using the min–max method and then passed to each of the three classifiers separately. The performance of each classifier is evaluated over each of the FG-NET subsets separately.

The classification results of the three classifiers show that recognition is difficult in small children (FG-NET-8) age group. In general, all the classifiers performed better over both FG-NET-18 and FG-NET-adult subsets. The KNN classifier yielded the best classification performance over the three subsets. However, its performance degraded by about 40% when tested over the FG-NET-8 subset. The performance of the KNN classifier over the FG-NET-18 and FG-NET-Adults subsets was relatively comparable. The same observations hold for the Naïve Bayes and SVM classifiers.

The best performance in terms of classification accuracy, FP rate, precision, recall and F-measure was achieved by the KNN classifier over the FG-NET-Adult subset. All the classification metrics agree on the relative ranking of the results, where the KNN achieved the best performance, the Naïve Bayes achieved a reasonable overall classification performance, followed by the SVM classifier.

In the next part of the experiments, the features vectors were normalized by using Z-score method. The goal of using such pre-processing method is to eliminate the unit of measurements by transforming the data into new scores with a mean of 0 and a standard deviation of 1. The classification results of using the Z-score for normalizing the dataset are illustrated in Table 5.

Table 5 shows that the ranking of the classifiers performances is the same as when min–max scaling was used for pre-processing. Again the KNN classifier evidently performed better than the other classifiers. Using Z-score normalization, the classification accuracy of the KNN classifier was increased to above 97% over the FG-NET-Adult subset and above 96% over the FG-NET-18 subset. Moreover, the KNN classifier achieved a relatively high accuracy over the FG-NET-8 subset. Such results

Table 5. Classification results using Z-score normalization.

| Classifier                  | KNN      |           |              | Naïve Bayes |           |              | SVM      |           |              |
|-----------------------------|----------|-----------|--------------|-------------|-----------|--------------|----------|-----------|--------------|
|                             | FG-NET-8 | FG-NET-18 | FG-NET-Adult | FG-NET-8    | FG-NET-18 | FG-NET-Adult | FG-NET-8 | FG-NET-18 | FG-NET-Adult |
| Classification accuracy (%) | 88.33    | 96.67     | 97.56        | 47.30       | 78.21     | 79.91        | 15.94    | 45.20     | 48.17        |
| FP rate                     | 0.002    | 0.001     | 0            | 0.013       | 0.003     | 0.004        | 0.018    | 0.012     | 0.011        |
| Precision                   | 0.886    | 0.966     | 0.977        | 0.497       | 0.785     | 0.804        | 0.181    | 0.476     | 0.506        |
| Recall                      | 0.882    | 0.962     | 0.976        | 0.473       | 0.780     | 0.799        | 0.182    | 0.457     | 0.482        |
| F-measure                   | 0.882    | 0.962     | 0.976        | 0.479       | 0.778     | 0.799        | 0.180    | 0.458     | 0.482        |

show that the KNN classifier was able to preserve adequate performance over the three subsets.

It can be observed from the classification results reported in Table 2 that the performance of both Naïve Bayes and SVM classifiers over the three subsets degraded when Z-score normalization was used for pre-processing the features vectors. However, the classification results related to the Naïve Bayes and SVM classifiers are in line with the results presented Table 1, where the performance of the two classifiers over the FG-NET-8 is low compared to their performance over the FG-NET-Adult and FGN-ET-18 subsets.

The classification results of using the column vector normalization for pre-processing are illustrated in Table 6. Once again the KNN classifier yielded an adequate overall classification performance over the three FG-NET subsets. However, its performance degraded by a relatively small margin compared to its performance when the Z-score was used for preprocessing the feature vectors. The KNN classifier achieved a maximum accuracy of above 94% over the FG-NET-Adult compared to a maximum accuracy of above 97% in the previous phase.

The classification results of both Naïve Bayes and SVM classifiers improved noticeably during this phase of the classification experiments compared to the two previous phases. In particular, The Naïve Bayes classifier achieved high accuracies of above 91% over both of the FG-NET-Adult and FG-NET-18 subsets. Yet, its

Table 6. Classification results using column vector normalization.

| Classifier                  | KNN      |           |              | Naïve Bayes |           |              | SVM      |           |              |
|-----------------------------|----------|-----------|--------------|-------------|-----------|--------------|----------|-----------|--------------|
|                             | FG-NET-8 | FG-NET-18 | FG-NET-Adult | FG-NET-8    | FG-NET-18 | FG-NET-Adult | FG-NET-8 | FG-NET-18 | FG-NET-Adult |
| Classification accuracy (%) | 79.07    | 93.33     | 94.26        | 44.50       | 91.03     | 91.8         | 24.60    | 70.30     | 72.95        |
| FP rate                     | 0.004    | 0.001     | 0.019        | 0.014       | 0.001     | 0.027        | 0.019    | 0.009     | 0.092        |
| Precision                   | 0.790    | 0.921     | 0.944        | 0.518       | 0.909     | 0.919        | 0.166    | 0.707     | 0.835        |
| Recall                      | 0.787    | 0.921     | 0.943        | 0.450       | 0.907     | 0.918        | 0.162    | 0.703     | 0.73         |
| F-measure                   | 0.787    | 0.919     | 0.942        | 0.445       | 0.905     | 0.918        | 0.147    | 0.70      | 0.656        |

Table 7. Classification results using scaling to unit length.

| Classifier                  | KNN          |               |                  | Naïve Bayes  |               |                  | SVM          |               |                  |
|-----------------------------|--------------|---------------|------------------|--------------|---------------|------------------|--------------|---------------|------------------|
|                             | FG-<br>NET-8 | FG-<br>NET-18 | FG-NET-<br>Adult | FG-<br>NET-8 | FG-<br>NET-18 | FG-NET-<br>Adult | FG-<br>NET-8 | FG-<br>NET-18 | FG-NET-<br>Adult |
| Classification accuracy (%) | 34.10        | 77.45         | 83.48            | 29.70        | 75.24         | 76.89            | 37.33        | 83.37         | 86.22            |
| FP rate                     | 0.015        | 0.003         | 0.002            | 0.015        | 0.005         | 0.003            | 0.013        | 0.002         | 0.002            |
| Precision                   | 0.344        | 0.781         | 0.849            | 0.215        | 0.757         | 0.804            | 0.383        | 0.828         | 0.884            |
| Recall                      | 0.340        | 0.773         | 0.835            | 0.219        | 0.750         | 0.769            | 0.382        | 0.824         | 0.862            |
| F-measure                   | 0.337        | 0.771         | 0.837            | 0.205        | 0.750         | 0.777            | 0.377        | 0.822         | 0.856            |

performance degraded rapidly when it was tested over the FG-NET-8 subset. The same observation holds for the SVM classifier, where its performance degraded by about 66% when it was tested over the FG-NET-8 subset.

Table 7 illustrates the classification results for the three classifiers over the FG-NET subsets when the features vectors were scaled by using the unit length scaling method. At variance of the classification results reported in the three previous classification phases, the SVM classifier achieved the best classification results over all the FG-NET subsets. It was able to achieve a maximum classification accuracy of more than 86% over the FG-NET-Adult subset and more than 83% over the FG-NET-18 subset. Yet, it was not able to maintain the same performance over the FG-NET-8 subset, where its classification accuracy dropped to 37%.

The performance of the KNN and Naïve Bayes classifiers over the FG-NET-Adult and FG-NET-18 degraded compared to their performance in the previous phase when the column vector normalization was used for preprocessing. Moreover, their performance degraded when they were tested over the FG-NET-8 subset.

The best classification results for all the four phases can be summarized as follows:

- The best classification performance over the three FG-NET subsets for the normalization of the features vectors by using the Z-score method was achieved by the KNN classifier. The KNN classifier was able to achieve a maximum accuracy of more than 97%.
- Although KNN classifiers was able to achieve overall classification accuracies of more than 90% when combined with different pre-processing methods, its performance degraded when the feature vectors were scaled to unit length prior to classification.
- The KNN classifier was able to preserve adequate performance over all of the FG-NET subsets when the Z-score and the column vector normalization methods were used for preprocessing. Nevertheless, its performance degraded rapidly over the FG-NET-8 subset when min-max scaling and scaling to unit length methods were used for preprocessing.

- The Naïve Bayes classifier achieved the best classification results over the FG-NET subsets when the column vector normalization method was used for preprocessing. It was able to achieve a maximum classification accuracy of more than 91%.
- The SVM classifier performed better than the other two classifiers over the FG-NET subsets when it was combined with the Unit length preprocessing method. It achieved a maximum classification accuracy of more than 86%.
- Both the Naïve Bayes and SVM classifiers yielded low classification performance over the FG-NET-8 subset compared to their performance over the FG-NET-Adult and FG-NET-18 subsets during all the classification phases.
- The proposed system maintained adequate and stable performance over all of the age groups (represented by the FG-NET subsets) when a combination of the KNN classifier and  $Z$ -score normalization was used.

Based on the above summary, it can be concluded that the KNN classifier and  $Z$ -score normalization method are the optimum preprocessing-classifier combination, thus to be used in conjunction with the proposed model. Accordingly, in all of the subsequent experiments, the proposed system feature vectors will be preprocessed by using  $Z$ -score normalization and the 1NN classifier will be used as the main classifier.

### 5.3. *Experiments comparing the performance of the proposed system with state-of-the-art face recognition systems and state-of-the-art age-invariant face recognition systems*

In the third set of the experiments, the performance of the proposed system (operating in the verification mode) is compared with that of some of the popular state-of-the-art face recognition methods including the local binary patterns (LBPs)<sup>44</sup> and Gabor<sup>45</sup> over the entire FG-NET database.

The Gabor filter (Gabor wavelet) represents a band-pass linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Thus, a dimensional Gabor filter constitutes a complex sinusoidal plane of particular frequency and orientation modulated by a Gaussian envelope.<sup>46</sup> It achieves an optimal resolution in both spatial and frequency domains. The optimum values for the Gabor filter bank parameters were determined empirically to have the following values: wave length = 8, orientations degree = 30, 90, 150, phase offsets degree = 0, 90, aspect ratio = 0.5, bandwidth = 1. The number of orientations and number of scales for a Gabor filter bank determine the number of desired features according to the expression (orientation  $\times$  scale). In our experiments, we considered 8 orientations and 5 scales as parameters for Gabor method (the popular Gabor parameters, 5 scales  $\times$  8 orientations, have been assumed to be the best choice in many studies<sup>45,47</sup>), which yields a feature vector of 40 relevant features for each image.

The LBP $_{P,R}$  is defined on a circular neighborhood allowing changes in the number of neighbors ( $P$ ) and the radius size ( $R$ ). However, increasing the number of

neighbors increases the information redundancy and the computational cost, which is not always resulting in a more discriminant LBP label. Thus, in our experiments we use the minimum possible values for  $P$  and  $R$  (i.e.  $P = 8$  and  $R = 1$ ).

Furthermore, five of the published age-invariant face recognition systems, namely: Geng *et al.*,<sup>10</sup> Singh *et al.*,<sup>18</sup> Park *et al.*,<sup>48</sup> Ling *et al.*,<sup>6</sup> Li *et al.*,<sup>49</sup> were implemented and tested over the entire FG-NET face aging database. The state-of-the-art age-invariant face recognition systems that were selected to be part of the validation experiments, have produced relatively adequate results over the FG-NET as reported by the authors in the literature review.

All the algorithms and systems involved in this part of the experiments were implemented and tested upon the same platform (i.e. MATLAB R2013a) and the same system with the following specifications: AMD APU 2.10 GHz, 4.00 GB RAM and  $\times 64$ -based processor.

The verification tests were run by using the leave-one-person-out (LOPO) scheme that was adopted by Li *et al.*<sup>49</sup> and Park *et al.*<sup>48</sup> (most commonly used for evaluation over the FG-NET database). Recognition was then performed by using 1NN classification to acquire the verification rate for the proposed approach. In the LOPO scheme, the entire set of images of a single subject was used as test images and the rest of the images were used for training the system. Such procedure was performed for all the subjects to ensure that for testing, each subject was used only once. Also, the purpose of this evaluation scheme was to make sure that the images of one subject were not in the training and testing set at the same time. The LOPO was combined with 1NN technique in a way that every single image of the 1002 (82 subject  $\times$  12 images) images of the FG-NET database is used as a query image and some similarity measure is computed between each query and each of the remaining 1001 images. Then, the identity of the subject presented in the query image is recovered by using the 1NN match. The result of the classification is considered true positive, if the 1NN match corresponds to any of the remaining 11 images of that particular subject instead of to any of the 891 images related to the remaining 81 subjects.

Statistical 95% confidence intervals (CI) for the EER and the verification rate (VR) of each algorithm are obtained empirically by bootstrap sampling the outputs of each algorithm.<sup>50,51</sup> The criterion employed for establishing statistical differences in performance while comparing two algorithms is that if the observed quantity for either algorithm falls within the 95% CI of the other, then the performances of the two algorithms are regarded as not statistically significantly different. Otherwise, they are regarded as significantly different. The verification results for the proposed system, LBP, Gabor and a number of the state-of-the-art age-invariant face recognition systems are illustrated in Table 8.

The verification results show that the proposed system outperformed the LBP and Gabor methods by a large margin in terms of EER and VR, followed by face recognition using Gabor method and LBP respectively. However, the performance of

Table 8. Verification results of the proposed system, state-of-the-art face recognition methods, and state-of-the-art age-invariant face recognition systems over the entire FG-NET database.

| Algorithm                         | Verification Accuracy<br>(%) at FAR = 1% | CI            | EER (%) | CI            | Processing<br>Time (s) |
|-----------------------------------|--|---------------|---------|---------------|------------------------|
| LBP <sup>44</sup>                 | 25.00                                    | [22.70 27.20] | 37.58   | [35.28 39.78] | 10.74                  |
| Gabor <sup>45</sup>               | 35.33                                    | [33.03 37.53] | 34.22   | [31.42 38.52] | 45.35                  |
| Geng <i>et al.</i> <sup>10</sup>  | 38.80                                    | [35.03 42.27] | 32.80   | [30.00 37.10] | 74.00                  |
| Li <i>et al.</i> <sup>49</sup>    | 48.00                                    | [45.70 50.20] | 27.50   | [25.20 29.70] | 162.80                 |
| Park <i>et al.</i> <sup>48</sup>  | 54.60                                    | [50.23 57.47] | 21.00   | [18.70 23.20] | 399.00                 |
| Ling <i>et al.</i> <sup>6</sup>   | 77.00                                    | [74.00 79.86] | 12.50   | [11.50 13.40] | 127.68                 |
| Singh <i>et al.</i> <sup>18</sup> | 85.28                                    | [81.86 88.09] | 8.33    | [7.16 10.40]  | 90.70                  |
| Proposed method                   | 94.66                                    | [92.69 96.59] | 6.51    | [4.92 7.22]   | 13.75                  |

LBP surpassed the performance of the proposed system in terms of processing speed by a relatively small margin. Such result is acceptable since LBP is known to be computationally effective. On the other hand, the speed of the proposed system surpassed that of the Gabor method. Gabor-based methods suffer from high computational complexity associated with extracting multiple scales-orientations filters from a single image, which makes it incompetent for speed sensitive scenarios.<sup>45</sup> Nevertheless, the proposed system has reported a minimum processing time for a complete round of leave-one-person verification test compared to the other state of the art age-invariant face recognition systems. The system developed by Park *et al.*<sup>48</sup> has reported a maximum processing time for a whole round of leave-one-person verification test. This is due to the need for building a separate 3D morphable model to represent the aging texture variations and another 3D morphable model to represent the facial aging shape variations for each 2D image in the database. In addition, the need for fitting the two 3D morphable models to perform verification between a test and gallery image, had a significant effect in increasing the overall processing time. The system presented by Geng *et al.*<sup>10</sup> was able to achieve comparable results to Park *et al.*<sup>48</sup> in terms of verification accuracy, precision, recall, and error rates, yet the processing time was 80% less than the processing time reported by Park *et al.* The system developed by Geng *et al.*<sup>10</sup> was built by using an AAM and subspaces which are cost effective approaches in terms of computational complicity. Singh *et al.*<sup>18</sup> system has shown the best performance in terms of verification accuracy and equal error rate compared to the other age-invariant face recognition systems. Nevertheless, the computational complexity associated with the system has led to a processing time of more than 90s for a complete LOPO round due to the computational complexity associated with the 2D-log Gabor used the developed age transformation algorithm for validation. The system introduced by Ling *et al.*<sup>6</sup> has also shown an adequate verification performance with 77% verification accuracy and moderate error rate compared to that of Geng *et al.*,<sup>10</sup> Park *et al.*,<sup>48</sup> and Li *et al.*<sup>49</sup> In terms of computational complexity, the system operated in a low processing speed and reported a high processing time of more than 127s for a complete LOPO round. This is due to the high computational complexity associated with the pyramid of

gradient orientation descriptor which divides the image into a grid of cells and for each of them; it computes a histogram of gradients. Finally, the cells are grouped into regions called blocks. The system introduced by Li *et al.*<sup>49</sup> has also shown low performance in terms of processing time compared to Singh *et al.*<sup>18</sup> and Geng *et al.*<sup>10</sup> The system reported an average processing time of about 162s for a complete LOPO round. The high computational complexity associated with the system introduced by Li *et al.*<sup>49</sup> is due to the extensive preprocessing which involves building both shape and texture models from the 2D face image, simulating both shape and texture aging patterns in 3D space, and finally, matching the 3D model with a gallery image.

The proposed system was able to achieve the best performance in terms of verification accuracy, equal error rate and processing time compared to the state of the art age-invariant face recognition systems that are subjected to the verification test. However, the system performance in terms of computational complexity and processing time needs to be improved to meet the requirements of real-time applications and to expand the scope of the proposed system applications.

Several physiological studies have demonstrated that facial aging appears in different patterns in different age groups.<sup>3</sup> Whilst facial aging appears as shape variations in human faces due to the craniofacial growth from babyhood to the teenage years, it appears more commonly in the form of textural variations such as changes in skin color, wrinkles and other skin artefacts on mature faces.<sup>12</sup> Thus, all the systems and methods that were involved in the above verification experiments were tested over the FG-NET-8, FG-NET-18, and FG-NET-Adult separately. Such tests assist in evaluating the ability of the systems in handling the variations that result from different aging patterns in different age groups. The recognition accuracies achieved by each method when tested over each of the FG-NET subsets are shown in Fig. 12.

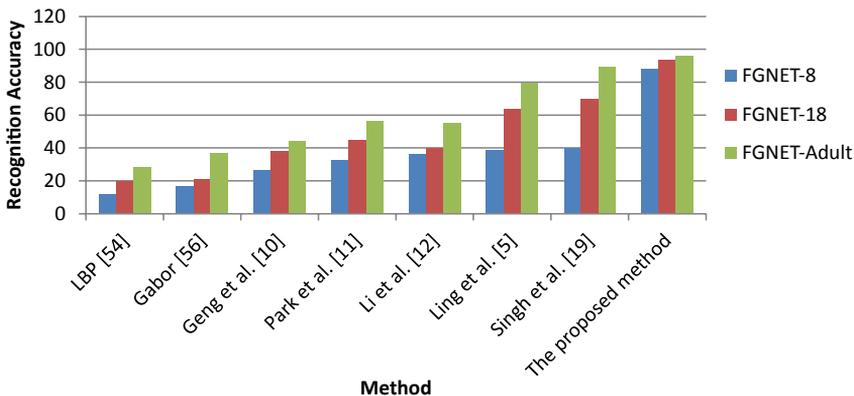


Fig. 12. Recognition accuracies reported by the proposed system, the different state-of-the-art face recognition and state-of-the-art age-invariant face recognition systems over the FG-NET subsets (FG-NET-8, FG-NET-18 and FG-NET-Adult).

The results depicted in Fig. 12 are in line with the previous results, the proposed system presents itself as performing evidently better than the other systems over all of the FG-NET subsets. The proposed system was able to achieve a maximum recognition accuracy of 95.68% over the FG-NET-Adult subset and 93.4% over the FG-NET-18 subset. Such results show that the proposed model is able to work effectively in teenage and adult age groups. However, the proposed system was not able to deliver similar remarkable performance over the FG-NET-8 subset. This can be explained by the fact that craniofacial growth occurs in a rapid rate during childhood than other age groups. This has a significant effect over the size of the facial features and the way facial features are correlated with each other. Such effects may have a large influence over the performance of the proposed system in which it attempts to model the changes in the size of the main facial features and the distances between them over the years. Nevertheless, the proposed system was able to maintain stable recognition performance over the different age groups, as it was able to achieve high recognition accuracies over the FG-NET-Adult and FG-NET-18 subsets and a relatively good accuracy over the FG-NET-8 subset.

The results achieved by the state-of-the-art face recognition methods and the state-of-the-art age-invariant face recognition systems are also in line with the LOPO verification results. The system introduced by Singh *et al.*<sup>19</sup> shows a maximum recognition accuracy of 89% over the FG-NET-Adult followed by Ling *et al.*<sup>6</sup> (79.66%), Park *et al.*<sup>48</sup> (56%), Li *et al.*<sup>49</sup> (55.4%), Geng *et al.*<sup>10</sup> (44%), Gabor<sup>45</sup> (37%) and LBP<sup>44</sup> (28.3).

In general, the performance of all the methods and systems subjected to the evaluation degraded rapidly when they were tested over the FG-NET-8 subset. Again, such results demonstrate that recognition is difficult in small children (FG-NET-8) age group.

## 6. Conclusions and Future Work

This work proposed a novel age-invariant facial recognition system based on a geometric approach. The proposed geometric model comprises triangular regions which surround and connect the main facial features (i.e. eyes, nose and mouth). A set of six triangular features were extracted from each sample's face image by connecting particular facial feature points. A total of twelve facial feature points which represent the main facial features corner and center points were considered in the proposed model. The coordinates of the facial features points are attached to the annotated version of the FG-NET database. Later, those coordinates were used to calculate the areas and perimeters of each of the triangular regions by using the standard Euclidean distance formula. Subsequently, a set of mathematical relationships called TSPs was developed to measure the degree of similarity between every two triangles.

A total of thirty features were created for each sample image in the FG-NET database by using the proposed TSPs mathematical relationships. The entire feature

vectors related to the same subject were used to form a class for each subject. Feature selection was performed by using the RF algorithm. Different set of the most significant features was produced for different age groups by using the RF algorithm. This indicates that there are different facial aging patterns for different age groups.

The proposed systems' performance was evaluated in terms of classification accuracy, FP rate, precision, recall and F-measure in the classification experiments. Different data pre-processing methods were used with different classifiers over the developed dataset. The system was tested over three FG-NET subsets representing childhood, teenage and adult age spans. A maximum classification accuracy of more than 97% was achieved when the Z-score normalization method was used for pre-processing the feature vectors and the 1NN classifier was used for classification over the FG-NET-Adult subset. The system preserved adequate classification performance over all the age groups. Moreover, the performance of the system in the verification mode was stable over all the age groups. The proposed system outperformed the LBP, Gabor and a number of the-state-of-the-art age-invariant face recognition systems with a maximum verification accuracy of more than 94% over the entire FG-NET database. The significance of the proposed geometric model exists in its ability of showing the unique craniofacial growth pattern for each subject. Additionally, such geometric model exhibits the way the main facial features are correlated with each other through different age spans. The system preserved simplicity since it does not require building aging models or any pre-processing step prior to performing the recognition task. Such system is promising towards real time applications that require high performance in addition to low processing time.

In our future work, we are planning to evaluate the proposed system using different feature selection methods in conjunction with multiple classification methods. Also, the system will be evaluated over the largest publicly available face aging database, namely MORPH face aging database, to test the ability of the system in handling large populations.

## References

1. R. Jafri and H. R. Arabnia, A survey of face recognition techniques, *JIPS* **5** (2009) 41–68.
2. H. Ling, S. Soatto, N. Ramanathan and D. W. Jacobs, A study of face recognition as people age, in *IEEE 11th Int. Conf. Computer Vision, 2007. ICCV 2007* (2007), pp. 1–8.
3. S. R. Coleman and R. Grover, The anatomy of the aging face: Volume loss and changes in 3-dimensional topography, *Aesthetic Surgery J.* **26** (2006) S4–S9.
4. C. Meng, J. Lu and Y.-P. Tan, A comparative study of age-invariant face recognition with different feature representations, in *2010 11th Int. Conf. Control Automation Robotics & Vision (ICARCV)* (2010), pp. 890–895.
5. Y. H. Kwon and N. da Vitoria Lobo, Age classification from facial images, in *Proc. CVPR'94., 1994 IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 1994* (1994), pp. 762–767.
6. H. Ling, S. Soatto, N. Ramanathan and D. W. Jacobs, Face verification across age progression using discriminative methods, *IEEE Trans. Inform. Forensics Security* **5** (2010) 82–91.

7. A. Lanitis, C. J. Taylor and T. F. Cootes, Toward automatic simulation of aging effects on face images, *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (2002) 442–455.
8. A. Lanitis, Facial biometric templates and aging: Problems and challenges for artificial intelligence, in *AIAI Workshops* (2009), pp. 142–149.
9. G. Xin, F. Yun and S.-M. Kate, Automatic facial age estimation, *Tutorial at PRICAI* (2010).
10. X. Geng, K. Smith-Miles and Z.-H. Zhou, Facial age estimation by nonlinear aging pattern subspace, in *Proc. 16th ACM Int. Conf. Multimedia* (2008), pp. 721–724.
11. G. Guo, G. Mu, Y. Fu and T. S. Huang, Human age estimation using bio-inspired features, in *IEEE Conf. Computer Vision and Pattern Recognition, 2009, CVPR 2009* (2009), pp. 112–119.
12. Y. Fu, G. Guo and T. S. Huang, Age synthesis and estimation via faces: A survey, *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (2010) 1955–1976.
13. N. Ramanathan and R. Chellappa, Modeling age progression in young faces, in *2006 IEEE Computer Society Conf. Computer Vision and Pattern Recognition* (2006), pp. 387–394.
14. A. Drygajlo, W. Li and K. Zhu, Q-stack aging model for face verification, in *Proc. 17th European Signal Processing Conference (EUSIPCO 2009)* (2009), pp. 3–4.
15. G. Mahalingam and C. Kambhamettu, Age invariant face recognition using graph matching, in *2010 Fourth IEEE Int. Conf. Biometrics: Theory Applications and Systems (BTAS)* (2010), pp. 1–7.
16. U. Park, Face recognition: Face in video, age invariance and facial marks, Michigan State University (2009).
17. J. Suo, S.-C. Zhu, S. Shan and X. Chen, A compositional and dynamic model for face aging, *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (2010) 385–401.
18. R. Singh, M. Vatsa, A. Noore and S. K. Singh, Age transformation for improving face recognition performance, in *Pattern Recognition and Machine Intelligence* (Springer, 2007), pp. 576–583.
19. B. Tiddeman, M. Burt and D. Perrett, Prototyping and transforming facial textures for perception research, *IEEE Comput. Graph. Appl.* **21** (2001) 42–50.
20. B. Tiddeman, M. Stirrat and D. I. Perrett, Towards realism in facial image transformation: Results of a wavelet MRF method, in *Computer Graphics Forum* (2005), pp. 449–456.
21. F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui and C. Y. Suen, Investigating age invariant face recognition based on periocular biometrics, in *2011 Int. Joint Conf. Biometrics (IJCB)*, (2011), pp. 1–7.
22. J. Yang, D. Zhang, J.-Y. Yang and B. Niu, Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics, *IEEE Trans. Pattern Anal. Mach. Intell.* **29** (2007) 650–664.
23. S. Fazli and L. A. Heidarlool, Wavelet based age invariant face recognition using gradient orientation.
24. D. Bouchaffra, Nonlinear topological component analysis: Application to age-invariant face recognition (2014).
25. S. Mitra and T. Acharya, Gesture recognition: A survey, *IEEE Trans. Syst. Man, Cybernet. Part C: Appl. Rev.* **37** (2007) 311–324.
26. M. J. Fehrenbach and S. W. Herring, *Illustrated Anatomy of the Head and Neck* (Elsevier Health Sciences, 2013).
27. M. M. Hom and L. Bielory, The anatomical and functional relationship between allergic conjunctivitis and allergic rhinitis, *Allergy Rhinol.* **4** (2013) e110.

28. L. Wang, Y. Zhang and J. Feng, On the Euclidean distance of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (2005) 1334–1339.
29. B. J. McCartin, *Mysteries of the Equilateral Triangle* (Hikari Limited, 2010).
30. J. Quackenbush, Microarray data normalization and transformation, *Nat. Genet.* **32** (2002) 496–501.
31. L. Bo, L. Wang and L. Jiao, Feature scaling for kernel fisher discriminant analysis using leave-one-out cross validation, *Neural Comput.* **18** (2006) 961–978.
32. D. M. Tax and R. P. Duin, Feature scaling in support vector data descriptions, Technical report, Technical report, American Association for Artificial Intelligence (2000).
33. L. E. Peterson, K-nearest neighbor, *Scholarpedia* **4** (2009) 1883.
34. K. P. Murphy, Naive bayes classifiers, University of British Columbia (2006).
35. M. A. Hearst, S. T. Dumais, E. Osman, J. Platt and B. Scholkopf, Support vector machines, *IEEE Intell. Syst. Their Appl.* **13** (1998) 18–28.
36. V. Svetnik, A. Liaw, C. Tong and T. Wang, Application of Breiman’s random forest to modeling structure-activity relationships of pharmaceutical molecules, in *Multiple Classifier Systems* (Springer, 2004), pp. 334–343.
37. J. Hua, Z. Xiong, J. Lowey, E. Suh and E. R. Dougherty, Optimal number of features as a function of sample size for various classification rules, *Bioinformatics* **21** (2005) 1509–1515.
38. E. Youn and M. K. Jeong, Class dependent feature scaling method using naive Bayes classifier for text datamining, *Pattern Recogn. Lett.* **30** (2009) 477–485.
39. B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich *et al.*, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, *BMC Bioinformat.* **10** (2009) 213.
40. H. J. Seo and P. Milanfar, Face verification using the lark representation, *IEEE Trans. Inform. Forensics Security* **6** (2011) 1275–1286.
41. T. F. Cootes, G. J. Edwards and C. J. Taylor, Active appearance models, *IEEE Trans. Pattern Anal. Mach. Intell.* **23** (2001) 681–685.
42. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, The WEKA data mining software: An update, *ACM SIGKDD Explorations Newsllett.* **11** (2009) 10–18.
43. Y. Yang and X. Liu, A re-examination of text categorization methods, in *Proc. 22nd Ann. Int. ACM SIGIR Conf. Research and Development in Information Retrieval* (1999), pp. 42–49.
44. T. Ojala, M. Pietikäinen and D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recogn.* **29** (1996) 51–59.
45. E. Bruno and D. Pellerin, Robust motion estimation using spatial Gabor-like filters, *Signal Process.* **82** (2002) 297–309.
46. Y.-l. Tian, T. Kanade and J. F. Cohn, Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity, in *Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition, 2002* (2002), pp. 229–234.
47. L. Shen, L. Bai and M. Fairhurst, Gabor wavelets and general discriminant analysis for face identification and verification, *Image Vis. Comput.* **25** (2007) 553–563.
48. U. Park, Y. Tong and A. K. Jain, Age-invariant face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (2010) 947–954.
49. Z. Li, U. Park and A. K. Jain, A discriminative model for age invariant face recognition, *IEEE Trans. Inform. Forensics Security* **6** (2011) 1028–1037.
50. B. Efron and G. Gong, A leisurely look at the bootstrap, the jackknife and cross-validation, *The Am. Statist.* **37** (1983) 36–48.

51. R. J. Micheals and T. E. Boulton, Efficient evaluation of classification and recognition systems, in *Proc. 2001 IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2001. CVPR 2001*, Vol. 1 (2001), pp. I-50-I-57.
- 



**Ms. Amal Seralkhatem Osman Ali** received her B.Sc. degree in Communication Engineering from Sudan University of Science and Technology Khartoum, Sudan in 2007, and her M.Sc. degree in Computer Architectures and Networks from University of Khartoum,

Sudan in 2009. She is currently working as a Ph.D. scholar in the Department of Electrical and Electronics Engineering at University Teknologi PETRONAS since Jan 2012. Her research focuses on developing an Age-Invariant Face Recognition system to reduce the effects of facial aging on face recognition performance.



**AP. Dr. Aamir S. Malik** received his M.Sc. degree in Information and Communication and Ph.D. in Information and Mechatronics from Gwangju Institute of Science and Technology, South Korea. He has more than 15 years of research experience and is currently Associate Professor at the Department of Electrical Engineering and Director of Biomedical Technology group at Universiti Teknologi PETRONAS. His research interests include bio-medical signal & image processing, visual surveillance, remote sensing and brain sciences.



**Dr. Mohamed Meselhy Eltoukhy** is with the Computer Science Department, Faculty of Computers and Informatics at Suez Canal University, Egypt.



**AP. Dr. Vijanth S. a/l Asirvadam** studied at the University of Putra, Malaysia for the Bachelor Science, B.Sc. degree (Hon) majoring in Statistics and graduated in April 1997. He received his M.Sc. degree in Engineering Computation with a Distinction in December 1998. He took up employment as a system administrator at the Multimedia University (MMU) Malaysia prior to joining the Intelligent Systems and Control Research Group at Queen's University Belfast in November 1999 where he completed his Doctorate (Ph.D.) in March 2003, researching on online and constructive neural learning methods.



**Dr. Azrina Aziz** is with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia.