Comparing Different Classifiers for Automatic Age Estimation

Andreas Lanitis, *Member, IEEE*, Chrisina Draganova, and Chris Christodoulou, *Member, IEEE*

***Abstract*—We describe a quantitative evaluation of the performance of different classifiers in the task of automatic age estimation. In this context, we generate a statistical model of facial appearance, which is subsequently used as the basis for obtaining a compact parametric description of face images. The aim of our work is to design classifiers that accept the model-based representation of unseen images and produce an estimate of the age of the person in the corresponding face image. For this application, we have tested different classifiers: a classifier based on the use of quadratic functions for modeling the relationship between face model parameters and age, a shortest distance classifier, and artificial neural network based classifiers. We also describe variations to the basic method where we use age-specific and/or appearance specific age estimation methods. In this context, we use age estimation classifiers for each age group and/or classifiers for different clusters of subjects within our training set. In those cases, part of the classification procedure is devoted to choosing the most appropriate classifier for the subject/age range in question, so that more accurate age estimates can be obtained. We also present comparative results concerning the performance of humans and computers in the task of age estimation. Our results indicate that machines can estimate the age of a person almost as reliably as humans.**

***Index Terms*—Aging, face recognition, image classification, neural networks.**

I. INTRODUCTION

FACE IMAGES convey a significant amount of information including information about the identity, emotional state, ethnic origin, gender, age, and head orientation of a person shown in a face image. This type of information plays a significant role during face-to-face communication between humans. The use of facial information during interaction is made possible by the remarkable ability of humans to accurately recognize and interpret faces and facial gestures in real time.

Current trends in information technology dictate the improvement of the interaction between humans and machines, in an attempt to upgrade the accessibility of computer systems. As part of this effort, many researchers have been working in the area of automatic interpretation of face images [5], [11], [22] so that contact-less human–computer interaction (HCI) based on facial gestures can be developed. In this context, systems capable of identifying faces [14], recognizing emotions [19], gender [17], and head orientation [16] have been developed. Despite the fact that the age of a person plays an important role during interaction, so far no researcher has been involved in designing automatic age estimation systems based on face images. With our work, we aim to produce a system which is capable for estimating the age of a person as reliably as humans.

Manuscript received January 23, 2003; revised June 23, 2003. This work was supported by a research grant from the Research Promotion Foundation of the Republic of Cyprus (Project PROSOPO—26/2001).

A. Lanitis is with the Department of Computer Science and Engineering, Cyprus College, Nicosia, Cyprus (e-mail: alanitis@cycollege.ac.cy).

C. Draganova is with the Department of Computing, Communication Technology and Mathematics, London Metropolitan University, London EC3 1JY, U.K. (e-mail: c.draganova@londonmet.ac.uk).

C. Christodoulou is with the School of Computer Science and Information Systems Birkbeck College, University of London, London WC1E 7HX, U.K. (e-mail: chris@dcs.bbk.ac.uk).

Digital Object Identifier 10.1109/TSMCB.2003.817091

A. Motivation

The motivation behind our work lies in the important real life applications of the proposed methodology. In summary, those applications include the following.

- *Age specific human computer interaction*: If computers could determine the age of the user, both the computing environment and the type of interaction could be adjusted according to the age of the user. Apart from standard HCI, such a system could be used in combination with secure internet access control in order to ensure that under-aged persons are not granted access to internet pages with unsuitable material.
- *Age-based indexing of face images*: Automatic age estimation can be used for age-based retrieval of face images from databases. The most common application of this technology is in e-photoalbums, where users could have the ability to retrieve their photographs by specifying a required age-range.
- *Development of automatic age progression systems*: Automatic age estimation systems rely on their ability to understand and classify changes in facial appearance due to aging. The methodology required in this task could form the basis of designing automatic age progression systems (i.e., systems with the ability to predict the future facial appearance of subjects). A description of our early work in this area is described elsewhere [15].
- *Understanding the process of age perception by humans*: Work in the area of automatic age estimation could provide invaluable help to psychologists who study the topic of age perception by humans.

B. Overview of Our Work

The basis of our approach is a statistical model of facial appearance [8], which models the variability in the facial appearance due to all systematic sources of variability. Face models of this type are generated based on a statistical analysis of the shape and intensity variation in a representative sample of face images. The most important feature of these models is the ability

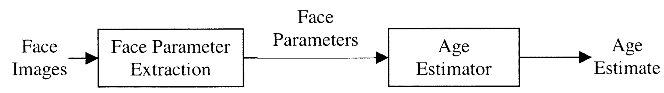

Fig. 1. Block diagram of the age estimation approach.

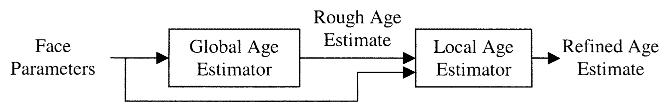

Fig. 2. Block diagram of the age specific age estimation approach.

to code faces into a small number of parameters—the "face parameters." This representation can be used as the basis for designing algorithms suitable for automatic face recognition, expression recognition, and determination of the head orientation [14]. In our work, we aim to use this representation for estimating the age of an individual as shown in the block diagram of Fig. 1.

For this purpose, we investigated the use of the following methods for designing an age estimator.

Quadratic Functions: Optimization methods have been used for defining the optimum coefficients of quadratic functions, which can best explain the relationship between the face parameters and the age of a person in the corresponding face image. Once the functions are established, we use them for transforming a given set of face parameters to an estimate of the age.

Shortest Distance Classifier: Based on the training data, the distributions of face parameters corresponding to a certain age are defined. Given a new set of face parameters, we assign it to the closest distribution in order to estimate the age.

Supervised Neural Networks: Supervised neural networks have been trained with a set of face parameters and their corresponding ages so that given an unknown set of parameters they produce at the output an estimate of the age of the person in the corresponding face image.

Unsupervised Neural Networks: The Kohonen Self-Organizing Map (SOM) [13], which is a clustering algorithm, has been utilized to train networks to classify a set of input vectors of face parameters in a number of clusters corresponding to different age groups. Given a new vector of face parameters, the trained networks determine the age group of the person corresponding to the face image.

We also investigated the use of hierarchical age estimation methods by using age specific and/or appearance specific classifiers. In the case of age specific classifiers, we used a global classifier for performing rough age estimation followed by the use of a classifier specific to the age-range indicated by the global age estimator, so that the age estimate is refined. A block diagram of this approach is shown in Fig. 2.

In order to perform age estimation using appearance specific classifiers, we train age classifiers for different clusters of subjects within our training set. The formation of the clusters is based on the appearance of the subjects and the similarity in the aging pattern adopted by each subject. We also train a classifier used for selecting the most suitable cluster given the parametric representation of a new face image. As a result, for different subjects we may use a different age classifier according to the facial appearance of the subject in question. A block diagram of this approach is shown in Fig. 3.

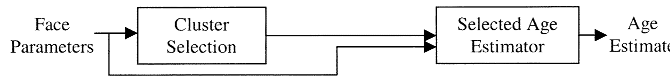

Fig. 3. Block diagram of the appearance specific age estimation approach.

A combination of both the appearance and the age specific age estimation algorithms was also investigated. Based on the methods presented earlier, we describe age estimation experiments and report comparative results between the performances of different algorithms. We have also asked a number of volunteers to estimate the age of subjects in the typical sample of face images from our test set so that we could compare the performance of each method against the performance of humans.

C. Paper Organization

The remainder of the paper is organized as follows. In Section II, we present a brief overview of the relevant bibliography. In Section III, we briefly describe how the face models are generated. In Section IV, we present the age estimation methods used in our experiments and, in Section V, we present the results obtained. Finally, in Section VI, we discuss the results and present our conclusions.

II. Literature Review

Although the general topic of face image processing received considerable interest [5], [11], [22], only a small number of researchers carried out research work in the area of modeling and/or simulating aging effects on face images.

D'Arcy Thompson [7] suggested that it is possible to use coordinate transformations for altering the shape of biological organizms in order to produce shapes belonging to different but similar biological organizms. Based on this idea, a number of researchers [1], [20], [21] investigated the use of coordinate transformations in an attempt to impose age-related changes on human faces. According to their experimental evaluation, the perceived age of transformed facial outlines can be altered according to the transformation factors used.

Burt and Perrett [2] investigated the process of aging using face composites from different age groups and caricature algorithms. In order to simulate age effects on previously unseen face images, they calculate the differences in shape and color between the composites for the 25–29 years and 50–54 years age groups. By superimposing those differences into the color and shape of subjects in new images, the perceived age of the subject is increased. According to their experimental evaluation, the proposed method produced a significant increase in the perceived age of subjects.

O'Toole *et al.* [18] use three-dimensional (3-D) facial information for building a parametric 3-D face model. They use a caricature algorithm in order to exaggerate or de-emphasize distinctive 3-D facial features; in the resulting caricatures the perceived age is increased or decreased according to the exaggeration level, suggesting that 3-D distinctive facial features are emphasized in older faces. Their findings were verified by a number of observers.

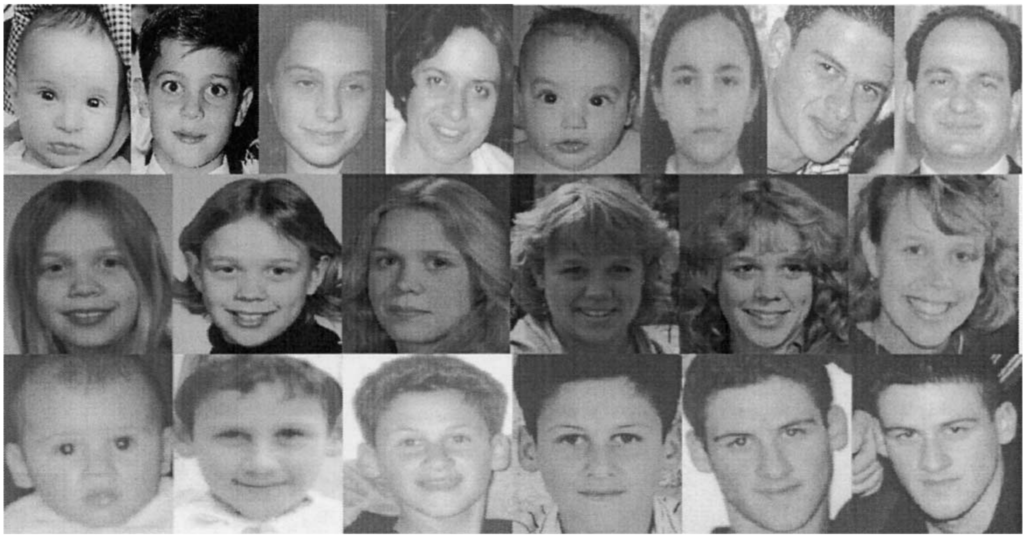

Fig. 4. Typical images used in our experiments. The top row shows images from different individuals, whereas the middle and bottom row show images of the same individual at different ages.

Wu and Thalmann [26] use a physically based face model where facial skin is modeled as a nonlinear and inelastic material. The appearance of the skin is controlled by the parameters of the model used—it is possible to simulate skin wrinkles by varying those parameters.

All the approaches described have been used for simulating age effects rather than estimating the age of subjects. Our work on automatic age estimation is one of the first attempts to solve the problem. In our work, we use statistical face models generated by applying Principal Component Analysis (PCA) on an ensemble of face images. Kirby and Sirovich [12] first used models of this form in an attempt to achieve low-bit coding of face images. Based on this methodology, Turk and Pentland [24] describe an automatic face identification system that uses the weights of the basis face images (or eigenfaces) as the feature vector during the classification procedure. As an extension to this technique, Craw *et al.* [6] suggest that before the PCA decomposition is applied the shape of faces should be normalized so that eigenfaces capture only variation in the intensities of face images. A complete representation of face images can be obtained by applying PCA based decomposition on shape and texture facial information independently [14], [25]. In this context, the shape and intensity of a face are coded in terms of the parameters of a shape model and an intensity model respectively. Edwards *et al.* [8] describe how a shape model and an intensity model can be combined in order to produce a combined shape-intensity face model capable of modeling effectively both shape and intensity variation—the face models used in our work are based on the work reported by Edwards *et al.* [8], [9]. Models of this form have been used widely in applications involving the location [4] and interpretation of face images [10].

III. Face Model

In this section, we briefly describe the process of building the face models used in our work, and describe the face database used in our experiments.

A. Building a Face Model

Cootes *et al.* [3] propose a method for generating statistical models from a set of training examples. During the process, we perform Principal Component Analysis (PCA) on the deviations of each example from the mean example. As a result of the analysis, training examples can be reconstructed/parametrized using

$$\mathbf{X} = \mathbf{X_m} + \mathbf{Pb} \tag{1}$$

where $\mathbf{X}$ is a vector describing the shape or intensity pattern of a training example, $\mathbf{X_m}$ is the mean example, $\mathbf{P}$ is the matrix of eigenvectors, and $\mathbf{b}$ is a vector of weights, or *model parameters*. Edwards *et al.* [8] describe in detail how this type of model can be used for modeling combined shape and intensity variation in face images. As an extension to the basic technique, Edwards *et al.* [9] describe how color models can be generated by including in the intensity pattern the red green and blue component of each pixel in the facial region.

B. Face Database

In our work, we use color face models generated based on the method proposed by Edwards *et al.* [9]. In order to train the models, we used images from a database containing 400 color face images of 40 individuals where each individual in the database supplied a collection of photographs taken over many years (average three-year intervals). The age of the individuals shown in the face images of the database, range between newborns up to 35 years. Typical images from the database used in our experiments are shown in Fig. 4. Due to the difficulty in obtaining suitable images for our database, the number of images per individual is not constant; i.e., there are individuals with less than ten images in the database whereas for some individuals it was possible to obtain more than ten images. For the needs of our experiments, the database was split into two parts, used for training and testing (and vice versa) the age estimation algorithms presented in subsequent sections of the paper. The division of the database in each part was done in such a way that a single individual does not have images in both sets.

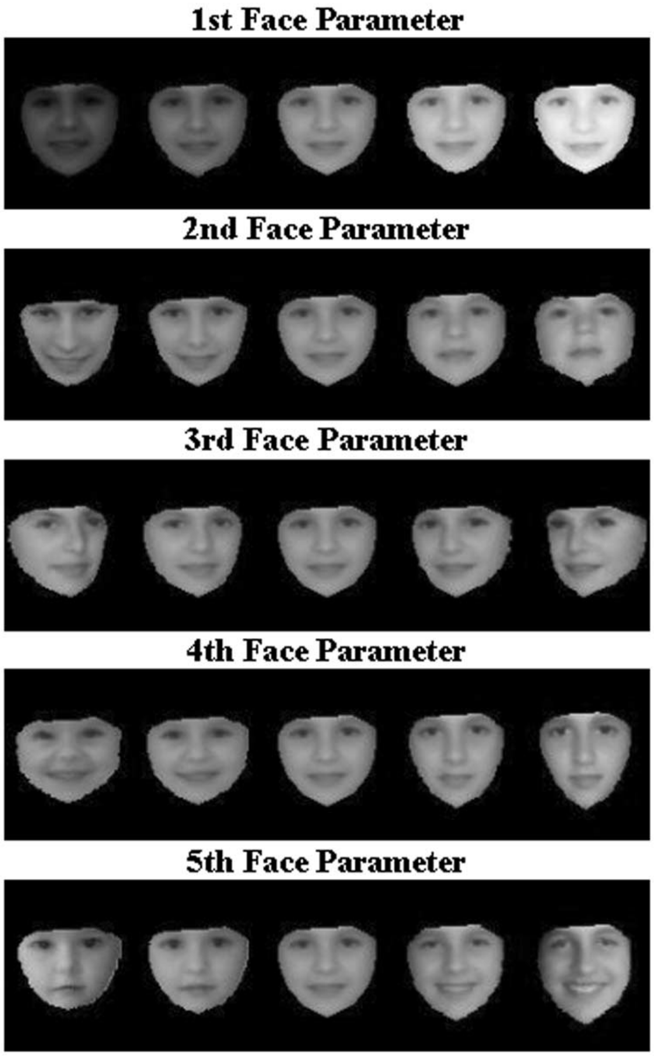

Fig. 5. Main modes of variation within the training set. Each row shows the effect of varying a single face model parameter.

C. Face Models Used in Our Experiments

We have generated two color face models of the internal facial region, using images from the two parts of the database. We have chosen to model only the internal region of the face, since due to the enormous variability of the hairline in face images, the introduction of hairline information deteriorates the ability of the model to explain adequately other important types of variation.

Each face model trained requires 22 model parameters for explaining most of the variation within the training set. Since those parameters can be used for describing the appearance of a face image we refer to them with the term "face parameters." It is important to note that given a face image, it is possible to code the appearance of the face into the face parameters [4], [10] and subsequently reconstruct the appearance based on the face parameters. Each of the model (or face) parameters controls a systematic way in which face images vary within the training set. As illustrated in Fig. 5, face parameters control changes in lighting, 3-D pose, expression, and individual appearance.

IV. Age Estimation

We use the face model-based coded representation of face images as the basis for estimating the age of a subject shown in a face image. In this context, we aim to define the relationship between the model-based parametric description of faces and age, so that given a set of face parameters, we will be able to estimate the age of the person shown in the image. Since face parameters describe the appearance of a face, in effect we aim to train a classifier that estimates the age of a person based on the facial appearance of the internal face.

In the remaining part of this section, we describe the single step age estimators and hierarchical age estimators used in our experiments. Also, a comparative experiment involving age estimation by human observers is described in Section IV-C.

A. Single Step Age Estimation

In the case of single step age estimation, the age estimation task is performed using a single classifier. In our experiments, we used four different types of classifiers: a classifier based on quadratic functions, a shortest distance classifier, a Multilayer Perceptron (MLP) based classifier, and a Self-Organizing Map (SOM) classifier.

1) Quadratic Models: We define the relationship between face parameters and age, based on quadratic functions of the form

$$age = c + \mathbf{W_1 b^T} + \mathbf{W_2 (b^2)^T} \qquad (2)$$

where

age actual age of an individual in a face image;
$\mathbf{b}$, $\mathbf{b^2}$ are vectors containing the 22 face parameters and the squares of the 22 parameters, respectively;
$\mathbf{W_1}$,$\mathbf{W_2}$ vectors containing weights for each element of $\mathbf{b}$ and $\mathbf{b^2}$, respectively;
c offset required.

We treat the problem of calculating the weights and offset as an optimization problem, where we seek to minimize the difference between the actual ages of individuals in a training set and the ages estimated using (2). Once the optimum parameters of the quadratic function are defined, we use (2) for estimating the age of a person, given the face parameters corresponding to his/her face. A more detailed description of our work in this area appears elsewhere [15].

2) Shortest Distance Classifier: All training faces are coded into the face parameters and the distribution of face parameters for each age within the age range of interest is defined. In effect we define the cloud of face parameters for each age in the 22-dimensional face parameter space. For example we establish the distribution of face parameters among all samples with age 35, the distribution ace parameters among all samples with age 34, etc. Given a new set of face parameters we calculate the Mahalanobis distance (d) between the given set of face parameters and the centroid of each distribution corresponding to each age, using (3):

$$\mathrm{d} = (\mathbf{b_n} - \mathbf{b_m})\mathbf{C^{-1}}(\mathbf{b_n} - \mathbf{b_m}) \qquad (3)$$

where

$\mathbf{b_n}$ vector containing the face parameters of an unseen face image;

$\mathbf{b_m}$ vector of the mean face parameters for a particular age;

$\mathbf{C}$ within-class covariance matrix of the face parameters derived from the training images.

The age of the new individual is estimated as the age corresponding to the distribution, which minimizes the Mahalanobis distance (d) between the parametric description and the centroid of that distribution. In effect, the Mahalanobis distance provides an indication of the probability that a certain set of face parameters belongs to any age group.

3) Supervised Neural Networks: We have investigated the use of Multilayer Perceptrons (MLPs) with the back propagation learning algorithm [23] for estimating the age of a subject given a set of face parameters. Based on the training set, each type of network is evaluated in order to establish the optimal architecture and optimal parameters. In each case, the generalization capability of the neural network is assessed as a function of the initial parameters of the respective network.

For our purpose, the 22 model parameters representing each face image are used as an input vector to the MLP. The output layer is a single neuron representing the corresponding age of each face scaled in the interval [0,1]. According to our experiments, the optimal network architecture and parameters are as follows: one hidden layer with 15 hidden neuronodes, learning rate equal to 0.2, and momentum equal to 0.7. The processing time for training the network is in the order of minutes.

4) Unsupervised Neural Networks: The Kohonen Self Organizing Map (SOM), which is an unsupervised type neural network [13], has been used in our experiments.

The aim of our experiments is to transform the input vectors of 22 parameters into a single two-dimensional (2-D) map organized in a number of clusters representing different age ranges. We perform several experiments, varying the number of clusters and the size of the 2-D map. In particular, we consider the cases where the number of clusters is seven, nine, and 12, where each cluster corresponds to an age range (age range five, four, and three, respectively). For example, in the case of seven clusters, the age range is five years (0:4, 5:9, 10:14, 15:19, 20:24, 25:29, and 30:35). The size of the 2-D map varies as follows: 5×5, 10×10, 15×15, 20×20, and 25×25 neurons. For each experiment, the initial learning rate is set to 0.5 and a Gaussian neighborhood function is used with initial width equal to the respective dimension of the grid (5, 10, 15, 20, or 25). All parameters mentioned above are adjusted during the process of training the network. The experiments are run for 5000 to 15 000 epochs requiring considerable computing power. The processing time for training increases with increasing the size of the grid and the number of clusters for classification—in the case that the grid size is greater than 10×10, the total processing time is in the order of several hours.

After training and labeling is performed with the SOM algorithm, the error (E) in years for this algorithm is calculated as follows:

- check in which age cluster each input is classified;
- let g be the target range age and a_g the target age for input vector $\mathbf{b}$ classified in the age range $c = (c_{min}, c_{max})$;
- if $g > c$ then $\mathrm{E} = a_g - c_{max}$;
- if $g < c$ then $\mathrm{E} = c_{min} - a_g$;
- if $g = c$ then $\mathrm{E} = 0$.

The best results are achieved for the network with grid size 20×20 neurons and classification in 12 clusters. Learning Vector Quantization (LVQ) [13] has also been utilized for further optimization.

B. Hierarchical Age Estimation

In the case of hierarchical age estimation, the classification process is performed by using a numbers of different classifiers instead of a single classifier. According to the way in which different classifiers are trained and used, we have investigated three variations of hierarchical age estimation. The first one involves age-specific classifiers, the second appearance specific, and the third is a combination of the two types of classifiers already mentioned. The actual classifiers used as part of hierarchical age estimation methods can be of any type among the ones described in Section IV-A, i.e., they can be classifiers based on quadratic functions, shortest distance classifiers, MLP classifiers and SOM classifiers. However, for reasons explained later (see Section V) we did not use SOM based classifiers for hierarchical age estimation.

1) Age-Specific Classifiers: Age specific classifiers are classifiers trained to estimate ages only between a certain age-range. By training classifiers for a specific age range, we aim to enhance the ability of the classifier to estimate accurately the age of persons in the age range in question. When using age specific classifiers apart from the age specific classifiers, we also need a global age classifier, used for defining the possible age range of a subject shown in a face image.

During the training procedure, we first train a global age classifier (similar to the classifiers described in the previous sections) using all the training data, which contains subjects with ages ranging from zero to 35 years. Apart from the global classifier, we also train a local age classifier using samples from a specific age range. For the experiments described in this paper, we have trained three age specific classifiers for the age ranges of 0–10 years, 11–20 years, and 21–35 years, respectively. During the process of age estimation, we use the global classifier for rough age estimation, followed by the use of the corresponding local classifier for the specific age range indicated by the results of the rough age estimation procedure. It is expected that the use of local age classifiers will produce more accurate age estimates than the global age estimator. A block diagram of age estimation using age specific classifiers is shown in Fig. 2.

2) Appearance Specific Classifiers: This approach is based on the observation that people who look similar tend to age in a similar way [15]. In order to take advantage of this observation, we train individual age classifiers for different clusters of subjects who look similar and/or exhibit similar aging pattern in our training set. We call those classifiers appearance specific, since they are specific to clusters of subjects resembling similarities in appearance. Appearance specific classifiers can be used for more accurate age estimation, since such classifiers associate more accurately facial appearance and age in the cluster in question.

When using appearance specific classifiers, we train a bank of appearance specific classifiers for different clusters of subjects in the training set. During the age estimation procedure, we also need a classifier for selecting the most appropriate appearance specific classifier to be used—we call this classifier the *cluster selection classifier*. In the following sections, we describe how appearance specific classifiers are trained, how the cluster-selection classifier is trained, and we also outline the main steps involved during the age estimation process.

Training appearance specific classifiers: The formation of the clusters is done during the training procedure by iteratively training age estimators, until the age estimation error within the training set is minimized. The algorithm used for this application is as follows:

1) Use the training set in order to train an age classifier based on any of the methods described in Section IV-A.
2) For each sample in the training set, calculate the error between the actual and estimated age. All examples for which the error is smaller than a threshold are eliminated from the training set, since the current classifier copes well with those examples. The training set now contains all samples for which the age estimation error is unacceptable.
3) Go back to step 1. The procedure terminates until most (usually 95%) of the training examples are classified correctly.

In our experiments, usually about five to seven age estimators are required in order to minimize the age estimation error in the training set, implying that that the use of appearance specific age estimators can deal with diversities in the aging pattern adopted by different subjects.

Training the cluster-selection classifier: Based on the training samples in each of the clusters defined earlier, we train a cluster selection classifier used for selecting one of the clusters, given a vector of face parameters presented to the system. In the case of quadratic age classifiers and shortest distance age classifiers, the cluster selection classifier is a shortest distance classifier, whereas in MLP based age classifiers, the cluster selection classifier is an MLP classifier.

Age estimation using appearance specific classifiers: Given a set of previously unseen face parameters we first use the cluster selection classifier in order to assign the new set of face parameters to one of the clusters. Once the most appropriate cluster is selected the age estimator for the selected cluster is used for producing an age estimate. A block diagram of the method is shown in Fig. 3.

3) Appearance-Specific and Age-Specific Classifiers: The methods described in Sections IV-B1 and IV-B2 were combined in order to perform age estimation using both appearance specific and age specific age classifiers. In this context, we first use the algorithm presented in the previous section for defining clusters of subjects within our training set. Instead of using a single age classifier for each cluster, we train classifiers for different age ranges in each cluster in the training set. Following the approach used earlier, for each cluster we train age specific classifiers for the age ranges of 0–10 years, 11–20 years, and 21 to 35 years, respectively.

For this approach, we need a cluster selection classifier in order to select the most appropriate cluster and also, for each cluster, we need a global age estimator in order to select the most appropriate age specific classifier in the selected cluster.

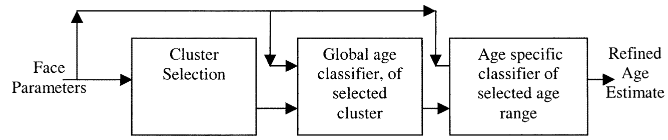

Fig. 6. Block diagram of the appearance and age specific age estimation approach.

During the age estimation process, we assign a set of given face parameters to one of the clusters defined earlier. The global age estimator for the cluster in question is used for selecting a specific age range, so that the age estimator for that age range, for the cluster in question, is used for estimating the age. A block diagram of the method is shown in Fig. 6.

In the case of using MLP based classifiers for hierarchical age estimation the optimal structure of each network was with one hidden layer and ten, 15, 20, or 25 hidden units. The learning rate varied between 0.2 and 0.3 and the momentum between 0.5 and 0.9. The processing time for training the neural networks with the MLP was in the order of minutes.

C. Age Estimation by Humans

Humans are not perfect in the task of estimating the age of subjects based on facial information. The accuracy of age estimation by humans depends on various factors, such as the ethnic origin of a person shown in an image, the overall conditions under which the face is observed, and the actual abilities of the observer to perceive and analyze facial information. The aim of this experiment was to get an indication of the accuracy in age estimation by humans, so that we can compare their performance with the performance achieved by machines.

As part of the experiments, a number of face images were presented to 20 volunteers, and each volunteer was asked to estimate the age of the subject shown. It should be noted that none of the volunteers knew any of the subjects whose images were included in the test set.

It is important to point out that, in the case of age estimation by humans, the whole face image (including the hairline) was presented to observers, whereas in the case of age estimation by machines, in effect only information from the internal facial area was utilized.

V. Experimental Investigation

In this section, we describe the experiments conducted in order to assess the suitability of the methods presented in Section IV in the task of automatic age estimation. In particular, we tested the performance of single step classifiers based on quadratic functions, shortest distance classifiers, MLP based classifiers, and SOM based classifiers. In the case of hierarchical age estimation, we used all types of classifiers apart from the SOM based classifier. Although the unsupervised Kohonen SOM network can theoretically be used in hierarchical age estimation, we decided not to use it because in effect the SOM performs the process of classification in a manner similar to the one utilized in hierarchical age estimation. In the case of the SOM based classifier, the output is an age range rather than the exact age estimate, and if it were to be used in a hierarchical age estimation the result would be a cluster within a cluster; in other words, smaller-based cluster giving a

TABLE I
RESULTS OF OUR EXPERIMENTS (ALL NUMBERS SHOW THE ABSOLUTE ERROR IN YEARS)

Method	Quadratic	Shortest Distance	MLP	SOM	Human Observers		
					Male	Female	All
Single Step	5.04	5.65	4.78	4.9	3.69	3.59	3.64
Age Specific	4.87	5.02	4.52	N/A			
Appearance Specific	4.61	5.58	4.64	N/A			
Appearance and Age Specific	3.82	4.92	4.38	N/A			

narrower age-range. This should theoretically give a better result. This is achieved, however, in an equivalent way by varying the number of the clusters we expect to end up with during training, which is what we do (see Section IV-A4); the higher the number of clusters, the narrower the age range. Therefore, in the case of SOM classifiers, we can improve the performance by increasing the number of clusters rather than employing hierarchical age estimation. In addition, the processing time for a SOM hierarchical age estimation would have increased enormously (currently is in the order of several hours, see Section IV-A4), which would have made the technique impractical and offset any possible improvement in the performance.

A. Experimental Setup

For our experiments, we used 400 images divided into two sets of 200 images in each set. Each set contains images showing 20 persons at ages ranging from zero to 35 years. During the experimental investigation, we train a face model using the images in the training set (200 images for each case), code the images of the training and test set into the corresponding face model parameters, and use the resulting representation for training and testing the age estimation classifiers described in Section IV. The whole process was repeated by interchanging the training and test set, so that we could obtain results on a larger number of test cases. The performance of each method was assessed by calculating the mean absolute error between the real and estimated ages among the test sets. During the tests, we assumed that the images were already coded into the face parameters, thus the results quoted refer only to the process of age estimation given a set of face parameters. Errors in the coding procedure, caused by mislocalization of faces in images, are not affecting our results.

A random subset of 32 images was used for testing the accuracy of humans in the task of age estimation. Those images were presented to 20 observers who were asked to estimate the age of the subjects shown in the images.

B. Results

The overall results obtained are shown in Table I.

C. Discussion

When a single step classifier is used the classifiers based on multilayer perceptrons, self-organizing maps and on quadratic functions achieved the best performance (4.78, 4.9, and 5.04 respectively)—the performance of the shortest distance classifier was not as good. We believe the shortest distance classifier is not suitable in our framework because it requires a large number of training samples in order to describe adequately the distribution of face parameters for each age group. In the case of using a relatively small training set, it is possible that outliers significantly influence the distributions resulting in the deterioration of the age estimation performance. In the case of using self-organizing maps, the performance is related to the size of the clusters in terms of age differences. The best results with the SOM (presented in Table I) are with clusters of three years (see Section IV-A4); with clusters of four years the mean error is 5.3, while with clusters of five years, the mean error is 5.5 years. The SOM algorithm as an unsupervised technique is the only one applicable for cases where we do not have explicit information about the age of each subject shown in the training face images.

The use of hierarchical age classifiers improves the age estimation accuracy. Age estimation based either on age specific classifiers or appearance specific classifiers produced improved results when compared with the results obtained when single step classifiers were used. This observation suggests that the process of aging causes different types of deformations in facial appearance according to the age-range of a subject and his/her facial appearance. The best overall performance was achieved when age specific classifiers were combined with appearance specific classifiers, because in that case, different classifiers were employed for different groups of subjects.

The results quoted earlier refer to a database containing subjects with ages ranging from newborns up to 35 years old, thus the validity range of the methods presented are within that range. It is expected that the same methods would apply for an extended age range, provided that appropriate samples are introduced in the training set. In the case of increasing the range of ages in the training/test set, it is expected that the advantage of hierarchical age estimation over single step age estimation will be enhanced, since the effects of aging display different types of deformations in older subjects.

The time required for training the classifiers used in our experiments is in the order of minutes (about 3–4 min for single step and about 15–20 min for hierarchical classifiers). The only exception is the SOM based classifiers, for which the training time is in the order of hours (for grid sizes greater than 10×10).

Age estimation using any of the methods presented earlier can take place in real time, provided that the model-based representation of face images is already available. In a real application, the time required for obtaining the face parameters of a new face is in the order of seconds, implying that based on this methodology we could have almost real time age estimation.

VI. CONCLUSIONS

We presented an experimental evaluation into the problem of automatic age estimation where the performance of a classifier based on quadratic functions, a shortest distance classifier, and neural network based classifiers were evaluated. The classifiers in question were evaluated when a single step classification method was used and in the cases where hierarchical age estimation approaches were utilized.

According to our experimental results, hierarchical age estimation can be used for achieving better age estimation results. The best results were obtained when classifiers based on quadratic function and neural network based classifiers are used.

In order to train age classifiers, it is essential to have available information about the exact age of each subject in the database. An exception to this requirement is the SOM classifier, which as an unsupervised technique does not require the exact age of the subject for training, but forms its own age group clusters automatically by detecting the common features of the facial information presented. Therefore, the SOM based classifier provides an alternative way of training an age classifier in the case that information about the ages of subjects in the training set is not available.

The results of automatic age estimation obtained compare favorably with the results achieved by humans on the task on age estimation based on face images. In particular, human observers achieved an age estimation error of 3.64 years when tested on a similar, but significantly smaller, database.

The experiments described in this paper rely on the compact parameterization of face images provided by applying Principal Component Analysis. This type of analysis tends to discard local and unsystematic sources of variability in favor of global and systematic sources in variability. Since face parameters derived form principal components have been used successfully for age estimation, we can conclude that the age of a person could be defined based on the holistic structure of a face. However, we believe that in order to improve even more the performance of automatic age estimation algorithms, information about the fine details of a face (i.e., wrinkles) must be incorporated in the age estimation procedure.

The results obtained so far prove that it is feasible to develop in the short-term systems that incorporate automatic age estimation in their functionality, resulting in more efficient human machine interaction systems.

ACKNOWLEDGMENT

The authors are grateful to all volunteers who provided face images for our experiments and all volunteers who took part in the age estimation experiments.

REFERENCES

[1] V. Bruce, M. Burton, T. Doyle, and N. Dench, "Further experiments on the perception of growth in three dimensions," *Percept. Psychophys.*, vol. 46, no. 6, pp. 528–536, 1989.

[2] D. M. Burt and D. I. Perrett, "Perception of age in adult caucasian male faces: computer graphic manipulation of shape and color information," in *Proc. R. Soc. Lond.*, vol. 259, 1995, pp. 137–143.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Graph. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 681–685, June 2001.

[5] R. Chellapa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–741, May 1995.

[6] I. Craw and P. Cameron, "Face recognition by computer," in *Proceedings of the British Machine Vision Conference 1992*, D. Hogg and R. Boyle, Eds. London, U.K.: Springer-Verlag, 1992, pp. 489–507.

[7] D. W. Thompson, *On Growth and Form*. Cambridge, U.K.: Cambridge Univ. Press, 1961.

[8] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. F. Cootes, "Statistical face models: Improving specificity," *Image Vis. Comput.*, vol. 16, no. 3, pp. 203–211, 1998.

[9] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Advances in active appearance models," in *Proceedings of the 7th International Conference of Computer Vision—ICCV99*. Los Alamitos, CA: IEEE Computer Society Press, 1999, vol. I, pp. 137–142.

[10] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting face images using active appearance models," in *Proc. 5th Eur. Conf. Computer Vision*, 1998, pp. 581–595.

[11] *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Special Issue on Face and Gesture Recognition, vol. 19, no. 7, 1997.

[12] M. Kirby and L. Sirovich, "Application of the Karhumen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, Jan. 1990.

[13] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Berlin/Heidelberg, Germany: Springer-Verlag, 2001.

[14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic identification and coding of human faces using flexible models," *IEEE Trans. Pattern Analy. Machine Intell.*, vol. 19, pp. 743–756, July 1997.

[15] ——, "Toward automatic simulation of aging effects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 442–455, Apr. 2002.

[16] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze detection measurement," in *Proc. 4th Int. Conf. Face and Gesture Recognition*, 2000, pp. 499–505.

[17] B. Moghaddam and M. Yang, "Gender classification with support vector machines," in *Proc. 4th Int. Conf. Face and Gesture Recognition*, 2000, pp. 306–313.

[18] A. J. O'Toole, T. Vetter, H. Volz, and E. Salter, "Three-dimensional caricatures of human heads: Distinctiveness and the perception of age," *Perception*, vol. 26, pp. 719–732, 1997.

[19] M. Pantic and J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1424–1445, 2000.

[20] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception," *J. Exper. Psychol.: Human Percept. Perf.*, vol. 1, no. 4, pp. 374–382, 1975.

[21] J. B. Pittenger, R. E. Shaw, and L. S. Mark, "Perceptual information for the age level of faces as a higher order invariant of growth," *J. Exper. Psychol.: Human Percept. Perf.*, vol. 5, no. 3, pp. 478–493, 1979.

[22] *Proc. 1st, 2nd, 3rd, and 4th Int. Conf. Face and Gesture Recognition*, 1995, 1996, 1998, 2000.

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[24] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[25] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 733–742, July 1997.

[26] Y. Wu and N. M. Thalmann, "A dynamic wrinkle model in facial animation and skin aging," *J. Visualiz. Computer. Animat.*, vol. 6, pp. 195–205, 1995.

Andreas Lanitis (M'97) received the B.Eng degree in electronic and electrical engineering in 1991 and the Ph.D degree in image processing and computer vision in 1995, both from the University of Manchester, U.K.

He was then a Research Associate in the Department of Medical Biophysics, University of Manchester, where he carried out research in the area of statistical appearance models. Currently, he is Associate Professor, Department of Computer Science and Engineering, Cyprus College, Cyprus. His research interests include real applications of face image interpretation, coding, and synthesis.

Chrisina Draganova received the M.Sc. degree and the Ph.D. degree, both in mathematics, from Sofia University, Sofia, Bulgaria, and the M.Sc. degree in computing science from Birkbeck College, University of London, London, UK.

She is currently a Senior Lecturer, Department of Computing, Communications Technology and Mathematics, London Metropolitan University, London, U.K. Her research interests are in the areas of numerical analysis, spline interpolation, function approximation, and neural network approximation and applications.

Chris Christodoulou (M'87) received the B.Eng. degree in electronic engineering from Queen Mary College, University of London, London, U.K., and the Ph.D. degree in neural networks/computational neuroscience from King's College, University of London.

He was a Postgraduate Research Assistant (1991–1995) and a Postdoctoral Research Associate (1995–1997) at the Centre for Neural Networks, King's College, where he was also a Visiting Research Fellow (1997–2001). Since 1997, he has been a Lecturer in computer science at Birkbeck College, University of London. His research interests focus in the areas of neuroinformatics and computational neuroscience, artificial neural networks (including applications), learning for dynamic applications, neurobiological modeling, and hardware neural implementations.