Author(s): Lanitis, Andreas; Draganova, Chrisina; Christodoulou, Chris.
Article title: Comparing different classifiers for automatic age estimation
Year of publication: 2004
Citation: Lanitis A., Draganova, C. and Christodoulou C. (2004) 'Comparing different classifiers for automatic age estimation' IEEE Transactions on Systems Man and Cybernetics, Part B Cybernetics 34 (1) 621-8.
Link to published version: http://dx.doi.org/10.1109/TSMCB.2003.817091
DOI: 10.1109/TSMCB.2003.817091

Comparing Different Classifiers for Automatic Age Estimation

A. Lanitis

Department of Computer Science and Engineering,
Cyprus College,
P.O. Box 22006,
Nicosia, Cyprus.

alanitis@cycollege.ac.cy

C. Draganova

Dept. of Computing, Communication technology and Mathematics,
London Metropolitan University, 100 Minories,
London EC3 1JY,UK
c.draganova@londonmet.ac.uk

C. Christodoulou

School of Computer Science and Information Systems
Birkbeck College,
University of London
Malet Street,
London WC1E 7HX, UK
chris@dcs.bbk.ac.uk

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.

1. 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 [4,10,21] so that contact-less Human Computer Interaction (HCI) based on facial gestures can be developed. In this context systems capable of identifying faces [13], recognizing emotions [18], gender [16] and head orientation [15] 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.

1.1 Motivation

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

- *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 [14].
- *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.

1.2 Overview of our work

The basis of our approach is a statistical model of facial appearance, 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 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 [15]. In our work we aim to use this representation for estimating the age of an individual as shown in the block diagram of Figure 1.

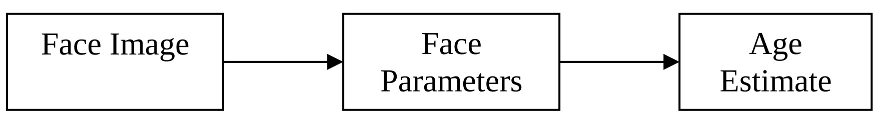

Figure 1: Block diagram of the age estimation approach

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 best 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-Organising Map (SOM) [12], 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 figure 2.

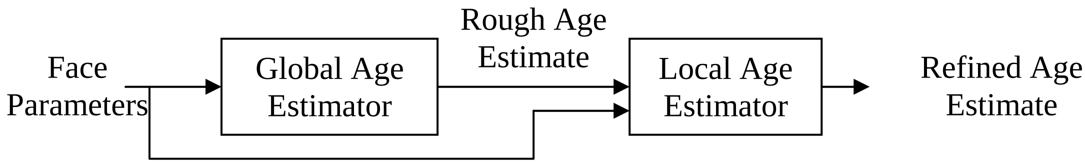

Figure 2: Block diagram of the age specific age estimation approach

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 figure 3.

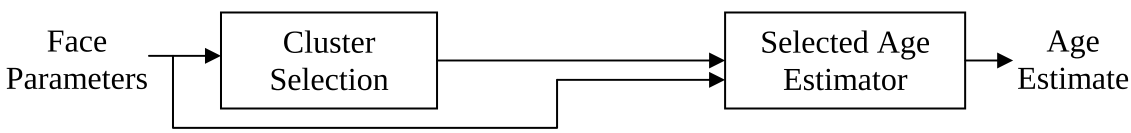

Figure 3: Block diagram of the appearance specific age estimation approach

A combination of both the appearance and the age specific age estimation algorithms was also investigated. Based on the methods presented above, we run age estimation experiments and report comparative results between the performances of different algorithms. Also we have 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.

1.3 Paper Organization

The remainder of the paper is organised as follows: In section 2 we present a brief overview of the relevant bibliography, in section 3 we briefly describe how the face model is generated, in section 4 we present the age estimation methods used in our experiments and in section 5 we present the results obtained. Finally in section 6 we discuss the results and present our conclusions.

2. Literature review

Although the general topic of face image processing received considerable interest [4, 10, 23], 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 [6] suggested that it is possible to use co-ordinate transformations for altering the shape of biological organisms in order to produce shapes belonging to different but similar biological organisms. Based on this idea a number of researchers [1,19,20] investigated the use of co-ordinate transformations in an attempt to impose age-related changes on human faces. According to their experimental evaluations 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 [17] use three-dimensional facial information for building a parametric 3D face model. They use a caricature algorithm in order to exaggerate or deemphasize distinctive 3D facial features; in the resulting caricatures the perceived age is increased or decreased according to the exaggeration level, suggesting that 3D distinctive facial features are emphasized in older faces. Their findings were verified by a number of observers.

Wu and Thalmann [25] 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 above 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 a coded representation of face images derived from a statistical face model generated by applying Principal Component Analysis (PCA) on an ensemble of face images. Kirby and Sirovich [11] first used this representation for low bit coding of face images. Based on this methodology Turk and Pentland [23] 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 [5] 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 [15,24]. 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. As an extension to this technique

Edwards et al [7,8,9] 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 model used in our work is based on the work reported by Edwards et al [7,8,9].

3. The 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:

$$\mathbf{X} = \mathbf{X_m} + \mathbf{Pb} \qquad \textbf{Eq. 1}$$

Where **X** is a vector describing the shape or intensity pattern of a training example, $\mathbf{X_m}$ is the mean example, **P** is the matrix of eigenvectors and **b** is a vector of weights, or *model parameters*. Edwards et al [7, 8, 9] describe in detail how this type of model can be used for modeling combined shape and color variation in face images.

We have generated a face model using 500 face images of 40 individuals. Each individual (see figure 4) supplied a collection of photographs taken over many years (average 3 year intervals). The age of the individuals in each image is known. For the experiments described in this paper we have focused our attention on age variation between zero and 35 years old. The combined face model trained using images from our database requires 22 model parameters for parametrizing/reconstructing face images. Each of those parameters controls a distinct way in which face images vary within the training set (see figure 5). For example face parameters control changes in expression, lighting, 3D pose and individual appearance.

Figure 4: Typical images used in our experiments

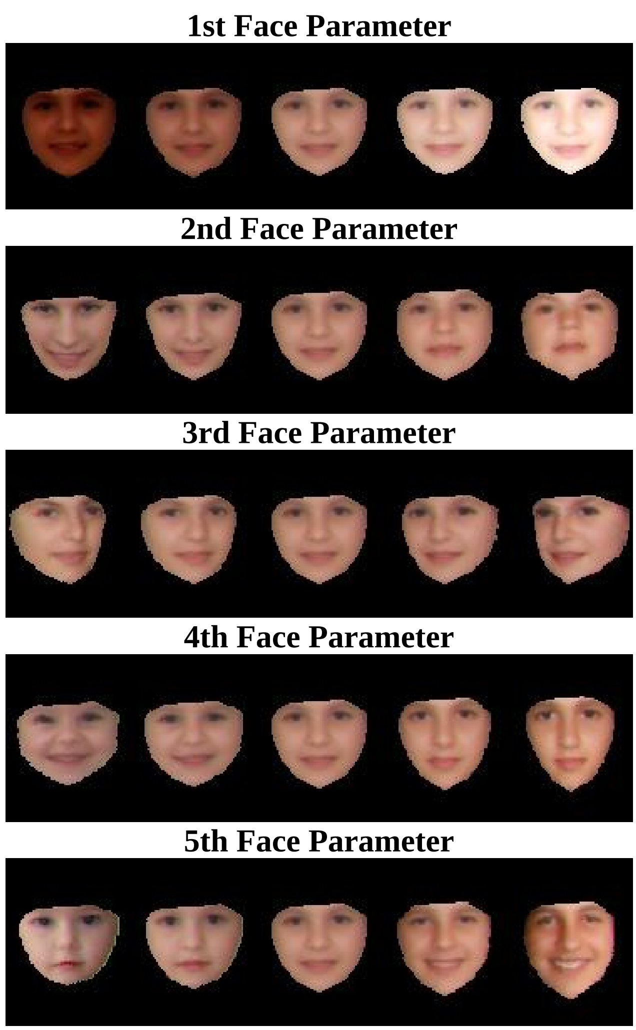

Figure 5: The main modes of variation within the training set. Each row shows the effect of varying a single face model parameter

4. Age Estimation

We use the 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 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. The simple age estimators used in our work are described in section 4.1 and hierarchical age estimation methods are presented in section 4.2.

4.1 Simple Age Estimation

4.1.1 Quadratic models

We define the relationship between face parameters and age, based on quadratic functions of the following form:

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

where

age is the 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.

$\mathbf{W_1}, \mathbf{W_2}$ are vectors containing weights for each element of $\mathbf{b}$ and $\mathbf{b^2}$ respectively.

c is the offset required.

We treat the problem of calculating the weights and offset, as an optimisation problem where we seek to minimise the difference between the actual ages of individuals in a training set, and the ages estimated using equation 2. Once the optimum parameters of the quadratic function are defined and given a set of face parameters corresponding to a previously unseen face image, it is possible to estimate the age of the person shown in the image. A more detailed description of our work in this area appears elsewhere [16].

4.1.2 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. 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 equation 3.

$$d = (\mathbf{b_n} - \mathbf{b_m})\, \mathbf{C}^{-1}\, (\mathbf{b_n} - \mathbf{b_m}) \qquad \textbf{Eq. 3}$$

Where

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

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

$\mathbf{C}$ is the 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 minimises 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.

4.1.3 Supervised Neural Networks

We have investigated the use of Multilayer Perceptrons (MLPs) with the back propagation learning algorithm [22] 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 generalisation 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.

4.1.4 Unsupervised Neural Networks

The Kohonen Self Organising Map (SOM), which is an unsupervised type neural network [12] 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 map organised in a number of clusters representing different age ranges. We perform several experiments varying the number of clusters and the size of the two-dimensional map. The number of clusters that we consider is 7, 9 and 12 for age range 5, 4 and 3 respectively. Each cluster corresponds to an age range, e.g. in the case of 7 clusters the age range is 5 years (0:4, 5:9, 10:14, 15:19, 20:24, 25:29 and 30:35). The size of the two-dimensional map varies as follows: 5x5, 10x10, 15x15, 20x20 and 25x25 neurons. For each experiment the initial learning rate is set to 0.5. A Gaussian neighbourhood function is used with initial width equal to the dimension of the grid (5, 10, 15, 20 or 25). These parameters are adjusted during the training. The experiments are run for 5000 to 15000 epochs. The processing time for training increases with increasing the size of the grid and the number of clusters for classification. The processing time for grid size greater than 10 is in the order of several hours.

The error for the SOM algorithm is calculated in the following way: after the final labelling takes place we check where each input is classified; for example assume that the input vector $\mathbf{x}$ is classified in the age range $\mathbf{c=(c_{min},c_{max})}$, the target range age for this vector is $\mathbf{g}$ and the target age is $\mathbf{a_g}$. If $\mathbf{g>c}$ we calculate the difference $\mathbf{a_g - c_{max}}$, if $\mathbf{g<c}$, $\mathbf{c_{min} - a_g}$ and 0 if $\mathbf{g=c}$. The best results are achieved for the network with grid size 20x20 neurons and classification in 12 clusters. Learning Vector Quantisation (LVQ) [12] has also been utilised for further optimisation.

4.2 Hierarchical Age Estimation

In this section we present the hierarchical age estimation algorithms used in our work.

4.2.1 Age-Specific Classifiers

For age specific age estimation we use a global age classifier (similar to the classifiers described in the previous sections) trained using all the training data, which contains subjects with ages ranging from 0 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 age specific classifiers for the age ranges of 0-10 years, 11-20 years and 21 to 35 years. 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.

4.2.2 Appearance Specific Classifiers

This approach is based on the observation that people who look similar, tend to age in a similar way [16]. 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. 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:

I. Use the training set for training an age estimator using any of the methods described in subsection 4.1.

II. For each training example 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.

III. Go back to step I. The procedure terminates until most (usually 95%) of the training examples are classified correctly.

Usually when about five to seven age estimators are used, the age estimation error is minimized, implying that that the use of appearance specific age estimators can be advantageous. Based on the training set used for each cluster we train a cluster classifier that is used for selecting one of the clusters given a vector of face parameters presented to the system. Once the given face parameters are assigned to a cluster, the age estimator for the particular cluster is used for producing an age estimate. We call this method appearance specific since the choice of the age estimator to be used is based on the appearance of the given face image (the face parameters).

4.2.3 Appearance-Specific and Age-Specific Classifiers

The methods described in sections 4.2.1 and 4.2.2 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 within our training set. We then train classifiers for a specific age range for each cluster in the training set. Given the face parameters of an unseen subject we assign it 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 Figure 6.

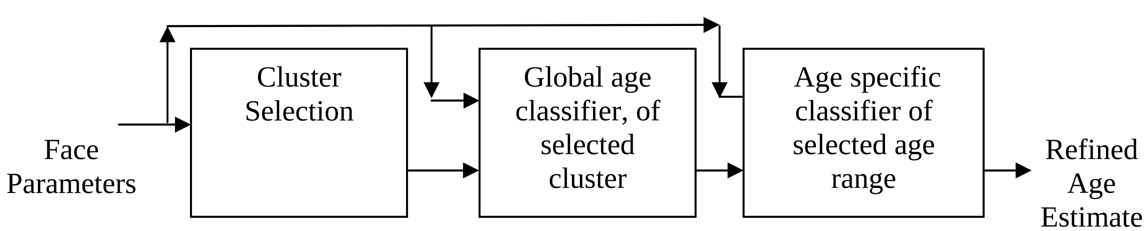

Figure 6: Block diagram of the appearance and age specific age estimation approach

Remark: The classifiers described in 4.2.1, 4.2.2 and 4.2.3 were tested with the Quadratic models, the Short Distance Classifier and the MLP with the back propagation learning. Although the unsupervised Kohonen Self Organising Map network can theoretically be used in this case as well, we decided not to use it because the age estimation is in an age range rather than exact (making it thus questionable on whether the results would improve) and in addition the processing time is much longer compared to the one of the MLP.

The optimal structure of each network in the case of the MLP for the age-specific, the age-appearance and the combined classifier was with one hidden layer and 10, 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 of the neural networks with the MLP was in the order of minutes.

4.3 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 analyse 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 no one of the volunteers knew any of the subjects whose images were included in the test set.

5. Experimental Investigation

In this section we describe the experiments conducted in order to assess the suitability of the methods presented in section 4, in the task of automatic age estimation.

5.1 Experimental Set Up

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. Images from different individuals are used in the two sets (i.e a single individual does not have images in both sets). For our experiments we used the two sets for training and testing and vice versa. The performance of each method was assessed by calculating the mean absolute error between the real and estimated ages among each set. 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 mis-localization 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.

5.2 Results

The overall results obtained are shown in the following table:

<table>
<tr><td rowspan="2">Method</td><td rowspan="2">Quadratic</td><td rowspan="2">Shortest Distance</td><td rowspan="2">MLP</td><td rowspan="2">SOM</td><td colspan="3">Human Observers</td></tr>
<tr><td>Male</td><td>Female</td><td>All</td></tr>
<tr><td>Global</td><td>5.04</td><td>5.65</td><td>4.78</td><td>4.9</td><td rowspan="4">3.69</td><td rowspan="4">3.59</td><td rowspan="4">3.64</td></tr>
<tr><td>Age Specific</td><td>4.87</td><td>5.02</td><td>4.52</td><td>N/A</td></tr>
<tr><td>Appearance Specific</td><td>4.61</td><td>5.58</td><td>4.64</td><td>N/A</td></tr>
<tr><td>Appearance and Age Specific</td><td>3.82</td><td>4.92</td><td>4.38</td><td>N/A</td></tr>
</table>

Table 1: The results of our experiments (all numbers show the absolute error in years)

6. 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 was 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.

When a single step classifier is used the classifiers based on multi-layer perceptrons, self-organising 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 for this application because it requires a large number of training samples in order to describe adequately the distribution of face parameters for each age. In the case of using a limited 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-organising 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 1) are with clusters of 3 years (see section 4.1.4); with clusters of 4 years the mean error is 5.3 while with clusters of 5 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 considerably 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 of automatic age estimation obtained compare favourably 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 parameterisation of face images provided by applying Principal Component Analysis. This type of analysis tends to discard local and unsystematic sources of variability in favour 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 rather than on isolated facial details. 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 produce in the short-term systems that incorporate automatic age estimation in their functionality, resulting in more efficient human machine interaction systems.

7. Acknowledgements

The work described in this paper was supported by a research grant from the Research Promotion Foundation of the Republic of Cyprus (project 26/2001). We are grateful to all volunteers who provided face images for our experiments and all volunteers who took part in the age estimation experiments.

8. Bibliography

1. V. Bruce, M. Burton, T. Doyle and N. Dench, "Further Experiments on the Perception of Growth in Three Dimensions", Perception and Psychophysics 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". Procs of the Royal Society London, Vol 259, pp 137-143, 1995.

3. T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham. Active Shape Models - Their Training And Application. Computer Vision Graphics and Image Understanding, 61, no 1, pp 38-59, 1995.
4. R. Chellapa, C.L. Wilson and S. Sirohey. “Human and Machine Recognition of Faces: A Survey”. Procs of the IEEE, 83, no 5, 1995.
5. I. Craw and P. Cameron. Face Recognition by Computer. Procs of British Machine Vision Conference 1992, pp 489-507, eds. David Hogg and Roger Boyle, Springer Verlag, 1992.
6. D’ Arcy Wentworth Thompson, “On Growth and Form”, Cambridge University Press, 1961.
7. G.J. Edwards, A. Lanitis, C.J. Taylor, and T.F.Cootes.” Statistical Face Models: Improving Specificity”. Image and Vision Computing, Vol 16, no 3, pp, 203-211, 1998.
8. G.J. Edwards, T.F.Cootes and C.J.Taylor. “Advances in Active Appearance Models”. Procs of the 7th International Conference of Computer Vision – ICCV99, Vol I, pp 137-142, IEEE Computer Society Press, Kerkyra, Greece, 1999.
9. G.J. Edwards, C.J. Taylor, and T.F.Cootes. “Interpreting Face Images Using Active Appearance Models”, Procs of the 5th European Conference on Computer Vision, pp 581-595, 1998.
10. IEEE Transactions of Pattern Analysis and Machine Intelligence, Special issue on Face and Gesture Recognition, Vol 19, no 7, 1997.
11. M. Kirby and L. Sirovich. Application of the Karhumen-Loeve procedure for the

12. T. Kohonen, Self-organising Maps, Springer-Verlag, 1997.
13. A. Lanitis, C.J. Taylor and T.F. Cootes, "Automatic Identification and Coding of Human Faces Using Flexible Models". IEEE Transactions of Pattern Analysis and Machine Intelligence, vol 19, no 7, pp 743-756, 1997.
14. A. Lanitis, C.J. Taylor and T.F. Cootes, "Toward Automatic Simulation of Aging Effects". IEEE Transactions of Pattern Analysis and Machine Intelligence, vol 24, no 4, pp 422-455, 2002.
15. Y. Matsumoto and A. Zelinsky. "An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Detection Measurement", Procs of the 4th International Conference on Face and Gesture Recognition, pp 499-505, 2000.
16. B. Moghaddam and M. Yang. " Gender Classification with Support Vector Machines", Procs of the 4th International Conference on Face and Gesture Recognition, pp 306-313, 2000.
17. 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.
18. M. Pantic and J.M. Rothkrantz., "Automatic Analysis of Facial Expressions: The State of the Art". IEEE Transactions of Pattern Analysis and Machine Intelligence, vol 22, no 12, pp 1424-1445, 2000.
19. J.B. Pittenger and R.E Shaw. "Aging Faces as Viscal-Elastic Events: Implications for a Theory of Nonrigid Shape Perception". Journal of Experimental

20. 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”. Journal of Experimental Psychology: Human Perception and Performance Vol 5, No. 3, pp 478-493, 1979.
21. Procs. of the 1st , 2nd, 3rd and 4th International Conference on Face and Gesture Recognition, (1995,1996,1998,2000).
22. D.E. Rumelhart, G.E. Hinton and R.J. Williams, Learning representations by back-propagating errors. Nature, 323, 533–536, 1986.
23. M. Turk and A. Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3, no 1, pp 71-86, 1991.
24. T. Vetter and T. Poggio: Linear Object Classes and Image Synthesis from a Single Example Image. IEEE Transactions of Pattern Analysis and Machine Intelligence, vol 19, no 7, pp 733-742, 1997.
25. Y. Wu and N.M. Thalmann. “A Dynamic Wrinkle Model in Facial Animation and Skin Aging”, Journal of Visualization and Computer Animation, Vol. 6, pp 195-205, 1995.