A Compositional and Dynamic Model for Face Aging

Jinli Suo , Song-Chun Zhu , Shiguang Shan and Xilin Chen

Abstract—In this paper we present a compositional and dynamic model for face aging. The compositional model represents faces in each age group by a hierarchical And-Or graph, in which And nodes decompose a face into parts to describe details (e.g. hair, wrinkles, etc.) crucial for age perception and Or nodes represent large diversity of faces by alternative selections. Then a face instance is a transverse of the And-Or graph—*parse graph*. Face aging is modeled as a Markov process on the parse graph representation. We learn the parameters of the dynamic model from a large annotated face dataset and the stochasticity of face aging is modeled in the dynamics explicitly. Based on this model, we propose a face aging simulation and prediction algorithm. Inversely an automatic age estimation algorithm is also developed under this representation. We study two criteria to evaluate the aging results using human perception experiments: (i) The accuracy of simulation: whether the aged faces are perceived of the intended age group, and (ii) preservation of identity: whether the aged faces are perceived as the same person. Quantitative statistical analysis validates the performance of our aging model and age estimation algorithm.

Index Terms—Face Aging Modeling, Face Age Estimation, Generative Model, And-Or Graph, ANOVA.

1 INTRODUCTION

THE objective of this paper is to study a statistical model for human face aging, which is then used for face aging simulation and age estimation. Face aging simulation and prediction is an interesting task with many applications in digital entertainment. In such applications, the objective is to synthesize aging effects that are visually plausible while preserving identity. This is distinguished from the task of face recognition in biometrics where two key considerations are to extract features stable over a long time span and learn the potential tendency of facial appearance in aging process. Building face recognition systems robust to age related variations[27][34][38] is a potential applications, but it is beyond the scope of this paper.

We adopt a hierarchical And-Or graph representation to account for the rich information crucial for age perception and large diversity among faces in each age group. A specific face in this age group is a transverse of the And-Or graph, and is called parse graph. Aging process is modeled as a Markov chain to describe the evolution of parse graphs across age groups and to account for the intrinsic stochasticity of the face aging process. The accuracy of simulation (i.e. whether the synthetic images are perceived of the intended age group) and preservation of face identity (i.e. whether aged faces are perceived as the same person) are two criteria used to evaluate our modeling results in human experiments.

Compared with other face modeling tasks, modeling face aging encounters some unique challenges. (i) There are large shape and texture variations over long period, say 20-50 years: hair whitens, muscles drop, the wrinkles appear, and so on. Traditional AAM model[10] is hard to describe all those variations. (ii) The perceived face age often depends on global non-facial factors, such as the hair color and style, the boldness of the forehead, et al., while these non-facial features are usually excluded in face modeling. (iii) It is very difficult to collect face images of the same person over a long time period, and the age related variations are often mixed with other variations (i.e. illumination, expression, et al.). (iv) There exist large variations of perceived age within each biologic face group due to external factors, such as health, life style, et al. (v) There lack quantitative measurements for evaluating the aging results in the literature. All these characteristics demand a sophisticated face aging model to account for rich face details related to age perception, intrinsic uncertainty in aging process and a criteria for evaluating the age simulation results.

1.1 Previous Work

Face aging modeling and face aging simulation have attracted growing research interest from psychology, graphics, and lately computer vision. Previous work on face aging can be divided into two categories: child growth and adult aging.

For child growth modeling, shape change of face profile is the most prominent factor. Most researchers adopted specific transformation on a set of landmarks[11][15][33] or statistical parameters [21][26][29] to model age related shape changes. Ramanathan and Challeppa[33] defined growth parameters over the landmarks to build a crania facial growth model, and anthropometric evidences are included to make the model consistent with the actual data. Lanitis et al.[21] built three aging functions to describe the relationships between facial age and the AAM parameters, by which they could estimate the age from a child image and predict face growth inversely. Some others[17][37] included texture parameters in their facial

*Jinli Suo is with the Graduate University of Chinese Academy of Sciences, and the Lotus Hill Research Institute, China. E-mail: jlsuo@jdl.ac.cn.
*Song-Chun Zhu is with the University of California, Los Angeles, and the Lotus Hill Research Institute, China. E-mail: sczhu@stat.ucla.edu.
*Shiguang Shan and Xilin Chen are with key Lab of Intelligent Information Processing of CAS, Institute of Computing Technology, CAS, Beijing, 100190, China. E-mail: {sgshan, xlchen}@ict.ac.cn.

growth model. All these methods showed the validness of modeling shape changes in growth prediction.

For adult aging, both appearance and shape were studied. In computer graphics, people built physical models to simulate aging mechanisms of cranium, muscles and skin. For example, Boissieux et al.[6] and Wu et al.[44] built layered skin models to simulate the skin deformation as age increases. Berg and Justo[4] simulated the aging process of obituaries muscles. Other similar work include Bando et al.'s[2], Lee et al.'s[25] and Ramanathan's work[35].

In computer vision, most aging approaches are example based and can be divided into three types. (i) Prototype method[7][41] computes average face image of each age group as prototype and defines the differences between prototypes as aging transformation. Wang et al.[43] applied this prototype approach in PCA space instead of on image directly and Park et al.[30] applied it to 3D face data. Prototype method is able to extract average pattern but many details (e.g. wrinkles, pigments, et al.) crucial for age perception are ignored. There are also work studying texture transfer from a specific senior face to young ones, such as [13][28]. (ii) Function based method describes relationships between a face image and its age label with an explicit function, such as quadratic function[31], support vector regression[42], kernel smoothing method[18] or an implicit function[5]. Jiang and Wang[19] directly built a mapping function between young faces and their appearances at later ages. All those functions need considerable real aging sequences to learn the function parameters. (iii) Distance based methods[22] formulate aging simulation as an optimization problem. They synthesize a face close to the images of intended age in age space and close to the input individual in the identity space simultaneously. The algorithm in [22] adopted global AAM model and simple similarity metrics, simulation results are not realistic enough.

Other related work is age estimation, which selects discriminative features to estimate face age. Primary studies on age estimation[20] coarsely divided human faces into groups based on facial landmarks and wrinkles. Most recent approaches considered the continuous and temporal property of face age and formulated age estimation as a regression problem. Researchers explored different features, including AAM coefficients[23], image intensities[12][14][46], features designed heuristically[40] and adopted various regression methods, such as quadratic function[23], piece-wise linear regression[23][40], multi-perceptron projection[12][23][40], et al. Different to aforementioned methods, Geng et al.[16] defined an aging sequence as an aging pattern and estimated age by projecting a face instance onto appropriate position of a pattern.

Despite of the progress, there are some problems in the existing work. Firstly, example based models need large number of image sequences of the same person across age groups to learn aging patterns, and existing dataset is far from being sufficient. Secondly, most of the existing models do not account for high resolution features, therefore they are insufficient for describing the large facial variations across age groups and the aging results lack crucial details (e.g. wrinkles, pigments, et al.) for age perception. Thirdly, hair features are usually not considered, despite its influence on the perception of face age. Fourthly, the ground truth for aging modeling is difficult to collect and appropriate performance measurement is not standardized, and a quantitative evaluation of face aging results is also needed.

1.2 Overview of Our Approach

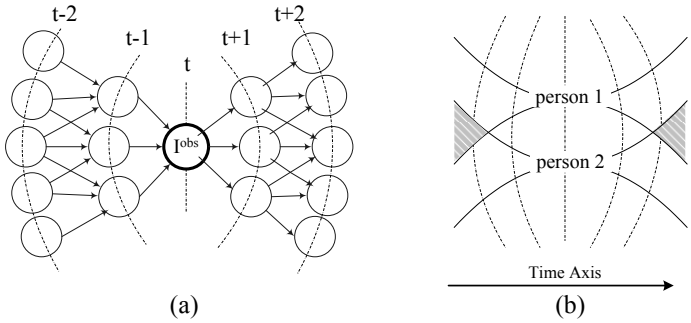


Fig. 1. Stochasticity of face aging. (a) The node I^{obs} is a face observed at time t, while the other nodes are the plausible faces before and after time t. Each dashed curve represents a space of possible face images at certain time. (b) The shadowed area means that two people may become unidentifiable after certain period, which reflects that the difficulty of preserving face identity increases as time evolves.

Motivated by the aforementioned problems, we propose a compositional and dynamic model to represent the face aging process. Our model represents faces in each age group by a three-level *And-Or graph*[8] (see Fig. 3), which consists of And-nodes, Or-nodes and Leaf nodes. The And nodes represent the decomposition, which divides a face into parts and primitives at three levels from coarse to fine. The first level describes face and hair appearance, the facial components are refined at the second level, and wrinkles and skin marks are further refined at the third level. Or nodes represent the alternatives to represent the diversity of face appearance at each age group, and leaf nodes are basic primitives. Spatial relations and constraints are imposed between the nodes at the same level to ensure the validness of the configurations (symmetry of eyes, spatial relationships among facial parts, et al.). By selecting alternatives at the Or-nodes, one obtains a hierarchic *parse graph* for a face instance, and the face image can be synthesized from this parse graph in a generative manner. Based on the And-Or graph representation, we represent the dynamics of face aging process as a first-order Markov chain on parse graphs (see Fig. 5), and learn the aging patterns from annotated faces of adjacent age groups at each level. To overcome the difficulty of collecting face images of the same person at different ages, our compositional model decomposes face into facial components and skin zones. The part-based strategy allows

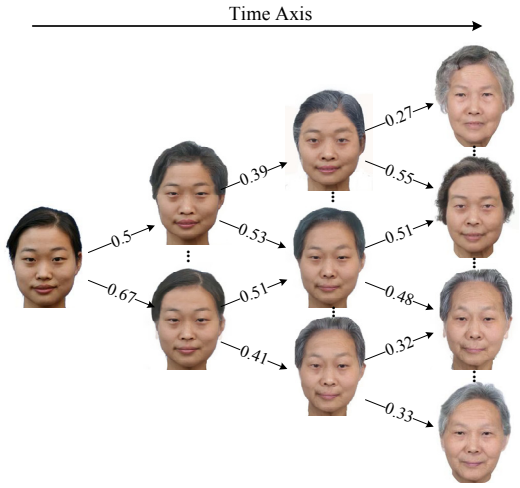


Fig. 2. The uncertainty of aging increases with time. Given an input face image (leftmost), the algorithm simulates a series of plausible aging results reflecting the stochasticity. The vertical column shows the plausible faces at certain age group. For each arrow, we show the transition probability computed by the dynamic model.

the aging pattern of each part across age groups to be learned from similar patches. Our dataset includes about 50,000 face images with large diversity in the age range of 20-80. The patterns learned from similar patches might be different from those learned from the aging data of the same person, thus we need to evaluate the results quantitatively as an important extension of work in the published short version[39].

A central issue in face aging modeling is to study the stochasticity of the aging process, as Fig. 1 illustrates. For an observed young face I^{obs}, the appearance changes over time is intrinsically a stochastic process. Like Brownian motion, the uncertainty increases along both directions of the time axis and confusion between two subjects increases as well, as Fig. 1(b) shows. As an example, Fig. 2 shows some plausible aging results of a young individual to illustrate the uncertainty of face aging. The value of each arrow is the transition probability computed by our dynamic model.

Since there is intrinsic uncertainty for face aging, we propose two criteria to evaluate the face aging results.

(i) **The accuracy of simulation.** For each age group we select 80 real images from our dataset and 80 simulated images synthesized using our algorithm. Then these images are given to 20 volunteers for age estimation. By analyzing the results with ANalysis of VAriance(ANOVA), we find no significant difference in age estimation performance between real images and synthetic images.

(ii) **Preservation of the identity**. We collect real aging sequences of 20 individuals from relatives and friends, for each individual we synthesize one aging sequence from the photo at the initial age group, then 20 volunteers are asked to identify the individuals in the two sets. The ANOVA analysis of recognition results shows that our face aging model preserves face identity effectively.

2 Representation and Formulation

We study adult faces in the age range of 20-80, and divide them into 5 groups: [20, 30), [30, 40), [40, 50), [50, 60) and [60, 80]. In this section we present the And-Or graph model for face representation, the dynamic model for aging, and the procedure of model learning.

2.1 Compositional And-Or graph for Face Modeling

We extend a multi-resolution face representation proposed by Xu et al.[45] with hair features and build age group specific face models. As Fig. 3 illustrates on the left column, a face image I_t at age group t is represented at three levels, from coarse to fine,

$$I_t = ((I_{\text{hair},t}, I_{\text{face},t}), I_{\text{cmp},t}, I_{\text{wkl},t}). \quad (1)$$

$(I_{\text{hair},t}, I_{\text{face},t})$ is the whole face image, where $I_{\text{hair},t}$ represents hair and $I_{\text{face},t}$ accounts for general face appearance. $I_{\text{cmp},t}$ refines the facial components (eyes, eyebrows, nose, mouth et al.). $I_{\text{wkl},t}$ further refines the wrinkles, skin marks, and pigments in 6 facial skin zones. All faces of age t are collectively represented by an And-Or graph $\mathcal{G}_t^{\text{AO}}$ (see the middle column of Fig. 3), where an And-node (in solid ellipse) represents the decomposition and an Or-node (in dashed ellipse) represents the alternatives to account for large diversity of faces, for example, different eye shapes. A dictionary Δ_t for each age group t is shown on the right side for various components over the three levels.

$$\Delta_t = ((\Delta_{\text{hair},t}, \Delta_{\text{face},t}), \Delta_{\text{cmp},t}, \Delta_{\text{wkl},t}) \quad (2)$$

The dictionary Δ_t is learned from a large number of faces at age group t. Fig. 4 shows the diversity of the examples in the dictionary at different age groups.

By choosing the alternatives at the Or-nodes, the And-Or graph $\mathcal{G}_t^{\text{AO}}$ is converted to an And-graph G_t as a specific face instance at age group t, called *parse graph*.

Generative model accounts for a large variety of faces, we denote the set of faces generated by $\mathcal{G}_t^{\text{AO}}$ as

$$\Sigma_t = \{G_t\}, \quad (3)$$

which is evidently much larger than the training set. A face instance is represented by

$$G_t = (w_{1,t}, w_{2,t}, w_{3,t}), \quad (4)$$

where $w_{i,t}, i = 1, 2, 3$ are the hidden variables controlling the generation of I_t at three resolutions and i indexes the three resolutions. They can be further decomposed as

$$w_{i,t} = (l_{i,t}, T_{i,t}^{\text{geo}}, T_{i,t}^{\text{pht}}). \quad (5)$$

In the above notation, $l_{i,t} = \{l_{i,t}(m) : m = 1, 2, ..., n_{i,t}^{\text{Or}}\}$ includes a vector representing all the "switch" variables for the alternatives in each Or-node m at resolution i and age group t, $T_{i,t}^{\text{geo}} = \{T_{i,t}^{\text{geo}}(m) : m = 1, 2, ..., n_{i,t}^{\text{And}}\}$ and $T_{i,t}^{\text{pht}} = \{T_{i,t}^{\text{pht}}(m) : m = 1, 2, ..., n_{i,t}^{\text{And}}\}$ are variables for the geometric and photometric attributes in each And-node m at resolution i and age group t respectively.

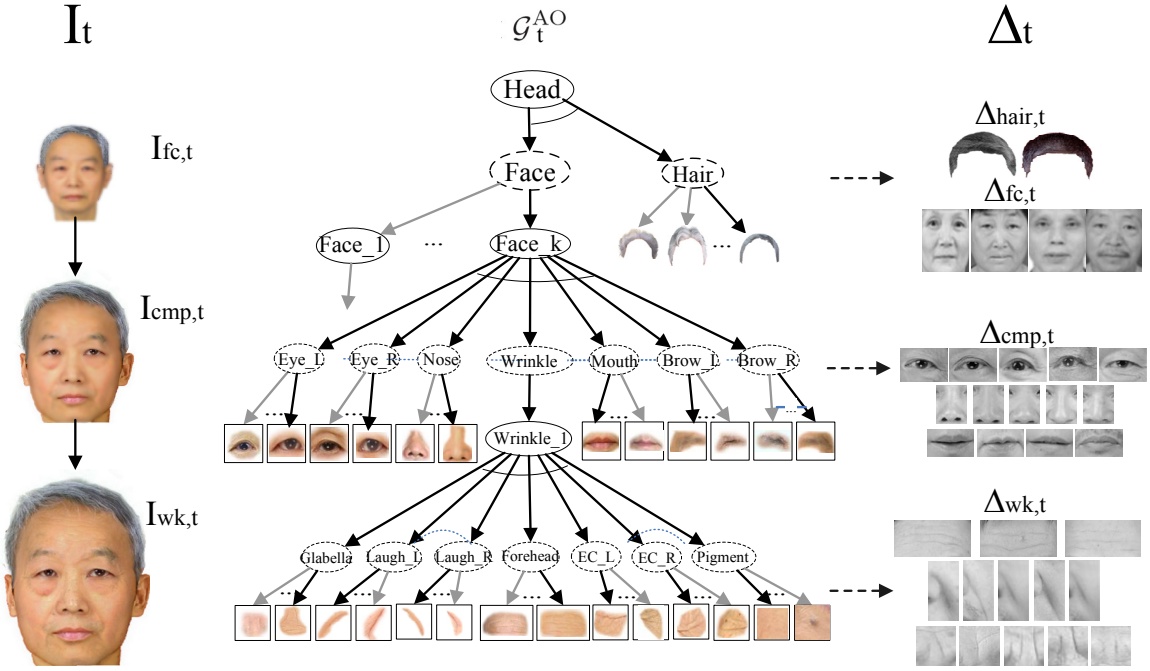


Fig. 3. Left: A high resolution face image I_t at age group t is represented at three resolutions – $I_{\text{face},t}$,$I_{\text{cmp},t}$ and $I_{\text{wkl},t}$. Middle: All face images at age group t are represented collectively by a hierarchic And-Or graph $\mathcal{G}_t^{\text{AO}}$. The And nodes (in solid ellipses) in the graph $\mathcal{G}_t^{\text{AO}}$ represent coarse-to-fine decomposition of a face image into its parts and components. The Or-nodes (in dashed ellipses) represent alternative configurations. By choosing the Or-nodes, we obtain a parse graph G_t for a specific face instance. Right: Dictionary Δ_t includes $\Delta_{\text{hair},t}$, $\Delta_{\text{face},t}$, $\Delta_{\text{cmp},t}$ and $\Delta_{\text{wkl},t}$ at three levels from coarse to fine.

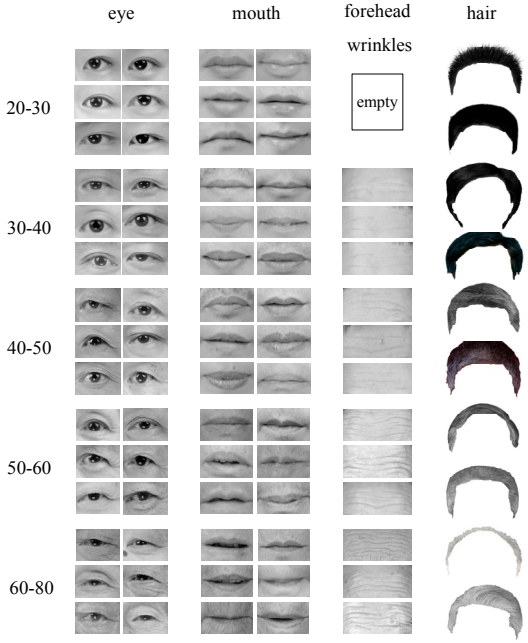


Fig. 4. Examples of the facial components and hairs from the dictionaries of different age groups.

We impose a prior probability for the hierarchical parse graph G_t,

$$\begin{aligned} & p(G_t;\Theta_{\text{AOG}}) \\ = \; & p(\omega_{3,t}|\omega_{2,t};\Theta_{3,\text{AOG}})p(\omega_{2,t}|\omega_{1,t};\Theta_{2,\text{AOG}})p(\omega_{1,t};\Theta_{1,\text{AOG}}), \end{aligned} \tag{6}$$

which accounts for the constraints of upper level to current level as well as the constraints among nodes at the same level, e.g. enforcing the same type of eyes. $\Theta_{\text{AOG}} = (\Theta_{1,\text{AOG}}, \Theta_{2,\text{AOG}}, \Theta_{3,\text{AOG}})$ includes the parameters. The above probability can further be decomposed into three factors:

$$\begin{aligned} & p(\omega_{i,t}|\omega_{i-1,t};\Theta_{i,\text{AOG}}) \\ = \; & p(l_{i,t}|l_{i-1,t};\Theta_{i,\text{AOG}}) \cdot p(T_{i,t}^{\text{geo}}|T_{i-1,t}^{\text{geo}};\Theta_{i,\text{AOG}}) \\ & \cdot p(T_{i,t}^{\text{pht}}|T_{i-1,t}^{\text{pht}};\Theta_{i,\text{AOG}}). \end{aligned} \tag{7}$$

G_t in turn generates image I_t in a generative manner.

$$G_t \overset{\Delta_t}{\Longrightarrow} I_t \tag{8}$$

The likelihood model specifies how $\omega_{i,t}$ generates image $I_{i,t}$ as in [45] and AAM[10].

$$I_{i,t} = J_i(w_{i,t};\Delta_{i,t}) + I_{i,t}^{\text{res}}, i = 1,2,3. \tag{9}$$

where J_i is the reconstruction function of human face at resolution i using the dictionary $\Delta_{i,t}$. $I_{i,t}^{\text{res}}$ is a residual image of the reconstruction at resolution i, which follows a Gaussian distribution. The likelihood model of the whole face can be written as

$$p(I_t|G_t;\Delta_t) = \prod_{i=1}^{3} p(I_{i,t}|w_{i,t};\Delta_{i,t}). \tag{10}$$

The parse graph is computed from an observed image by Bayesian inference from coarse to fine in a way similar to [45]. By denoting $w_{0,t}^* = \emptyset$, for $i = 1,2,3$ we have

$$w_{i,t}^* = \arg\max\nolimits_{\omega} p(I_{i,t}|\omega_{i,t};\Delta_{i,t})p(\omega_{i,t}|\omega_{i-1,t}^*;\Theta_{i,\text{AOG}}). \tag{11}$$

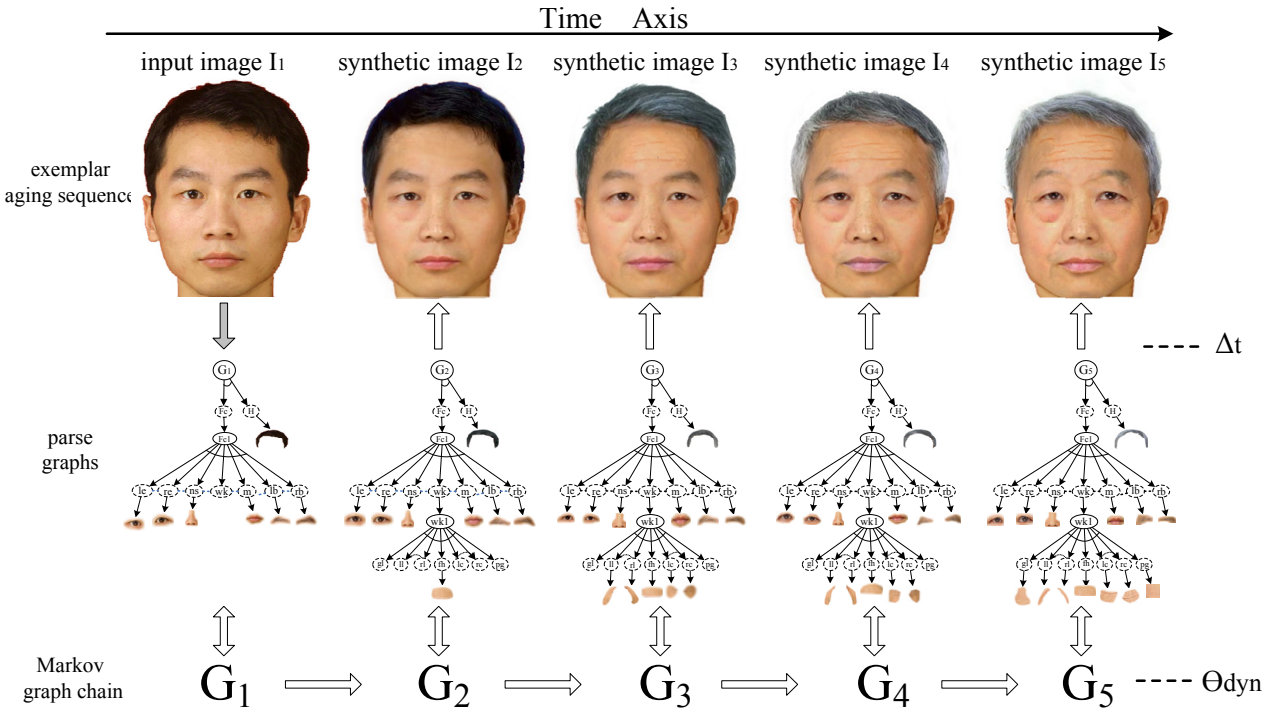


Fig. 5. Modeling the aging process as a Markov chain on parse graphs. Top row is a face image sequence at different ages, with the leftmost one being the input image and the other four being synthetic aged images. The second row is the parse graphs of the image sequence. The third row shows the Markov chain and Θ_{dyn} includes the parameters for Markov chain dynamics.

2.2 Modeling Aging Procedure as A Markov Chain on Parse Graphs

Based on the above graph representation, the face aging process is modeled as a Markov chain on the parse graphs. We denote by $I[1,\tau]$ and $G[1,\tau]$ the sequence of images and parse graphs respectively for a period $[1,\tau]$. Therefore, our probabilistic model is a joint probability,

$$\begin{aligned} & p(I[1,\tau], G[1,\tau]; \Theta) \qquad (12) \\ = & \prod_{t=1}^{\tau} p(I_t|G_t;\Delta_t) \cdot p(G_1) \cdot \prod_{t=2}^{\tau} p(G_t|G_{t-1};\Theta_{\text{dyn}},\Theta_{\text{AOG}}) \end{aligned}$$

Here $\Theta = \{\Delta_t, \Theta_{\text{dyn}}, \Theta_{\text{AOG}}\}$ denotes the parameters. $p(I_t|G_t;\Delta_t)$ is the image model in Eq. 10 generating an image I_t from a parse graph G_t. $p(G_t|G_{t-1};\Theta_{\text{dyn}},\Theta_{\text{AOG}})$ is the dynamic model for the evolution from one parse graph G_{t-1} to the next G_t with Θ_{dyn} being the aging parameters.

Fig. 5 is an illustration of our dynamic model for face aging. I_1 is an input young face image and G_1 is its parse graph representation. By sampling from the dynamic model $p(G_t|G_{t-1};\Theta_{\text{dyn}},\Theta_{\text{AOG}})$ we can simulate a series of parse graphs G_2, G_3, G_4 and G_5. Then new face images I_2, I_3, I_4, I_5 are synthesized in four consecutive age groups with dictionaries Δ_2 to Δ_5.

In the dynamic model, we factorize the transition probabilities of $l_{i,t}$, $T^{\text{geo}}_{i,t}$ and $T^{\text{pht}}_{i,t}$ separately over time t and resolution i. Each component $\omega_{i.t}$ depends on its upper level $\omega_{i-1,t}$ and previous age group $\omega_{i,t-1}$.

$$\begin{aligned} & p(G_t|G_{t-1},\Theta_{\text{dyn}},\Theta_{\text{AOG}}) \qquad (13) \\ = & \prod_{i=1}^{3} p(l_{i,t}|l_{i,t-1},l_{i-1,t}) \cdot p(T^{\text{geo}}_{i,t}|T^{\text{geo}}_{i,t-1},T^{\text{geo}}_{i-1,t}) \\ & \cdot p(T^{\text{pht}}_{i,t}|T^{\text{pht}}_{i,t-1},T^{\text{pht}}_{i-1,t}). \end{aligned}$$

Here Θ_{dyn} is learned from a large training data. In the following, we discuss the two types of variations in the dynamic model above: (i) abrupt changes for the emergence of new age related features; and (ii) continuous changes of the geometric and photometric attributes.

(1) *Abrupt changes*. The aging process may change the topology of the graph. For example, inserting new nodes (e.g. wrinkles emerge, et al.) or switching the alternatives in the Or-nodes (e.g. change of hair style, the type of eyes, et al.). We use the transition probabilities of $l_{i,t}$ to represent this type of variation.

$$\begin{aligned} & p(l_{i,t}|l_{i,t-1},l_{i-1,t}) \qquad (14) \\ \propto & \prod_{m=1}^{n^{\text{Or}}_{i,t}} \lambda_{i,t}(l_{i,t}(m), l_{i,t-1}(m)) \cdot p(l_{i,t}|l_{i-1,t}),\ i=1,2,3. \end{aligned}$$

In the above model, m indexes the corresponding Or-nodes between two adjacent graphs G_t and G_{t-1} at resolution i, and $\lambda_{i,t}()$ is a stochastic transition matrix for how likely a node of type $l_{i,t-1}(m)$ ages to a node of type $l_{i,t}(m)$. $p(l_{i,t}|l_{i-1,t})$ is the hierarchy model from the And-Or graph and accounts for the frequency of $l_{i,t}(m)$ and constraints for symmetry between nodes.

(2) *Continuous changes*. Some variations in aging only change the attributes of leaf-nodes, such as skin color, facial part shape, wrinkle length et al.. We represent them by the transition probabilities of $T^{\text{geo}}_{i,t}$ and $T^{\text{pht}}_{i,t}$. The continuous variation transitions are represented in the following model at three resolutions, $i = 1, 2, 3$:

$$p(T^{\text{geo}}_{i,t}|T^{\text{geo}}_{i,t-1}, T^{\text{geo}}_{i-1,t}) \tag{15}$$
$$\propto \exp\{-\sum_{m=1}^{n^{\text{And}}_{i,t}} \psi(T^{\text{geo}}_{i,t}(m), T^{\text{geo}}_{i,t-1}(m))\} \cdot p(T^{\text{geo}}_{i,t}|T^{\text{geo}}_{i-1,t}),$$
$$p(T^{\text{pht}}_{i,t}|T^{\text{pht}}_{i,t-1}, T^{\text{pht}}_{i-1,t}) \tag{16}$$
$$\propto \exp\{-\sum_{m=1}^{n^{\text{And}}_{i,t}} \psi(T^{\text{pht}}_{i,t}(m), T^{\text{pht}}_{i,t-1}(m))\} \cdot p(T^{\text{pht}}_{i,t}|T^{\text{pht}}_{i-1,t}).$$

In the above formula, m indexes the And-node at resolution i between two adjacent groups t and $t-1$. $T^{\text{geo}}_{i,t}(m)$ and $T^{\text{pht}}_{i,t}(m)$ denote the geometric and photometric attributes of an And-node m respectively. $\psi()$ is a potential which favors the transitions between similar parts, and penalizes large variations of the same part between adjacent groups. For geometric distance, we adopt the thin-plate spline (TPS) model after aligning the landmark points on the parts. Although large variations may occur in real data (e.g. the scars caused by injury, the change of hair styles, the variations introduced by expression, illumination, et al.), we try to penalize these effects of external unpredictable factors and learn only the natural aging patterns. The probabilities $p(T^{\text{geo}}_{i,t}|T^{\text{geo}}_{i-1,t})$ and $p(T^{\text{pht}}_{i,t}|T^{\text{pht}}_{i-1,t})$ are parts of the original prior model of the parse graph in Eq. 8.

2.3 Automatic Learning of Face Aging Model

The image model and dynamic model can both be learned automatically from a large labeled dataset, we summarized the procedure in Alg. 1. For clarity of presentation, we shall discuss the implementation details in Sec. 4.

Algorithm 1. Learning of face aging model

input : Dataset of face images at 5 age groups
output: Hierarchical face model and dynamic face aging model

for $t = 1$ **to** 5 **do**
1. Label facial landmarks and wrinkle lines for:
1.1 Learn the parameters of hierarchical face model $\Theta_{i,\text{AOG}}$
1.2 Build the dictionary $\Delta_{i,t}$
2. Compute parse graphs of faces in the dataset from Eq. 11;
3. Learn the probabilistic image model by MLE;

for $t = 2$ **to** 5 **do**
1. Define similarity metrics between images of the same part from adjacent age groups;
2. Learn the dynamics of aging model—transition probabilities;

3 Face Aging: Analysis and Synthesis

Following the compositional face representation and the dynamic model, we propose a multi-level face aging algorithm, which is implemented in three steps: (i) Computing the parse graph representation from an input young face by Bayesian inference in Eq. 11; (ii) Sampling the parse graphs of other age groups from the dynamic model in Eq. 13; (iii) Generating the aging image sequence by the generative model in Eq. 10.

3.1 The Overall Algorithm

Given a young face image I_1 at age group 1, our objective is to infer the parse graph G_1 by maximizing a Bayesian posterior probability, and then synthesize the parse graphs G_2, G_3, G_4, G_5 by sampling the dynamic model. These parse graphs then generate the face images I_2, I_3, I_4, I_5 at consecutive age groups. We summarize the flow of our face aging algorithm as below:

Algorithm 2. Inferring the face aging sequences

input : A young face image I_1
output: A sequence of aged faces I_2 **to** I_5

1. Compute G_1 as parse graph of I_1;
$G_1 = \arg\max p(G_1 \mid I_1; \Delta_1)$
2. Sample the graphs at consecutive age groups from Eq. 13;
$G_t \sim p(G_t|G_{t-1}; \Theta_{\text{dyn}}, \Theta_{\text{AOG}}), t = 2, 3, 4, 5.$
3. Synthesize the aged image I_t from the generative model;
$I_t = J(G_t; \Delta_t)$

3.2 Details of the Algorithm

In this subsection, we present the details for the three steps in the algorithm above.

3.2.1 Computing G_1 from I_1

The process of computing the parse graph representation of the input face image is to infer the hidden variables generating the image, as in Eq. 9 and Eq. 11. This part of work is the integration and extension of the grammatical face model[45] and generative hair model[9] in our group, for self-containment, we briefly discuss step 1 in the following three subsections:

(1.a) Computing the hair representation

Following Chen's generative hair model[9], the geometric attributes $T^{\text{geo}}_{\text{hair}}$ of hair can be represented by its sketch, which includes a set of curves C_k and corresponding directions d_k. After extracting hair image as Fig. 6(a) illustrates, the sketch can be computed by a sketch pursuit algorithm. The photometric attributes $T^{\text{pht}}_{\text{hair}}$ describe the hair texture and include three variables: I_{flow} I_{UV} and I_{shd}. I_{flow} is the vector flow in the hair region, which controls the generation of high frequency hair texture. It can be computed using the hair sketches with prior knowledge of hair direction by a diffusion method. I_{UV} accounts for the hair color, and $I_{\text{shd}} = \{x_i, y_i, \theta_i, \sigma_{x,i}, \sigma_{y,i}\}$ is a set of Gaussian basis simulating the lighting and shading of hair image. Based on $T^{\text{geo}}_{\text{hair}}$

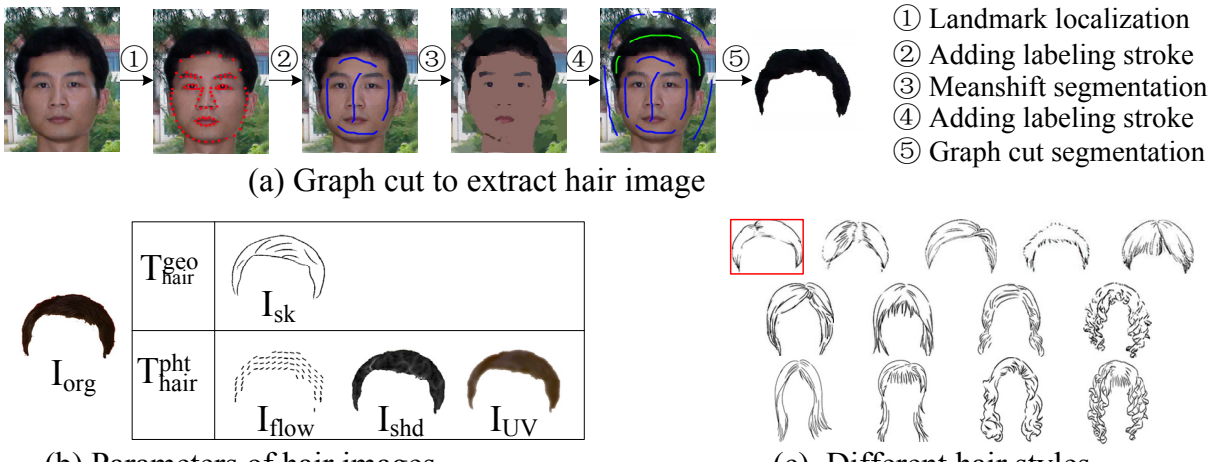


Fig. 6. Computing hair parameters. (a) illustrates the procedure of extracting hair image from complex background. (b) explains the parameters for a hair image I_{org}. The geometric attributes are described by the directed curves in sketch image I_{sk}. Photometric attributes are described by three components: I_{flow} is the vector flow accounting for hair directions, I_{shd} represents the lighting and shading in the hair, and I_{UV} is the color channel of hair image. (c) lists the hair styles in our hair dictionary, the one with boundary is the hair type of I_{org} in (b).

and $T^{\text{pht}}_{\text{hair}}$, we classify hair into a number of styles, which are listed in Fig. 6(c) and indexed by l_{hair}.

(1.b) Computing parameters of face and facial components

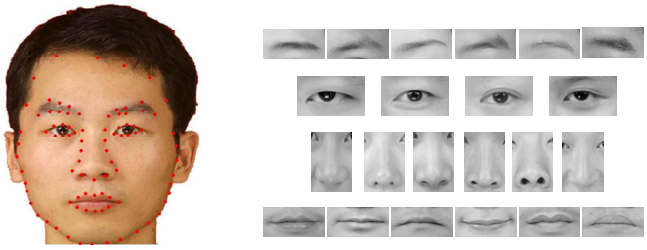


Fig. 7. AAM models of face and facial components. (a) The 90 landmarks defined for the global AAM model. (b) We cluster various images of each facial component into sub classes and build a local AAM model for detailed representation.

We represent the face and facial components with AAM [10] models. First we train a traditional AAM model for the first level face image with 90 landmarks as shown in Fig. 7(a). Because there exist large variations for each facial component (e.g. single-lid eyes and double-lid eyes, etc. in Fig. 7(b)) and a global AAM model is not sufficient for presenting all these details, we build local AAM models to refine these component regions at second level. After clustering facial components into prototypes indexed by l_{cmp}, we train a local AAM model for each prototype. For face and facial components, T^{geo}_i and T^{pht}_i are coefficients of shape eigenvectors and texture eigenvectors respectively, which are both computed by minimizing the reconstruction error.

(1.c) Computing parameters of wrinkles and pigments

In the third level representation, we divide the face skin into 6 wrinkle zones as Fig. 8(a) shows. The wrinkles (curves or sketches) in each zone are located with matching pursuit algorithm using two types of filters:

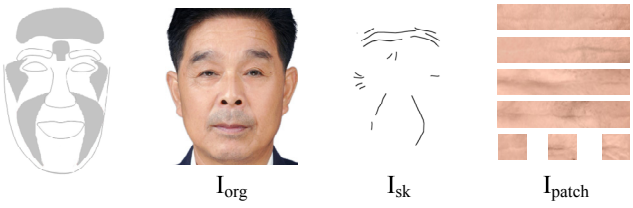


Fig. 8. Parameters for wrinkles and skin marks at the third level. (a) The skin is divided in 6 wrinkle zones, our algorithm adds wrinkles in each zone separately. (b) I_{org} is the input image. The curves and marks in I_{sk} and image patches I_{patch} account for the geometric and photometric attributes of wrinkles respectively.

Gabor wavelets and blobs. The geometric variables $T^{\text{geo}}_{\text{wkl}}$ describe the position, length, orientation of the traced curves, and the position and scale of the marks. The photometric variable $T^{\text{pht}}_{\text{wkl}}$ is represented directly by the straighten wrinkle intensity profiles perpendicular to the wrinkle curves and the skin mark patches in Fig. 8(b). Mostly there is no wrinkle for faces of age under 30, so the initial parse graph G_1 usually has only two levels.

3.2.2 Simulating the Evolution of Parse Graphs

(2.a) Learning the dynamic parameters

To overcome the difficulty of collecting photos of the same person across all age groups, our model decomposes face into parts, and learns the aging transition probabilities for each part separately, which can be cropped from faces of different persons. Fig. 9(a) gives a subset of the training data in three groups for learning dynamics of eye aging and illustrates the aging process of eye, where the thickness of the arrows reflects the transition probability.

The transition of a face component across age groups is allowed only between images of the same prototype, i.e. the same number of landmarks. The similarity mea-

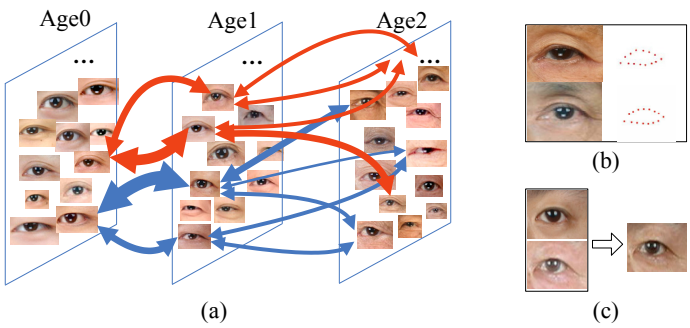


Fig. 9. Learning aging pattern for each part. (a) Eye examples in three age groups. The thickness of the arrows between two eye images indicates the transition probability between the two images in consecutive age groups. (b) The labeled landmarks describing the contours of one pair of selected eyes from two adjacent age groups. (c) An aging result of an eye.

surement over the geometric and photometric attributes, i.e. $\psi()$ in eqns 16 and 17 follows the TPS model and AAM model respectively.

(2.b) Probabilistic sampling to simulate evolution of parse graphs from the dynamic model

For aging simulation, we use probabilistic sampling instead of maximizing the conditional probability $p(G_t|G_{t-1})$ to preserve the intrinsic stochasticity. In our algorithm we adopt widely applicable Gibbs sampling technology as in Alg. 3. For each parse graph G_{t-1}, we can sample a variety of G_t from the probability with different attributes, which in turn generates different aged images. This process is similar to the Brownian motion. The longer the time period, the larger variance can be observed in the sampled results. Fig. 2 illustrates some simulation results over four age groups and we often need to sample more examples for longer time period to account for the large diversity.

Algorithm 3. Gibbs sampling algorithm for evolution of Markov chain

input : $l_{i,1}, T^{\text{geo}}_{i,1}, T^{\text{pht}}_{i,1}$
output: $l_{i,t}, T^{\text{geo}}_{i,t}, T^{\text{pht}}_{i,t}, t = 2, 3, 4, 5$

for $t = 2$ **to** 5 **do**
  for $i = 1$ **to** 3 **do**
    for $loop = 1$ **to** T **do**
      for $m = 1$ **to** $n^{Or}_{i,t}$ **do**
        $l_{i,t}(m) \sim$
        $p(l_{i,t}(m)|l_{i,t-1}(1), ..., l_{i,t-1}(n^{\text{Or}}_{i,t-1}), l_{i-1,t}(m))$
      for $m = 1$ **to** $n^{And}_{i,t}$ **do**
        $T^{\text{geo}}_{i,t}(m) \sim$
        $p(T^{\text{geo}}_{i,t}(m)|T^{\text{geo}}_{i,t-1}(1), ..., T^{\text{geo}}_{i,t-1}(n^{\text{And}}_{i,t-1}), T^{\text{geo}}_{i-1,t}(m))$
      for $m = 1$ **to** $n^{And}_{i,t}$ **do**
        $T^{\text{pht}}_{i,t}(m) \sim$
        $p(T^{\text{pht}}_{i,t}(m)|T^{\text{pht}}_{i,t-1}(1), ..., T^{\text{pht}}_{i,t-1}(n^{\text{And}}_{i,t-1}), T^{\text{pht}}_{i-1,t}(m))$

3.2.3 *Synthesizing Image I_t from G_t.*

By the generative model, we synthesize face image I_t from its parse graph $G_t = (\omega_{1,t}, \omega_{2,t}, \omega_{3,t})$. The image generation process proceeds in three steps from coarse to fine[45]. Firstly, it generates the face and hair image $I_{1,t}$ from $\omega_{1,t}$ based on the AAM model for face and the hair model in [9]. Secondly, it refines the five face components based on $\omega_{2,t}$ and $I_{1,t}$. Each component is again an AAM model with landmarks and appearance. This step leads to higher resolution details and diverse appearance for these components. Thirdly, it generates wrinkles and marks in the 6 skin zones based on $\omega_{3,t}$.

4 Implementation Details

In this section, we discuss some implementation details for the representation and aging of each part—hair, face, components, wrinkles in the dynamic model.

4.1 Level 1: Global Appearance Aging

4.1.1 *Hair Aging*

We annotated $10,000$ face images across the 5 age groups in the Lotus Hill dataset[47], thus a large set of hair images are collected for each age group. For an observed hair image I^{obs}_{t-1} in group $t-1$, we select a similar hair image I^{obs}_t at group t according to two metrics: geometric similarity and texture similarity. The geometric similarity between hair contours is computed using a Thin Plate Spline(TPS) warping energy between two contours, while the texture similarity is computed by KL distance between vector flow histograms of two hair textures. Then the selected hair of group t is warped to fit the face shape of I^{obs}_{t-1} under constraints from the skull structure. Finally we get the final result I^{syn}_t. Fig. 10(b) shows an example of hair aging.

4.1.2 *Face Aging*

At level one, the face aging effects reflect the change of global face shape, skin color darkening and drop of muscles. We select aging patterns based on geometric and photometric similarities. For each face, we have 90 facial points describing the facial geometry $T^{\text{geo}}_{\text{face},t}$. TPS warping energy measuring the cost for aligning two face geometries is used as a natural shape distance. The appearance distance is computed as the KL distance between histograms of corresponding filter responses (mean, variance, et al.) of two aligned faces. As are studied in [1][3][48], there occur certain noticeable bony and soft tissue changes in shape, size and configuration during adult aging, and the shape changes in muscular regions is larger than in bony regions. We compute the differences between mean face shapes of different age groups as is illustrated in Fig. 10(c) and adopt the mean shape changes as soft constraints during warping of face shape as age increases. Fig. 10(d)(e) show the process of first level face aging.

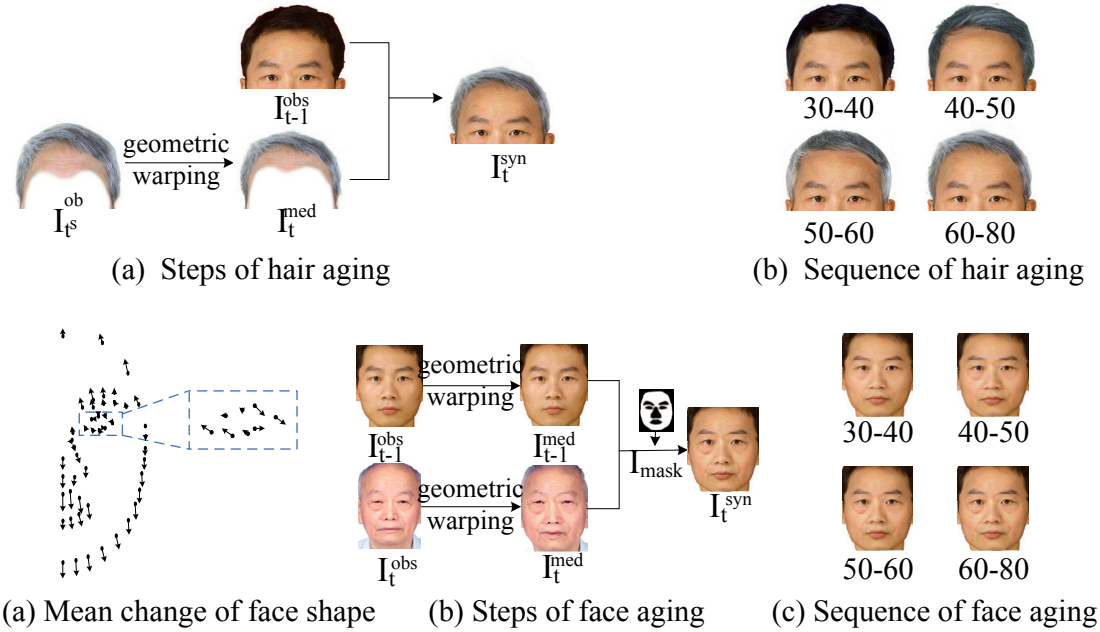


Fig. 10. Steps of hair aging and global face aging. (a) illustrates the aging process of a hair image in age group $t-1$, denoted as I_{t-1}^{obs}. I_t^{obs} is a similar hair image in age group t selected based on similarity metrics. After applying geometric transformation to meet the shape of I_{t-1}^{obs}, we get the intermediate result I_t^{med}. The final aging result is I_t^{syn}. (b) shows a resulted hair aging sequence. (c) illustrates the mean shape changes of face shape in adult aging. Here the length of line-segments denotes change magnitude and orientation describes the moving direction. (d) shows the face aging process of I_{t-1}^{obs}, which is a young face in age group $t-1$. Here I_t^{obs} is the selected similar face image in age group t. I_t^{med} and I_{t-1}^{med} are the intermediate results after applying geometric transformations under Anthropometric constraints. With a mask image excluding the facial components, we can synthesize an aged image as I_t^{syn}. (e) shows an aging sequence synthesized for I_t^{obs} in (d).

4.2 Level 2: Facial Component Aging

Different variations occur to different facial components during face aging. In general, variations include changes in both geometry and photometry. The aging pattern of eyes is the most complex and most important for the final results, therefore, we take the eye aging as an example to explain the component aging approach.

The evolution parameters for eye aging are learned from the dataset of eye patches across age groups, as is shown in Fig. 9(a). By applying AAM searching with the local eye model, we can locate the landmarks of the components accurately as shown in Fig. 9(b). Then the transition probability (thickness of arrows) is computed following Eq. 16,17. The geometric distance in Eq. 16 is measured by TPS bending energy between two eye shapes with the same topology, while the photometric distance in Eq. 17 is computed by summing over the squared intensity difference in the Gaussian window around the matched points. For a given eye image $I_{2,t-1}$, after selecting a similar aged image $I_{2,t}$, we perform two transformations to $I_{2,t-1}$. (1) Warping it to the target shape by applying a set of affine transformations $\mathcal{T}$ to $I_{2,t}$ to minimize the geometric distance between the landmarks of $\mathcal{T}(I_{2,t})$ and $I_{2,t-1}$. (2) Using Poisson image editing[32] techniques to transfer high frequency information in skin region of $\mathcal{T}(I_{2,t})$ to $I_{2,t-1}$ and perform color histogram specification to the non-skin area texture. An aging result of eye is shown in Fig. 9(c).

Symmetry of facial components, such as the left and right eyes, eyebrows, is represented by imposing constraints on the transformations mentioned above. The aging pattern of facial components should also be constrained by the upper level face aging. Fig. 11 gives an aging sequence for each facial component.

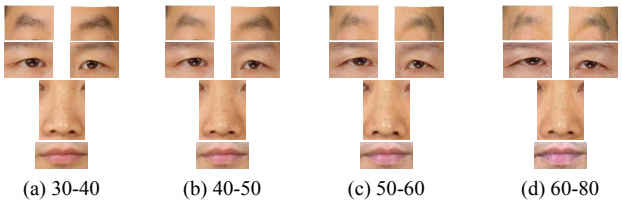
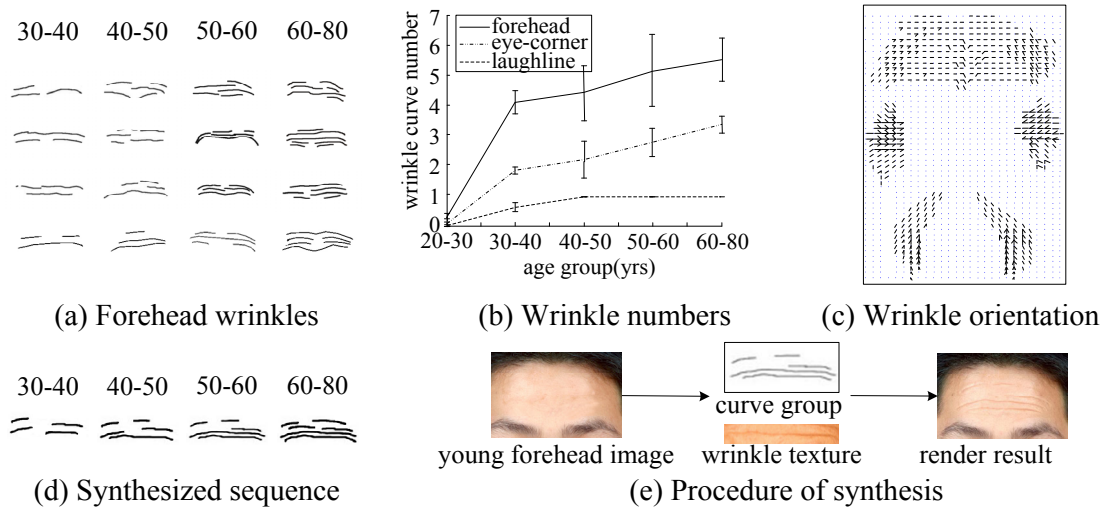


Fig. 11. Intermediate results of facial component aging.

4.3 Level 3: Wrinkle Addition

At level three, we model the aging effects of the 6 wrinkle zones (see Fig. 8). For each age group we labeled 200 images randomly selected from our dataset to learn the statistics of wrinkles. Fig. 12(a) shows some labeled forehead wrinkles collected from the dataset. According to the generative model, the wrinkle addition is completed in two steps: (i) Generating curves in various wrinkle zones. The number of curves and their positioning follow some prior probability densities, as is shown in Fig. 12(b)(c). (ii) Rendering the curves with wrinkle intensity profiles in the dictionary. Given a wrinkle curve and intensity profile, the wrinkle image can be synthesized according to Eq. 8. Fig. 12(d) shows a series of generated wrinkle curves over four age groups and Fig. 12(e) shows an example of generating the wrinkle image from the wrinkle curves.



(a) Forehead wrinkles (b) Wrinkle numbers (c) Wrinkle orientation

(d) Synthesized sequence (e) Procedure of synthesis

Fig. 12. Prior learning and wrinkle synthesis in different skin zones. (a) Some examples of wrinkle curves in the forehead zone. A large number of wrinkle curves are collected from the annotated dataset. (b) The statistics of wrinkle numbers in three wrinkle zones over different age groups. (c) The prior distribution of wrinkle curve orientation in the 6 wrinkle zones, where the length of the arrow reflects the strength and the orientation describes the directions. (d) A sequence of synthetic wrinkle curves. (e) The process of rendering photo-realistic wrinkle images.

4.3.1 Learning Prior of the Wrinkles from Labeled Data

For wrinkle zone m, we model the number of wrinkles with a Poisson distribution:

$$p(n_t(m) = k; t) = \frac{\exp(-\lambda_t(m))(\lambda_t(m))^k}{k!} \tag{17}$$

Here $n_t(m)$ is the number of wrinkles in zone m at age group t and $\lambda_t(m)$ is the parameter learned from the training data.

$$\lambda_t(m) = \frac{1}{M_t}\sum_{k=1}^{M_t} N_t^l(m), \tag{18}$$

in which M_t is the number of training images at age group t and $N_t^l(m)$ is the wrinkle number in zone m of the lth sample at age group t. $\lambda_t(m)$ equals to the mean value in Fig. 12(b).

Similarly, we compute priors of curve length, distance between two adjacent curves. Prior distributions of curve position and the orientation are also learned from the labeled data, as Fig. 12(c) shows.

4.3.2 Generating Wrinkle Curves

In our algorithm, the transition probability of $n_t(m)$ between two consecutive age groups is modeled by a bi-gram model.

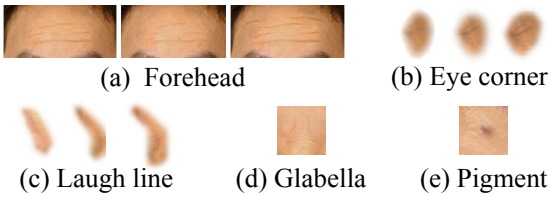
$$p(n_t(m) = k | n_{t-1}(m) = j) = \begin{cases} 0 & \text{k} < \text{j} \\ \frac{1}{z} p(n_t(m) = k; t) & \text{k} \geq \text{j} \end{cases} \tag{19}$$

Here we force $p(n_t(m) < n_{t-1}(m)) = 0$ to ensure that the wrinkle number increases as time goes and z is a normalization factor.

From the statistics of the wrinkle curves, we compute the geometric parameters of the wrinkle curves. Wrinkle number is computed from bi-gram model in Eq. 19. The other variables (length, position and orientation) can be sampled from the corresponding prior distribution. With these geometric parameters we can generate a sequence of curve groups as are shown in Fig. 12(d).

4.3.3 Generating Realistic Wrinkle Images

For the initial wrinkles, we select the wrinkle intensity profile randomly from the dictionary. After warping the profiles to the shape of wrinkle curves, we use Poisson image editing techniques to render realistic wrinkle images (shown in Fig. 12(e)). Because the wrinkle texture across age groups will not change much, we select similar wrinkle profile in the next age group based on the photometric distance. Fig. 13 shows some intermediate results of wrinkle addition in different skin zones.



(a) Forehead (b) Eye corner

(c) Laugh line (d) Glabella (e) Pigment

Fig. 13. Intermediate results of wrinkles and marks emerge at consecutive age groups.

After the aging process at all the three levels, we integrate them together to generate the final results. Our face model is an additive model and Poisson image editing[32] techniques are adopted to obtain seamless fusion results.

5 Experiments: Aging Simulation, Age Estimation, and Human Evaluation

5.1 Dataset Collection and Organizations

One of the widely used datasets for face aging is the FG-Net aging database [49]. It includes 1,002 photos of 82 subjects, whose ages are between 0 and 69. As

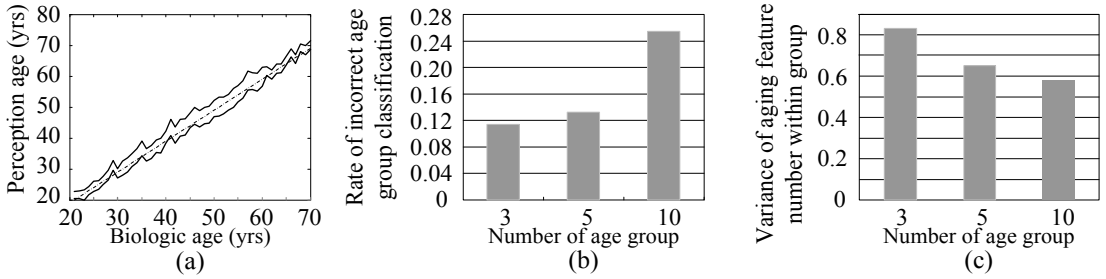


Fig. 14. Experiments for selecting the number of age groups. (a) Plot of the appearance age against the biologic age averaged over 500 faces in human experiment. The two solid lines illustrate the standard deviation of subjective age estimation results and dashed line is the ground truth. (b) Vertical axis is the rate of images with age group being incorrectly estimated, and the horizontal axis is the age group number. (c) The within-group appearance variations for different group numbers. Here the appearance variations is described by the standard deviation of certain age related feature number.

many images in the FG-Net dataset are not of very high resolution and about 60% are children, we did not use it for aging simulation, instead it is used for a comparative study on age estimation. We collected a database with about 50,000 ID photos of Asian adults in the age range of [20, 80], and the statistics of the database is shown in Table 1. All these face images have high resolutions with the between-eye distance being about 100 pixels. We train our algorithm and perform face aging simulation on this dataset, some results are shown in Fig. 15. Another publicly available aging dataset is the MORPH database[36], an extended version of which includes 16,894 face images of 4,664 adults, among whom 13,201 images from African-Americans, 3,634 from Caucasian descents and 59 are of other groups. There are 2,505 females and 14,389 males in this dataset. The average age is 40.28 years and maximum age is 99 years. we reorganize the MORPH database for face aging (see Table 1) and synthesize several aging sequences on this dataset to validate the generality of our algorithm. We also collected real aging sequences from 20 people (friends and relatives) for the evaluation experiments.

As life experiences affect face appearance, we must distinguish the appearance age from biologic age. Biologic age is the actual age of the subject while the appearance age is the perceived age. Often appearance age needs to be estimated through human experiments, the biologic age is not completely a sure thing either. In our dataset, we know the birth dates of the people in the ID photos and the time when the photo was taken. The latter is recorded at the time when the file was created.

In our first human experiment, we use 500 face images of different ages and asked 20 volunteers(college students) to estimate the appearance age. Fig. 14(a) plots the results. The two solid lines illustrate the standard deviation of difference between appearance age and biologic age. In general, the estimated age can be different from the biologic age by 3-5 years older or younger.

Due to the intrinsic ambiguities, we divide the age range into 5 age groups: [20,30), [30,40), [40,50), [50,60) and [60,80] based on the following reasons. (i) The difference between biologic age and appearance age is about 3-5 years. Thus the appearance ages between two individuals in a certain age have an uncertainty

TABLE 1
Data Distribution

Database & ethnic group	Gender group	Age group				
		1	2	3	4	5
LHI DB Asian	Male	8599	5731	3639	2846	2216
	Female	7725	6324	3747	2893	1989
MORPH DB Caucasian	Male	465	4578	4570	1373	128
	Female	116	768	777	180	11
MORPH DB African American	Male	180	1135	1182	358	69
	Female	32	269	257	41	4

interval of 6-10 years. (ii) As we increase the number of age groups, the perceptual errors among these groups increase (see Fig. 14(b)), thus it is hard to evaluate the synthesis results. On the other hand, when the number of age groups increases, the feature variance within each group decreases, and make the model more accurate (see Fig. 14(c)). As a tradeoff, we select 5 groups. (iii) The number of images within group [60, 80] is relatively small, because less senior people took ID photos.

5.2 Experiment I: Face Aging Simulation

We take 10,000 images from the Asian dataset and annotate these images by decomposing them into 3 levels to build the compositional and dynamic model. For each face image, we label 90 landmarks on the face and about 50 landmarks for hair contour. Based on the annotation, our algorithm parses the face into parts and primitives, and then builds the hierarchic dictionaries for each age group automatically. We learn the dynamic model as discussed in Section 4. Based on the learned model, we test our inference and simulation algorithms using a number of young faces in the [20, 30) age range, and generate images for the other 4 age groups. Fig. 15 shows some of the aging results synthesized by our algorithms. Fig. 2 shows an example of simulating multiple plausible aging sequences for a person following the Markov chain model, as Fig. 1 specifies. Note that people shown in Fig. 15 and Fig. 2 are not in the training set, as we cannot show the ID photos for privacy reasons.

We also synthesize a series of aging results from MORPH database to test the generality of our algorithm. Since aging pattern has large variations for subjects from different ethnic groups, we label 1,000 images of African-Americans and 1,000 of Caucasians from MORPH database, and learn two aging models for two

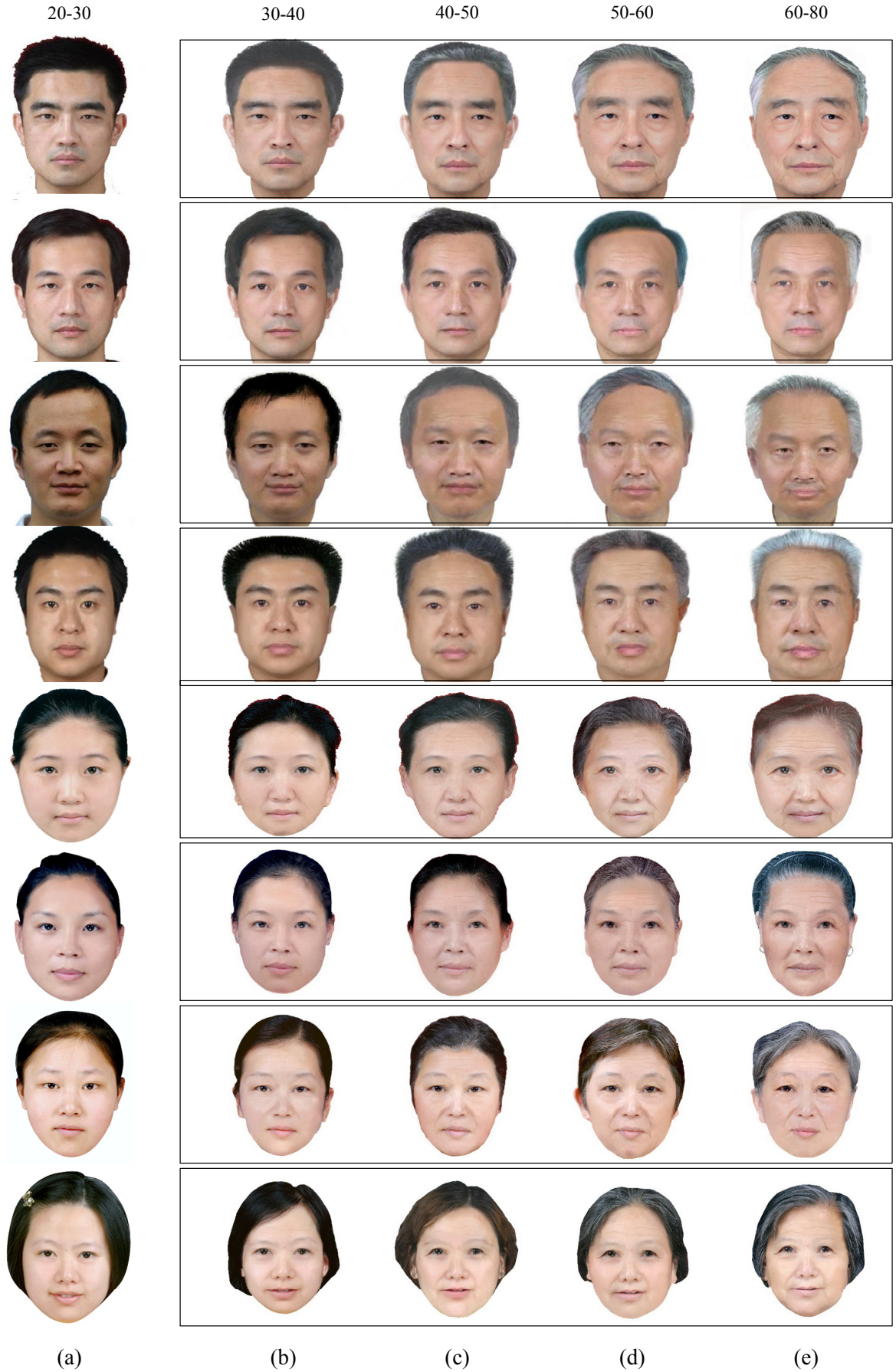


Fig. 15. Some aging simulation results. The leftmost column is the original images of the individuals in group 1. The 2nd to 5th columns are synthetic aged images at 4 consecutive age groups.

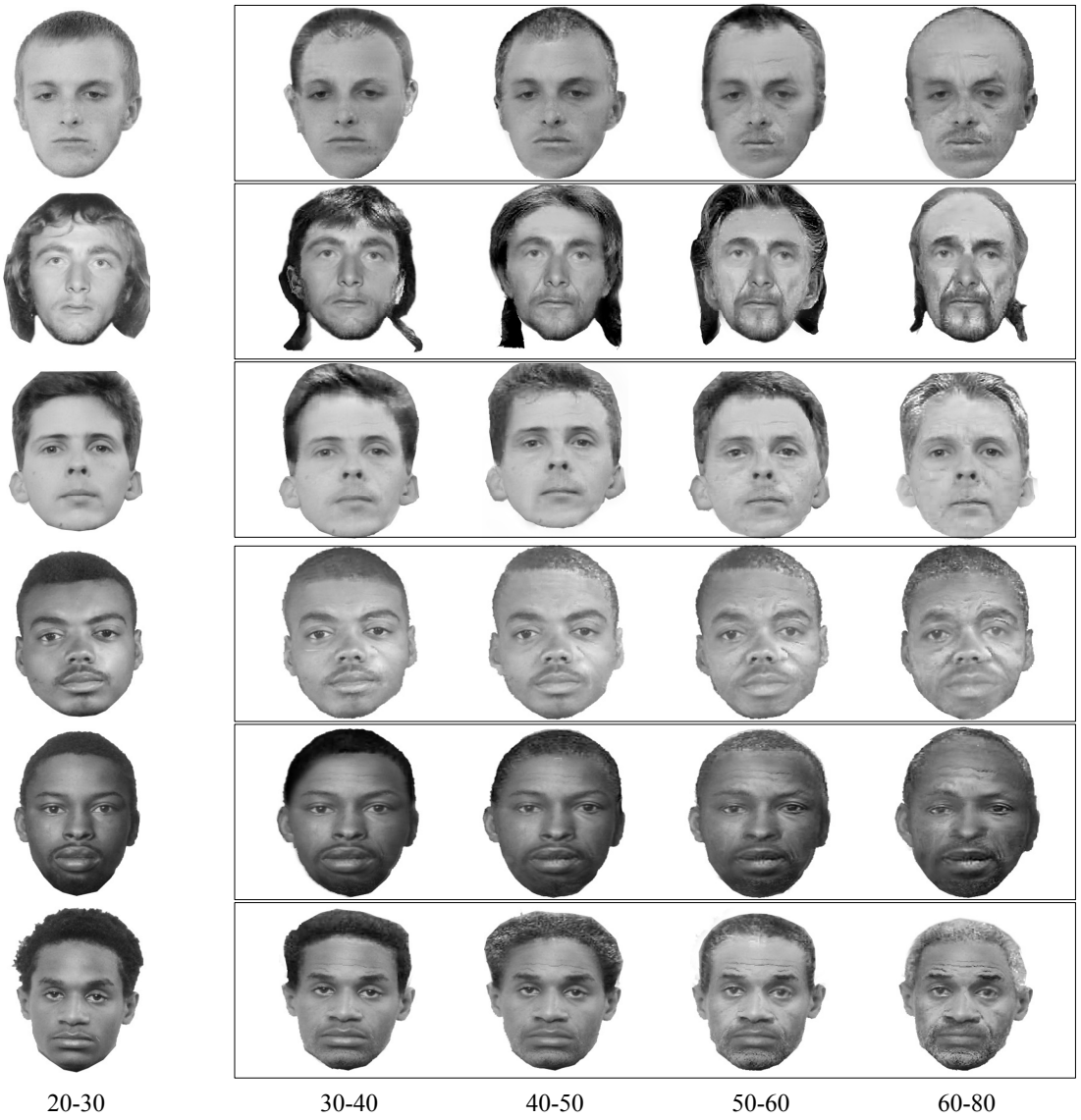


Fig. 16. Some aging simulation results on MORPH database. First three rows are Caucasian faces and next three are African-American faces. The leftmost column is the original input images of the individuals in group 1. The 2nd to 5th columns are synthetic aged images at 4 consecutive age groups.

ethnic groups separately. Female aging sequences are not synthesized because the number of females in age group 4 and 5 is too small for learning the dynamics. The simulation results are shown in Fig. 16, in which top three rows and bottom three rows are separately for Caucasian and African-American faces.

5.3 Experiment II: Contributions of Facial Parts to Subjective Age Estimation

Our aging algorithm uses part based strategy and we notice that some features influence the age perception significantly more than others. This observation inspires us to study the relative contribution of each aging feature to age perception quantitatively. The features considered in our experiment include both the internal factors (e.g. brow, eyes, nose, mouth, skin zones) and the external factor (mainly the hair).

We select 100 mid-resolution images from our database with 20 images for each of the five age groups.

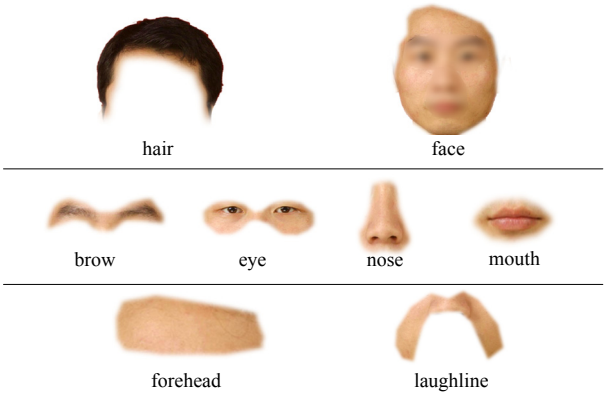


Fig. 17. Eight masks are designed to extract different parts for the experiment of relative contributions.

As Fig. 17 displays, we extract eight sub images for face, hair, brow, eye, nose, mouth, forehead, and laughline. Volunteers are presented the masked images and asked to estimate the age of each part. Then we apply

TABLE 2
Relative Contribution of Each Facial Part to Subjective Age Perception

Subjective estimation	Facial part	face	eye	hair	forehead	laughline	mouth	nose	brow
	β	0.357	0.205	0.179	0.09	0.083	0.058	0.041	0.013
Objective estimation	Facial part	laughline	hair	forehead	face	eye	mouth	brow	nose
	β	0.480	0.408	0.373	0.257	0.181	0.017	0.006	0.002

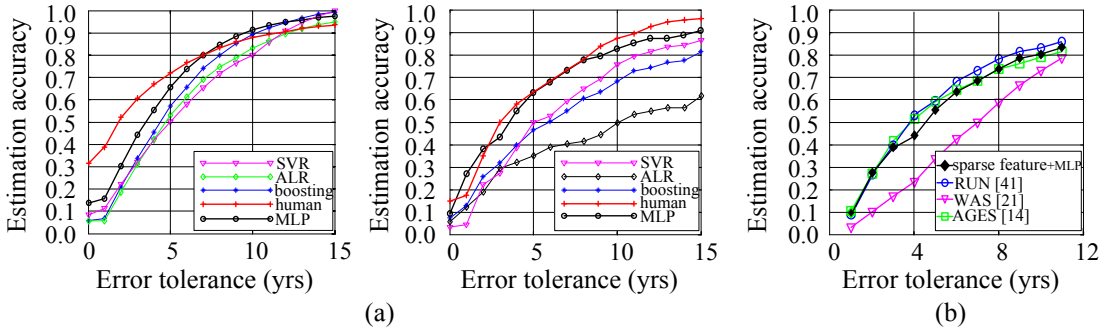


Fig. 18. Cumulative scores of automatic age estimation algorithms. (a) Performances of proposed regressors on our dataset(left) and FG-NET database(right). (b) Comparison between the performances of our estimation algorithm and the-state-of-art algorithms on FG-NET database.

Multivariate Regression Analysis(MRA) to measure the contributions of each component to the perception age of the whole face. The R square value is 0.907, this indicates that our model accounts for most of the age related changes. β values of different features are shown in Table 2.

From the β value of each feature, we can see clearly that there are 5 features contribute most to the age perception. The large contribution of the hair confirms the effectiveness of the hair feature in face age perception which was missing in previous studies.

5.4 Experiment III: Automatic Age Estimation

In this experiment we use an age estimation algorithm[40] to test the accuracy of synthetic images. The estimation approach formulates age estimation as a repression problem on features extracted from the hierarchical face model and tests performances of various regressors, including Age specific Linear Regression(ALR), Support Vector Regression(SVR), Multi-layer Perceptron(MLP) and logistic regression(boosting). Among these regressors, MLP performances best in our experiment.

Here we conduct age estimation experiments on two datasets: First we selected a set of 8,000 face images (4,000 males and 4,000 females) from our dataset and denote it as set A, four-folds cross validation is conducted for performance measurement. Then we conduct comparative study on 1,002 photos from FG-net (denoted as set B) to validate the effectiveness of our algorithm. On set A, mean absolute error(MAE) of our algorithm is about 4.68 years and $CS_{\leq 10} = 91.6\%$ averagely. Performance on FG-NET dataset is relative lower, with MAE $= 5.97$ years and $CS_{\leq 10} = 82.7\%$ due to resolution limitations and affects from other variations, while it is still comparative to the state-of-art algorithms (see Fig. 18(b)), with MAE being 5.78 years in Geng's[16] and 6.22 years in Yan's[46].

Similar to Experiment II, we perform MAR to measure relative contributions of different facial parts in our algorithm, the R square value is 0.95 and β values are shown in Table 2. From the rank of contributions, one can see that for adult age estimation, wrinkles in laughline, forehead and around-eye region provide plenty of information and hair is also an important cue for age perception. Here wrinkles in the laughline region and hair features display larger significances than in the experiment II (Table 2) subjective experiment, this maybe due to that other features (e.g. wrinkles in eyecorner region, etc.) can be more easily affected by illumination.

5.5 Experiment IV: Evaluating Face Aging Results

Similar to [24], We use two criteria to evaluate the goodness of the aging model. (i) *The accuracy of simulation*, i.e. whether the synthetic faces are indeed perceived to be of the intended age. (ii) *Preservation of the identity*, i.e. whether the synthetic faces are still recognized as the original person. In this subsection, we conduct both subjective(human) experiments and objective(algorithmic) experiments as quantitative measurement for these two criteria. Twenty volunteers are recruited to evaluate our aging results subjectively and age estimation algorithm[40] is adopted as objective evaluations to measure the accuracy of aging simulation. Corresponding to the hierarchical face representation and three-level aging algorithm, we conduct evaluation experiments on facial images at three resolutions. The quantitative analysis in following two subsections are performed on face images of Asians.

1. Experiment IV.a: evaluating the accuracy of simulation. We compare set C and D in this experiment. For set C, we select randomly 20 real face images from the ID photo dataset for each of the age groups 2-5 respectively. For set D, we select 20 young faces in age group 1 and synthesize one aging sequence for each person as Fig. 15 shows. Thus, set D has 80 synthetic images with 20 images in each of the age groups 2-5. We normalize the

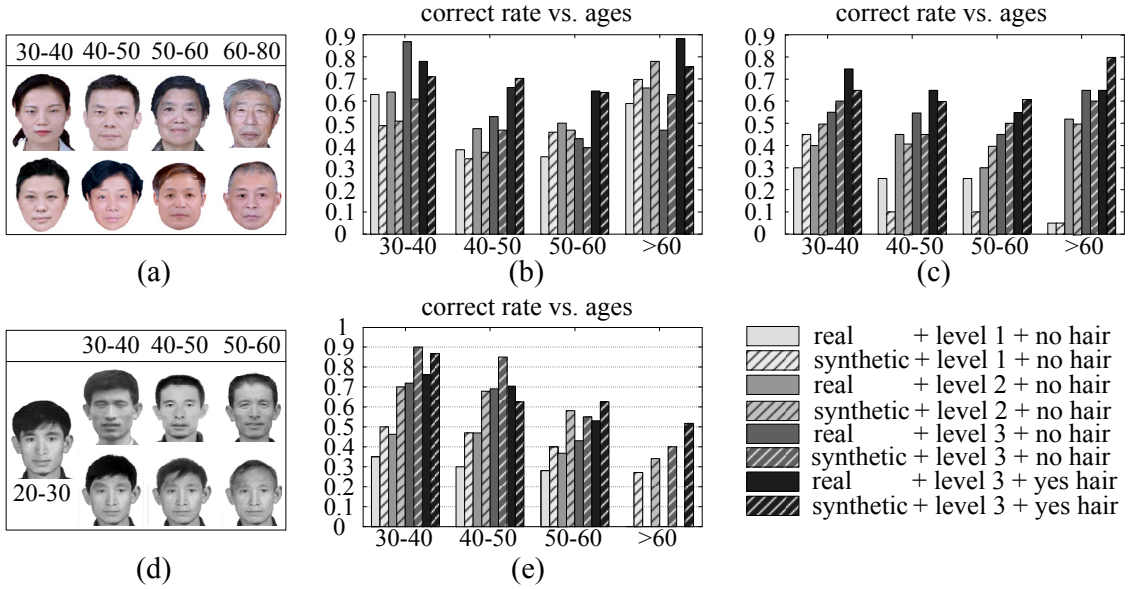


Fig. 19. The accuracy of age perception and face identification. (a) The first row shows real images from set C and second row shows synthetic images from set D. (b) and (c) are separately performance of subjective age perception and algorithmic age estimation on set C and D. In (d), the first row is a real aging sequence from set E, and the second row is a sequence synthesized by our algorithm for the same individual. (e) plots the performances of subjective face recognition on set E and F.

images in set C to the same resolution and intensity level as the set D images. Fig. 19(a) gives some example images from set C (1st row) and set D (2nd row).

In the human perception experiment, the volunteers are asked to estimate the age of each face in the two sets. Fig. 19(b) plots the human estimation results on the two sets. From the plot, we can see following phenomena: (1) The accuracy improves with resolution increasing, because details at middle and high resolutions indeed provide informations for facial age estimation. The high performance at high resolution also validates the adopted hierarchical face model. (2) The age estimation results of the synthetic images are mostly consistent with those of real images. (3) Estimation result with hair cropped out is a little lower, this shows that hair is an effective feature for age perception. (4) For subjective evaluation hair has negative influences on estimation performance in group 30-40 and helps estimation a lot in group 60-80, maybe because that large intersection occurs in hair styles in group 30-40 and 40-50, whereas in group 60-80 hair appearance is informative for age estimation.

We analyze the age estimation results on high resolution face images by ANOVA. The large main effects of age group on age estimation ($F_{3,156} = 216.511, p = 0.000$ with hair included and $F_{3,156} = 49.142, p = 0.000$ without hair) indicate that our model accounts for the aging related variations mostly and the small main effects ($F_{1,158} = 0.080, p = 0.295$ with hair and $F_{1,158} = 1.415, p = 3.885$ without hair) of image set on age estimation show that the estimation accuracies on two sets do not differ significantly.

At the same time, we perform objective age estimation on both sets using age estimation algorithm in experiment III and obtain similar results, as Fig. 19(c) shows. The plot indicates that synthetic images include appropriate aging related variations and consist with real images in age perception accuracy. Performance is improved about 15% with hair features included. ANOVA analysis result is also similar to that of subjective experiment: Age group shows significance ($F_{3,156} = 235.39, p = 0.000$ for images with hair and $F_{3,156} = 167.368, p = 0.000$ for images without hair) and there is no apparent difference between estimation accuracies of two sets ($F_{1,158} = 0.006, p = 0.023$ with hair and $F_{1,158} = 0.225, p = 0.613$ without hair).

2. Experiment IV.b: evaluating the preservation of face identity. We compare set E and F. For set E, we use 20 real aging sequences from friends and relatives (they are all Asians and the images are different from the ID photo dataset) with images in group 5 missing. For each young face at age group 1 in Set E, we synthesize one aging sequence as Fig. 15. Thus we have 80 synthetic images and denote them as Set F. Fig. 19(d) shows some examples from set E (1st row) and set F (2nd row).

We then add 50 faces in age group 1 as "distracting background". Since the resolution of some old photos is relatively low, we down-sample the images in F to the same resolution with images in Set E. Thus we randomly draw an image from set E or set F in the age groups 2-5, and ask the volunteers to identify the image to the 70 candidates (20 real and 50 distractors) in age group 1.

Fig. 19(e) shows the recognition rates by humans on both sets in the four age groups. From the result, we can see that recognition rate improves as resolution increases in each age group. In accord with our model, it has lower recognition rate for longer aging period, and recognition rate after three decades is only around 50%. One can also see that the recognition performance on synthetic images is slightly higher than on real aging sequences, this indicates that our algorithm preserves face identity

of the input face very well. The lower performances on real aging sequences is partially due to the effects from non-age related variations(i.e. illumination, pose, et al.)

In the same way as in Experiment 1, we apply ANOVA to the recognition results on synthetic high resolution faces, and find that recognition rate is affected significantly by age group ($F_{2,117} = 0.839, p = 0.000$ with hair and $F_{2,117} = 6.291, p = 0.003$ with hair excluded). Different to age perception results, image set also shows some significance for the intrinsic variations between two image sets ($F_{1,118} = 0.104, p = 0.031$ with hair included and $F_{1,118} = 2.739, p = 0.101$ without hair).

6 Conclusions

We present a compositional and dynamic face aging model, based on which we develop algorithms for aging simulation and age estimation. Results synthesized by our algorithm are evaluated for the accuracy of age simulation and the preservation of identity. Our estimation algorithm obtains performances comparative to the state-of-art algorithms. Our results are attributed to two factors: a large training set and the expressive power of the compositional model, including external appearance (e.g. hair color and hair style) and high resolution factors (e.g. wrinkles, skin marks, etc.).

Although our work on modeling adult face aging achieved promising visual results, more work remains to be explored in the future. (i) When more image aging sequences from the same individuals become available, our model should be extended by assigning more weights to these samples, and hopefully our model may be also suitable for face recognition applications besides entertainment ones. (ii) Objective evaluation on identity preservation is not conducted for lacking of real face aging sequences over a long period, i.e. 3-4 decades, and effective recognition algorithms. With more and more aging databases become available as well as the progress of face recognition technologies, this kind of evaluation will be conducted on time.

Acknowledgments

This work is done at the Lotus Hill Institute and the data used in this paper were provided by the Lotus Hill Annotation project [47]. This project is supported by grants from NSFC China under contract No. 60672162, No. 60728203 and two 863 programs No. 2006AA01Z121 and No. 2007AA01Z340.

References

[1] A. M. Albert, K. Ricanek Jr. and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," Journal of Forensic Science International, vol. 172, no. 1, pp. 1-9, Apr. 2007.

[2] Y. Bando, T. Kuratate and T. Nishita, "A simple method for modeling wrinkles on human skin," Proc. Tenth Pacific Conf. Computer Graphics and Applications, pp. 166-175, 2002.

[3] R. G. Behrents, *An atlas of growth in the aging cranio facial skeleton*, Center for Human Growth and Development, University of Michigan Press, 1985.

[4] A. C. Berg and S. C. Justo, "Aging of orbicularis muscle in virtual human faces," Proc. Seventh Int'l Conf. Information Visualization, pp. 164-168, Jul. 2003.

[5] A. C. Berg, F. J. Perales Lopez and M. Gonzalez, "A facial aging simulation method using flaccidity deformation criteria," Proc. Tenth Int'l Conf. Information Visualization, pp. 791-796, Jul. 2006.

[6] L. Boissieux, G. Kiss, N. M. Thalmann and P. Kalra, "Simulation of skin aging and wrinkles with cosmetics insight," Proc. Eurographics Workshop on Animation Computer Animation and Simulation, pp. 15-27, 2000.

[7] D. M. Burt and D. I. Perrett, "Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and color information," Proc. Royal Society of London, vol. 259, pp. 137-143, Feb. 1995.

[8] H. Chen, Z. Xu, Z. Liu and S. C. Zhu, "Composite templates for cloth modeling and sketching," Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. 943-950, Jun. 2006.

[9] H. Chen and S. C. Zhu, "A generative sketch model for human hair analysis and synthesis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1025-1040, Jul. 2006.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, Jun. 2001.

[11] D. DeCarlo, D. Metaxas and M. Stone, "An anthropometric face model using variational techniques," Proc. Twenty-fifth Int'l Conf. Computer Graphics and Interactive Techniques, pp. 67-74, 1998.

[12] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," IEEE Trans. Multimedia, vol. 10, no. 4, pp. 578-584, Jun. 2008.

[13] Y. Fu and N. Zheng, "M-Face: An appearance-based photorealistic model for multiple facial attributes rendering," IEEE Trans. Circuits and Systems for Video Technology, vol. 16, no. 7, pp. 830-842, Jul. 2006.

[14] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," IEEE Trans. Image Processing, vol. 17, no. 7, pp. 1178-1188, Jul. 2008.

[15] M. Gandhi, "A method for automatic synthesis of aged human facial images," Master's thesis, McGill University, 2004.

[16] X. Geng, Z. Zhou and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

[17] C. M. Hill, C. J. Solomon and S. J. Gibson, "Aging the human face-a statistically rigorous approach," IEE Symposium on Imaging for Crime Detection and Prevention, pp. 89-94, Jun. 2005.

[18] T. J. Hutton, B. F. Buxton, P. Hammond, and H. W. W. Potts, "Estimating average growth trajectories in shape-space using kernel smoothing," IEEE Trans. Medical Imaging, vol. 22, no. 6, pp. 747-753, Jun. 2003.

[19] F. Jiang and Y. Wang, "Facial aging simulation based on super resolution in tensor space," Proc. Fifteenth Int'l Conf. Image Processing, pp. 1648-1651, 2008.

[20] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," Computer Vision and Image Understanding, vol. 74, no. 1, pp. 1-21, Apr. 1999.

[21] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442-455, Apr. 2002.

[22] A. Lanitis, "Comparative evaluation of automatic age-progression methodologies," EURASIP Journal on Advances in Signal Processing, vol. 8, no. 2, Jan. 2008.

[23] A. Lanitis, C. Dragonova and C. Christondoulou, "Comparing different classifiers for automatic age estimation," IEEE Trans. Systems, Man and Cybernetics, Part B, vol. 34, no. 1, pp. 621-628, Feb. 2004.

[24] A. Lanitis, "Evaluating the performance of face-aging algorithms," Proc. Eighth Int'l Conf. Automatic Face and Gesture Recognition, 2008.

[25] W. S. Lee, Y. Wu and N. M. Thalmann, "Cloning and aging in a VR family," Proc. Virtual Reality, pp. 61-68, Mar. 1999.

[26] F. R. Leta, A. Conci, D. Pamplona and I. Itanguy, "Manipulating facial appearance through age parameters," Proc. Ninth Brazilian Symposium on Computer Graphics and Image Processing, pp. 167-172, 1996.

[27] H. Ling, S. Soatto, N. Ramanathan and D. W. Jacobs, "Study of face recognition as people age," Proc. Eleventh Int'l Conf. Computer Vision, pp. 1-8, 2007.
[28] Z. Liu, Z. Zhang, and Y. Shan, "Image-based surface detail transfer", IEEE Computer Graphics and Applications, vol. 24, no. 3, pp. 30-35, May. to Jun. 2004.
[29] S. Mukaida, H. Ando, K. Kinoshita, M. Kamachi and K. Chihara, "Facial image synthesis using age manipulation based on statistical feature extraction," Proc. Visualization, Imaging, and Image Processing, pp. 12-17, 2002.
[30] U. Park, Y. Tong, and A. K. Jain, "Face recognition with temporal invariance: A 3D aging model," Proc. Eighth Int'l Conf. Automatic Face and Gesture Recognition, 2008.
[31] E. Patterson, K. Ricanek, M. Albert and E. Boone, "Automatic representation of adult aging in facial images," Proc. Sixth IASTED Int'l Conf. Visualization, Imaging, and Image Processing, pp. 612, 2006.
[32] P. Perez, M. Gangnet and A. Blake, "Poisson image editing," ACM Trans. Graphics, vol. 22, no. 3, pp. 313-318, Jul. 2003.
[33] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," Proc. Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 387-394, 2006.
[34] N. Ramanathan and R. Chellappa, "Face verification across age progression," IEEE Trans. Image Processing, vol. 15, no. 11, pp. 3349-3361, 2006.
[35] N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," Proc. Eighth Int'l Conf. Automatic Face and Gesture Recognition, 2008.
[36] K. Ricanek Jr. and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognition, pp. 341-345, 2006.
[37] C. M. Scandrett, C. J. Solomon and S. J. Gibson, "A person-specific, rigorous aging model of the human face," Pattern Recognition Letters, vol. 27, no. 15, pp. 1776-1787, Nov. 2006.
[38] R. Singh, M. Vatsa, A. Noore, and S. K. Singh, "Age transformation for improving face recognition," Proc. Second Int'l Conf. Pattern Recognition and Machine Intelligence, pp. 576-583, 2007.
[39] J. Suo, F. Min, S. C. Zhu, S. Shan and X. Chen, "A multi-resolution dynamic model for face aging simulation," Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2007.
[40] J. Suo, T. Wu, S. C. Zhu, S. Shan, X. Chen and W. Gao, "Design sparse features for age estimation using hierarchical face model," Proc. Eighth Int'l Conf. Automatic Face and Gesture Recognition, 2008.
[41] B. P. Tiddeman, M. R. Stirrat and D. I. Perrett, "Towards realism in facial prototyping: results of a wavelet MRF method", Proc. Twenty-fourth Theory and Practice of Computer Graphics, pp. 105-111, 2006.
[42] J. Wang and C. Ling, "Artificial aging of faces by support vector machines," Proc. Seventeenth Canadian Conference on Artificial Intelligence, pp. 499-503, 2004.
[43] J. Wang, Y. Shang, G. Su and X. Lin, "Age simulation for face recognition," Proc. Eighteenth Int'l Conf. Pattern Recognition, vol. 3, pp. 913-916, 2006.
[44] Y. Wu and N. M. Thalmann, "A dynamic wrinkle model in facial animation and skin aging," Journal of Visualization and Computer Animation, vol. 6, pp. 195-205, Oct. 1995.
[45] Z. Xu, H. Chen, S. C. Zhu and J. Luo, "A hierarchical compositional model for face representation and sketching," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 955-969, Jun. 2008.
[46] S. Yan, H. Wang, X. Tang and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," Proc. Eleventh Int'l Conf. Computer Vison, pp. 1-8, 2007.
[47] Z. Yao, X. Yang, and S. C. Zhu, "Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks," Proc. Sixth Int'l Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 169-183, 2007.
[48] M. S. Zimbler, M. S. Kokosk and J. R. Thomas, "Anatomy and pathophysiology of facial aging," Facial Plastic Surgery Clinics of North America, vol. 9, no. 2, pp. 179-187, 2001.
[49] "Face and Gesture Recognition," Network: FG-NET aging database. [online] Available:(http://sting.cycollege.ac.cy/alanitis/fgnetaging/).

Jinli Suo received the B.S. degree in Department of Computer Science from Shan Dong University in 2004, Shandong, China. She is currently a PhD candidate in Graduate University of Chinese Academy of Sciences(GUCAS), Beijing, China. Her research interest mainly include face aging modeling, face recognition, and perception of human faces.

Song-Chun Zhu received the BS degree from the University of Science and Technology of China in 1991 and the MS and PhD degrees from Harvard University in 1994 and 1996, respectively. He is currently a professor with the Department of Statistics and the Department of Computer Science at the University of California, Los Angeles (UCLA). Before joining UCLA, he was a postdoctoral researcher in the Division of Applied Math at Brown University from 1996 to 1997, a lecturer in the Department of Computer Science at Stanford University from 1997 to 1998, and an assistant professor of computer science at Ohio State University from 1998 to 2002. His research interests include computer vision and learning, statistical modeling, and stochastic computing. He has published more than 100 papers in computer vision. He has received a number of honors, including the David Marr Prize in 2003, the J.K. Aggarwal prize from the Int'l Association of Pattern Recognition in 2008, the Marr Prize honorary nominations in 1999 and 2007, a Sloan Fellowship in Computer Science in 2001, a US National Science Foundation Early Career Development Award in 2001, and an US Office of Naval Research Young Investigator Award in 2001. In 2005, he founded, with friends, the Lotus Hill Institute for Computer Vision and Information Science in China as a nonprofit research organization (www.lotushill.org).

Shiguang Shan (M'04) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, in 2004. He has been with ICT, CAS since 2002 and has been an Associate Professor since 2005. He is also the Vice Director of the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition, ICT, CAS. His research interests cover image analysis, pattern recognition, and computer vision. He is focusing especially on face recognition related research topics. He received the China's State Scientific and Technological Progress Awards in 2005 for his work on face recognition technologies.

Xilin Chen Xilin Chen (M00) received the BS, MS and PhD degrees in Computer Science from Harbin Institute of Technology (HIT), China, in 1988, 1991 and 1994 respectively. He was a Professor with the HIT from 1999 to 2005 and was a Visiting Scholar with Carnegie Mellon University from 2001 to 2004. He was selected into the one hundred talent program of Chinese Academy of Sciences (CAS) in 2004, and as a professor with Institute of Computing Technology (ICT), CAS. Prof. Chen is now the director of the Intelligent Information Processing Division, ICT, CAS, and the director of the Key Lab of Intelligent Information Processing, CAS. He also leads the ICT-ISVision joint lab for face recognition. He has served as a program committee member for more than twenty International Conferences in these areas, including ICCV, CVPR, ICIP, ICPR, etc. His research interests are Image Understanding, Computer Vision, Pattern Recognition, Image Processing, Multimodal Interface, and Digital Video Broadcasting. He received several awards, including thrice of the China's State Scientific and Technological Progress Award in 2000, 2003 and 2005 respectively, for his academic researches. He is the (co-)author of more than 150 papers.