

Learning from Facial Aging Patterns for Automatic Age Estimation

Xin Geng¹, Zhi-Hua Zhou², Yu Zhang², Gang Li¹, Honghua Dai¹

¹School of Engineering and Information Technology
Deakin University, Victoria 3125, Australia
{xge, gang.li, hdai}@deakin.edu.au

²National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
{zhouzh, zhangy}@lamda.nju.edu.cn

ABSTRACT

Age Specific Human-Computer Interaction (ASHCI) has vast potential applications in daily life. However, automatic age estimation technique is still underdeveloped. One of the main reasons is that the aging effects on human faces present several unique characteristics which make age estimation a challenging task that requires non-standard classification approaches. According to the speciality of the facial aging effects, this paper proposes the AGES (AGing pattErn Subspace) method for automatic age estimation. The basic idea is to model the aging pattern, which is defined as a sequence of personal aging face images, by learning a representative subspace. The proper aging pattern for an unseen face image is then determined by the projection in the subspace that can best reconstruct the face image, while the position of the face image in that aging pattern will indicate its age. The AGES method has shown encouraging performance in the comparative experiments either as an age estimator or as an age range estimator.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (*e.g.*, HCI)]: User Interfaces; I.4.8 [Image Processing and Computer Vision]: Scene Analysis; I.2.6 [Artificial Intelligence]: Learning

General Terms

Algorithm, Design, Experimentation

Keywords

Age Specific Human-Computer Interaction, Automatic Age Estimation, Face Image, Aging Pattern

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'06, October 23-27, 2006, Santa Barbara, California, USA.
Copyright 2006 ACM 1-59593-447-2/06/0010 ...\$5.00.

1. INTRODUCTION

Studies [22] show that only 7% meanings in human communication are conveyed via words, which can be further translated into computer commands. The remaining 93% are nonverbal meanings, quite amount of which are expressed through the face. As such, rather than simply responding to user's commands, a face-based Human-Computer Interaction (HCI) system is able to understand various information conveyed by human faces or display results through facialized interface, just like the human-to-human communication.

As a matter of fact, the face is a prolific information source. People can effortlessly extract many kinds of useful information from a face image, such as identity, expression, emotion, gaze, gender, age, *etc.* The automatic extraction of most of these information has been extensively studied [5, 8, 18, 24, 32, 33] in several research areas including multimedia, HCI, computer vision, pattern recognition, machine learning and neural networks. However, so far there is relatively few literature concerning age estimation, which is one of the basic abilities required for smooth communication between humans. People at different ages have different requirements and preferences in various aspects, such as linguistics, aesthetics and consumption habit. Consequently Age Specific Human-Computer Interaction (ASHCI) is widely demanded by numerous applications in daily life. With ASHCI, for example, an enquiry terminal can automatically choose the vocabulary, interface, and services that suit the customer's age; A web browser can determine by itself whether the user satisfies the age limitation to view certain web pages; A vending machine will refuse to sell alcohol or cigarettes to the underage people. It is the vast potential applications and the still underdeveloped state of the art that inspires the work described in this paper.

The difficulty in automatic age estimation is mainly due to the speciality of aging effects on the face compared with other facial variations. The unique characteristics of aging variation mainly include:

1. *The aging progress is uncontrollable.* Sad but true, no one can age at will. The procedure of aging is slow and irreversible. Thus the collection of sufficient training data for age estimation is extremely laborious.
2. *Personalized aging patterns.* Different persons age in

different ways. The aging pattern of each person is determined by his/her gene as well as many external factors, such as health, living style and weather conditions, *etc.*

3. *The aging patterns are temporal data.* The aging progress must obey the order of time. The face status at a particular age will affect all older faces, but will not affect those younger ones.

The difficulty in data collection is now partly alleviated thanks to the share of the FG-NET Aging Database [1]. However, most subjects in this database only have face images at a few ages, *i.e.* the data set is highly incomplete in the view of aging pattern. Indeed, a ‘complete’ aging face database is *unnecessary* since human beings also learn to perceive facial ages from incomplete aging patterns. Thus the learning algorithm applied to the aging patterns must be able to handle highly incomplete data. The last two characteristics also make age estimation different from typical classification problems. In detail, first, the mapping from features (face images) to class labels (ages) is not unique, but complicatedly depends on personalized factors. Thus the personal identity should be simultaneously taken into consideration. Second, each class (age) has a unique rank in the time sequence. Thus the set of classes must be a totally ordered set. To sum up, the learning procedure for age estimation can be described as: *Learning from sparse temporal aging patterns which are determined by personalized factors.*

This paper proposes a novel age estimation method named AGES (AGing pattErn Subspace), which addresses the issues mentioned above. Instead of using isolated face images as training samples, AGES regards each aging pattern as a sample, which naturally integrates the personal identity and the time concept. This changes the learning task from an *image-class* problem based on *complete* data into an *(image sequence)-(class sequence)* problem based on *incomplete* data. To solve this problem, an iterative learning algorithm is specially developed to learn a representative subspace from the highly incomplete aging patterns. Then a two-step age estimation approach geared to the characteristics of facial aging variation is proposed based on the learned aging pattern subspace.

The rest of this paper is organized as follows. Some related works are briefly introduced in Section 2. Then the data representation of the aging pattern is explained in Section 3. After that, the AGES algorithm is proposed in Section 4. In Section 5, the experimental results are reported and analyzed. Finally in Section 6, conclusions are drawn and several issues for the future work are indicated.

2. RELATED WORKS

Face tracking [4, 16, 27] and eye gaze [17, 24, 31] has occupied a great share of the attention of HCI researchers. Both of them have been widely used in the studies of interface usability, machine-mediated human communication, and alternative input devices for disabled users. The technology has become sufficiently advanced that several companies offer working commercial systems. This area can therefore be viewed as a relatively mature sub-field of face-based HCI.

Another relatively successful sub-field of facial information processing is face recognition [5, 15, 33], mostly studied by the pattern recognition community, where it has become a widely recognized challenging problem. Because of its

prospective social value, the U.S. government regularly conducts the Face Recognition Vendor Test (FRVT, previously known as FERET) to evaluate the commercially available and prototype face recognition technologies and promote the research in this area.

There are several existing works on facial aging progress, originating from psychological and biological studies. However, most of them aim to simulate the aging effects on human faces (given an age, simulate a face at that age), which is the inverse procedure of age estimation (given a face, estimate its age). For example, Pittenger and Shaw [20] proposed the shear & strain transformations to model the changes in the shape of face profiles due to growth. Burt and Perrett [3] simulated aging variations by superimposing typical aging changes in shape and color on face images. Wu et al. [30] came up with a dynamic model to simulate wrinkles in 3D facial animation and skin aging. O’Toole et al. [19] described how aging variations can be made by applying a standard facial caricaturing algorithm to the three dimensional models of faces. Tiddeman et al. [25] developed prototype face models for aging face images by transforming facial textures. Although these works did not target to age estimation, they did reveal some of the important facts in the relationship between ages and face images, such as the personalized age transformations [19], which are irradiative for the design of age estimation algorithms. Some other works tried to partly reveal the the mapping from face images to ages. For example, Ramanathan and Chellappa [21] proposed a method for face verification across age progression based on Bayesian classifier, which is able to roughly estimate the age difference between a pair of face images of the same individual. Kwon and Da Vitoria Lobo [11] proposed an age classification method based on cranio-facial development theory [2] and skin wrinkle analysis. But this method relies on well controlled high-quality face images and can only classify faces into one of three groups (babies, young adults, and senior adults), which greatly restricts its usage in real applications.

The first ‘real’ human age estimation algorithm was proposed by Lanitis *et al.* [13] in a regression way. In their work, the aging pattern is represented by a quadratic function called *aging function*

$$age = f(\mathbf{b}), \quad (1)$$

where \mathbf{b} is the feature vector of the face image. During the training procedure, a quadratic function is fitted for each individual in the training set as his/her aging function. As for age estimation, they proposed four approaches to determine the suitable aging function for the unseen face image, among which the Weighted Person Specific (WPS) approach achieved the best performance in the experiments. However WPS is based on the ‘lifestyle profiles’ which contain information about various external factors, such as gender, health, living style, weather conditions, *etc.* Such information is very difficult to obtain in real applications, which makes WPS impractical. The Weighted Appearance Specific (WAS) approach is the best one among the methods that do not use the ‘lifestyle profiles’. In WAS, the aging function for the unseen face is calculated by the weighted sum of the known aging functions, where the weights are determined by the Mahalanobis distance from the test image to the training images. Later Lanitis *et al.* [12] proposed three hierarchical architectures to further improve their aging-function-

based method. Among them, the Appearance-Age Specific architecture (AAS) achieved the best performance because it handles the face image clusters separately according to both the appearance and the age group.

The aging-function-based approaches regard age estimation as a conventional regression problem without special design for the unique characteristics of aging variation. This prevents them from getting more satisfying results. In detail, there might be four weaknesses in such approaches. First, the formula of the aging function is empirically determined. There is no evidence suggesting that the relation between face and age is as simple as a quadratic function. Second, the temporal characteristic cannot be well utilized by the aging function. The dependence relation among the aging faces is monodirectional, *i.e.* the status of a certain face only affects those older faces. However, the relation revealed by the aging function is bidirectional: any changes on a particular face will affect all other faces. Third, the learning of one person’s aging pattern is solely based on the face images of that person. Although people age in different ways, there must be some commonality among all aging patterns, *i.e.* the general trend of aging. Such commonality is also crucial in age estimation, especially when the personal training data is insufficient. Fourth, the new aging function for the unseen face image is simply a linear combination of the known aging functions, rather than being generated from a certain model of aging patterns. All of these problems can be solved, from an entirely new point of view, by the AGES algorithm proposed in this paper. Changes start from the very beginning: data representation.

3. AGING PATTERN

If age estimation is regarded as a conventional classification problem, then one straightforward way is to model face images at each age. The problem is that different persons age in different ways. ‘Age’ is a relative concept specified to each person. Thus a face at a particular age is more related to the same person’s faces at different ages than to other person’s faces at the same age. Consequently, the aging pattern, rather than the separate ages, should be modelled. The concept of *aging pattern* is defined by

Definition 1. An aging pattern is a sequence of personal face images sorted in time order.

Two keywords in this definition are ‘personal’ and ‘time’. All face images in an aging pattern must come from the same person, and they must be arranged by time. Suppose a gray-scale face image is represented by a two-dimensional matrix \mathbf{I} , where $\mathbf{I}(x, y)$ is the intensity of the pixel (x, y) , then an aging pattern can be represented by a three-dimensional matrix \mathbf{P} , where $\mathbf{P}(x, y, t)$ is the intensity of the pixel (x, y) in the face image at the time t . Take the aging pattern shown in Figure 1 as an example. Along the t axis, each age (from 0 to 8 in this example) is allocated one position. If face images are available for certain ages (in this case 2, 5 and 8), they are filled into the corresponding positions. If not, the positions are left blank (dotted squares). If all positions are filled, the aging pattern is called a *complete aging pattern*, otherwise it is called an *incomplete aging pattern*.

Before the aging pattern can be further processed, the face images in it are first transformed into feature vectors. In this paper, the feature vector is extracted by the Appearance Model [7]. The main advantage of this model is that

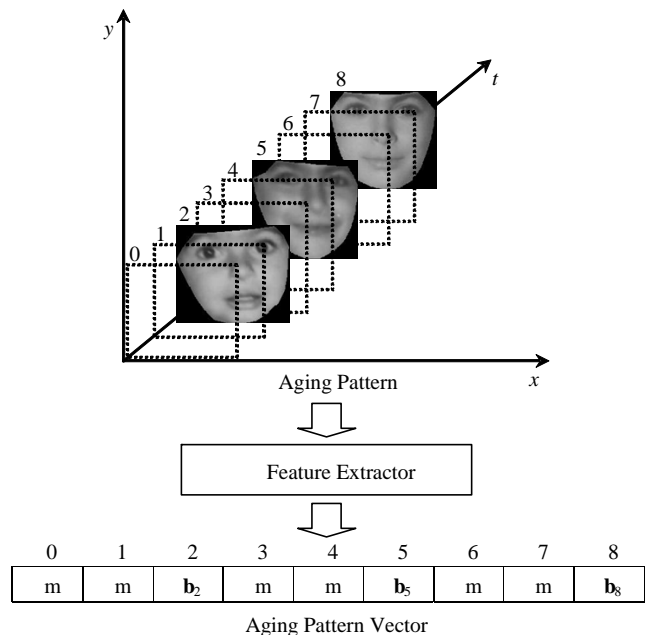


Figure 1: Vectorization of the aging pattern. The ages (0-8) are marked at the top-left to the corresponding positions and above the corresponding feature vectors. The missing parts in the aging pattern vector are marked by ‘m’.

the extracted feature combines both the shape and the intensity of the face images. Suppose the extracted feature vector \mathbf{b} is n -dimensional, the number of interested ages is p . Then the aging pattern can be represented by an $(n \times p)$ -dimensional feature vector \mathbf{x} . Each age is allocated n consecutive elements in \mathbf{x} . If the face image at a particular age is not available, then the corresponding part in \mathbf{x} is marked as missing features. Figure 1 gives an example of the vectorization of the aging pattern when the interested ages are from 0 to 8. \mathbf{b}_2 , \mathbf{b}_5 and \mathbf{b}_8 represent the feature vectors of the face images at the ages 2, 5 and 8, respectively.

By representing aging patterns in this way, the concepts of identity and time are naturally integrated into the data without any pre-assumptions, consequently the personality and the temporal characteristic of aging variation can be well utilized. So long as the aging patterns are well sampled, *i.e.* the training aging patterns are enough to represent the real distribution of all aging patterns, a proper model of aging pattern can be learned for age estimation. However this brings two other challenges: (1) The learning task has been changed from a conventional *image-class* problem into a novel (*image sequence*)-(class sequence) problem; (2) The learning algorithm applied to the aging patterns must be able to handle highly incomplete training samples. The following section mainly tackles these two problems.

4. THE AGES ALGORITHM

4.1 Learning the Aging Pattern Subspace

A representative model for the aging patterns can be built up by the information theory approach of coding and decoding. One widely adopted technology is using PCA [10]

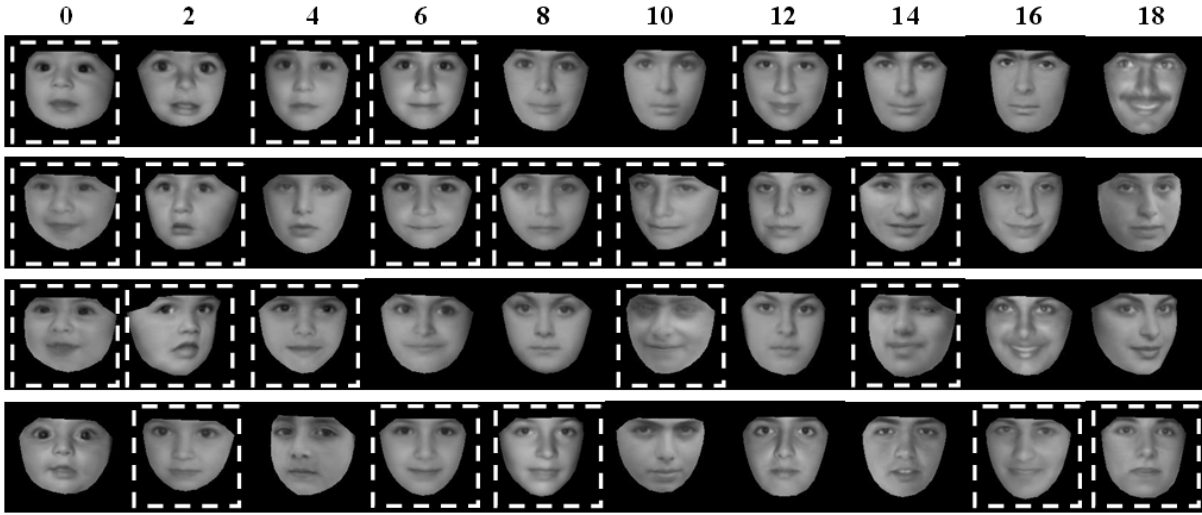


Figure 2: ‘Full-filled’ aging patterns. Each line shows one aging pattern from the age 0 to 18. The ages are marked above the corresponding faces. The faces learned by the algorithm are surrounded by the dashed squares.

to construct a subspace that captures as much as possible variation in the data set. The projection in the subspace is computed by

$$\mathbf{y} = \mathbf{W}^T(\mathbf{x} - \mu), \quad (2)$$

where μ is the mean vector of \mathbf{x} , and \mathbf{W}^T is the transpose of \mathbf{W} , which is composed by the eigenvectors of the covariance matrix of \mathbf{x} . The problem is that the aging pattern vector \mathbf{x} is highly incomplete. There exist several approaches to PCA with missing data [10]. However, the statistical distribution of the aging patterns is unlikely to be normal, thus those methods based on the assumption of normal distribution [23, 26] are not suitable for this problem. Moreover, the aging pattern vector is highly incomplete, thus those methods dealing with minor missing data [23, 28] are also not suitable. Based on the characteristics of aging patterns, an EM-like iterative learning algorithm is proposed here to learn a representative subspace.

Suppose the training set has N aging pattern vectors $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Any sample in this set can be written as $\mathbf{x}_k = \{\mathbf{x}_k^a, \mathbf{x}_k^m\}$, where \mathbf{x}_k^a is the available features and \mathbf{x}_k^m is the missing features of \mathbf{x}_k . Once the transformation matrix \mathbf{W} is determined, the reconstruction of \mathbf{x}_k is calculated by

$$\hat{\mathbf{x}}_k = \mu + \mathbf{W}\mathbf{y}_k. \quad (3)$$

$\hat{\mathbf{x}}_k$ can be written as $\hat{\mathbf{x}}_k = \{\hat{\mathbf{x}}_k^a, \hat{\mathbf{x}}_k^m\}$, where $\hat{\mathbf{x}}_k^a$ is the reconstruction of \mathbf{x}_k^a , and $\hat{\mathbf{x}}_k^m$ is the reconstruction of \mathbf{x}_k^m . Then the mean reconstruction error of the available features is

$$\bar{\varepsilon}^a = \frac{1}{N} \sum_{k=1}^N (\mathbf{x}_k^a - \hat{\mathbf{x}}_k^a)^T (\mathbf{x}_k^a - \hat{\mathbf{x}}_k^a). \quad (4)$$

Now the goal is to find a \mathbf{W} that minimizes $\bar{\varepsilon}^a$. When initializing, \mathbf{x}_k^m is replaced by the mean vector $[\mu_k^m]$, calculated from the samples for which the features at the same positions as \mathbf{x}_k^a are available. Then standard PCA is applied to the full-filled data set to get the initial transformation matrix \mathbf{W}_0 and mean vector μ_0 . In iteration i , the projection of \mathbf{x}_k in the subspace spanned by \mathbf{W}_i is estimated first. Since there are many missing features in \mathbf{x}_k , the projection cannot

be computed directly by Equation (2). Note that the aging patterns are highly redundant, it is possible to estimate \mathbf{y}_k only based on part of \mathbf{x}_k [14], *i.e.* \mathbf{x}_k^a . Instead of using inner product, \mathbf{y}_k is solved as the least squares solution of

$$[\mathbf{W}_i^{(a)}]\mathbf{y}_k = \mathbf{x}_k^a - [\mu_i^{(a)}], \quad (5)$$

where $[\mathbf{W}_i^{(a)}]$ is the part in \mathbf{W}_i and $[\mu_i^{(a)}]$ is the part in μ_i that correspond to the positions of \mathbf{x}_k^a . Suppose the subspace is d -dimensional, then there are d unknowns in \mathbf{y}_k . Thus at least d available features (elements in \mathbf{x}_k^a) are needed to over-constrain the problem. After getting the estimation of \mathbf{y}_k , $\hat{\mathbf{x}}_k$ is calculated by Equation (3) and \mathbf{x}_k^m is updated by $\hat{\mathbf{x}}_k^m$. Then, standard PCA is applied to the updated data set to get the new transformation matrix \mathbf{W}_{i+1} and mean vector μ_{i+1} . The whole procedure repeats until the maximum iteration τ is exceeded or $\bar{\varepsilon}^a$ is smaller than a predefined threshold θ . The convergence of this algorithm is proved as follows.

PROOF. Suppose in iteration i , the training data is $\mathbf{x}_k^{(i)}$, the reconstruction of $\mathbf{x}_k^{(i)}$ by \mathbf{W}_i is $[\hat{\mathbf{x}}_k^{(i)}(\mathbf{W}_i)]$, the reconstruction error of $\mathbf{x}_k^{(i)}$ by \mathbf{W}_i is $\varepsilon(\mathbf{x}_k^{(i)}, \mathbf{W}_i)$, and the reconstruction error of the available features is $\varepsilon^a(\mathbf{x}_k^{(i)}, \mathbf{W}_i)$. Note that $[\hat{\mathbf{x}}_k^{(i)}(\mathbf{W}_i)]$ and the data of the next iteration, $\mathbf{x}_k^{(i+1)}$, share the same values at the positions of missing features, so

$$\varepsilon^a(\mathbf{x}_k^{(i)}, \mathbf{W}_i) = U([\hat{\mathbf{x}}_k^{(i)}(\mathbf{W}_i)], \mathbf{x}_k^{(i+1)}), \quad (6)$$

where $U(\mathbf{v}_1, \mathbf{v}_2)$ denotes the squared Euclidean distance between \mathbf{v}_1 and \mathbf{v}_2 . Consequently, $\bar{\varepsilon}_i^a = \bar{U}$, where $\bar{\varepsilon}_i^a$ is the $\bar{\varepsilon}^a$ of iteration i and \bar{U} is the mean value of $U([\hat{\mathbf{x}}_k^{(i)}(\mathbf{W}_i)], \mathbf{x}_k^{(i+1)})$. If $\mathbf{x}_k^{(i+1)}$ is also reconstructed by \mathbf{W}_i , then

$$\begin{aligned} \varepsilon(\mathbf{x}_k^{(i+1)}, \mathbf{W}_i) &= U([\hat{\mathbf{x}}_k^{(i+1)}(\mathbf{W}_i)], \mathbf{x}_k^{(i+1)}) \\ &\leq U([\hat{\mathbf{x}}_k^{(i)}(\mathbf{W}_i)], \mathbf{x}_k^{(i+1)}) \end{aligned} \quad (7)$$

because the line between $[\hat{\mathbf{x}}_k^{(i+1)}(\mathbf{W}_i)]$ and $\mathbf{x}_k^{(i+1)}$ is orthogonal to the subspace spanned by \mathbf{W}_i so that they have the

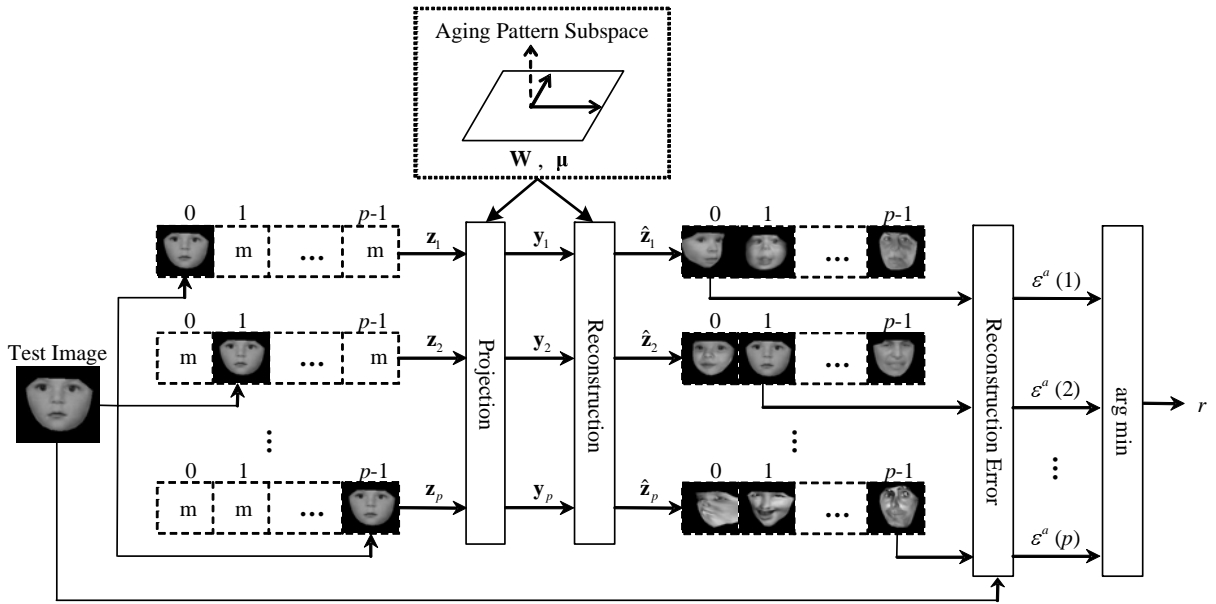


Figure 3: Flowchart of age estimation in AGES. The missing parts in the aging patterns are marked by ‘m’. The test image is from a 1 year old baby and the interested ages are from 0 to $p - 1$.

minimum Euclidean distance. Consequently, $\bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_i) \leq \bar{U}$, where $\bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_i)$ is the mean reconstruction error of $\mathbf{x}_k^{(i+1)}$ by \mathbf{W}_i . After applying PCA on $\mathbf{x}_k^{(i+1)}$, the new transformation matrix \mathbf{W}_{i+1} minimizes the mean reconstruction error, thus

$$\bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_{i+1}) \leq \bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_i). \quad (8)$$

Obviously, $\bar{\varepsilon}_{i+1}^a \leq \bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_{i+1})$. So

$$\bar{\varepsilon}_{i+1}^a \leq \bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_{i+1}) \leq \bar{\varepsilon}(\mathbf{x}_k^{(i+1)}, \mathbf{W}_i) \leq \bar{U} = \bar{\varepsilon}_i^a. \quad (9)$$

Thus the algorithm will converge to minimize $\bar{\varepsilon}^a$. \square

During the learning procedure of AGES, the missing faces in the training aging patterns can be simultaneously learned by reconstructing the whole aging pattern vectors through Equation (3). Figure 2 gives some typical examples of the ‘full-filled’ aging patterns when AGES is applied to the FG-NET Aging Database [1]. For clarity, only the faces in the most changeable age range from 0 to 18 with 2 years as interval are shown. It can be seen that the learned faces inosculate with those known faces very well in the aging patterns. Thus this learning algorithm can also be used to simulate aging effects on human faces. It is noteworthy that in some cases, such as the 2nd line of Figure 2, the learned faces might be the majority part of the aging pattern, which indicates the ability of the proposed algorithm to learn from highly incomplete (sparse) data.

The procedure of the learning algorithm is actually a procedure of interaction between the global aging pattern model and the personalized aging patterns. In each iteration of the learning, the missing part of the personal aging pattern is first estimated by the current global aging pattern model. Then, the global model is further refined by the updated personal aging patterns. In this way, the commonality and the personality of the aging patterns are alternately utilized to learn the final subspace.

4.2 Age Estimation

The aging pattern subspace is a model for image sequences, each of which corresponds to a sequence of class labels. But the task of age estimation is usually based on a single face image input, and requires a single age output. This section will describe how this can be done on the basis of aging pattern subspace.

When an unseen face image \mathbf{I} is presented, its feature vector \mathbf{b} is first extracted by the feature extractor. According to the last two characteristics of aging effects mentioned in Section 1, the procedure of age estimation should involve two steps. Step 1 is to select an aging pattern suitable for \mathbf{I} (the 2nd characteristic), and step 2 is to find a proper position for \mathbf{I} in the selected aging pattern (the 3rd characteristic). Step 1 can be achieved by finding a projection in the aging pattern subspace that can reconstruct \mathbf{b} with minimum reconstruction error. However, without knowing the position of \mathbf{I} in the aging pattern, the reconstruction error cannot be actually calculated. Thus \mathbf{I} is placed at every possible position in the aging pattern, getting p aging pattern vectors $\mathbf{z}_j (j = 1 \dots p)$ by placing \mathbf{b} at the position j in \mathbf{z}_j . Noticing that \mathbf{b} is the only available feature in \mathbf{z}_j , the projection \mathbf{y}_j of \mathbf{z}_j can be estimated by Equation (5), and the reconstruction error of \mathbf{y}_j can be calculated by

$$\varepsilon^a(j) = (\mathbf{b} - \mu_{(j)} - \mathbf{W}_{(j)}\mathbf{y}_j)^T (\mathbf{b} - \mu_{(j)} - \mathbf{W}_{(j)}\mathbf{y}_j), \quad (10)$$

where $\mu_{(j)}$ is the part in μ and $\mathbf{W}_{(j)}$ is the part in \mathbf{W} that correspond to the position j . Then the projection \mathbf{y}_r that can reconstruct \mathbf{b} with minimum reconstruction error over all the p possible positions is determined by

$$r = \arg \min_j (\varepsilon^a(j)). \quad (11)$$

Thus the suitable aging pattern for \mathbf{I} is \mathbf{z}_r . Step 2 afterward becomes trivial because r also indicates the position of \mathbf{I} in \mathbf{z}_r . Finally the age associated to the position r is returned as the estimated age of \mathbf{I} .

Table 1: Pseudocode of the AGES Algorithm

<i>Learning Procedure</i>	<i>Test Procedure</i>
<p>Input: $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ Output: \mathbf{W}, μ</p> <p>$i \leftarrow 0; \mathbf{x}_k^m \leftarrow [\mu^{(k)}]$; Apply PCA to the full-filled D to get \mathbf{W}_0; $\mu_0 \leftarrow$ the mean vector of the full-filled D; while $i < \tau$ and $\bar{\varepsilon}^a > \theta$ Estimate \mathbf{y}_k by Equation (5); Reconstruct \mathbf{x}_k by Equation (3); $\mathbf{x}_k^m \leftarrow \hat{\mathbf{x}}_k^m$; Apply PCA to the updated D to get \mathbf{W}_{i+1}; $\mu_{i+1} \leftarrow$ the mean vector of the updated D; $i \leftarrow i + 1$; end $\mathbf{W} \leftarrow \mathbf{W}_i; \mu \leftarrow \mu_i$</p>	<p>Input: $\mathbf{b}, \mathbf{W}, \mu$ Output: <i>age</i></p> <p>for $j \leftarrow 1$ to p do Place \mathbf{b} in the j-th position of \mathbf{z}_j; Estimate \mathbf{y}_j by Equation (5); Calculate $\varepsilon^a(j)$ by Equation (10); end $r \leftarrow \arg \min_j (\varepsilon^a(j))$; <i>age</i> \leftarrow the age associated to the r-th position;</p>

The flowchart of age estimation in AGES is shown in Figure 3. To make it more understandable, the face images (both the original and reconstructed ones) instead of feature vectors are shown in the aging patterns. It is interesting to note that when the test image is placed at a wrong position in the aging pattern, such as placing the 1 year old face at the position for the $p - 1$ years old face, the reconstructed faces become ghost-like twisted faces. On the other hand, if the test image is placed at the right position, the aging pattern subspace can not only reconstruct the original face very well, but also reasonably conjecture all other faces in the aging pattern. This vividly proves the ability of the aging pattern subspace to distinguish different facial ages. Moreover, as a byproduct of age estimation, the faces at different ages of the subject in the test image can be simulated at the same time without additional computation.

During the age estimation procedure of AGES, the proper aging pattern for the test image is generated based on both the aging pattern subspace and the face image feature. The subspace defines the general trend of aging, and the face image feature indicates the personalized factors relevant to facial aging. By placing the feature vector at different positions, candidate aging patterns specified to the test face are generated. Among these candidates, only one is consistent with the general aging trend, which can be detected via minimum reconstruction error by the aging pattern subspace. In this way, the personality of aging patterns can be utilized as much as possible without referring to extra profiles. The pseudocode of the entire AGES algorithm, including both the learning and test procedures, is summarized in Table 1.

4.3 Age Range Estimation

In real applications, sometimes the age range is more meaningful than the exact age of a person. For instance, the police are more often to describe an unknown suspect as “21 to 25 years old” rather than “exactly 23 years old”. There might be two ways to transform AGES from an age estimation algorithm to an age range estimation algorithm. The first way is straightforward, i.e. binning the estimated age into the corresponding age range. It is nice that the

estimator does not need to be re-trained. However, it fails to utilize the main advantage of age range estimation that it is possible to obtain a better aging pattern training set in both quantity and quality (completeness). This advantage can be exploited in the second way, i.e. re-labelling the training images by the age ranges and then feeding them to AGES. In this approach, each age range is represented by an integer ar . Suppose different age ranges have no intersection and each age range contains w different continuous ages. If the age of a face image is *age*, then the new label of this image is derived by

$$ar = \lfloor age/w \rfloor + 1, \quad (12)$$

where $\lfloor \cdot \rfloor$ is the floor operator. Such age range labels share similar characteristics with ages. The difference mainly lies in that each age range rather than the exact age is allocated a position along the t axis in the new aging pattern. This brings the advantage that more images might be available for each position. On the one hand, aging patterns with less or even no blank positions can be composed. On the other hand, large number of aging patterns for each individual can be generated through the combination of multiple images at different positions. After the new aging pattern training set for age range estimation is generated, the learning and test procedures of AGES can be executed as those used in age estimation.

5. EXPERIMENTS

5.1 Methodology

The FG-NET Aging Database [1] is used in this experiment. There are totally 1002 face images from 82 different subjects in this database. Each subject has 6-18 face images labeled with ground truth ages. The ages are distributed in a wide range from 0 to 69. The age distribution in either the number of images or the number of subjects is highly uneven. Lanitis *et al.* tested their methods in the age range from 0 to 30 [13] and later from 0 to 35 [12], where the data is relatively abundant. In this experiment, the algorithms are tested on all available ages ($p = 70$). This is a much

more difficult case because the additional age range (31-69) only accounts for less than 15% of the data set, but the possible classes (ages) increase by 126% ((70-31)/31). The images are collected under totally uncontrolled conditions. Besides the aging variation, most of the age-progressive image sequences display other types of facial variations, such as significant changes in 3D pose, illumination, expression, *etc.* This greatly increases the difficulty of age estimation on these face images. Some typical aging face sequences in this database are shown in Figure 4. Only the face region (not including hair) is used as training data. Typical face regions are shown in the first line of Figure 5

As mentioned in Section 2, among the methods proposed in [13], WAS is the best one without using the ‘lifestyle profiles’, and among the methods proposed in [12], AAS achieved the best performance. Thus in this experiment, AGES is compared with both WAS and AAS in age estimation as well as age range estimation.

The face feature extractor used in the experiments is the Appearance Model, which is a combined model of shape and intensity. The shape model is trained on 68 manually labeled key points on each face image. For the intensity model, each face image is aligned to the mean face shape (i.e. the mean of the key points) of the training set. It has been shown that the Appearance Model is robust against many facial variations such as illumination, view angle, and facial hair. Refer to [7] for more details about this face model. The extracted feature requires 200 ($n = 200$) model parameters in order to explain about 95% of the variance in the training data. All of the algorithms are tested based on these 200-dimensional feature vectors. In AGES, the dimension of the aging pattern subspace is set to 20 ($d = 20$). For WAS [13] and AAS [12], the original algorithms use Genetic Algorithm (GA) to calculate the optimum parameters of the quadratic aging functions. However, it is discovered in this experiment that if using Least Square Fit (LSF) to optimize the aging functions, the algorithms will always perform better than using GA (for example, when using GA, the mean absolute error of WAS is 8.24, much higher than that shown in Table 2) and the training procedure will be much faster. Thus in this experiment, LSF is used instead of GA in WAS and AAS to determine the quadratic aging functions. It is worth mention that the output of the aging function is a real number, but the ages in this paper are represented by non-negative integers. So the final output of the aging functions in WAS and AAS is rounded to the nearest non-negative integer. There are some additional parameters in AAS. In the experiment, several configurations of them are tested and the best result is reported. When the best performance is observed, the error threshold in the appearance cluster training step is set to 3, and the age ranges for the age specific classification are set as 0-9, 10-19, 20-39 and 40-69.

The algorithms are tested through the Leave-One-Person-Out (LOPO) mode, *i.e.* in each fold, the images of one person are used as the test set and those of the others are used as the training set (about $(1002/82) \times 81 \approx 990$ images). After 82 folds, each subject has been used as test set once, and the final results are calculated based on all the estimations. In this way, the algorithms are tested in the case similar to real applications, *i.e.* the subject for whom the algorithms attempt to estimate his/her age is previously unseen in the training set. Meanwhile, the relatively limited data set can be adequately utilized.



Figure 4: Typical aging face sequences in the FG-NET Aging Database. Each row shows part of the aging sequence from younger to older of one subject.



Figure 5: Typical samples used in HumanA and HumanB. The first line shows the gray-scale face regions used in HumanA. The second line shows the corresponding original color images used in HumanB.

As an important baseline, the human ability in age estimation is also tested. 51 face images are randomly selected from the database and presented to 29 observers. Each observer is asked to assign an age from 0 to 69 to each image based on their estimation. The results are calculated based on the $51 \times 29 = 1479$ estimations. The same image set is presented to the observers in two different ways. First, only the gray-scale face regions are shown to them. After they have made their estimations, the whole color images are then presented to them. The first test is denoted by HumanA, and the second one is denoted by HumanB. Figure 5 gives some examples used in HumanA and HumanB. In reality, people usually estimate age based on multiple cues, such as face, skin, clothes, hair, body build, voice and movement. HumanA intends to test the human ability of age estimation purely based on face, while HumanB intends to test the human ability of age estimation based on multiple cues including face, hair, skin color, clothes, and background. To avoid mutual hints between the two tests, HumanA (which is more difficult) is conducted before HumanB, and in both tests, the order of image showing is random. Note that the information provided in HumanA is the same as that provided to the algorithms, while in HumanB, additional information is available to the observers.

5.2 Performance Evaluation

5.2.1 Age Estimation

First the algorithms and the human ability are evaluated by the criterion used in [13] and [12], *i.e.* the Mean Absolute

Table 2: Mean Absolute Error of the Algorithms and Humans in Age Estimation

Method	MAE
AGES	6.77
WAS	8.06
AAS	14.83
HumanA	8.13
HumanB	6.23

Error (MAE). Suppose there are M test images, the real age of the image \mathbf{I}_k is age_k , and the estimated age of \mathbf{I}_k is \widehat{age}_k , then the MAE is calculated by

$$MAE = \sum_{k=1}^M |\widehat{age}_k - age_k| / M. \quad (13)$$

The MAEs of the algorithms and the humans are tabulated in Table 2. With additional information provided to the observers, HumanB gets remarkably lower MAE than HumanA does. It can be seen that no algorithm gets an MAE lower than that of HumanB. But both AGES and WAS achieve lower MAE than HumanA does. In detail, The MAE of AGES is about 17% $((8.13 - 6.77) / 8.13)$ lower than that of HumanA, and WAS also gets a slightly 0.9% $(8.13 - 8.06) / 8.13$ lower MAE than that of HumanA. It is a bit surprising that AAS gets a very poor result since it was proposed as an improved version of WAS. The reason might be that in this experiment the possible ages (0-69) are much more than those tested in [12] (0-35). Thus the cluster specific training scheme of AAS is much easier to overfit the training set and performs poorly on the test set. This justification will be further verified in the later experiment on age range estimation, where the possible classes are much less. Recall that in HumanA, the observers are provided with the same information as that fed into the algorithms. So the comparison between the algorithms and HumanA is more meaningful. But can the conclusion be drawn from Table 2 that both AGES and WAS perform better than humans in age estimation by face regions? Perhaps not.

MAE is only an indicator of the average performance of the age estimators. It does not provide enough information on how accurate the estimators might be. When considering accuracy of age estimation, people care more about how many estimations make an absolute error lower than an acceptable level, rather than the mean absolute error. Those estimations with higher absolute error than the acceptable level are usually regarded as the ‘wrong estimations’ and the exact error values of them do not matter much any more. Thus the performance of age estimation methods should be compared at different error levels. Since the acceptable error level is unlikely to be very high, the performance at lower error levels is more important than that at higher error levels. Here the error levels are defined as the possible values of the absolute error. At each level, a *cumulative score* can be used to evaluate the performance. Suppose there are M test images, $M_{e \leq l}$ is the number of test images on which the age estimation makes an absolute error no higher than l (years), then the cumulative score at error level l is calculated by

$$CumScore(l) = M_{e \leq l} / M \times 100\%. \quad (14)$$

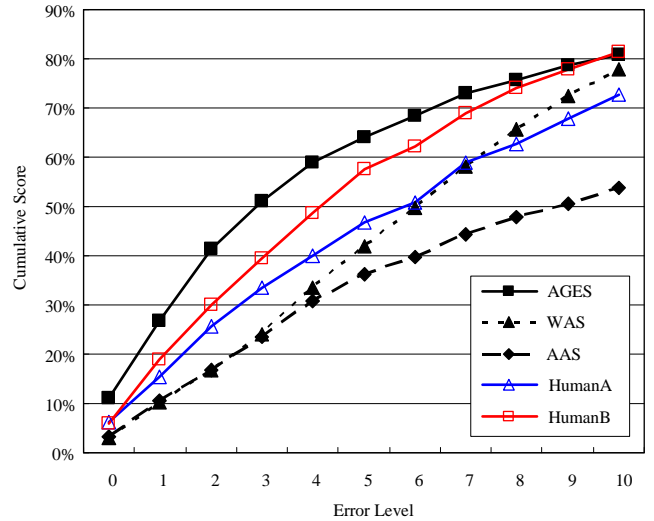


Figure 6: Cumulative scores of the algorithms and humans in age estimation at error levels from 0 to 10 (years).

If the ‘correct estimation’ is defined as the estimation with an absolute error no higher than l , then $CumScore(l)$ is actually the correct rate.

The cumulative scores of the algorithms and humans at the error levels from 0 to 10 (years) are compared in Figure 6. The situation at higher error levels is not shown because in general, age estimation with an absolute error higher than 10 (a decade) is not acceptable. As expected, the cumulative scores of HumanB are significantly higher than those of HumanA at all error levels. But even HumanB loses its leading position this time. Instead, AGES is the most accurate age estimator at almost all error levels. This is impressive since more information is provided in HumanB than that fed into AGES. Considering that even human beings cannot achieve very high accuracy in age estimation, the performance of AGES is reasonably good for applications that currently rely on human estimation. Although the MAE of WAS is lower than that of HumanA, its cumulative scores are worse than those of HumanA at most error levels. Not until the error level increases to 8 does the performance of WAS begin to exceed that of HumanA. WAS performs better at higher error levels than at lower error levels. But at least at the relatively important low error levels from 0 to 10, the average accuracy of WAS is worse than that of HumanA. Due to overfitting, again AAS gets the worst result.

5.2.2 Age Range Estimation

As mentioned in Section 4.3, there are two different ways to transform an age estimator into an age range estimator, i.e. binning and re-labeling. In this experiment, the performances of the algorithms (AGES, WAS and AAS) in age range estimation are tested through the re-labeling method. The age estimations of the humans are transformed into age range estimations through the binning method (because obviously we cannot ‘re-train’ the observers).

The ages (0-69) are divided into 14 ranges each of which contains 5 continuous ages, i.e. 0-4, 5-9, ..., 65-69. The age range label is calculated by Equation (12), where $w = 5$.

Table 3: Performance of the Algorithms and Humans in Age Range Estimation

Method	MAE	Hit Rate (%)
AGES	1.26	40.92
WAS	1.61	18.16
AAS	1.35	26.65
HumanA	1.64	25.70
HumanB	1.23	31.78

The performance of age range estimation can be evaluated by the same measurements used in age estimation, *i.e.* MAE as an indicator of average performance, and cumulative score as an indicator of accuracy. However, since the class label itself represents a range, the cumulative scores are only compared at the error level 0, *i.e.* the percentage of the exactly correct estimations (hit rate). If wider age range (corresponding to higher error levels) is preferred, the data set can be re-labeled by the preferred age ranges and the algorithms can still be evaluated at the error level 0. It is worth mention that the hit rate of the age range estimation with 5 continuous ages in each range is much tougher a criterion than the cumulative score at the error level 4 of the age estimation. If the estimated age range is correct, then the absolute age error is definitely no higher than 4. But contrarily if the difference between the estimated and real ages is no more than 4, the age ranges are not necessarily the same.

The performances of the algorithms and the humans in age range estimation are compared in Table 3. It can be seen that AGES achieves the highest hit rate and the second best MAE which is very close to the lowest MAE obtained by HumanB. When provided with the same information (AGES, WAS, AAS and HumanA), AGES is the best age range estimator in sense of both criterions. WAS is able to get an MAE slightly better than that of HumanA, but its hit rate is much lower than HumanA's. One remarkable difference with the results of age estimation is that AAS performs much better this time than WAS does in either MAE or hit rate. This supports the justification in the analysis of Table 2 that AAS tends to overfit the training data when the number of different classes is relatively large (70 ages), but it performs reasonably well when there are not many classes (14 age ranges).

In summary, the experiments reveal that AGES performs very well either as an age estimator or as an age range estimator. It significantly exceeds the state-of-the-art algorithms in both average performance and accuracy. Under the experimental configuration, AGES is even comparable to human ability. If provided with the same information, AGES can achieve better results than those of humans (HumanA). When the humans are provided with additional information (HumanB), AGES is still comparative: although the average performance of AGES is worse than that of HumanB, it has more chance to get accurate (low error levels) estimations.

6. CONCLUSIONS

This paper mainly concerns one barely touched problem in face-based HCI, *i.e.* automatic age estimation. Research

on this topic will midwife the age specific human-computer interaction systems to satisfy the increasing demand of the world for more humanized, friendly and intelligent computer systems. A novel automatic facial age estimation method named AGES is proposed in this paper. In the experiments, AGES achieves very good performance compared with the state-of-the-art algorithms and the human ability in either age estimation or age range estimation. The success of AGES is attributed to a number of new techniques and ideas that are consistent with the characteristics of facial aging variation. In particular, the following three are the most important ones: (1) The aging patterns are regarded for the first time as training samples and represented as sparse temporal data which naturally integrate the concept of identity and time; (2) An effective learning algorithm is developed to learn a representative subspace from the highly incomplete aging patterns; (3) A two-step age estimation approach geared to the characteristics of the aging variation is proposed based on the aging pattern subspace.

As illustrated by Figure 2 and Figure 3, AGES can be easily used to simulate aging effects on faces. It even does not need to know the age of the input image, which is required by most other aging face simulation technologies. The performance of AGES in aging face simulation still needs to be compared with the existing methods [3, 19, 20, 25, 30] in the future. Also the morphing technology might be integrated with AGES for better visual result.

In real life, people usually estimate age based on multiple cues. The advantage of multi-cue age estimation has been illustrated in the experiments by the remarkably better performance of HumanB over that of HumanA. The work described in this paper focus on age estimation by face images. In the future, AGES could be integrated, through multimodal fusion technology [29], with the algorithms from other biometric research, such as speaker recognition [6] and gait recognition [9], for multi-cue age estimation.

The aging face sequences used in this paper contains many types of variation other than the aging variation, such as changes in pose, illumination and expression. If the feature extractor in Figure 1 could filter out all facial variations except for the aging variation, then better performance could be expected. Thus the design of special feature extractor for age estimation is another important future work.

7. ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their valuable comments and suggestions, and Dr. M. Liao for carefully reading the manuscript. Part of the work was done when Xin Geng was at the LAMDA Group, Nanjing University. Z.-H. Zhou was partially supported by NSFC (60325207) and FANEDD (200343).

8. REFERENCES

- [1] The FG-NET Aging Database: <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm>.
- [2] T. R. Alley. *Social and Applied Aspects of Perceiving Faces*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.
- [3] M. Burt and D. Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and color information. *Proc. Royal Society of London, Series B: Biological Sciences*, 259(1355):137–143, 1995.

- [4] S. Daigo and S. Ozawa. Automatic pan control system for broadcasting ball games based on audience's face direction. In *Proc. 12th ACM Int'l Conf. Multimedia*, pages 444–447, New York, NY, 2004.
- [5] M. Davis, M. Smith, J. F. Canny, N. Good, S. King, and R. Janakiraman. Towards context-aware face recognition. In *Proc. 13th ACM Int'l Conf. Multimedia*, pages 483–486, Singapore, 2005.
- [6] G. Doddington, M. Przybocki, A. Martin, and D. Reynolds. The nist speaker recognition evaluation overview, methodology, systems, results, perspective. *Speech Communication*, 31:225–254, 2000.
- [7] G. J. Edwards, A. Lanitis, and C. J. Cootes. Statistical face models: Improving specificity. *Image and Vision Computing*, 16(3):203–211, 1998.
- [8] B. Fasel and J. Luetttin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1):259–275, 2003.
- [9] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.
- [10] I. T. Jolliffe. *Principal Component Analysis, 2nd Edition*. Springer-Verlag, New York, 2002.
- [11] Y. H. Kwon and N. da Vitoria Lobo. Age classification from facial images. *Computer Vision and Image Understanding*, 74(1):1–21, 1999.
- [12] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. Systems, Man, and Cybernetics - Part B*, 34(1):621–628, 2004.
- [13] A. Lanitis, C. J. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.
- [14] A. Leonardis and H. Bishof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.
- [15] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(2):1357–1362, 1999.
- [16] I. Moon, K. Kim, J. Ryu, and M. Mun. Face direction-based human-computer interface using image observation and emg signal for the disabled. In *Proc. IEEE Int'l Conf. Robotics and Automation*, pages 1515–1520, Taipei, Taiwan, 2003.
- [17] P. Olivieri, J. Gips, and J. McHugh. EagleEyes: Eye controlled multimedia (video). In *Proc. 3rd ACM Int'l Conf. Multimedia*, pages 537–538, San Francisco, CA, 1995.
- [18] A. J. O'Toole, K. A. Deffenbacher, D. Valentin, K. McKee, D. Huff, and H. Abdi. The perception of face gender: The role of stimulus structure in recognition and classification. *Memory and Cognition*, 26(1):146–160, 1998.
- [19] A. J. O'Toole, T. Vetter, H. Volz, and E. Salter. Three-dimensional caricatures of human heads: Distinctiveness and the perception of age. *Perception*, 26:719–732, 1997.
- [20] J. B. Pittenger and R. E. Shaw. Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1(4):374–382, 1975.
- [21] N. Ramanathan and R. Chellappa. Face verification across age progression. In *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pages 462–469, San Diego, CA, 2005.
- [22] G. Rebel. *Mehr Ausstrahlung durch Körpersprache*. Gräfe und Unzer Verlag GmbH, München, 1997.
- [23] S. Roweis. EM algorithms for PCA and SPCA. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 626–632. MIT Press, Cambridge, MA, 1998.
- [24] Y. Takemae, K. Otsuka, and N. Mukawa. Video cut editing rule based on participants' gaze in multiparty conversation. In *Proc. 11th ACM Int'l Conf. Multimedia*, pages 303–306, Berkeley, CA, 2003.
- [25] B. Tiddeman, M. Burt, and D. Perret. Prototyping and transforming facial texture for perception research. *IEEE Computer Graphics and Applications*, 21(5):42–50, 2001.
- [26] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 61:611–622, 1999.
- [27] X. Wei, L. Yin, Z. Zhu, and Q. Ji. Avatar-mediated face tracking and lip reading for human computer interaction. In *Proc. 12th ACM Int'l Conf. Multimedia*, pages 500–503, New York, NY, 2004.
- [28] T. Wiberg. Computation of principal component when data are missing. In *Proc. 2nd Symp. Computational Statistics*, pages 229–236, Berlin, Germany, 1976.
- [29] Y. Wu, E. Y. Chang, K. C. Chang, and J. R. Smith. Optimal multimodal fusion for multimedia data analysis. In *Proc. 12th ACM Int'l Conf. Multimedia*, pages 572–579, New York, NY, 2004.
- [30] Y. Wu, N. Thalmann, and D. Thalmann. A dynamic wrinkle model in facial animation and skin aging. *Journal of Visualization and Computer Animation*, 6:195–205, 1995.
- [31] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pages 246–253, Pittsburgh, PA, 1999.
- [32] L. Zhang, Y. Hu, M. Li, W. Ma, and H. Zhang. Efficient propagation for face annotation in family albums. In *Proc. 12th ACM Int'l Conf. Multimedia*, pages 716–723, New York, NY, 2004.
- [33] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399–459, 2003.