

Human face aging with guided prediction and detail synthesis

Ming-Han Tsai · Yen-Kai Liao · I-Chen Lin

Published online: 2 March 2013

© Springer Science+Business Media New York 2013

Abstract In this paper, we present an example-based method to estimate the aging process of a human face. To tackle the difficulty of collecting considerable chronological photos of individuals, we utilize a two-layer strategy. Based on a sparse aging database, an EM-PCA-based algorithm with the personal guidance vector is first applied to conjecture the temporal variations of a target face. Since the subspace-based prediction may not preserve detailed creases, we propose synthesizing facial details with a separate texture dataset. Besides automatic simulation, the proposed framework can also include further guidance, e.g., parents' impact vector or users' indication of wrinkles. Our estimated results can improve feature point positions and user evaluation demonstrates that the two-layer approach provides more reasonable aging prediction.

Keywords Face aging · Image generation · Pattern analysis

1 Introduction

Aging human faces is a distinctive task in facial image processing and pattern analysis. It aims at generating an elder face image from a young one, and has many important applications, such as pursuing criminals, seeking missing children or face recognition.

If a child has been missing or a criminal has fled for a long period. Her or his facial appearance usually changes due to the aging process. For the law enforcement agencies, differences between the aged person and the early-taken photos make face identification difficult. The same problem occurs in automatic face recognition systems. Portrait samples in a recognition system were established at early days, and it is troublesome to update the database every several years. While applying face aging techniques, the maintenance of a portrait database can be easier and less frequent. It also improves the recognition accuracy.

M.-H. Tsai · Y.-K. Liao · I.-C. Lin (✉)

Department of Computer Science, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu City 30010, Taiwan
e-mail: ichenlin@cs.nctu.edu.tw

On the other hand, more and more online games or internet communities provide personalized avatars as ones' representatives. Applying aging simulation can increase realism of virtual characters.

Facial aging is an inevitably biological process and makes plenty of changes on appearance features. For instance, due to the elasticity loss of facial tissue, creases and wrinkles gradually appear. Zimble et al. [23] described the anatomy of facial aging. Recent research shows that the bone structure, e.g., mandible, also changes significantly with age [19] and it results in the changes of facial geometry. Besides, the aging process is also influenced by various factors, such as inheritance, lifestyle, environment, and so on. The complicated variations make it difficult to predict one's aged face by a deterministic function or simulation by a simplified physics model. Therefore, we take a statistical approach as the basis and further adjust the prediction with examples.

It is challenging to collect qualified image examples for human aging analysis. The aging database should have sufficient numbers of subjects. Each subject has to provide complete photo samples from her/his childhood to elder ages, so called the age pattern. Fortunately, a public human database, FG-NET (Face and Gesture Recognition Research Network) Aging Database [21], helps us lighten the burden. We make use of this database, where 1002 photos of 82 Caucasian subjects (35 females and 47 males) at different ages (0 to 69) are recorded. Several examples are shown in Fig. 1.

Even with such ample contents, aging analysis is still difficult. In FG-NET, most subjects' aging patterns are incomplete and the distributions of the "missing" ages are uneven. Besides, most of the photos, taken several decades ago, were contaminated due to color fading, digitization by scanning, and other noises. Hence, in addition to FG-NET, we include

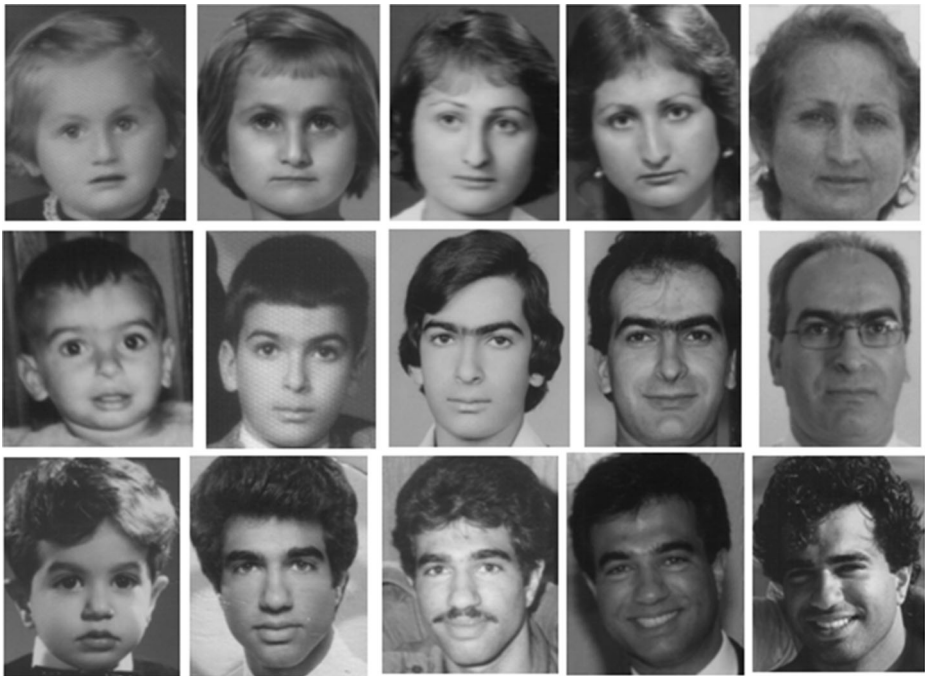


Fig. 1 Example sequences in FG-NET [21]. Each row shows a portrait sequence of the same person from the young to elder ages

a supplementary facial texture dataset and propose a novel two-stage method for simulating the age process.

In the first stage, we approximate the geometric and appearance variations with the FG-NET database. Our approximation method is inspired by Aging Pattern Subspace (AGES) algorithm by X. Geng et al. [6]. They used an Expectation-maximization (EM) algorithm under Principal Component Analysis (PCA) subspace to full-fill the incomplete aging database for improving the accuracy of an age recognition system.

However, our goal is to generate elder faces for art or entertainment usage, instead of recognizing age numbers, and our inputs are only one or few pictures at child or youth ages. When we directly apply AGES algorithm for elder data prediction, the results are unstable and of bare personal characteristics. To solve this problem, we propose setting guidance faces, according to the target's characteristics. Our experiments show that the aged face can preserve more personal characteristics and will not be dominated by few samples at that age. Moreover, our system can further include enhancement vectors, such as the influence of the parents by the similarities between the child and parents. That is not mentioned in related articles.

Although our algorithm at the first stage can predict the variations of facial geometries and appearance, the facial appearance are blurry due to two issues. First, the original image resolutions in the FG-NET are about 5,000 pixels and a large portion of samples are in gray-scale. Second, during the PCA-based estimation, facial textures are projected onto the subspace composed by principal axes. High-frequency facial details are usually discarded, but these wrinkles and creases are important visual cues for one's age.

In the second stage, we propose using a texture synthesis approach to enhance detailed facial appearances. Since the chronological variations have been approximated in the first stage, in this stage, we do not need to carefully collect age patterns, but require abundant face images of various ages in a higher resolution. Our facial texture database consists of 133 photos collected from the internet. After aligning and normalizing the portraits by 68 feature points defined in FG-NET database, we search the most appropriate skin details according to the estimated low-frequency aged face, and then transfer the details by patch-based texture synthesis. Furthermore, based on our controllable synthesis method, users can also accentuate the possible wrinkles, creases or disguise by intuitive graphical interfaces.

We compare our results with photos generated by related articles through user evaluations. The experiments show that the proposed method can generate reasonable aged faces and keep personal characteristics.

This paper is organized as follows. In Section 2, we introduce related articles and several age recognition and prediction methods. In Section 3, we introduce how we utilize guided aging patterns for aging simulation and how to apply parents' effects in prediction. The facial appearance synthesis and transfer are described in Section 4. Section 5 demonstrates our results and comparisons. We conclude this paper in the last section.

2 Related work

This section describes related articles about age recognition and simulation of aging effects. As mentioned above, the aging process changes both facial geometry (shape) and appearance (texture). Researches in face aging usually focus on modeling and approximation of these two points.

For facial image manipulation, Burt and Perrett [2] evaluated the facial colors and shape in the same age bracket and extracted the difference between the means of the elder and

young faces. Ramanathan and Chellappa [15] represented the growth-related shape variations of young faces by a craniofacial growth model. Hubball et al. proposed a data-driven approach of aging simulation [8]. They used non-uniform radial basis function (NURBF) and genetic algorithms for facial image parameterization and aging regression. The results look reasonable but due to the native weakness of data regression, salient or irregular data like wrinkles are under-fitted. Furthermore, an evolutionary computing method makes the training process non-deterministic for different setting, and difficult for further user intervention.

Suo et al. [20] proposed a compositional model for face aging. They divided face samples into several components, e.g., hairs, eyes and mouths, from coarse to fine levels. They represented faces of the same age group by a three-level graph. Then, they learned the probabilities of face aging with Markov chains. This method generated impressive results but required a large number of high-quality samples with precise alignment.

Cootes et al. proposed an Active Appearance Model (AAM) [4] that represented a face image by subspace-projected parameters of facial shape and texture. This representation is popularly used in human face research and is applied in several studies on aging recognition. Wang et al. [22] propose formulating the age estimation on face images as a Bayesian estimation problem. Lanitis et al. [9] projected the landmarks and grayscale image of a face onto 50 parameters in the subspace. Then, they approximated one's aging variations by polynomial functions. For an unseen individual, they used weighted-blending aging functions to estimate the aging effects. Geng et al. [6] proposed a method called Aging Pattern Subspace (AGES). They first projected the shape and image data to the PCA subspace with 200 parameters and sorted these data in time order, named aging patterns. Then, they tackled the problem of incomplete aging patterns by least squares fitting of subspace parameters.

Fu and Huang [5] demonstrated that manifold of conformal embedding analysis and quadratic regression can improve the accuracy of age estimation. In the same year, Guo et al. [7] presented another method for improvement by locally adjusting curves of support vector regression. Ramanathan et al. [16] provide a thorough analysis of various aging modeling methods.

Other researches simulated the aging effect on 3D face model. Based on a 3D morphable face model [1], Scherbaum et al. [17] reconstructed a 3D face model from the input image, and computed the individual aging trajectories of the 3D model. They rendered the aged 3D face model back to the 2D image for the final composition.

In our work, besides prediction with guidance vectors, we also present a patch-based method for detail transfer. A closely related work is Visio-lization algorithm proposed by Mohammed et al. [13]. It was used for generating unseen facial images. They perturbed parameters at principal eigenface subspace as a preliminary face, and seamlessly concatenated closest facial patches from a database with high-resolution portraits.

3 Aging effects by PCA-based prediction

3.1 Preparing training data

We use FG-NET aging database [21] as our primary training database. Here, we briefly introduce how we arrange the training data for later evaluation.

Each sample in the FG-NET database contains the face image, identity, gender, age at the taken year, and 68 points indicating facial features. To make the data comparable and grasp the variations on geometry and appearance, we normalize color intensity and shape of each

face images. At the first step, we rotate tilted faces to upright positions and scale the face by the width of face features. Since a large portion of training samples are in grayscale, for consistence, all images are transformed into grayscale ones. Then, we project the textures about 5,000 pixels and geometry variations of 68 feature points to 200 parameters of PCA subspace. The feature points we marked are show in Fig. 2. The subspace projection still preserves more than 95 % variance of the original data. To keep the aging characteristics of each gender, we process male and female data separately. Note that we separate male and female only in the age-pattern reconstruction stage. The cross-gender operations in later sections, such as similarity to parents, are performed on the identical space.

After normalizing the face data, we construct the aging patterns for all subjects in the database. An aging pattern is a sequence of an identity's face data sorted from the young to elder ages. However, the age range of subjects in FG-NET database is large (0 to 69), and the age distribution of a subject's samples is uneven and sparse (12 samples in average). If we took one year as a slot, the aging patterns would become too specific, and plenty of slots would be empty, so called missing data. In this case, the prediction procedure would be unreliable and over-fitted for such sparse aging patterns.

Instead, we empirically divide 70 ages into 17 age groups according to visual distinction. If there are more than one sample data of an identity in a certain group, we evaluate their average as the representative. Table 1 shows the mapping between an original age and its designated age group. Merging the age slots into 17 groups can effectively reduce the missing data amount in aging patterns. An example is shown in Fig. 3. On the other hand, discrete age groups may result in the boundary effect. For instance, an image of 40 year old shall be close to both 35–39 and 40–49 groups, but it can affect only the 40–49 group. To alleviate this problem, we can either put redundant samples (but of half influence) to nearby



Fig. 2 The 68 feature points defined in FG-NET database. More dense feature points are used to indicate the month/lips, and fewer points are used for the face contour

Table 1 The mapping table between ages and age groups

Original age	Age group
0 to 1	0
2 to 3	1
4 to 5	2
6 to 7	3
8 to 9	4
10 to 12	5
13 to 15	6
16 to 18	7
19 to 21	8
22 to 24	9
25 to 27	10
28 to 30	11
31 to 34	12
35 to 39	13
40 to 49	14
50 to 59	15
60 to 69	16

groups or weighted-combine the simulated results from the designated and nearby groups. We take the first strategy in our system.

3.2 Aging pattern subspace

In this subsection, we briefly describe how to use AGES algorithm [6] to full-filled incomplete aging patterns. We separate group data of an aging pattern into two types: the

a

0	1	2	3	4	5	6	7	8
m			m					
9	10	11	12	13	14	15	16	
	m			m		m	m	

b

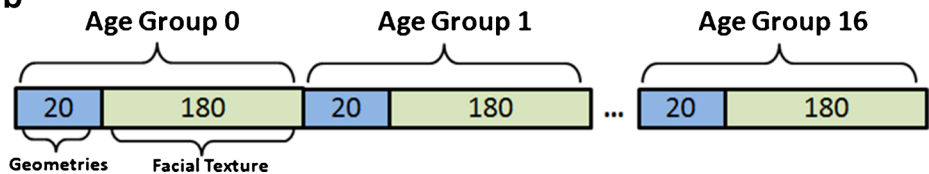


Fig. 3 **a** The aging pattern of image data. The missing data in the aging pattern are marked “m”. **b** The aging pattern data. 20 and 180 PCA parameters are used for geometry and appearance data, respectively

available data and the missing data. The purpose of AGES algorithm is to conjecture appropriate values of missing data according to available data in the database.

The AGES algorithm utilizes PCA to project the original data onto the subspace that represents the major variability of the aging patterns. The projection from original data to principal component subspace is

$$y = W^T(x - \mu), \tag{1}$$

where x is an aging pattern from the training data and μ is the mean vector of all x . W is composed of the principal eigenvectors of covariance matrix of x . If there are N aging patterns in the database, the database can be represented as $D = \{x_1, x_2, \dots, x_N\}$. For an identity k , $1 \leq k \leq N$, the parameters y_k in the PCA subspace can be calculated by projecting x_k as (1) and the reconstruction of x_k from y_k is thus

$$\hat{x}_k = \mu + W y_k. \tag{2}$$

However, plenty of slots are without data. We indicate an aging pattern x_k as two parts: the available data x_k^a and the missing data x_k^m . We calculate the mean value of available data of target group i from all aging patterns, and set the mean value (it can include small perturbations) as the initial guess of missing data $x_k^m(i)$. After full-filling missing data by initial guesses, an Expectation-Maximization like algorithm can be used to learn the appropriate values of age pattern from available data.

The EM-like algorithm iteratively utilizes the Estimation step and the Maximization step to maximize expected likelihoods of missing data. In the Estimation step, y_k cannot be calculated by direct projection of x_k . It should be solved from the available data in the aging pattern as the least squares solution of

$$[W_t \{x_k^a\}] y_k = x_k^a - [\mu_t \{x_k^a\}] \tag{3}$$

where $[W_t \{x_k^a\}]$ and $[\mu_t \{x_k^a\}]$ are the corresponding parts in W_t and μ_t , and the t is the iterative time stamp. The new parameters \hat{x}_k can be calculated by (2). The missing data x_k^m is updated by \hat{x}_k^m , but the available data x_k^a are constant since they are real data for reference. In the Maximization step, we evaluate the transformation matrix W_{t+1} and mean vector μ_{t+1} by the updated data with standard PCA. The AGES algorithm uses the mean reconstruction error between the available data x_k^a and the reconstructive available data \hat{x}_k^a as the objective function.

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

The EM-like algorithm performs iteratively until mean reconstruction error ε or the iterative time reaches their thresholds. Figure 4 shows the results that missing data in the aging pattern are full-filled by AGES algorithm. Figure 5 shows an example of iterative refinement of missing data and the ground-truth image.

3.3 Guided aging simulation for a single or few images

In most applications, such as aged face identification, the input is not an aging pattern with sparse data, but few images at young ages. In this subsection, we describe how to use the full-filled FG-NET database to predict aging effects by few or even one input image.

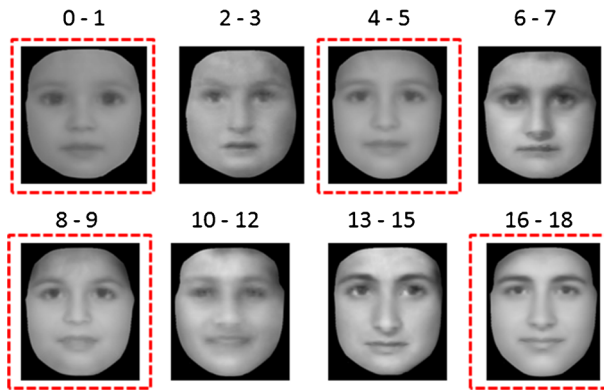


Fig. 4 The full-filled aging pattern. The images with dashed line frames are the missing data full-filled by AGES method. The ages are marked above the image

Assume that the age of the input image is given and it belongs to the r -age group. We first extend the concept of EM-like data filling for an unseen image, and the subspace-projected y_{new} of aging pattern is solved by

$$[W_t\{r_{new}\}]y_{new} = x_{input}^r - [\mu_t\{r_{new}\}], \quad (5)$$

where $[W_t\{r_{new}\}]$ and $[\mu_t\{r_{new}\}]$ are the parts corresponding to r -th age group in W_t and μ_t .

After iteratively minimizing the reconstruction error, we can use y_{new} to reconstruct the whole aging pattern. However, the estimated results are unstable, and the personal characteristics may not be preserved.

The unsatisfactory prediction results from two issues: loss of temporal guidance and insufficient variety of subjects. When we use multiple and sufficient input images of ages across decades, the EM-like method can estimate aging effects from training data according to both temporal (age groups) and spatial (geometry and appearance) variations. By contrast, when using a single or few young images as the input, we can only rely on spatial similarity. If the input subject is unlike to training subjects at that age, the unexpected prediction occurs. Figure 6b is a prediction result from 17 to 41 years old by AGES method.

For more reasonable prediction, we propose generating a guide aging pattern at the initial stage. We regard the difference between a subject and the mean face as the personal characteristic of a certain age. The difference between the mean faces of two ages can be considered as the main trend of aging. We can generate a new aging pattern by propagating

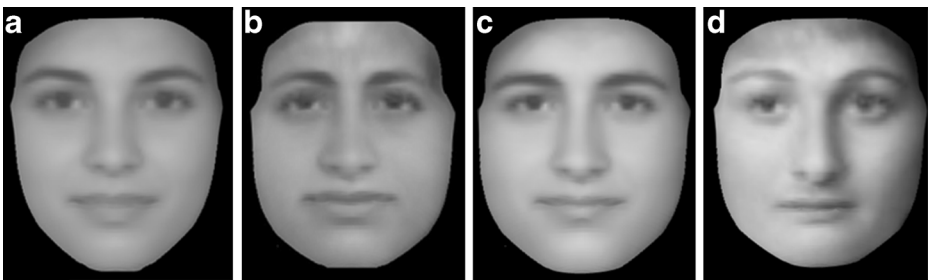


Fig. 5 Iterative refinement of missing data. **a** The initial guess face image at 17 years old. **b** The image at 2nd iteration **c** The full-filled missing image. **d** The ground truth image at 17 years old

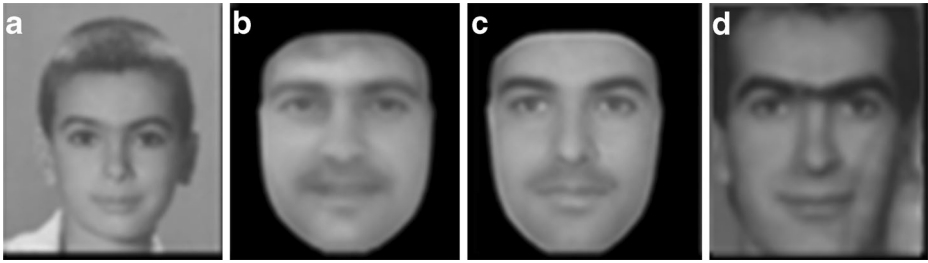


Fig. 6 Aging simulation by a single input image. **a** The input image at 14 years old. **b** The 41-years-old image simulated by AGES. **c** The 41-years-old image simulated by the proposed method. **d** The ground truth image at 41 years old

the personal characteristic of the input age to other age groups. We evaluate the guided aging pattern $X_{guide} = x_{guide}^i \quad 1 \leq i \leq 17$ through minimizing an objective function:

$$\begin{aligned} \arg \min_{x_{guide}^i, i \notin input} \Sigma & \left(\left(x_{guide}^i - \mu_{Tgt_gender}^i \right)^2 - \left(x_{input}^r - \mu_{Tgt_gender}^r \right)^2 \right)^2 \\ & + W_t \left(2x_{guide}^i - x_{guide}^{i-1} - x_{guide}^{i+1} \right)^2 \end{aligned} \tag{6}$$

where Tgt_gender is the gender of the input subject for aging prediction; $\mu_{Tgt_gender}^i$ is the mean vector of the target gender at age group i and r is the index of age group which input age belongs to. The first term of (6) preserves the personal face characteristics to the mean faces, and the second temporal-coherence term keeps the aging process smooth weighted by W_s and W_t , separately. With the new pattern x_{guide} as the initial vector, we can then apply the EM-PCA process as Eq. (5).

Taking the differences among mean faces for aging prediction was also discussed in [2]. Nevertheless, instead of a deterministic assignment, we take the trend as a guide and iteratively approximate the aging effects with the learned spatial and temporal variations from the training data. Figure 6 shows the simulated face from 17 to 41 years old and the ground truth. Other examples are shown in Fig. 12. The proposed aging with the guidance pattern can apparently make more reasonable prediction than that by the AGES method.

As mentioned in subsection 3.1, we use gray-scale images in training for consistence. For a color input image, we first transfer the data to YUV domains, and estimate the aging effects based on the Y data. To synthesize a colored aged image, we can transfer the UV data from the input to the aged image according to their pixel correspondences.

3.4 Adding Parents’ effects

In this subsection, we present how to guide the aging prediction by extending training data for the input subject. In our paper, we take parents’ guidance vectors as an optional input. We would like to emphasize that the aging relationship between parents and children is still a complicated and worth-investigating topic. Here, we only demonstrate that our system has the possibility to include various guided effects.

According to the Mendelian theorem, a child’s facial features may partially inherit from his/her parents. We propose using such highly influential aging patterns in the training data. If it is difficult to collect the parents’ aging patterns, we can generate guided aging patterns of the parents by only one or few images.

In our prediction, we separate male and female subjects to grasp the aging effects of the corresponding gender, but a child may also resemble her/his parent of the opposite gender. To utilize the resemblance and retain the gender-dependent aging estimation, we remove the gender characteristics of parents in the opposite gender by compensating the difference between mean faces of two genders at the age group. Similar to the personal guidance vector, parents' guided aging patterns: $X_{father_guide} = \{x_{father_guide}^i \mid 1 \leq i \leq 17\}$ and $X_{mother_guide} = \{x_{mother_guide}^i \mid 1 \leq i \leq 17\}$ can be formulated by minimizing gender difference as follows:

$$\begin{aligned}
 & \arg \min_{x_{father_guide}^i \notin input} \sum W_s \left(\left(x_{father_guide}^i - \mu_{Male}^i \right)^2 - \left(x_{father_input}^{father_age} - \mu_{Male}^{father_age} \right)^2 \right)^2 \\
 & + W_t \left(2x_{father_guide}^i - x_{father_guide}^{i-1} - x_{father_guide}^{i+1} \right)^2 \\
 & \arg \min_{x_{mother_guide}^i \notin input} \sum W_s \left(\left(x_{mother_guide}^i - \mu_{female}^i \right)^2 - \left(x_{mother_input}^{mother_age} - \mu_{female}^{mother_age} \right)^2 \right)^2 \\
 & + W_t \left(2x_{mother_guide}^i - x_{mother_guide}^{i-1} - x_{mother_guide}^{i+1} \right)^2
 \end{aligned} \tag{7}$$

where *father_age* and *mother_age* are the indices of father's and mother's age group, respectively. x_{father_input} and x_{mother_input} are the parameters of parents' input images, respectively.

One intuitive thought is directly included these new guidance vectors in the training data as aging patterns of two new subjects. On the contrary, we can accentuate the parents' effects according to their resemblances to the target child.

Assume that the target child can be approximated by linear combination of parents' faces at the input child age *r*. We solve the impact ratios of father R_f and mother R_m by minimizing the energy function E_p :

$$E_p = \left(R_f x_{father_guide}^r + R_m x_{mother_guide}^r - x_{input}^r \right)^2 \tag{8}$$

A new aging pattern can be generated according to similarity ratio to parents as shown in (9). The new training database become

$$D_{Parent_ext} = D_{origin} \cup \left(R_f \cdot x_{Father_guide} + R_m \cdot x_{Mother_guide} \right) \tag{9}$$

Another idea is to fit the child's initial guidance vector for the parents' pattern. The original guided aging pattern in (6) is to propagate the personal characteristic of the input age to other age groups. To propagate the parents' characteristics to whole age groups, we may use the combined parent aging pattern in (9) as additional guidance. We require the new guided vector in (6) not only keep similar distance to mean of all age groups, but also be close to parents' aging pattern.

Both methods perform well in our experiments. The direct guided method keeps more parents' characteristic on child's face, but the first method is moderate and not overemphasized. Since a child may be affected by her/his parents more or less, the faithfulness of these two methods differs from person to person. In some cases, children's faces are affected more by environment, or life style, an alternative guidance vector should be designed. Therefore, we set this part as an optional tool for users. Figure 7 shows an example by these two methods.

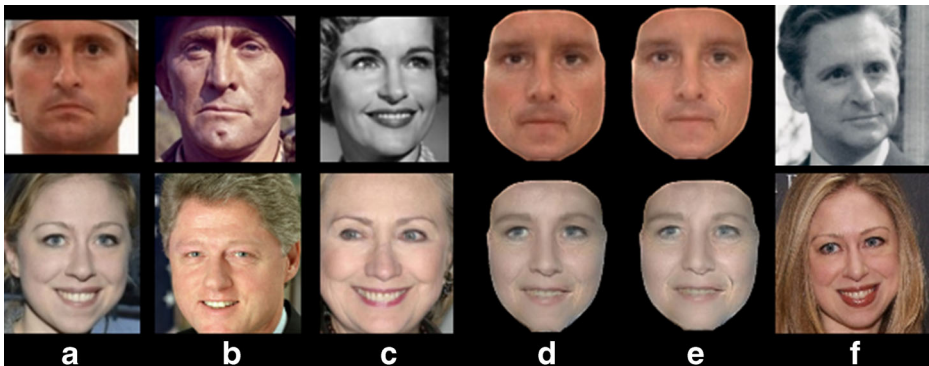


Fig. 7 Adding Parents Effects. **a** The input image. **b** Father's image. **c** Mother's image. **d** Aged images by our method without parent's effects. **e** Aged image (to 40–49 years old) with parents' effect. Upper (**e**): Including parents' mixed pattern into training database. Lower (**e**): Fitting the guidance vector to parents' one. **f** Real images (upper 40, lower 32 years old)

4 Facial detail enhancement

In the above sections, we utilize PCA subspace from FG-NET aging patterns to approximate reasonable aging effects. Under the projective subspace with blurry training samples, it is difficult to model most of facial details. In this section, we describe how to enhance the detailed wrinkles or creases, which are important cues in visual identification of an elder face.

4.1 Elder skin texture synthesis

The first step of detail enhancement is to generate a detailed face close to the guided prediction result, since most images in FG-NET are without sufficient details. We make use of the flourishing Internet search techniques and collected 71 male and 62 female images of age 20 to 70. Most portraits have distinct ages, and some were indicated empirically.

An intuitive thought is to find the closest face image and transfer the details to the predicted face. This image can be close to the target in average, but it may not be appropriate to every parts of the target face. For example, the nasolabial folds and forehead wrinkles are fit for the target image, but the shape and shade of the noses are different.

To fit for all parts of the face and preserve the image resolution as well, we propose using patch-based texture synthesis to generate the closest face texture. The patch-based texture synthesis is inspired by Visio-lization method [13]. Their purpose is to generate faces of inexistent identities by stitching patches from database. On the contrary, we use patches to piece up the image closest to a low-frequency reference face.

After alignment and normalization, we divide all images in the texture database into regular patches. For a reference face (target) image, it is also divided into patches, and each reference patch is then compared with all of the patches at the same position in the texture database. In the comparison process, we aim at finding the patch with the smallest sum of the square difference (SSD) in pixel values to the reference patch. We can also find the patch with the closest gradient values and adjust

their gains. Therefore, we define the cost function between two patches P_a and P_b as the combination of intensity and gradient differences:

$$Cost(P_a, P_b) = \sum_i \|P_a(i) - P_b(i)\|^2 + \lambda \sum_i \|\nabla P_a(i) - \nabla P_b(i)\|^2, \tag{10}$$

where i is the pixel index; ∇ represents the gradient operator; λ is the weight for the gradient term. In our implementation, $\lambda=1$.

Several constraints are further considered to improve the stitched result. To keep smooth stitching, we require the cost at boundary of adjacent patches should also be under a user-defined threshold. To keep the symmetry of human faces, we restrict the two symmetric patches in the right and left parts of a face should be selected from the same image. The process of patch-based synthesis is shown in Fig. 8.

Even with the minimum intensity and gradient cost and boundary constraints, obvious gaps may still occur between two adjacent patches selected from different texture sources. We use Poisson image editing [14] to remove the artifact boundary between two adjacent patches and preserve gradients within two patches. The image of directly stitched patches and the adjusted image are shown in Fig. 9b and c, respectively.

4.2 Transferring appearance details

After stitching the most appropriate skin patches, we have to transfer the creases and wrinkles of the texture image to the predicted face. One popularly-used approach for image transfer is Expression Ratio Image (ERI) by Liu et al. [12]. Given a pair of expressive and neutral faces of an individual, their deduction shows that the surface variations of an aligned pixel can simply be approximated by their intensity ratio if Lambertian reflectance assumption is applied.

In our case, we only have a low-frequency target face T_{base} (e.g., Fig. 9a) and a composite face with details S_{detail} (e.g., Fig. 9c). To acquire the intensity ratios, we have to create a corresponding face without details. Thus, we apply Gaussian filter to S_{detail} and denote the smoothed face as S_{base} . Under the Lambertian reflectance model, the intensity of an image S at pixel p can be formulated as

$$S(p) = k(p) \sum_{i=1}^{LNum} I_i(n(p) \cdot l_i) \tag{11}$$

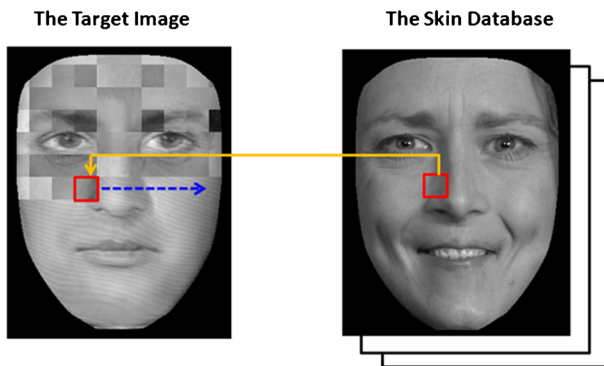


Fig. 8 The process of patch-based skin texture synthesis. The most appropriate patch is retrieved from the skin texture database according to gradient and intensity differences, boundary and symmetric constraints

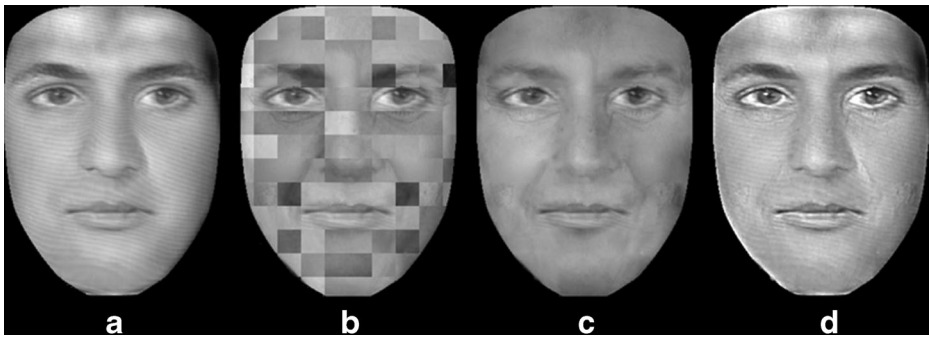


Fig. 9 **a** The original target image. **b** The original patches of a stitched face. **c** The face image adjusted by Poisson image editing. **d** The target image enhanced by detail ratios

where I_i and l_i are the intensity and direction of light source i and $LNum$ is the light source amount; $k(p)$, $n(p)$ are the reflectance coefficient and surface normal at pixel p . Our detail ratio of a pixel p can be represented as

$$\frac{S_{detail}(p)}{S_{base}(p)} = \frac{\sum_{i=1}^{LNum} I_i(n_{s_detail}(p) \cdot l_i)}{\sum_{i=1}^{LNum} I_i(n_{s_base}(p) \cdot l_i)}. \tag{12}$$

Assume that n_{s_base} and n_{t_base} , the surface normal of the S_{base} and T_{base} are nearly identical, and the lighting conditions are also similar. We can therefore approximate the detail-enhanced image of T_{base} as

$$T_{detail}(p) \approx T_{base}(p) \frac{S_{detail}(p)}{S_{base}(p)}. \tag{13}$$

An example of detail transfer is shown in Fig. 9, where image a and c are T_{base} , S_{detail} and d is the detail-transferred image T_{detail} .

4.3 Interactive detail enhancement

We have introduced the images predicted by our guided prediction algorithm and enhanced by appearance detail transfer. However, certain parts of the predicted face can be too blurry for search of wrinkle patches. In other cases, a user or witness would like to suggest or enhance the predicted results.

Interestingly, since the patch-based detail enhancement (4.1) searches the most appropriate patches according to gradients and intensities. Users can intuitively paint on the predicted faces and our system can automatically extract the possible wrinkles and creases from texture database. Figure 10 shows two examples of interactive detail enhancement.

5 Experiment and discussion

We performed two types of experiments to evaluate the effectiveness of our method. First, we compared feature-point accuracy of our results with comparative methods and the ground-truth data. Second, we performed user studies to evaluate the responses from users.

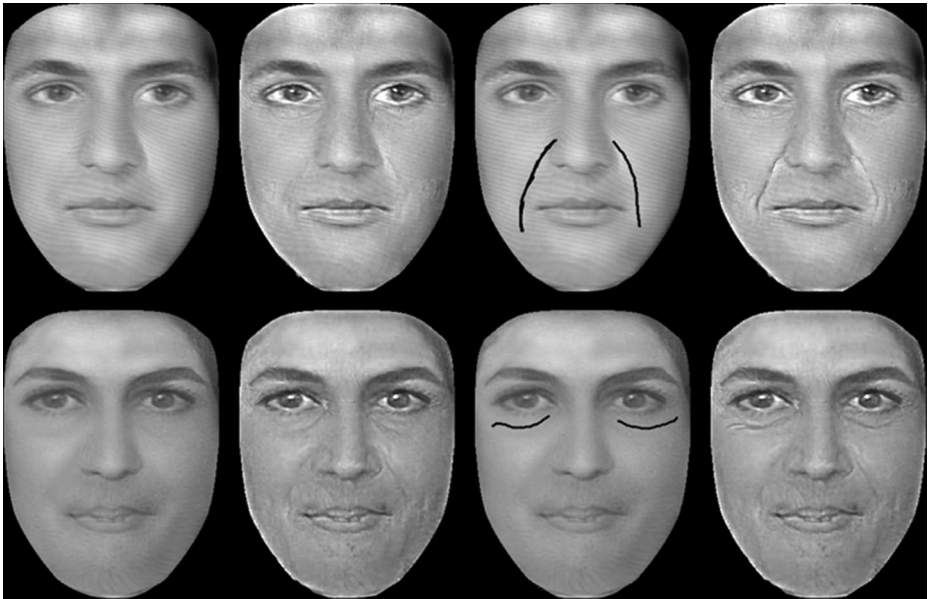


Fig. 10 Two examples of interactive aging enhancement. The 1st column shows the original reference or predicted faces; the 2nd column shows the detailed face by patch-based transfer; the 3rd column shows indication curves assigned by users; the 4th column shows the results by interactive enhancement

We used FG-NET database for training and testing. The data of a testing subject are excluded during the training stage. We also filter out images in side view, with glasses or conspicuous beard. To validate the aging effects, we chose three age gaps: from 5 to 30, from 10 to 40, and from 20 to 50, in our experiment. All compatible image pairs fit these three gaps are included in our experiments. For examples, we took a 12-years-old image of an identity as the testing image, and her/his 43-years-old image was regarded as the ground-truth. The EM optimization runs 10 iterations and the poisson image editing runs 180 iterations.

To analyze the advantages and weakness of different aging methods, we reproduced the AGES [6] as our primary comparative method. That is because this algorithm is one of the few aging prediction method that performs with sparse training data, like FG-NET. Our system is implemented by C++ and Matlab C library. When performing on a PC of i7-2600 K CPU with 8 GB memory, the synthesis time of each image is less than 2 min.

5.1 Evaluation of feature-point accuracy

We used leave-one-out approach to compare the feature-point accuracy of the AGES and our guidance method with the ground-truth. As above-mentioned, we took FG-NET identities with valid data in three age gaps as our testing subjects. For a testing subject, her/his aging pattern was excluded from the training data, and only one image at the young age was used as the input.

After aligning eyes (FG-NET feature point no.29 and 34) of aged images to the ground-truth, we calculate the average errors of feature points estimated by AGES and the proposed methods. The results are listed in Table 2. The average error of a region (e.g., face contour, eyes, nose, mouth) evaluates the average point distances between the regions at the ground

Table 2 Evaluation of feature-point accuracy

Avg. error (pixels)	All point positions		Face contour		Eyes		Nose		Mouth	
	AGES	Ours	AGES	Ours	AGES	Ours	AGES	Ours	AGES	Ours
5 to 30 (yrs. old)	4.51	3.42	4.06	3.23	1.12	1.06	2.89	1.95	2.04	1.95
10 to 40	4.33	3.44	5.48	3.82	0.91	0.89	2.05	1.71	1.77	1.56
20 to 50	3.48	3.67	3.06	3.18	0.88	0.78	1.42	1.45	1.53	1.61
Avg. error (pixels)	Weighted combination of normalized regions		Norm. face region		Norm. eyes region		Norm. nose region		Norm. mouth region	
	AGES	Ours	AGES	Ours	AGES	Ours	AGES	Ours	AGES	Ours
5 to 30 (yrs. old)	2.27	1.65	2.90	1.98	0.98	0.94	2.81	1.72	1.59	1.49
10 to 40	2.25	1.88	3.18	2.58	0.84	0.81	1.82	1.46	1.47	1.31
20 to 50	1.69	1.70	2.26	2.34	0.82	0.71	1.35	1.36	1.26	1.21

Table 3 Average scores of user evaluations

Experiment 1 (<i>reasonableness of aged images</i>)			
General dataset (30 identities)	AGES	Our guided prediction	Our full method
Grade	4.51	5.03	6.68
Standard Dev.	1.08	0.78	0.54
Experiment 2 (<i>the best aging image</i>)			
General dataset (30 identities)	AGES	Our guided prediction	Our full method
Approval percent	8 %	7 %	84 %
Experiment 3 (<i>similarity to the ground-truth elder image</i>)			
General dataset (30 identities)	AGES	Our guided prediction	Our full method
Grade	4.34	5.14	6.10
Standard Dev.	0.65	0.80	0.81

truth and the predicted face (regions are aligned by their region centers). The average error of a normalized region evaluates the average point distances between the normalized regions at the ground truth and the predicted face (regions are aligned by their region centers; the region width on a predicted face is scaled to fit the ground truth), and the errors of weighted combination of normalized regions evaluate the sum of the normalized region errors multiplied by the corresponding region width.

Notice that we not only compare the whole face shape but also feature contours. That's because the magnitude of position errors may not always consist with user perceptions. For instance, the FG-NET has dense feature points around the mouth. If the predicted mouth shape looks similar but is lightly shifted or scaled, the average position error will greatly increase. Therefore, we further compare the facial regions, including face contour, eye, nose, and mouth, respectively. Moreover, we also calculate the normalized region errors, where regions are normalized according to the regions widths. The weighted combination errors can be analogous to the weighed sum of region dissimilarities according to users' views.

In the two younger groups 5-to-30 and 10-to-40, our proposed method apparently outperforms the AGES. But, in the 20-to-50-group, the average position errors of the proposed method and AGES are similar. That is because 1) we have only few data between this age gap in the FG-NET. 2) In this age, human face geometry (feature point positions) changes relatively less from aging, but more due to living style and environment. Fatness and other age-invariant issues may dominate the position errors, and this makes two prediction methods fall into similar feature position errors.

5.2 User studies

Since we presented two techniques for aging prediction, to clarify their effects, we separated our method as *our guided prediction* and *our full method*. The former one made use of the guidance vectors in subspace prediction; the later one employed both the guided prediction and detailed texture synthesis.

Twenty-two volunteers, from 20 to 25 years old, participated in the user evaluation. We introduce three sets of experiments: *reasonableness of aging*, *the best aging*, and *similarity to the ground-truth*, to measure responses from users' perception. The first evaluation has 107 testing data, the second one has 37 testing groups, and the third evaluation has 80 testing data. The evaluation time for each volunteer was around one to one and a half hours.

In the first evaluation reasonableness, we simultaneously displayed a real young face image and a synthesized aged image by one of the testing algorithms. We also informed users of the ages of these two images but did not provide any information about the way of synthesis. According to the source image and the indicated ages, users had to rate satisfaction and reasonableness of the aged image by a one-to-nine score (1: very poor; 3: poor; 5: average/acceptable; 7: good; 9: satisfactory). Each user sequentially evaluated all aged images of testing subjects, and the aged images of different methods occurred in a random order. The general dataset we used includes 30 identities from the above-mentioned three training age gap sets.

As shown in Table 3 and Figure 11, in the general data set (FG-NET identities conforming to the three age gaps), the guided prediction surpass AGES about a half point, and the full method reach 6.68 points in average. This means our aging prediction is reasonable and more consistent. Figure 12 shows examples in the test.

In the second evaluation, we simultaneously displayed a real young face image and three synthesized aged images by different algorithms. These aged images were placed in a random order. We also informed users of the young and target age. Users were asked to choose the most reasonably aged image. Our results got 84 % of approvals in the general dataset.

In the third evaluation *similarity*, we simultaneously displayed a real elder face image and a synthesized image at the same age by one of the algorithms. Similar to the first evaluation, users had to evaluate the similarity of an aged image by a one-to-nine score. Our full method got the best scores 6.1 as shown in Table 3. Figure 13 shows the results.

5.3 Discussion

In this subsection, we discuss the advantages and disadvantages of two methods according to experiments and user evaluations.

The AGES method, an iterative least-square approximation, has demonstrated its success of filling missing data in age patterns [6]. However, for the most circumstances, only few or even a single young image can be provided as the input. AGES method is possibly trapped around the mean face of the target age. In our user studies, it got median scores and volunteers reported that it sometimes cannot preserve personal characteristics. The original AGES was designed for age estimation of facial images, and worked in PCA or other subspace, where high-frequency details like creases or wrinkles cannot be properly represented. Nevertheless, this seminal

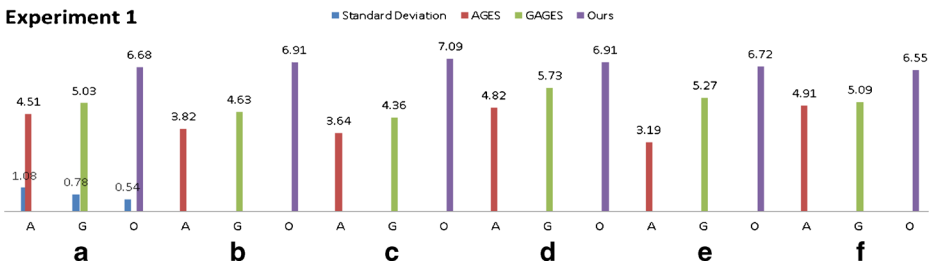
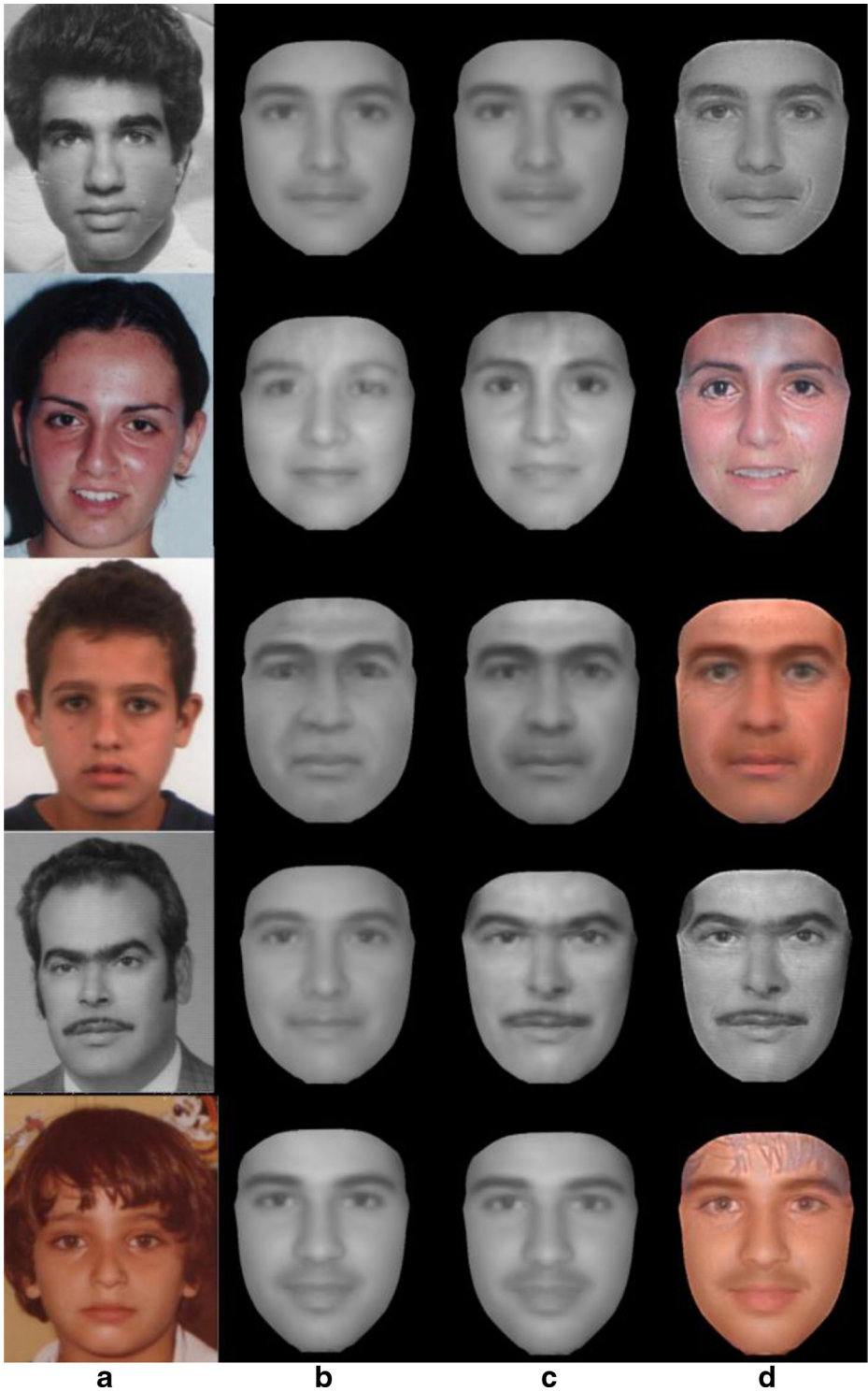


Fig. 11 The comparison of A: AGES, G: our guided prediction and O: our full method in experiment 1 and 3. **a** The first three bars show the average and standard deviation of three methods. **b, c, d, e, f** Exemplar five aging examples by three methods as shown in Fig. 12



◀ **Fig. 12** Examples in experiment 1 (reasonableness). Each row shows an example. **a** Input image. **b** Aging image of AGES. **c** Aging image of guided prediction. **d** Aging image of our full method. NOTE: for fair comparison, all images were shown in gray-scale in user evaluation

algorithm helps our system fill the initial aging-pattern database and the EM-like framework is extended for our base of aging simulation with personal guidance.

The proposed method takes a hybrid strategy. We utilize EM-like data filling with guidance vectors to predict the bases of aged faces, where more personal characteristics can be preserved. Patch-based facial detail synthesis is further applied to compensate the high-frequency wrinkles.

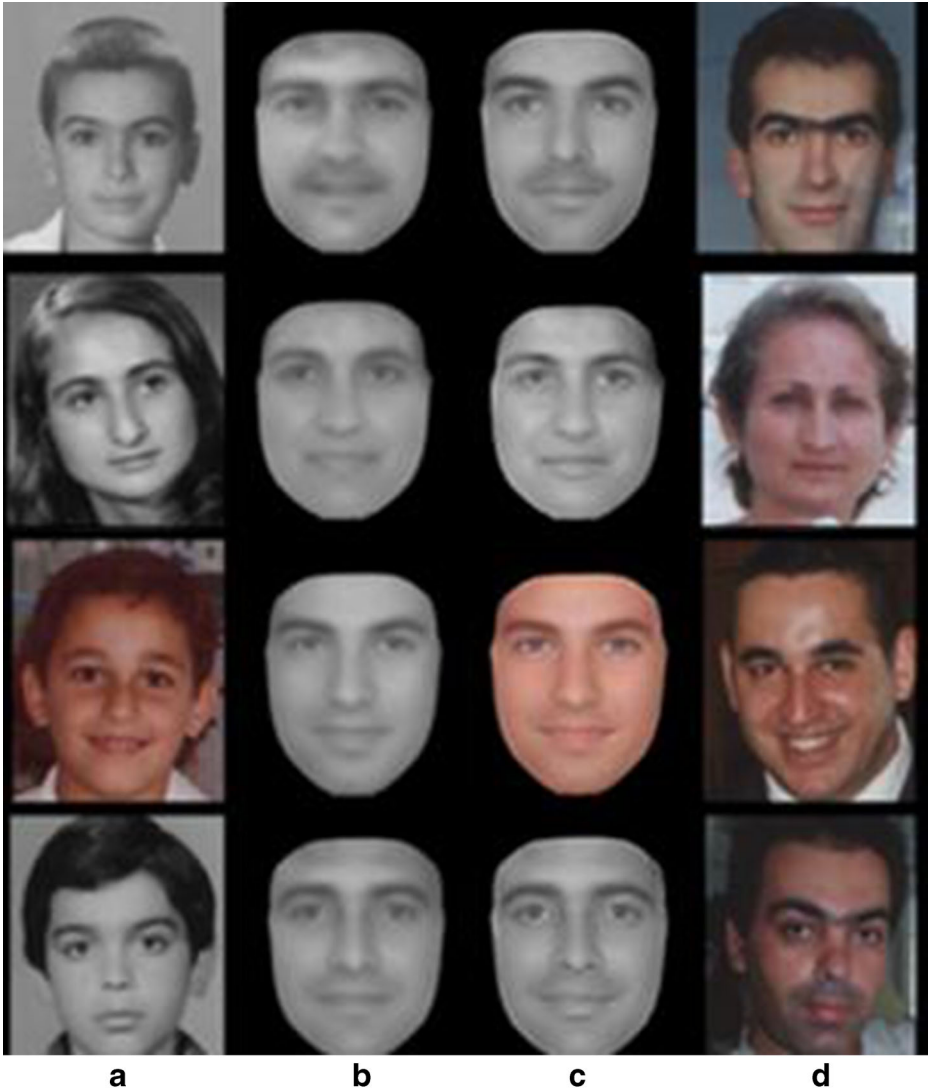


Fig. 13 Examples of experiment 3 (similarity). Each row shows an example. **a** Input image. **b** Aged images of AGES. **c** Aged image of our full method. **d** Real aged image. NOTE: for fair comparison, all images were shown in gray-scale in user evaluation

We conduct experiments to evaluate the effects of our method. We found that with our guidance vector, the geometry (feature point) prediction for young to middle ages can be more accurate. Using our synthetic aged images as training data can improve the recognition rate of elder faces. Moreover, our results get the highest point among three methods in user study “*reasonableness*”, “*the best aging*” and “*similarity to the ground-truth elder image*”. Our proposed method performs not only more detailed aging face, but also more reasonable personal characteristics. The small standard deviations show that our results are of more consistent qualities.

There are two major limitations of the proposed method. First, our method is based on statistical analysis and pattern matching. The prediction is feasible within the convex combination of training data. For instance, identities in FG-NET are Caucasian; the prediction may be biased when applying to Asian or African descents. Figure 14 shows an example. The limitation can be alleviated by extending datasets with more diverse aging patterns and textures. The other limitation is regarding the features. Currently, we do not exploit any prior domain knowledge and take all facial pixels of even weights for training. Our results may not be directly applicable to specific recognition systems, where special features are accentuated. More aged results by our methods are shown in Figure 15.

6 Conclusion and future work

A novel two-layer approach for facial image aging is proposed in this paper. At first, we approximate aged face in the aging-pattern subspace with the personal guidance vector. Detailed aged features, such as wrinkles and creases, can be synthesized and transferred with an additional facial appearance dataset.

The proposed method, considering personal characteristics and aging trends from examples, can generate reasonable aging prediction. It also overcomes the lack of details in

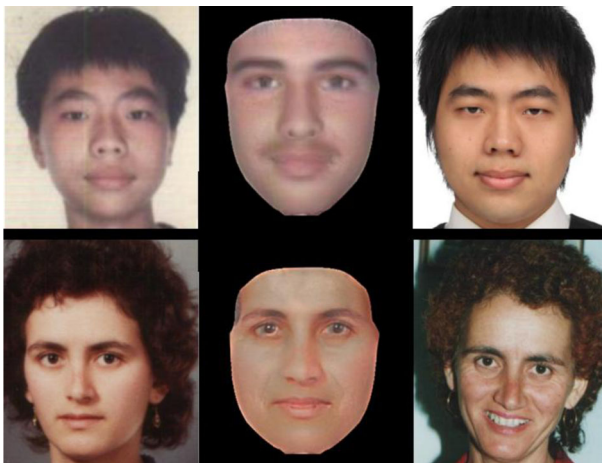


Fig. 14 Examples of two failure results. The left column is input face image. The 1st row shows an Asian male aged from his childhood. Since the training data are all Caucasians, the aged result (*mid*) tends to grow like a Caucasian, but not Asian. The 2nd row shows a reasonable aging result (*mid*), but do not resemble to the real aged face (*right*). That is because the real person is thinner in cheeks and there are no other age patterns in database like that. Besides, the oblique view of a face causes inaccuracy of feature point location, it also disturb the prediction

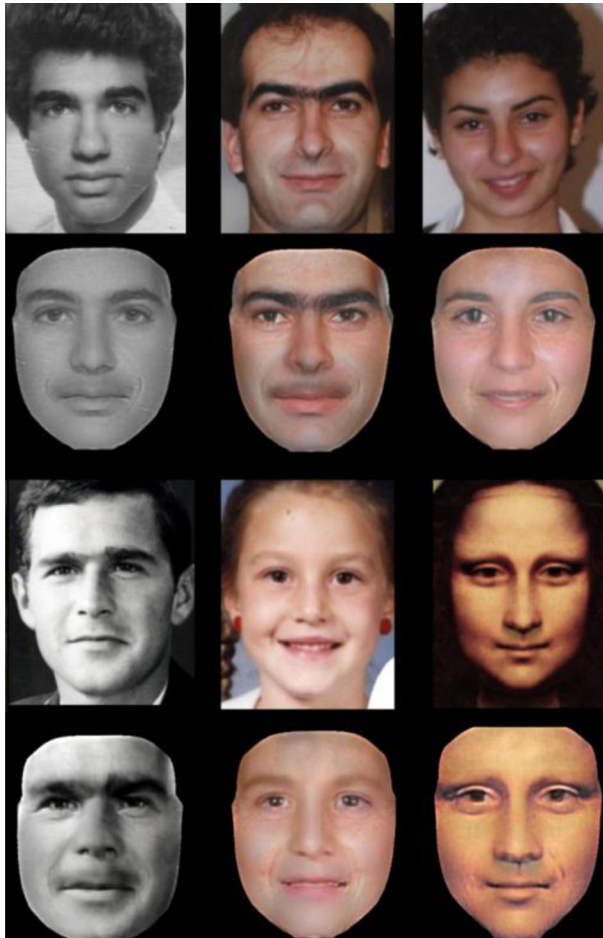


Fig. 15 More examples by our results. The 1st rows: input images from the FG-NET. The 2nd rows: aged images of the 1st rows by our method. The 3rd rows: input images not included in FG-NET. The 4th rows: aged images of the 3rd rows by our method. NOTE: The input of Mona Lisa was first rectified by View Morphing method [18]

previous aging methods. The objective experiments and user studies show that our aged results are comparable to those generated by related methods. Moreover, the proposed framework is flexible and extendable for user controls. It can easily include parents' effect vectors or users' sketch indication for guidance in the aging process. The proposed method is the first few articles that provide such controllable properties during aging prediction. The results can be applied to Art and Entertainment, increasing realism of virtual characters.

One possible extension is to combine with face or hair swapping techniques [3], a user can arbitrarily generate aged portraits from single young image. Currently, we regard the parents' effect as an optional tool, and preliminarily use weighted-blending for parents' effect vectors. We think it is worth to further investigate how parents' features influence the aging prediction of a child. Besides, it is also possible to collect

3D aging patterns by model-based [1] or scanning techniques [10]. For facial wrinkles and creases, image-based 3D modeling [11] improves the level of details. These techniques will benefit the side-view or 3D face aging.

Acknowledgement Authors would like to appreciate Hui-Ping Liu for her initial trials on face aging, and thank the volunteers that participated in our experiments. Ming-Han Tsai and Yen-Kai Liao are the co-first authors and I-Chen Lin is the corresponding author. This paper was partially supported by the National Science Council, Taiwan under grant no. NSC 100-2221-E-009-148.

References

1. Blanz V and Vetter T (1999) A morphable model for the synthesis of 3D faces. Proc. SIGGRAPH'99, pp. 187–194
2. Burt M, Perrett D (1995) Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and color information. Proc Roy Soc Lond B Biol Sci 259(1355):137–143
3. Chou J-K and Yang C-K (2011) Simulation of Face/Hairstyle Swapping in Photographs with Skin Texture Synthesis. Multimedia Tools and Applications (MTAP). doi: 10.1007/s11042-011-0891-1
4. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
5. Fu Y, Huang T-S (2008) Human age estimation with regression on discriminative aging manifold. IEEE Trans Multimed 10(4):578–584
6. Geng X, Zhou Z-H, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. IEEE Trans Pattern Anal Mach Intell 29(12):2234–2240
7. Guo G, Fu Y, Dyer CR, Huang TS (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Trans Image Process 17(7):1178–1188
8. Hubball D, Chen M, Grant PW (2008) Image-based aging using evolutionary computing. Comput Graph Forum 27(2):607–616
9. Lanitis A, Taylor CJ, Cootes TF (2002) Toward automatic simulation of aging effects on face images. IEEE Trans Pattern Anal Mach Intell 24(4):442–455
10. Li W, Zhang D, Lu G, Luo N (2012) A novel 3-D palmprint acquisition system. IEEE Trans Syst Man Cybern Syst Hum 42(2):437–451
11. Lin I-C, Chang W-H, Lo Y-S, Peng J-Y, Lin C-Y (2010) Image-based detail reconstruction of non-Lambertian surfaces. Comput Animat Virtual Worlds (CAVW) 21(1):55–68
12. Liu Z, Shan Y, Zhang Z (2001) Expressive expression mapping with ratio images. Proc. ACM SIGGRAPH'01, pp. 271–276, Aug
13. Mohammed U, Prince SJD and Kautz J (2009) Visio-lization: generating novel facial images. ACM Trans. Graphics, 28(3): 57:1–57:8
14. Pérez P, Gangnet M, Blake A (2003) Poisson image editing. ACM Trans Graph 22(3):313–318
15. Ramanathan N, Chellappa R (2006) Modeling age progression in young faces. Proc IEEE Conf Comput Vis Pattern Recogn 1:387–394
16. Ramanathan N, Chellappa R, Biswas S (2009) Age progression in human faces: a survey. J Vis Lang Comput 20(3):131–144
17. Scherbaum K, Sunkel M, Seidel H-P, Blanz V (2007) Prediction of individual non-linear aging trajectories of faces. Comput Graph Forum 26(3):285–294
18. Seitz SM and Dyer CR (1996) View Morphing. Proc. ACM SIGGRAPH' 96, pp. 21–30, Aug
19. Shaw RB, Katzel EB, Koltz PF, Kahn DM, Giroto JA, Langstein HN (2010) Aging of the mandible and its aesthetic implications. Plast Reconstr Surg 125(1):332–342
20. Suo J, Zhu S-C, Shan S, Chen X (2010) A compositional and dynamic model for face aging. IEEE Trans Pattern Anal Mach Intell 32(3):385–401
21. The FG-NET Aging Database (2002) <http://www.fgnet.rsunit.com/>
22. Wang C-C, Su Y-C, Hsu C-T, Lin C-W and Mark Liao H-Y (2009) Bayesian Age Estimation on Face Images. Proc. IEEE International Conference on Multimedia and Expo, New York, USA, pp. 282–285, June
23. Zimble MS, Kokoska MS, Thomas JR (2001) Anatomy and pathophysiology of facial aging. Facial Plast Surg Clin N Am 9(2):179–187



Ming-Han Tsai received the B.S. and M.S. degree in computer science from National Chiao Tung University, Taiwan, in 2007 and 2009, respectively. He is now a Ph.D. student in the Department of Computer Science, National Chiao Tung University, Taiwan. His research interests include image-based modeling, image synthesis, and human-computer interaction.



Yen-Kai Liao received the B.S. degree from National Chung Hing University in 2008. He received the M.S. degree in computer science from and National Chiao Tung University, Taiwan, in 2010. His research interests include character motion, image-based modeling and interactive image synthesis.



I-Chen Lin received the B.S. and Ph.D. degree in computer science from National Taiwan University, Taiwan, in 1998 and 2003, respectively. In 2005, he joined Department of Computer Science, National Chiao Tung University, Taiwan. He is currently an assistant professor. His research interests include computer graphics, image-based modeling and multimedia interface. He is a member of ACM SIGGRAPH and IEEE.