Expert Systems With Applications 72 (2017) 383–394

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A composite spatio-temporal modeling approach for age invariant face recognition

Fahad Bashir Alvi*, Russel Pears

Knowledge Engineering and Discovery Research Institute (KEDRI), Auckland University of Technology, New Zealand

ARTICLE INFO

Article history:
Received 6 June 2016
Revised 14 August 2016
Accepted 17 October 2016
Available online 29 October 2016

Keywords:
Anthropometric model
Local model
Personalized model
Integrated model
K nearest neighbor
Decision tree
Naive Bayes
Adaline Neural Network

ABSTRACT

In this research we propose a novel method of face recognition based on texture and shape information. Age invariant face recognition enables matching of an image obtained at a given point in time against an image of the same individual obtained at an earlier point in time and thus has important applications, notably in law enforcement. We investigate various types of models built on different levels of data granularity. At the global level a model is built on training data that encompasses the entire set of available individuals, whereas at the local level, data from homogeneous sub-populations is used and finally at the individual level a personalized model is built for each individual. We narrow down the search space by dividing the whole database into subspaces for improving recognition time. We use a two-phased process for age invariant face recognition. In the first phase we identify the correct subspace by using a probabilistic method, and in the second phase we find the probe image within that subspace. Finally, we use a decision tree approach to combine models built from shape and texture features. Our empirical results show that the local and personalized models perform best when rated on both Rank-1 accuracy and recognition time.

1. Introduction

Face image recognition is an important and significant part of the domain of biometric research (Cootes, Edwards, & Taylor, 1999; Gong, Li, Tao, Liu, & Li, 2015; Jain, Nandakumar, & Ross, 2016; Li, Park, & Jain, 2011; Park, Tong, & Jain, 2010; Ramanathan & Chellappa, 2006). It has important applications in real life in the fields of missing children identification, passport verification, security, animation and business intelligence. For example, in case of law enforcement, an image of a suspect is available and we need to find out whether the image of the same person exists and can be obtained from our crime database or not. If the suspect does actually appear in the crime database at a previous age then a match should be made and information on the last known address and other associated information of the suspect can be retrieved from the crime database.

Inspite of extensive research in face recognition (Gong, Li, Lin, Liu, & Tang, 2013; Gong et al., 2015; Li et al., 2011; Ramanathan & Chellappa, 2006) much ground has not been covered in the field of age invariant face recognition (Gong et al., 2013; Gong et al., 2015; Li et al., 2011; Park et al., 2010). Face recognition has two main areas of research namely, face verification and face identification. There are two types of approaches used in age invariant face recognition, generative and non generative. Generative approaches assume prior knowledge of human age, given an image. On the other hand, non-generative approaches concentrate on finding discriminative features of the face and the changes therein throughout the face aging lifespan. From another perspective we can identify three main directions of research related to face recognition across age: viz, age invariant face recognition, age simulation and age estimation. Most research has concentrated on age estimation and age simulation. Research into age invariant face recognition is still at a nascent stage.

In this paper we will concentrate on age invariant face recognition across different ages with special emphasis on the identification problem (see illustration in Fig. 1).

Aging is a complex problem because at different age points different types of changes occur in the human face. From childhood to teenage the changes are mostly related to craniofacial growth. At maturity the changes are mostly related to the skin color changes and texture effects, with facial skin starting to become slack and less smooth. So aging is a mixture of all of these components. Moreover, aging is a slow, irreversible, and a process that is unique to every human being. Many factors affect the aging process. For example every person has different genes, blood group, life style and belongs to a particular ethnic group. In order to resolve all

* Corresponding author.
E-mail addresses: falvi@aut.ac.nz, fahadalvi@hotmail.com (F.B. Alvi), russel.pears@aut.ac.nz (R. Pears).

http://dx.doi.org/10.1016/j.eswa.2016.10.042
0957-4174/

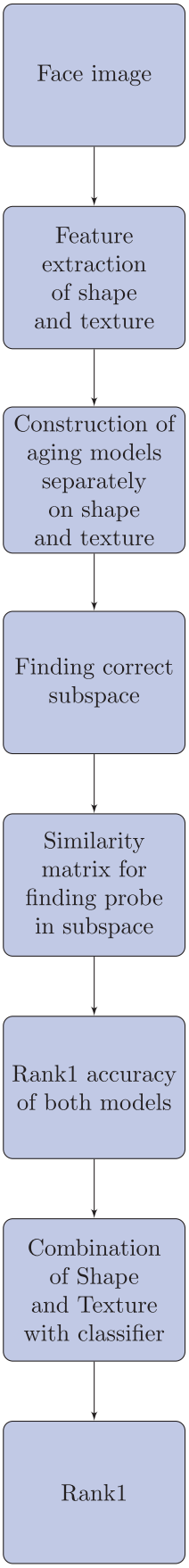


Fig. 1. Proposed methodology.

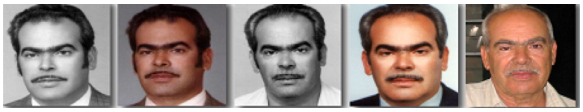


Fig. 2. Same individual at different ages from FG-NET (FG-NET, 2002).

Fig. 3. Same individual at different ages from MORPH (Ricanek Jr & Tesafaye, 2006).

these issues we need to deal with shape and texture features separately and finally combine them to exploit the natural synergy between these two types of features.Figs. 2 and 3 show images of two persons at different ages.

There are five major challenges which affect the performance of face recognition systems: pose, illumination, expression, occlusion and aging. It has been observed in past research that for resolving one challenge we need to compromise on others (Abdullah et al., 2014; Geng, Zhou, & Smith-Miles, 2007; Pujol & García, 2012) Thus, there is a need to find a solution that compromises on the minimum number of factors or challenges.

We use a novel approach in this paper making a two pronged attack on these challenges by using both shape and texture features. We build an Anthropometric model which is based on geometrical ratios and distances of a fixed number of fiducial landmarks. Such a model is not affected by occlusions such as moustaches or spectacles because landmarks do not alter their position. Likewise, facial expression and illumination also do not significantly alter the position of landmarks. However, different poses could create problems. To mitigate this problem, we take an average frontal image and warp all our images to that image through Procrustes analysis (Cootes et al., 1999). That leaves the aging aspect which is our main area of research and our study fully focuses on that area. We use edges on the face texture to extract wrinkles information and together with the Anthropometric features we build and analyze the aging process. We have used these same features in our previous studies (Alvi & Pears, 2015a; 2015b) as well. The contributions we make in this research are:

- We partition an image gallery into sub galleries by time segment before applying a probabilistic Bayesian method to identify the segment containing the desired (target image). Our empirical results show that such partitioning helps to improve recognition accuracy.
- We use a local modeling approach whereby we partition the global population of images into homogeneous sub-populations of clusters and build localized models on each of the clusters. This partitioning is orthogonal to the time based partitioning mentioned above.
- We combine shape and texture models in order to exploit any synergy that exists between them. In certain cases, both types of models point to the same image which happens also to be the right one. In such cases having two types of models enhances the robustness of the matching process as agreement exists. However in certain other cases, each model could identify different images. Conflict resolution is then necessary and we make use of a decision tree classifier for resolving such conflicts. Our choice of the decision tree classifier was based on the fact that it had the highest conflict resolution success rate out of the classifiers that we experimented with, which included

the Support Vector Machine (SVM), Decision Tree, Random Forest and Naive Bayes.

2. Literature review

Geng et al. (2007) worked on the assumption that similar faces age in similar ways across a population. He introduced the method of Aging pattern Subspace (AGES) for modeling aging across time. Geng et al constructed a representative subspace by utilizing a sequence of images of a given individuals face sorted in chronological order. Minimum reconstruction error was achieved by projecting the unseen face image into the proper aging subspace. Thus, it would be possible to determine the age of the image from the aging pattern.

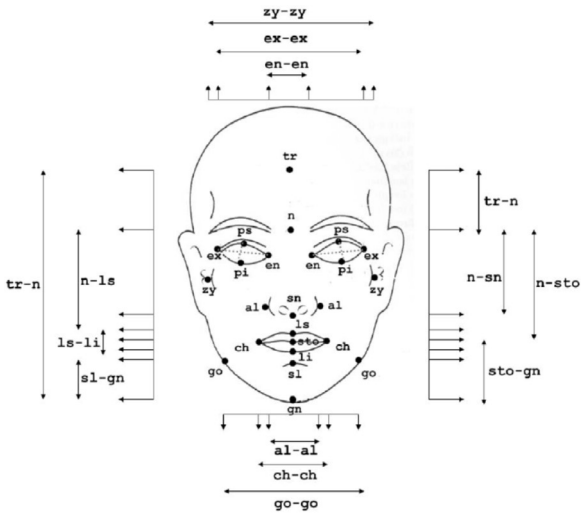
Syambas and Purwanto (2012) used a combination of Active Appearance Models (AAMs), Support Vector Machines (SVMs) and Monte-Carlo simulation to build a high accuracy aging model. Two types of experiments were reported. In the first experiment, they showed empirically that as the probe face progresses in age, the face recognition rate decreases. In order to overcome this problem, in a second experiment, they artificially aged probe faces so as to increase the likelihood of a match with a gallery image that would then be closer in time terms to the aged version of the probe image. The artificial aging was accomplished by the use of active appearance models while the matching process was based on the use of PCA. A rank 1 accuracy of 32 percent was obtained by Sethuram et al.

Lanitis, Taylor, and Cootes (1999) also used Active Appearance Models (AAM). He used a statistical face model for studying the age estimation problem. AAM parameters were extracted from facial images which were marked with 68 points. A Genetic Algorithm was then applied to build and optimize an aging function.

Ling, Soatto, Ramanathan, and Jacobs (2007) introduced the concept of coordinate patches and Gaussian Mixture Models (GMMs). He used images of individuals and encoded them as ensembles of overlapped spatially flexible patches (SFPs). These SFPs were modeled with Gaussian Mixture Models (GMMs). Local features were extracted with the help of the 2D discrete cosine transform (DCT). The local features were integrated with the patches and an estimate of the age of the individual was then obtained based on maximum likelihood estimators built from all the SFPs of the hypothetical age.

3. Anthropometric model

Face anthropometry, the science of measuring sizes and proportions on human faces, has the potential to play a crucial role in developing facial aging models. Such studies provide a quantitative description of the craniofacial growth at different ages and hence provide a plethora of options for learning based approaches to be adopted to characterize facial aging. Face anthropometric studies provide dense measurements taken between key landmarks on human faces across different ages and have played a critical role in surgical procedures employed on the faces of growing children (Farkas & Munro, 1987). Farkas provides a comprehensive overview of face anthropometry and its many significant applications. He defines face anthropometry in terms of measurements taken from 57 carefully selected landmarks on human faces spread across 6 regions in the craniofacial area (head, face, orbits, nose, lips and mouth, ear). The facial measurements are of three kinds: (i) projective measurements (shortest distance between two landmarks); (ii) tangential measurements (distance between two landmarks measured along the skin surface) and (iii) angular measurements. Fig. 4 illustrates the kind of data that is collected in face anthropometric studies and further illustrates the different fiducial features across



- Facial index $(\frac{n-gn}{zy-zy})$
- Mandibular index $(\frac{sto-gn}{go-go})$
- Intercantal index $(\frac{en-en}{ex-ex})$
- Orbital width index $(\frac{ex-en}{en-en})$
- Eye fissure index $(\frac{ps-pi}{ex-en})$
- Vermilion height index $(\frac{ls-sto}{sto-li})$
- Mouth Face width index $(\frac{ch-ch}{zy-zy})$

Fig. 4. Anthropometric model and seven features (Farkas, 1994) .

which such data is collected (Farkas, 1994). In the absence of age-based Anthropometric measurements they collected facial growth data by extracting facial features on the passport database.

Such growth data was collected on five different age groups: 21–30 years, 31–40 years, 41–50 years, 51–60 years and 61–70 years. The facial growth data collected in this manner was found to be effective in characterizing facial growth based on age, gender, ethnicity etc. and in cases where individuals gain or lose weight.

4. Texture features

In the case of texture, as a part of preprocessing, we first register(Goshtasby, 2012; Štruc & Pavešić, 2010) images on the basis of eye coordinates so that all images transform to the same size and that each point on any given face refers to the same point on all of the images. We then divide the facial image horizontally into five slices. The slices cover the forehead, eyes, nose, mouth and the area below the mouth. Edges are marked on the face after converting the image into grayscale. A histogram is prepared for these edges. The frequency of the edges in each slice is taken as a feature, thus resulting in five features across the entire face as shown in Fig. 5. We used two parameters for Canny edge detection: sigma and threshold. The threshold values were obtained after experimentation, to minimize the noise and obtain optimum results. The values of threshold are FG-NET, minimum=0.0375, maximum=0.0938 and for MORPH minimum = 0.0500 and maximum = 0.1250. The sigma value used was sqrt(2) for both databases.

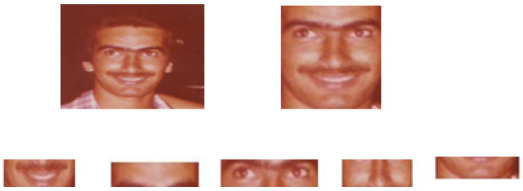


Fig. 5. Five slices of the face (FG-NET, 2002)

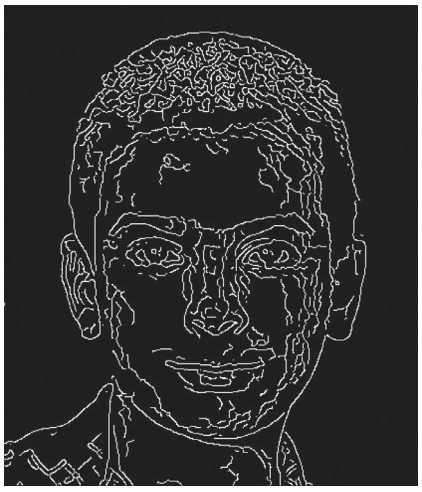


Fig. 6. Canny edge detection of one image (FG-NET, 2002).

4.1. Edge detection

We used the Canny edge detector to find edges. The Canny operator was designed to be an optimal edge detector (according to particular criteria — there are other detectors around that also claim to be optimal with respect to slightly different criteria). It takes as input a gray scale image, and produces as output an image showing the positions of tracked intensity discontinuities The Canny edge detector operator works as a multi-stage process. First of all the image is smoothed by Gaussian convolution. Then a simple 2D first derivative operator is applied to the smoothed image to highlight regions of the image with high values for the first spatial derivative. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output. This process is known as non-maximal suppression. The tracking process exhibits hysteresis and is controlled by two thresholds: $t1$ and $t2$, with $t1 > t2$. Tracking can only begin at a point on a ridge higher than t1. Tracking then continues in both directions out from that point until the height of the ridge falls below $t2$. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments. (Canny, 1986). Fig. 6 shows canny edge detection for a selected image.

5. Computational models for aging

In this research we treat the age recognition problem as a time series problem with each texture feature and anthropometric feature giving rise to a separate time series variable. In the standard time series problem a trajectory at several different time points from the past is available and the problem is to project into the future. The time series problem that we tackle differs from the standard time series problem in one fundamental aspect: in the face recognition scenario only one time point is available, which is the image (feature vector) at the time point at which the image was captured. No other training data is available. This would make the problem virtually unsolvable if not for the fact that we can explore the data spatially to obtain more training data. Thus at any given time point we have f data points obtained from the f different anthropometric features that we extract. However, since the time series trajectory of each feature is largely independent of the others, a simple approach of combining them into a single series would be highly ineffective. As such, our solution is to first build trajectories for each of the features separately and then combine predicted feature values at each band into a single vector which is used to determine whether a match exists between probe and gallery images. The other relatively minor difference is the direction in which the trajectory is built; in our case it is backwards in time instead of forwards.

We investigated four different types of models built at different levels of data granularity. At the global level a model is built on training data that encompasses the entire set of available individuals. At the local level data from homogeneous sub-populations were used to prepare a local model. At the individual level a personalized model was built for each individual. Finally, all these models were combined into a single, integrated model. We used a similarity matrix to compute Rank 1 accuracy.

5.1. Global model

5.1.1. Model construction

A global model is induced over the entire problem space that contains the entire population of available individuals. Such models serve to capture useful general trends across the population over a spectrum of age bands used. After grouping images into their respective age bands, we determine discriminative features and compute the values of 7 unique indexes (i.e. ratio of Euclidean distances for a given pair of Anthropometric features) for each image. The indexes were calculated after warping the image to a mean image using the Procrustes algorithm (Cootes et al., 1999). In the case of texture, as a part of preprocessing, we first register the images, so that all images transform to the same size and each point on any given image refers to the same point on all of the images. Then we divide the facial image into five horizontal slices. The slices cover the forehead, eyes, nose, mouth and the area below the mouth. Edges(Canny) are marked on the face after converting the image into grayscale. A histogram is prepared for these edges. The frequency of edges in each slice is taken as a feature, thus resulting in five features for the whole face. The centroid of each age band is determined and an nth order polynomial (we experiment with values of n in the range [1..3]) function is developed that spans all the age bands using a least square based non-linear regression method. Eq. (1) represents the Global model for the Facial index feature, where c_i represents the value of each coefficient and x_i represents its age band index value. Fig. 7 shows the trajectory of the Global Model across the agebands for the Facial index.

$$y_i^{(global)} = c_1 x_i^n + c_2 x_i^{n-1} + ... + c_n x_i + c_{n+1} \quad (1)$$

5.2. Personalized model

5.2.1. Model construction

The transductive or personalized approach, in contrast to the inductive approach, models each point in the problem space. It was defined by Vapnik in Vapnik and Vapnik (1998) and used by Kasabov in Kasabov (2007) and Pears in Pears, Widiputra, and Kasabov (2013). The intuition behind personalized modeling is that the aging process differs from person to person and hence modeling at the level of individuals can be expected to yield more accuracy. The k-NN (k-nearest neighbor) is one of the well-known

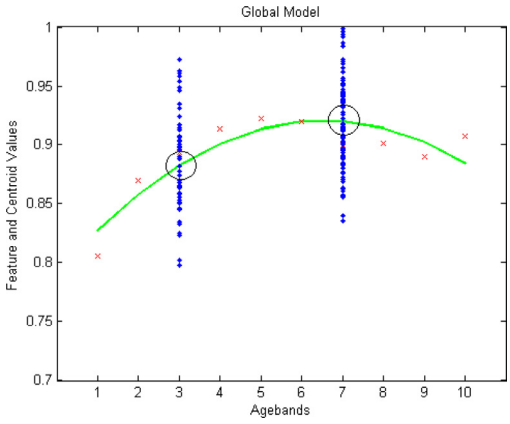


Fig. 7. Global model for feature 1 (Facial index)

transductive techniques and is the most widely used form of personalized modeling.

The model is constructed as follows. We take an image, determine its age band (in closed set evaluation mode the age for the image is known; in real-world applications, a human expert can be used to estimate the age) and obtain its *k*-nearest neighbors (the optimal value of *k*=5 is set by experimentation). When we get an Anthropometric sample of an image we de-age that sample as follows:

$$DF_i = M_i - F_i \tag{2}$$

We introduce a deviation factor *(DF)* as defined in Eq. (2) above. The deviation factor captures the degree to which an individual's feature differs from the population as a whole. In this research we assume that the DF_i is independent of age which amounts to assuming that the rate of change in a feature value that is constant over time. In Eq. (2), M_i represents the mean of feature i at the age band j estimated by the human expert and F_i is the feature value of the image in that same age band.

$$P_{ij} = G_{ij} + DF_i \tag{3}$$

The de-aging process is done in Eq. (3) by adding the deviation factor DF_i of the image in the given age band to the value G_{ij} returned by the global model at age band j which in turn is obtained by applying Eq. (1) on feature i. The resultant pivot value P_{ij} is used to find the *k*-nearest neighbors and determine the centroid a_i of these neighbors. This process is repeated for each age band to complete the construction of the personalized model for the image. We use these centroids a_i to fit a nonlinear function across the age bands; once again we fit an *nth* degree polynomial curve using the method of least squares. Eq. (4) represents the Personalized model where y_i is the feature value for feature i; x_i is the age band index and n is the degree of the polynomial.

$$y_i^{(personalized)} = c_1 {x_i}^n + c_2 {x_i}^{n-1} + ... + c_n x_i + c_{n+1} \tag{4}$$

Eq. (4) embodies the aging trajectory for an individual and provides insights into how one individual's aging trajectory differs from another in the given population. In our experimentation in Sections 8 and 9 we demonstrate the effectiveness of the personalized model against the global model.

5.3. Local model

While a personalized model captures aging effects that are tailored to an individual and a global model captures trends across a population, scope exists for an intermediate level modeling approach. Given a probe image the local modeling approach constructs a model from a homogeneous subset of the population that is most closely aligned with the image rather than from the entire population that in general is heterogeneous.

Essentially, the local approach constructs subpopulation through the use of a clustering scheme. For each probe image, the cluster that is closest to it is identified and thereafter the same model construction process as in personalized modeling is applied. Thus in terms of performance, the local approach is expected to be faster than personalized modeling as the construction and fitting of the aging function is restricted to a subset of instances rather than from the population as a whole.

In a real world situation, the population may be heterogeneous with several different sub-populations made up of different ethnicities, different lifestyles, etc, all of which have a bearing on the aging process. However, in this research we do not assume that the image gallery is annotated with such explicit feature information. Thus in order to segment the population we need to make use of a clustering algorithm that could be applied on the time series trajectories obtained from the fiducial features taken over time. The intuition is that individuals who age similarly will produce similar trajectories and hence will be clustered together in the same segment. However, it can also happen that any two given individuals age similarly but have pair-wise differences in the time series variable (fiducial feature) over the age bands that we track. In time series analysis this corresponds to two time series that are highly correlated but are out of phase with each other. A standard clustering algorithm such as *K*-means will fail to cluster such correlated but out of phase sequences. This problem can be avoided by applying Dynamic Time Warping (DTW) (Sakoe & Chiba, 1978) before applying *K*-means.

As mentioned earlier the construction follows the same procedure as for the personalized model but with segmentation performed as a pre-processing step and the use of cluster means in place of global means at each age band. The algorithmic details follow.

5.3.1. Model construction

Step 1: For each feature f, apply the *K*-means clustering algorithm on the time warped vector space and create K clusters that span the N age bands. Note that we use K to distinguish it from the neighborhood parameter k used in the k-NN search. This will result in a total of $K \times N$ centroids, one for each cluster and one for each age band.

Step 2: For each probe image its age band is determined. For that age band the closest cluster C, that has the maximum probability of containing the target image given feature X_j is determined from Eq. (5).

$$Pr(C_i|X_j) = Pr(C_i)Pr(X_j|C_i)/Pr(X) \tag{5}$$

where $Pr(X)$ is a scaling factor computed across all features. X_j, $1 \leq j \leq n$ where n is total number of features.

Step 3: Use Eq. (2) with M_i representing the mean of feature i across the cluster that the image belongs to, instead of the mean across all images in the database. Thereafter, Eq. (3) is applied to find the pivot points P_i and n polynomial aging functions of the same form as in Eq. (4) are developed for each of the K clusters.

5.4. Integrated model

Our Integrated model is inspired by the Integrated Multi Model Frame Work(IMMF) proposed by Widiputra (2011) which exploits synergy that may exists between models at the global, local and personalized data granularity levels.

The main component of the IMMF is the accumulator module which is implemented through the use of an Adaline Neural Network (Widrow & Stearns, 1985) that decides the contribution (weighting) contributed by each type of model. The accumulator

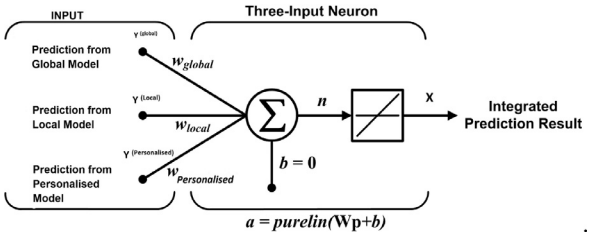


Fig. 8. Illustration of the Adaline for integrated model

module takes as input the set of training images and then for each training image will feed prediction outcomes (in the form of predicted feature value for feature j) from each of the global, local and personalized models in order to assign weights that will be associated with each type of model. The output of the accumulator represented by X in Fig. 8 is the optimized final set of weights over the entire set of training images. Eq. (6) defines the prediction for feature j for the Integrated model in terms of a linear weighted function of predictions from the global, local and personalized models.

$$(X)_j = W_{j,g}Y^{(Global)} + W_{j,l}Y^{(Local)} + W_{j,p}Y^{(Personalised)} \quad (6)$$

where $W_{j,g}$ is the weight attributed to the global model for fiducial feature j. $W_{j,l}$ is the weight attributed to the local model for fiducial feature j. $W_{j,p}$ is the weight attributed to the personalized model for fiducial feature j. $Y^{(Global)}$, $Y^{(Local)}$ and $Y^{(Personalised)}$ are values returned from Eq. (4).

6. Use of subspace for target image identification

In order to improve rank 1 accuracy, we aggregated adjacent age bands into subspaces. In effect, we combine two adjacent age bands into one subspace, thus building I subspaces from N age bands. This has two advantages. Firstly, we expect rank 1 accuracy to improve as there is less uncertainty in locating subspaces as opposed to age bands. Secondly, image recognition time is reduced as the search is restricted to a smaller number of time segments.

Naive Bayes is used to identify a subspace having a probe feature X_j. Naive Bayes is an effective and computationally efficient technique for constructing a classifier. The classifier assumes that features are independent of each other which implies that the conditional dependence of class on a given feature is independent of the corresponding conditional dependencies across the other features. Although this assumption may not hold strictly in practice several empirical studies in the machine learning literature have shown that this type of classifier is robust in practice (Witten & Frank, 2005). We thus select the Naive Bayes classifier on account of its robustness and computational efficiency.

Each subspace is constructed as shown in Tables 2 and 4. Two bins, *high* and *low*, spanning the higher and lower end of the median boundary range, were used for each subspace after normalizing each feature value to the [0..1] range. When a probe image is taken, models at the global, local and personalized data granularity levels are constructed. The feature values are de-aged to the age band of the probe image. Then a predicted value for all of the features are produced by the model. These values are compared with the bin values and a bin is chosen on the basis of the closest match with the predicted value.

A posterior probability for each subspace is computed by using Eq. (7) where $Pr(A_i)$ is a prior probability for subspace i. The conditional probability $Pr(X_j|A_i)$ is then found for each feature j by applying our aging models. The probability $Pr(X)$ is a scaling factor that is computed across the entire feature set.

$$Pr(A_i|X_j) = Pr(A_i)Pr(X_j|A_i)/Pr(X) \quad (7)$$

Table 1
FG-NET age bands

Ages	0–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–69
Images	233	178	164	155	81	62	38	31	26	34
Subjects	75	70	71	68	46	38	30	24	19	10

Table 2
FG-NET subspaces

Ages	0–10	11–20	21–30	31–40	41–69
Images	411	319	143	69	60
Subjects	77	81	50	35	22

where

$$Pr(X_j|A_i) = Pr(X_1|A_i)Pr(X_2|A_i).....Pr(X_n|A_i)$$

$$Pr(X_j|A_i) = \prod_{j=1}^{n} Pr(X_j|A_i)$$

$$Pr(A_i|X_j) = Pr(A_i)\prod_{j=1}^{n} Pr(X_j|A_i)$$

$$Pr(X) = \sum_{i=1}^{k} Pr(A_i)\prod_{j=1}^{n} Pr(X_j|A_i)$$

This process is repeated for every subspace. The subspace that yields the highest posterior probability $Pr(A_i|X_j)$ is chosen and is assumed to contain the desired probe image. We used a leave one person out strategy to train the classifier.

For recognition, given a probe image, its age band is obtained from the human expert. We use this age band estimate to compute predicted feature values for the probe image by using aging models. We then determine the target subspace of the probe image. Finally, a similarity matrix is built from all images in our search space and rank 1 accuracy is then computed.

7. Empirical study

Experiments were performed on the publicly available FG-NET and MORPH Album 2 (the largest publicly available face aging dataset) (Ricanek Jr & Tesafaye, 2006), both of which are used for benchmarking new methods. The lack of a large publicly available face aging database until recently limited research on age invariant face recognition.

There are two desired attributes of a face aging database: (i) large number of subjects, and (ii) large number of face images per subject captured at many different age points (Li et al., 2011). In addition, it is desired that these images should not have large variations in pose, expression, and illumination.

Each of the two datasets that we experimented with has their own challenges. The MORPH dataset has a large number of subjects with images taken across a narrow age timeline while FG-NET database has a smaller number of subjects and images but the average age gap between images from the same person is much larger than with the MORPH dataset.

The MORPH dataset contains about 55,000 face images from 13,000 different people. The FG-NET database on the other hand contains 1002 color and gray face images of 82 persons across a range of different ethnicities. There is a large variation in lighting, expression and pose across the different images. The image size is 300 × 400 in pixel units, on the average. The ages vary from 0 to 69 years. There are on the average 12 images per person across different ages. The database was divided into ten different age bands as shown in Tables 1 and 3.

Table 3
Morph age bands

Ages	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55	56–60	61–65
Images	2189	2113	1642	1760	1825	1657	1018	549	183	64
Subjects	973	1037	821	903	971	850	536	287	92	30

Table 4
Morph subspaces

Ages	16–25	26–35	36–45	46–55	56–65
Images	2283	1759	1680	675	103
Subjects	2283	1759	1680	675	103

Table 5
Classification accuracy of subspace identification

Models	Database, subjects and images in probe	Accuracy
Shape	FG-NET(82,1002)	96.4%
Texture	FG-NET(82,1002)	94.6%

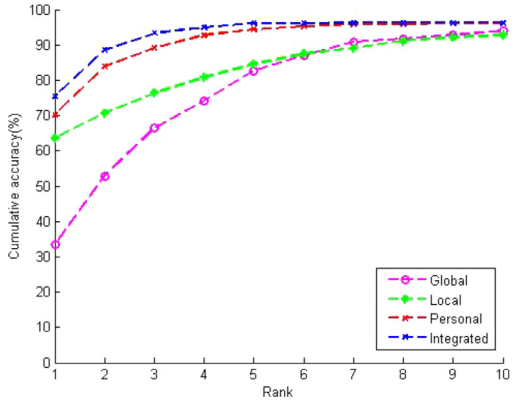


Fig. 9. Cumulative matching characteristic curves (CMC) for FGNET database.

8. Experiment with FG-NET database

For evaluation we chose all 1002 images taken from the entire set of 82 individuals. A leave-one person-out (LOPO) strategy was used for evaluation (Gong et al., 2013; Gong et al., 2015; Li et al., 2011).

Our experimentation focused on a three-way comparison between models at different levels of data granularity. We compared the performance of the Global, Local and Personalized models against each other and finally the effectiveness of the Integrated approach was investigated. The rationale behind this was to establish the merits of each method operating on its own in order to assess: (1) which one of the models on their own performed the best and (2) whether the Integrated approach offers better accuracy over and above each of the individual model approaches.

8.1. Success rate of subspace selection

Table 5 shows that the Bayesian classifier was highly effective at finding the correct subspace for both shape and texture models, with both types of models returning accuracy rates around the mid 90% mark as measured according to the LOPO strategy mentioned earlier.

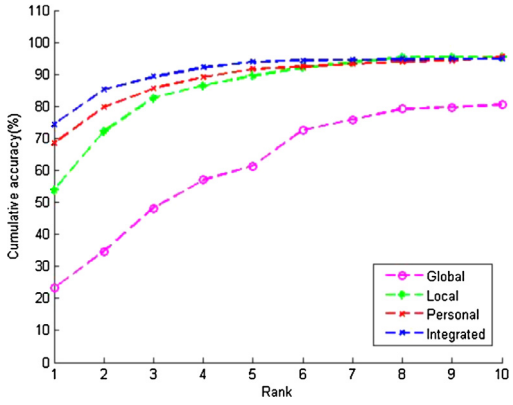


Fig. 10. Cumulative matching characteristic curves for FGNET database.

Fig. 11. The first row shows the probe images and the second row shows the gallery images correctly identified at the rank-1 level.

8.2. Comparison of models across data granularity level for shape features

The CMC curves in Fig. 9 confirm that the trends displayed in shape models hold across rank orders from 1 to 10, with the Integrated model emerging the winner across the range.

8.3. Comparison of models across data granularity level for texture features

The CMC curves in Fig. 10 confirm that the trends displayed in texture models hold across rank orders from 1 to 10, with the Integrated model emerging the winner across the range.

Fig. 11 shows that gallery images of people in their youth can be extracted successfully, given probe images of the same people at adulthood.

We note that the models developed can not only be used for gallery image extraction but could be used to simulate the aging process as well. The age trajectory functions that we build can be used to predict their features values by performing a forward traversal (rather than a backward one). This a further advantage of our aging models over the one proposed by Gong et al. (2015).

9. Morph experimentation

In order to check generality of our models on different types of databases, we used MORPH Album 2 (the largest publicly available face aging dataset) database. We annotated the fiducial landmarks with Stasm as shown in Figs. 12, a method based on Active shape Models (Milborrow & Nicolls, 2014). We used all 13,000 subjects for conducting the experiment. We used the same methodology as used in state of art algorithms (Gong et al., 2013; Gong et al., 2015; Li et al., 2011) in which the MORPH album 2 data set was split into three sets, viz training set, a probe set and a gallery set.

Fig. 12. Fiducial landmark detection results. Blue points represent the 68 landmark points.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

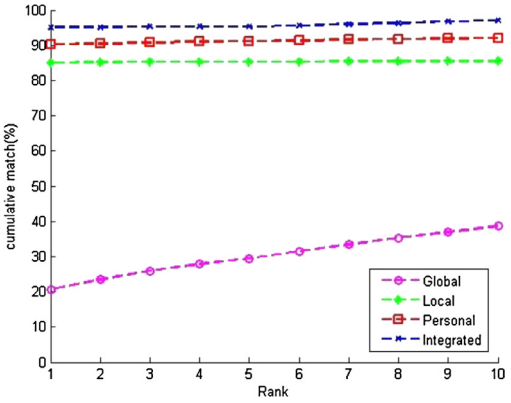


Fig. 13. Cumulative matching characteristic curves for MORPH database

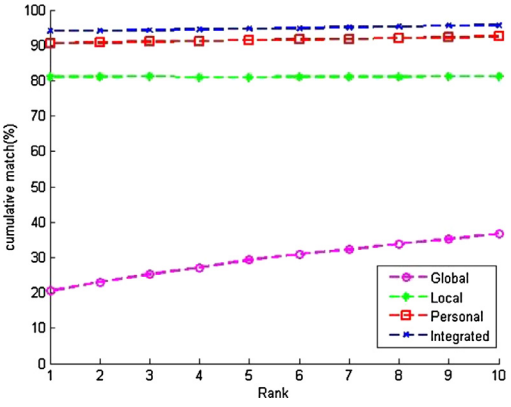


Fig. 14. Cumulative matching characteristic curves for MORPH database

The persons in the training set were entirely different from those in the probe and gallery sets. For the training dataset, we selected a subset of 13,000 face images from 6500 subjects, with two images per subject. These two images were selected such that they had the largest age gap between them. The gallery set was composed of 6500 face images corresponding to the youngest age of these 6500 subjects. The probe set was composed of 6500 face images corresponding to the oldest age of these 6500 subjects.

The CMC curves in Figs. 13 and 14 confirm the trends shown in Table 8 across rank order, thus confirming that the Integrated model is once again the overall winner on the MORPH dataset as well.

9.1. Success rate of subspace selection

Table 7 shows that the Bayesian classifier was highly effective at finding the correct subspace for both shape and texture models, with both types of models returning accuracy rates above 95% as measured according to the LOPO strategy mentioned earlier.

Fig. 15. The first row shows the probe images and the second row shows the gallery images correctly identified at the rank-1 level.

Table 6
Comparative results of Rank-1 score with FG-NET database.

Models	shape	Texture
Global	33.7%	23.3%
Local	63.7%	53.9%
Personalized	70.4%	68.6%
Integrated	75.6%	74.5%

Table 7
Comparative results of classification of different models

Models	Database,subjects and images in probe	accuracy
Shape	Morph(6500,6500)	98.8%
Texture	Morph(6500,6500)	99.0%

Table 8
Comparative results of Rank-1 score with MORPH database

Models	Shape	Texture
Global	21%	20%
Local	85%	81%
Personalized	91%	90.5%
Integrated	95.3%	95%

9.2. Comparison of models across data granularity level for shape features

The CMC curves in Fig. 13 confirm that the trends displayed in shape models hold across rank orders from 1 to 10, with the Integrated model emerging the winner across the range.

9.3. Comparison of models across data granularity level for texture features

The CMC curves in Fig. 14 confirm that the trends displayed in texture models hold across rank orders from 1 to 10, with the Integrated model emerging the winner across the range.

Fig. 15 above shows that gallery images of people in their youth can be extracted successfully, given probe images of the same people at adulthood.

Tables 6 and8 show that the Integrated model outperforms the Global, Local and personalized Model. This indicates that synergy exists between models at different levels of data granularity. Models built from shape features slightly outperform their texture counterparts although evidence exists that the texture features enhance face recognition as models built from a composite set of texture and shape features outperform each model type on its own.

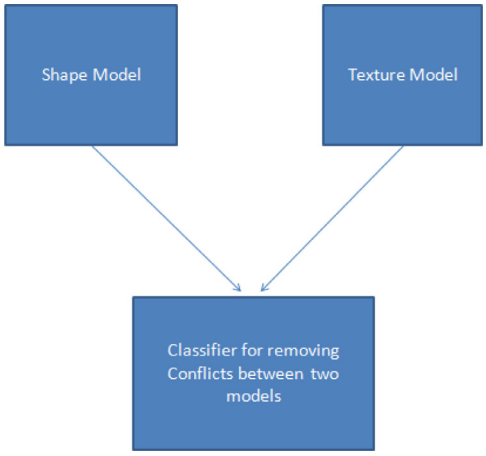


Fig. 16. Composite model

10. Composite aging modeling

Shape and texture models are built from different types of features and it can thus be expected that there will be differences in model outcomes. The question is: Does sufficient synergy exist between the two types of model formulations that can be exploited to obtain even higher accuracy than one type of model operating on its own.

There are two types of scenarios possible with both types of models operating in tandem. The first is that both model types return the same predicted image for a given probe image. This scenario increases confidence in the prediction process, and thus in this case the returned image is selected. The other scenario occurs when both types of models point to different images. Fig. 16 showing illustration of Composite model. This is a non trivial case to deal with as it requires a conflict resolution mechanism. The resolution mechanism that we adopt is a classifier. The conflict scenario can be represented as a two class problem: class 1 corresponds to the case when the belief is that the shape model type has made the right decision while class 2 corresponds to the belief that the texture model has made the right decision. In principle, any type of robust classifier can be utilized here, including the Bayesian that we used for subspace identification. We experimented with different classifiers such as the support vector machine (SVM), Decision tree, Random Forest and Naive Bayes, and found that the Decision Tree returned the highest accuracy amongst all.

The predictor used were: the 7 shape features, 5 texture features and 2 deviation factors df_s and df_t, one each for shape and texture respectively. In addition, we use two confidence measures c_s and c_t for shape and texture respectively, which we define below.

The confidence factor is a measure of the predictive power of the model type on identifying the correct image in the gallery. The closer the distance between the predicted image returned by the model to the desired image the larger should be the confidence. However, it is desirable that the confidence diminishes rapidly in a non-linear manner with distance; a distance of $2d$ between the images should be considered much more than twice as weak as a confidence obtained with a distance of d. Hence our decision to choose an inverse exponential function to measure confidence. In principle, any monotonically decreasing nonlinear function can be used, but our experimentation with different types of such functions has shown that the inverse exponential function defined in Eq. (8) performs best.

$$conf_m = 1/(1 + e_m^d) \tag{8}$$

Table 9
Comparative results of Rank-1 score with FG-NET database

Composite models	Rank-1
Global	41.7%
Local	65.1%
Personalized	71.2%
Integrated	77.6%
(Geng et al., 2007)	38.1%
(Park et al., 2010)	37.4%
(Li et al., 2011)	47.5%
(Gong et al., 2013)	69%
(Gong et al., 2015)	76.2%

Table 10
Comparative results of Rank-1 score with MORPH database

Composite models	Rank-1
Global	21.2%
Local	81.2%
Personalized	90.6%
Integrated	96.7%
(Klare & Jain, 2011)	79.08%
(Park et al., 2010)	79.80%
(Li et al., 2011)	83.90%
(Gong et al., 2013)	91.14%
(Gong et al., 2015)	94.59%
(Li et al., 2016)	94.87%

where subscript m refers to the model type, taking values of either "shape" or "texture" and e is the natural logarithm base

For either of the model types, the distance can be represented as a Euclidean distance between the predictor image feature vector and the actual image feature vector and is defined in expression (9) below.

$$E(P, A) = \sqrt{\sum_i (P_i - A_i)^2} \tag{9}$$

The confidence measures for both model types were generated from the gallery dataset.

11. Integration vs composition

In this set of experiments we investigated the effects of combining shape and texture models at different levels of granularity. We benchmarked our models against the current state-of-the-art one proposed by Gong et al. (2015).

The results in Tables 9 and 10 clearly indicate that the Composite Local and Personalized approaches significantly outperform the Global approach. In fact, the Composite Local and Personalized models are capable of good accuracy by themselves. This is to be expected as aging is a personalized process and its trajectory is very different for different individuals, depending on a range of factors such as lifestyle, genetic disposition and others. At the same time we observe that the Composite Integrated approach also provides further improvement to models operating on their own, indicating that a certain amount of synergy existed between the two modeling approaches. We also note that the Composite Integrated approach significantly outperformed the models proposed by Gong et al. (2015) and Li, Gong, Li, and Tao (2016). Our composite model results are statistically significant with P value for FG-NET is 0.00001 and for MORPH database P value is 0.0000001, which are determined through the use of the t-test with a confidence level of 95% are used.

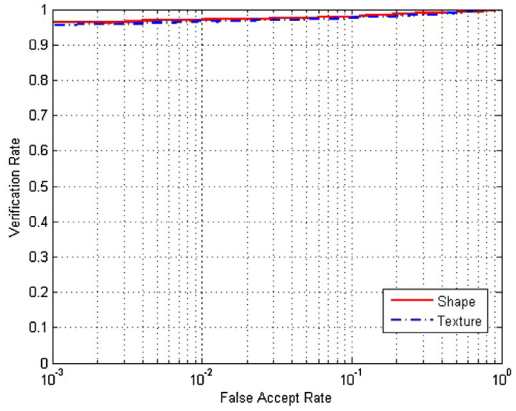


Fig. 17. ROC curve for FG-NET database

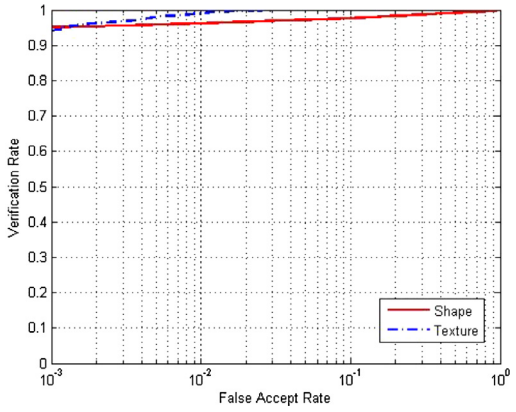


Fig. 18. ROC curve for MORPH database

Table 11
Comparative results of verification with personalized model for shape and texture at the 1% level for FAR and EER with FG-NET database

Models	EER	HTER	VR
Shape	2.62%	1.79%	98.36%
Texture	3.09%	2.04%	98.01%

Table 12
Comparative results of verification with state of art algorithms in terms of EER with FG-NET database

Models	EER
Shape(Personalized)	2.62%
Texture (Personalized)	3.09%
(Du & Ling, 2015)	19.4%
(Ali et al., 2015)	6.51%

Table 13
Comparative results of verification with personalized model for shape and texture at the 1% level for FAR and EER with MORPH database

Models	EER	HTER	VR
Shape	2.38%	3.02%	98.16%
Texture	1.01%	0.94%	99.75%

Table 14
Comparative results of verification with state of art algorithms in terms of EER with MORPH database

Models	EER
Shape(Personalized)	2.38%
Texture (Personalized)	1.01%
Lu, Zhou, Tan, Shang, and Zhou (2014)	7.5%
(Du & Ling, 2015)	5.5%

12. Verification experiments

Face verification is the usual process carried out where it is desired to restrict entrance to some ceratin categories of persons about whom information is already present in the database. It compares the probe image against an image present in the gallery. The probe image claims an identity and the system process verifies whether the stated claim is true and that it corresponds to the claimed identity image in the database. The system must then make a decision about truth or falsehood of the claim. Because, based on this decision, the claimant will be granted or refused access to the sensitive information or place. In the verification experiments, three performance metrics are computed. The first metric is called equal error rate or crossover error rate (EER or CER) which is the rate at which both acceptance and rejection error rates are equal. The second metric is the minimal half total error rate $(HTER) = (FAR + FRR)/2$, where FAR is the false accept rate and FRR is the false reject rate as used in Mihelič and Žibert (2006). We carried out verification experiments with the most common strategy of three fold cross validation for FG-NET and MORPH database as used in Ali, Sagayan, Saeed, Ameen, and Aziz (2015); Du and Ling (2015) . The graphic representation of results is made through a receiver operating characteristic (ROC) curve. It plots the false accept rate against the verification rate. The verification rate is the likelihood of correctly accepting a genuine match. The probability of verification (VR) was calculated and plotted against different FARs. The reference threshold was varied along ROC curve to study the effects. The ROC curves for texture and shape are shown in Figs. 17 and 18.

Tables 11 and 13 show comparative results of verification with Personalized Model for shape and texture at the 1% level for FAR and EER FG-NET and MORPH databases.

Table 12 summarizes the performances of the proposed algorithm along with other state-of-the-art approaches. It shows that our Personalized models for shape and texture an ERR of 2.62% and 3.09% ,which represents a significant improvement over the best published result of (6.51%) for the FG-NET database.

Table 14 summarizes the performances of the proposed algorithm along with other state-of-the-arts. It shows that our Personalized models for shape and texture with ERR rates of 2.38% and 1.01%; both of which significantly improves over the best published result (5.5%) for the MORPH database.

13. Conclusion and future work

In this research we presented shape and texture models at different levels of data granularity. The Personalized model, where a model is built for each individual at the lowest level of granularity yielded the best rank 1 recognition rate. This indicates that aging to large extent is an individualized process and is in agreement with other studies (Lanitis, Taylor, & Cootes, 2002).

Our method of dividing the whole database into subspaces and using the Naive Bayesian classifier to narrow down the search space also proved to be a success. This limits the search space and improves our recognition accuracy.

In the case of the Local model we further reduced the search space by first dividing the whole database into sub populations

Table 15
Comparison of different models

Models	Complexity	Effectiveness
Global	Low	Low
Local	Low	High
Personalized	High	High
Integrated	High	High

Table 16
Aging models when to use

Models	When to used
Global	When results required are required in very short time
Local	When good accuracy is required in short time
Personalized	When high accuracy is required and recognition time is not a problem
Integrated	When very high accuracy is required and recognition time is not a problem

(clusters) and then considered subspaces within each subpopulation. The Local model trades-off accuracy with recognition time. Reduction in recognition time follows directly from restricting the search space to a cluster rather than the entire dataset. Experimentation with the FG-NET database resulted in a reduction of recognition time by a factor of 60% in relation to the Personalized model approach but a reduction in rank 1 accuracy also resulted due to the fact that image recognition decisions are made on localized data rather than on the entire dataset.

The Integrated model which consisted of a combination across granularity level (Global, Local, Personalized) resulted in the highest accuracy because of the synergy that existed between the models. The increased accuracy did come at the cost of extra computation time, increasing the recognition time 35% in relation to the Personalized model approach.

Another idea that we explored in this research was model composition between models built from shape and texture features. It turned out that model composition, similar to model integration resulted in better accuracy. The final conclusion of the study was that the optimal modeling approach was to compose shape models built by integrating across data granularity levels with their corresponding texture counterparts using the same data granularity scheme.

A brief comparison in tabular form is given in Table 15. Table 16 compares the models on two factors, complexity and effectiveness. Complexity relates to computation time and effort involved in model generation. The effectiveness relates to accuracy. Based on Table 15, Table 16 was prepared, which suggests which of the models to be used depending on the end user's requirements. For highest accuracy the Integrated Model is suggested and when time is at a premium, the Local Model is a good compromise as it has a lower recognition time than the Personalized model while having better accuracy than the Global model.

Our aging models are also capable of forward traversal through time. Forward traversal has two attractive benefits. The first is automatic update of the image database. Over a period of time an image gallery containing older images of a person becomes less useful for recognition purposes as research has shown that the longer the time gap between the probe image and the target image, the lower is the rank 1 accuracy in general (Park et al., 2010). The solution to this problem either involves obtaining newer images from the subjects concerned which may not be practicable as some subjects may not be available. A more practical alternative would be an automatic refresh based on a computed trajectory provided by the aging functions proposed in this research. The second benefit lies in age simulation whereby the appearance of given person at a future point in time is required for applications as face animation or for making decisions at a future point in time on plastic surgery alternatives.

References

Abdullah, M. F. A., Sayeed, M. S., Muthu, K. S., Bashier, H. K., Azman, A., & Ibrahim, S. Z. (2014). Face recognition with symmetric local graph structure (slgs). *Expert Systems with Applications, 41*(14), 6131–6137.

Ali, A. S. O., Sagayan, V., Saeed, A. M., Ameen, H., & Aziz, A. (2015). Age-invariant face recognition system using combined shape and texture features. *IET Biometrics, 4*(2), 98–115.

Alvi, F. B., & Pears, R. (2015a). An integrated modeling approach to age invariant face recognition. *Sixth international conference on graphic and image processing (icgip 2014)*. International Society for Optics and Photonics. 94430A–94430A

Alvi, F. B., & Pears, R. (2015b). Texture modelling for age invariant face recognition. In *Proceedings of the international conference on image processing, computer vision, and pattern recognition (ipcv)* (p. 24). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 679–698.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1999). Comparing active shape models with active appearance models.. In *Bmvc: vol. 99* (pp. 173–182).

Du, L., & Ling, H. (2015). Cross-age face verification by coordinating with cross-face age verification. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2329–2338).

Farkas, L. G. (1994). *Anthropometry of the head and face*. Raven Pr.

Farkas, L. G., & Munro, I. R. (1987). *Anthropometric facial proportions in medicine*. Charles C. Thomas Publisher.

FG-NET (2002). Fg-net database. http://www-prima.inrialpes.fr/FGnet/.

Geng, X., Zhou, Z.-H., & Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(12), 2234–2240.

Gong, D., Li, Z., Lin, D., Liu, J., & Tang, X. (2013). Hidden factor analysis for age invariant face recognition. In *Proceedings of the ieee international conference on computer vision* (pp. 2872–2879).

Gong, D., Li, Z., Tao, D., Liu, J., & Li, X. (2015). A maximum entropy feature descriptor for age invariant face recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 5289–5297).

Goshtasby, A. A. (2012). *Image registration: Principles, tools and methods*. Springer Science & Business Media.

Jain, A. K., Nandakumar, K., & Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*.

Kasabov, N. (2007). Global, local and personalised modeling and pattern discovery in bioinformatics: An integrated approach. *Pattern Recognition Letters, 28*(6), 673–685.

Klare, B., & Jain, A. K. (2011). Face recognition across time lapse: On learning feature subspaces. In *International joint conference on biometrics (ijcb)* (pp. 1–8). IEEE.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1999). Modeling the process of ageing in face images. In *The proceedings of the seventh ieee international conference on computer vision: Vol. 1* (pp. 131–136). IEEE.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (2002). Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(4), 442–455.

Li, Z., Gong, D., Li, X., & Tao, D. (2016). Aging face recognition: A hierarchical learning model based on local patterns selection. *IEEE Transactions on Image Processing, 25*(5), 2146–2154. doi:10.1109/TIP.2016.2535284.

Li, Z., Park, U., & Jain, A. K. (2011). A discriminative model for age invariant face recognition. *IEEE Transactions on Information Forensics and Security, 6*(3), 1028–1037.

Ling, H., Soatto, S., Ramanathan, N., & Jacobs, D. W. (2007). A study of face recognition as people age. In *Ieee 11th international conference on computer vision (iccv)* (pp. 1–8). IEEE.

Lu, J., Zhou, X., Tan, Y.-P., Shang, Y., & Zhou, J. (2014). Neighborhood repulsed metric learning for kinship verification. *IEEE transactions on pattern analysis and machine intelligence, 36*(2), 331–345.

Mihelič, F., & Žibert, J. (2006). Robust speech detection based on phoneme recognition features. In *International conference on text, speech and dialogue* (pp. 455–462). Springer.

Milborrow, S., & Nicolls, F. (2014). Active shape models with sift descriptors and mars.. In *Visapp (2)* (pp. 380–387).

Park, U., Tong, Y., & Jain, A. K. (2010). Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(5), 947–954.

Pears, R., Widiputra, H., & Kasabov, N. (2013). Evolving integrated multi-model framework for on line multiple time series prediction. *Evolving Systems, 4*(2), 99–117.

Pujol, F. A., & García, J. C. (2012). Computing the principal local binary patterns for face recognition using data mining tools. *Expert Systems with Applications, 39*(8), 7165–7172.

Ramanathan, N., & Chellappa, R. (2006). Modeling age progression in young faces. In *Ieee computer society conference on computer vision and pattern recognition: Vol. 1* (pp. 387–394). IEEE.

Ricanek Jr, K., & Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. In *7th international conference on automatic face and gesture recognition (fgr)* (pp. 341–345). IEEE.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing, 26*(1), 43–49.

Syambas, N. R., & Purwanto, U. H. (2012). Image processing and face detection analysis on face verification based on the age stages. In *7th international conference on telecommunication systems, services, and applications (tssa)* (pp. 289–293). IEEE.

Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory*: 1. Wiley New York.

Štruc, V., & Pavešić, N. (2010). The complete gabor-fisher classifier for robust face recognition. *EURASIP Advances in Signal Processing, 2010*, 26.

Widiputra, H. (2011). *Integrated multi-model framework for adaptive multiple time-series analysis and modelling*. Auckland University of Technology Ph.D. thesis..

Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing* p. 1. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 491.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.