

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 dynamic model for face aging. By augmenting the high resolution hierarchic face model studied in Xu et al [40], [41] with aging and hair features, this compositional model represents all face images by a hierarchical And-Or graph[5]. The And nodes decompose face into parts and primitives at three levels from coarse to fine. The first level describes the low resolution face image including the hair-style and face appearance. The second level refines the first level by modeling shape, intensity and other variability of facial components. The third level accounts for the skin marks and wrinkles in different facial zones. Correspondingly three prominent aspects of aging related changes are integrated in this model: global appearance changes in hair style and face shape, deformations and aging effects of facial components, and wrinkles in various facial zones. Among the nodes at the same level, spatial relationships and constraints are modeled to force the validness of the configurations. At each level, the Or nodes describe the alternative sub-configurations, which model the diversity of human face across all ages. After choosing alternative of each Or nodes, the And-Or graph turns into a parse graph, which represents a specific instance of human face. Then face aging is modeled as a dynamic Markov process on this graph representation. The evolution of a graph includes not only the changes of its topology representing the emergence of features related to a new age but also the changes of its nodes attributes accounting for the extent of the aging status. The parameters of the dynamic model are learned from a large annotated dataset in a supervised way. Based on this model, we present an algorithm of synthesizing aged face images from a young face image: given an input image, we firstly compute its parse graph representation, and then sample the graph parameters over various age groups according to the learned dynamic model, including abrupt changes and continuous changes. Our sampling process accounts for the uncertainty in face aging by generating multiple plausible aged faces according to the stochastic model. Finally we generate new face images with the sampled parse graphs. We study two criteria to evaluate the synthesized aging results and conduct two human experiments on the aged image sequences synthesized by our algorithm. The results validate the performance of our model and aging algorithm.

Keywords: Face Aging Modeling, Generative Model, And-Or Graph, ANOVA.

Email contacts: Jinli Suo jlsuo.lhi@gmail.com, Song-Chun Zhu sczhu@stat.ucla.edu. This version is submitted to PAMI. A short version appeared in CVPR 2007 [33].

I. INTRODUCTION

Face aging is an interesting and challenging problem in computer vision. Aging can cause significant alterations in face appearances and deteriorate the performance of face recognition systems. Modeling face aging process and simulation of aging effects have the potential of building robust face recognition systems in a variety of real world applications, such as person identification based on old ID photos, recognition of fugitives, looking for missing children, and entertainment in digital art applications. 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: the hair whitens, the shape of head changes, the muscle drops, the skin relaxes, the wrinkles appear, and so on. Traditional AAM model[7] is not expressive enough to describe all those variations in an unified model. (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, such as illumination, expression, pose, et al. Traditional learning strategy is not able to tackle the problem of lacking real aging sequence of the same person as training data. (iv) There are no quantified measurements for evaluating the aging results.

A. Previous Work

Face aging modeling, face aging simulation and their wide applications 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 coordinate transformation on a set of landmarks [29], [25], [8], [13] or a parametric approach [24], [22], [14] to model age related shape changes. For example, Pittenger

and Shaw[29] introduced coordinate transformations to impose age related changes on human faces. Their experiment showed that the shear and cardioidal strain transformations are most effective for altering the perceived age of facial profiles. Lanitis et al. [22] represented face with parameters of AAM model and built three aging functions to describe the relationships between age and the parameters, by which they could estimate the age from a child image and predict face growth. Ramanathan and Challeppa [30] used the growth parameters defined over the landmarks to build a craniofacial growth model for child growth, and they include anthropometric evidences[9], [10] to make the model consistent with the actual data. Later on they also studied face verification problem across age progression[31]. Some others[17] studied both shape and texture in child growth.

For adult aging, both appearance and shape change. In computer graphics, people built physical models to simulate the facial changes during aging progress. For example, Wu et al. [38] built a multi-layer biomechanics skin model to simulate the wrinkles on face as age increases. Their model approximated the physics of facial skin tissue through incorporating a set of anatomically-motivated facial muscle actuators and the underlying skull structures. By considering large strain deformation and wrinkle formation, the age of skin was estimated[39]. Boissieux et al.[3] simulated the wrinkle using a model-based technique. In their method, skin was considered as a volumetric substance having layers of different materials, which was an extension to an elastic membrane model, and a finite element method was used for computing the deformation. Berg et al.[2] simulated the aging process of obituaries muscles. Other similar works include [14] and [3]. The disadvantage of these physical models is their complexity, which makes it time consuming and difficult to generate photo-realistic aging effects.

In computer vision, most aging approaches are example based. Burt and Perrett[4] studied two factors affecting age perception of adult faces—shape and color. From a dataset with face images in the 20 to 62 years old range, they generated average face image at different age groups, which are named *composite faces*. From these composite faces, they extract average

aging patterns and perform face aging by superimposing the color and texture difference onto the young face. In a similar approach, Tiddeman et al.[34], [35] studied the prototype images for different age groups. The differences between prototypes at different age groups were defined as the axis of age transformation. By introducing MRF wavelet method, they retained some missing face details (such as stubble and wrinkles) in the generation of *face prototype*. Similarly, Wang et al. [36] extracted aging pattern in PCA space instead of image space. There are also works studying the techniques that perform texture transfer from senior faces to young ones [32].

Some vision approaches have also been proposed for age estimation. Mark et al. [26] proposed that both shape and appearance can be used as the features for age classification. Based on the feature landmarks and wrinkle analysis, Kwon et al.[19], [18], [20] divided human faces coarsely into three groups: babies, young adults and senior adults. Lanitis et al. [21] represented human faces with a set of parameters and performed a series of automatic age estimation experiments using different classifiers. A quantitative evaluation of the performance indicated that computers can estimate the age rather accurately. Some researchers performed age estimation in other spaces. Geng et al. [15] learned aging patterns from image sequences, and estimated age by selecting proper patterns from them. Fu et al. [11], [12] proposed that the age manifold is a low dimensional subspace. After projecting images onto low dimensional manifold, age could be predicted through the learned manifold mapping and the regression model. These work on age classification inspired us to consider the features that are important for age perception during face aging simulation.

There are some drawbacks in the previous models. First, example based models are often limited by the database. A large training database of real aging sequences is needed for the face aging simulation. Secondly, most of the previous models did not account for high resolution features, which made them insufficient for describing the large facial variations across age groups. Consequently, these aging results lack critical details (e.g. wrinkles, pigments, et al.) for age perception. Thirdly, in the previous studies of face aging, hair features are usually not considered,

despite its large influence on the perception of face age. Fourthly, different from the other tasks, the ground truth for aging modeling is difficult to collect and appropriate performance measurement is not standardized yet. Therefore, it is necessary to build a more sophisticated model accounting for all the factors involved in face aging, and present new learning strategies to alleviate the dependence on database. A quantitative evaluation of face aging results is also needed.

B. Overview of our approach

Motivated by the aforementioned four problems, we propose a compositional dynamic model for face representation and a face aging algorithm for synthesizing aged faces. This hierarchical compositional model represents all human faces by a three-level *And-Or graph* [5], which consists of And-nodes, Or-nodes and leaf nodes (see Fig.2). The And nodes represent the decomposition, which divides a face into parts and primitives in 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, which model the diversity of face appearances at different age groups. Spatial relations and constraints are used between the nodes at the same level to ensure the validness of the configurations. By selecting the 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 model.

Based on the And-Or graph representation, we model the dynamics of face aging procedure as a first-order Markov chain on parse graphs (see Fig.4), and learn the aging patterns at each level from annotated faces in adjacent age groups. The Markov chain models both abrupt changes (hair style changes, wrinkle appears etc.) and continuous changes in the parse graph. Fig.1 illustrates the characteristics of this stochastic process and the intrinsic uncertainty associated with estimating the age of an face, preserving the identity of the subject, and confusion between

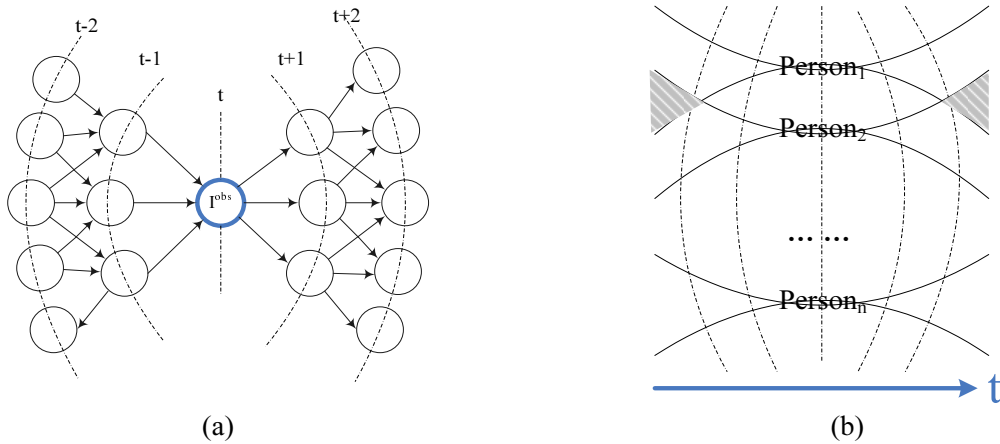


Fig. 1. The dynamic model of face aging is a Markov process. (a) is an illustration of face aging process, the blue node is the face observed at time t , while the other nodes are the stochastic status at before and after time t . The dashed curve represents the space of possible face images at each time and thus uncertainty. (b) shows that the difficulty of retaining face identity increases as time evolves. The shadowed area means that two people may become unidentifiable after a long period.

subjects over time.

For the difficulty of collecting face sequences of the same person at different ages, our compositional model decomposes face into facial components and facial zones. This part-based strategy allows learning the aging pattern of each part across age groups from similar sub-images cropped from face images of different persons. The patterns learned from the image sequences of similar persons might be different from those learned from the same person aging data, thus we need to evaluate the results with human experiments. Here we propose two criteria for aging results evaluation.

(i) *The accuracy of age prediction*, i.e. whether the synthesized faces are indeed perceived to be the intended age group.

(ii) *Preservation of the identity*, i.e. whether the synthesized faces are still recognized as the original person.

In this paper we make the following contributions:

(1) We adopt a high resolution compositional face model [40], [41] to account for the large variations of the face appearance in the aging process. This model represents all face images by

a multi-level And-Or graph and integrates three most prominent aspects of age related changes: global appearance changes in hair style and shape, deformations and aging effects of facial components, and wrinkle emergence in various facial zones.

(2) We augment the traditional face model with non-facial features – hair, since large variations emerge with time going and hair appearance influences the perception of face age in a large extent.

(3) By decomposing face into parts, we can learn dynamics across age groups from similar patches cropped from different persons. In the dynamic model, we have abrupt changes to allow for the emergence of new aging related features besides the continuous appearance changes.

(4) For face aging, people do not have many real aging sequences as ground truth, so we propose two criteria for evaluating the effect of the aging simulation and introduce some analysis approaches from psychology to validate the aging results quantitatively.

The paper is organized in the following way. In Sec.II we formulate the compositional and dynamic face aging model. The flow of the face aging simulation is presented in Sec.III. In Sec.IV, we describe the details of the implementation of the face aging simulation algorithm. Sec.V presents two human experiments and some quantitative study on the synthesized aging images. We conclude the paper with some discussion in Sec.VI.

II. REPRESENTATION AND FORMULATION

We study adult faces from 20 to 80 years old, and divide them in 5 groups $\{[20, 30), [30, 40), [40, 50), [50, 60), [60, 80]\}$. In this section we present the the And-Or graph model for face and the dynamic model for aging.

A. Compositional And-Or graph for face modeling

We extend a multi-resolution face representation proposed by Xu et al[40], [41] with aging and hair features. As Fig.2 illustrates on the left column, a face image I_t at age group t is

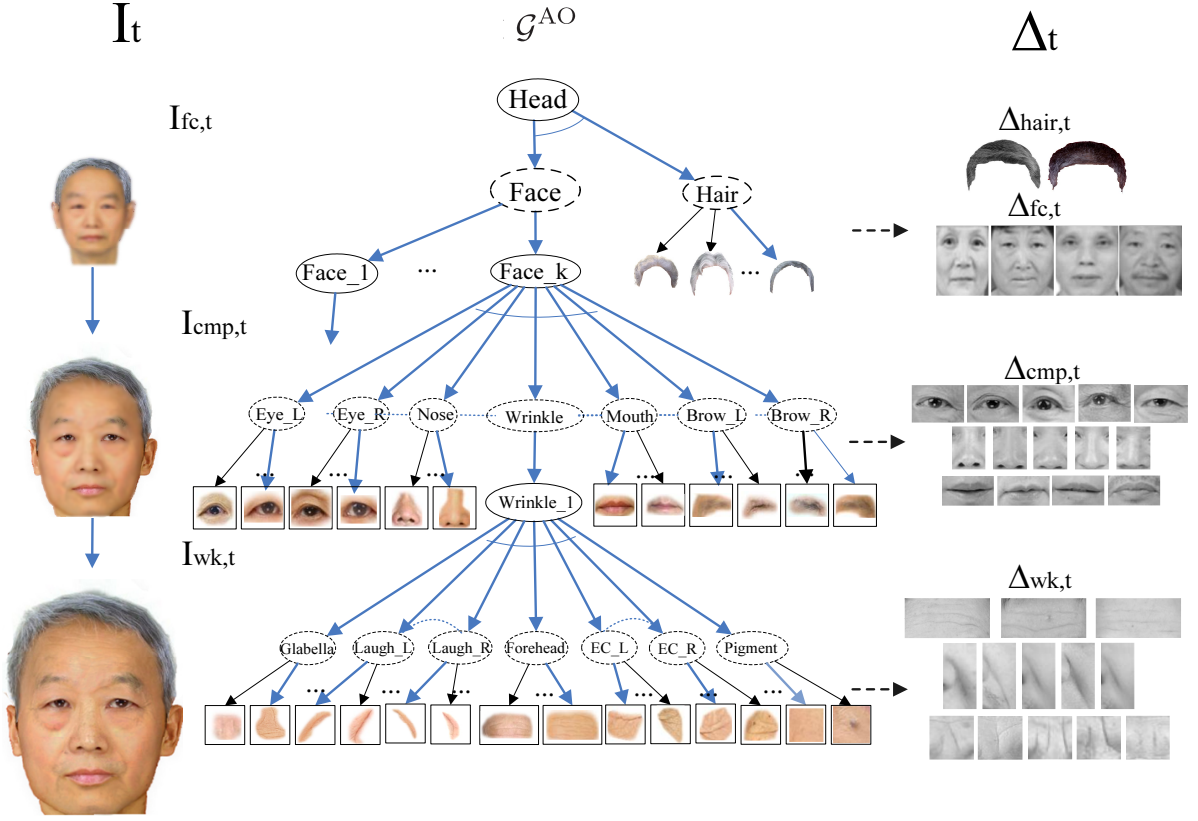


Fig. 2. A high resolution face image I_t at age group t is represented in three resolutions – $I_{fc,t}$, $I_{cmp,t}$ and $I_{wk,t}$ (left). All face images at an age group are represented collectively a hierarchic And-Or graph \mathcal{G}^{AO} (middle). The And nodes (in solid ellipses) in the graph \mathcal{G}^{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 an parse graph G_t for a specific face instance. On the right side are their corresponding dictionaries $\Delta_{fc,t}$, $\Delta_{hair,t}$, $\Delta_{cmp,t}$ and $\Delta_{wk,t}$.

represented in three resolutions, from coarse-to-fine,

$$I_t = ((I_{hair,t}, I_{fc,t}), I_{cmp,t}, I_{wk,t}). \quad (1)$$

$(I_{hair,t}, I_{fc,t})$ is the whole face image, with $I_{hair,t}$ describes hair style and $I_{fc,t}$ accounts for general face appearance. $I_{cmp,t}$ refines the facial components (eyes, eyebrows, nose, mouth et al.). $I_{wk,t}$ further refines the wrinkles, skin marks, and pigments in 9 facial skin zones. All faces are collectively represented by an And-Or graph \mathcal{G}^{AO} (see the middle column of Fig.2), where an And-node (in solid ellipse) represents the decomposition and an Or-node (in dashed ellipse) represents the alternatives accounting for large diversity of faces, for example, different eye

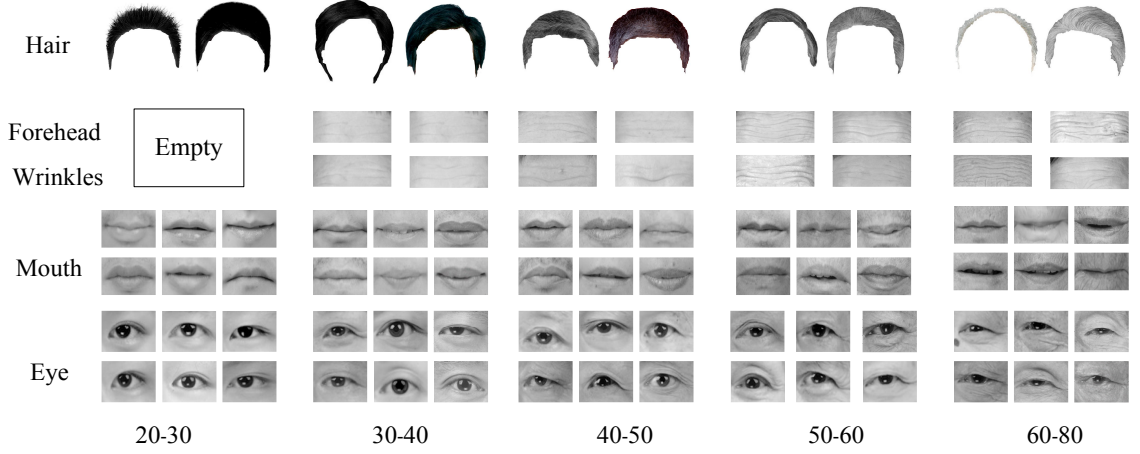


Fig. 3. Examples of the facial components and hairs from the dictionaries of different age groups.

shapes. A dictionary Δ_t for each age group t is shown on the right side for various components over the three levels.

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

where both $\Delta_{\text{hair},t}$ and $\Delta_{\text{fc},t}$ belong to the first level. This dictionary is learned from a large dataset for each individual age group separately. Fig.3 shows some examples from these dictionaries.

By choosing the alternatives at the Or-nodes, the And-Or graph \mathcal{G}^{AO} is converted to an And-graph G_t as a specific face instance at age group t . we call it *parse graph*. G_t in turn generates image I_t in a generative way [40], [41].

$$G_t \xrightarrow{\Delta_t} I_t \quad (3)$$

We denote the set of faces generated by \mathcal{G}^{AO} as

$$\Sigma = \{G_t : G_t \subset \mathcal{G}^{\text{AO}}, t = 1, 2, \dots, 5\}, \quad (4)$$

which is evidently much larger than the training set. The And-Or graph is a generative model that accounts for a large variety of faces, and a face instance is represented by

$$G_t = (w_{1,t}, w_{2,t}, w_{3,t}), \quad (5)$$

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

$$w_{i,t} = (l_{i,t}, T_{i,t}^{\text{geo}}, T_{i,t}^{\text{pht}}). \quad (6)$$

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

The prior probability for the hierarchic parse graph G_t is,

$$p(G_t; \Theta_{\text{AOG}}) = p(\omega_{1,t}; \Theta_{1,\text{AOG}})p(\omega_{2,t}|\omega_{1,t}; \Theta_{2,\text{AOG}})p(\omega_{3,t}|\omega_{2,t}; \Theta_{3,\text{AOG}}) \quad (7)$$

which accounts for the constraints of upper level to the current level as well as the constraints between nodes in the same level, e.g. enforcing the same types of eyes. $\Theta_{\text{AOG}} = (\Theta_{1,\text{AOG}}, \Theta_{2,\text{AOG}}, \Theta_{3,\text{AOG}})$ is the parameters. The above probability could further be decomposed to

$$p(\omega_{i,t}|\omega_{i-1,t}; \Theta_{i,\text{AOG}}) = p(l_{i,t}|l_{i-1,t})p(T_{i,t}^{\text{geo}}|T_{i-1,t}^{\text{geo}})p(T_{i,t}^{\text{pht}}|T_{i-1,t}^{\text{pht}}). \quad (8)$$

The likelihood model specifies how $\omega_{i,t}$ generates image $I_{i,t}$ [40], [41].

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

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

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

In [40], the parse graph was computed from an observed image by Bayesian inference from coarse to fine. By setting $w_{0,t}^* = \emptyset$,

$$w_{i,t}^* = \arg \max_{\omega} p(I_{i,t}|\omega_{i,t}; \Delta_{i,t})p(\omega_{i,t}|\omega_{i-1,t}^*; \Theta_{i,\text{AOG}}), \quad i = 1, 2, 3. \quad (11)$$

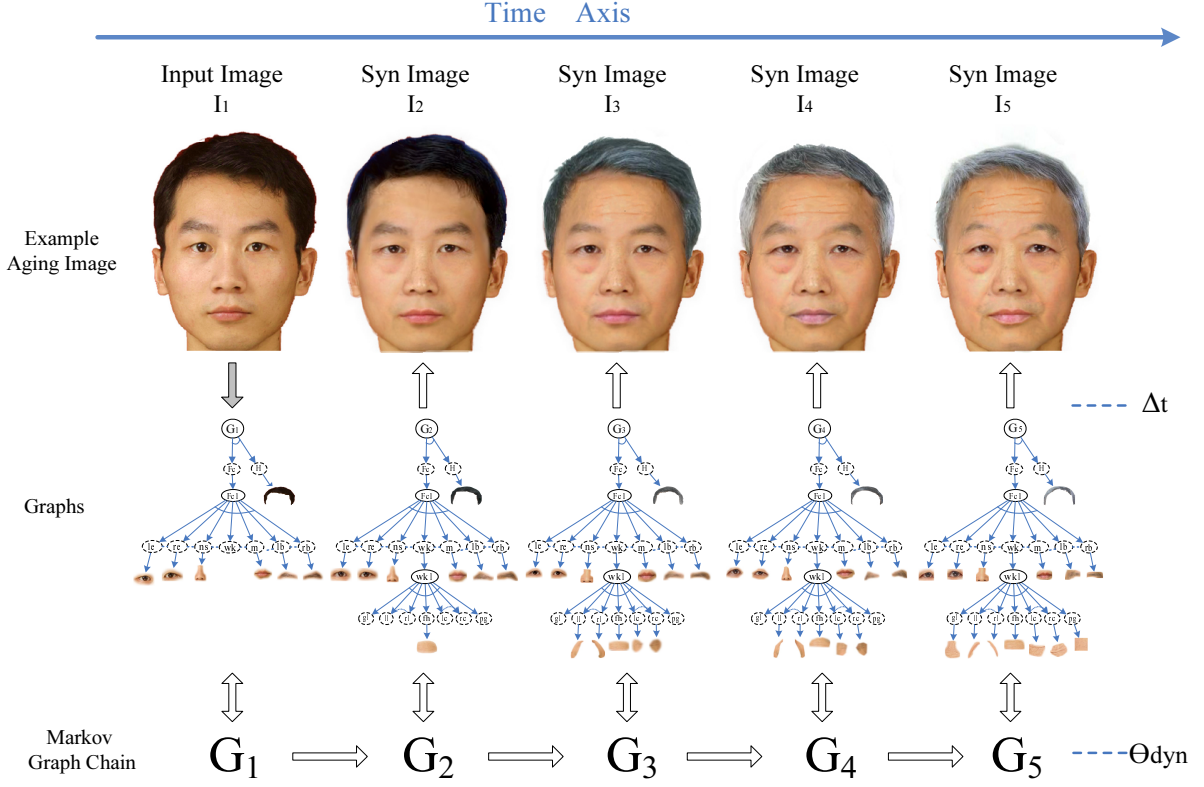


Fig. 4. An aging process can be modeled by a Markov Chain on the parse graphs G_t . The first row is a sequence of face images at different ages, the leftmost one is the input image, and the other four are synthesized aged images. The second row is the graph representations of the image sequence. Third row is the corresponding And-graphs G_t , which form a Markov Chain. Δ_t is the parameters for generating the images from G_t and Θ_{dyn} is the parameters for aging progression.

B. Modeling aging procedure as a Markov Chain on parse graphs

Based on the graph representation of human face, aging procedure is modeled as a first order Markov Chain on the parse graphs. We denote by $I[1, \tau]$ and $G[1, \tau]$ the image and graph sequences respectively for a period $[1, \tau]$. Therefore, our probabilistic model is a joint probability,

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

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

Fig.4 is an illustration of our dynamic model for graph evolution. I_1 is an input young face im-

age and G_1 its 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 . $\{G_1, G_2, \dots, G_5\}$ is a chain of parse graphs describing face aging procedure.

In the dynamic model for parse graph evolution, we model the transition probability of $l_{i,t}$, $T_{i,t}^{\text{geo}}$ and $T_{i,t}^{\text{pht}}$ separately over time t and resolutions i , and each $\omega_{i,t}$ depends on its upper level $\omega_{i-1,t}$ and previous group $\omega_{i,t-1}$.

$$p(G_t|G_{t-1}, \Theta_{\text{dyn}}, \Theta_{\text{AOG}}) = \prod_{i=1}^3 p(l_{i,t}|l_{i,t-1}, l_{i-1,t}) \cdot p(T_{i,t}^{\text{geo}}|T_{i,t-1}^{\text{geo}}, T_{i-1,t}^{\text{geo}}) \cdot p(T_{i,t}^{\text{pht}}|T_{i,t-1}^{\text{pht}}, T_{i-1,t}^{\text{pht}}), \quad (13)$$

where Θ_{dyn} is the parameters accounting for dynamics of hidden variables in the evolution of the graph chain and is learned from a large training data. In the following, we discuss the two types of variations in the Markov chain model above: (i) abrupt changes for the emergence of new features related to age; and (ii) continuous changes for the geometric and photometric attributes.

1) *Abrupt changes*. Some variations in the graph motion will change the topology of the graph. For example, inserting new nodes (e.g. wrinkles emerge, et al.) or switching the alternatives in the Or-node (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. The abrupt variation transition model is,

$$p(l_{i,t}|l_{i,t-1}, l_{i-1,t}) \propto \prod_{m=1}^{n_{i,t}^{\text{Or}}} \lambda_{i,t}(l_{i,t}(m), l_{i,t-1}(m)) \cdot p(l_{i,t}|l_{i-1,t}), \quad i = 1, 2, 3. \quad (14)$$

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

2) *Continuous changes*. Some variations in aging only change the attributes (skin color, facial part shape, wrinkle length et al.) of some leaf-nodes. We represent them by the transition

probability of $T_{i,t}^{\text{geo}}$ and $T_{i,t}^{\text{pht}}$. In Fig.4, the global skin aging is a continuous change and the changes on the nodes at the second level also belong to this type of variations.

The continuous variation transition model can be represented as follows for all the three resolutions.

$$p(T_{i,t}^{\text{geo}} | T_{i,t-1}^{\text{geo}}, T_{i-1,t}^{\text{geo}}) \propto \exp\left\{-\sum_{m=1}^{n_{i,t}^{\text{And}}} \psi(T_{i,t}^{\text{geo}}(m), T_{i,t-1}^{\text{geo}}(m))\right\} \cdot p(T_{i,t}^{\text{geo}} | T_{i-1,t}^{\text{geo}}), \quad (15)$$

$$p(T_{i,t}^{\text{pht}} | T_{i,t-1}^{\text{pht}}, T_{i-1,t}^{\text{pht}}) \propto \exp\left\{-\sum_{m=1}^{n_{i,t}^{\text{And}}} \psi(T_{i,t}^{\text{pht}}(m), T_{i,t-1}^{\text{pht}}(m))\right\} \cdot p(T_{i,t}^{\text{pht}} | T_{i-1,t}^{\text{pht}}). \quad (16)$$

In the above formula, m indexes the And-node at resolution i between two adjacent groups t and $t-1$. $T_{i,t}^{\text{geo}}(m)$ and $T_{i,t}^{\text{pht}}(m)$ denote the geometric and photometric attributes of an And-node m . $\psi(\cdot)$ is a potential which favors the transitions between similar patterns, 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 and it accounts for the geometric deformations. Although large variations may occur in real data too, for example, 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 $p(T_{i,t}^{\text{geo}} | T_{i-1,t}^{\text{geo}})$ and $p(T_{i,t}^{\text{pht}} | T_{i-1,t}^{\text{pht}})$ are from the original prior model of the parse graph in Eq. 8.

For clarity of presentation, we shall discuss implementation details in Sec.IV.

III. 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.

A. 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 below

input : A young face image I_1
output : A sequence of aged faces I_2 to I_5
1.1 begin
1.2 <i>Step 1</i>) Computing G_1 as parse graph of I_1 ;
1.3 $G_1 = \arg \max p(G_1 I_1; \Delta_1)$
1.4 <i>Step 2</i>) Sampling the graphs at the consecutive age groups from Eq.13;
1.5 Sampling $G_t \sim p(G_t G_{t-1}; \Theta_{\text{dyn}}, \Theta_{\text{AOG}}), t = 2, 3, 4, 5.$
1.6 <i>Step 3</i>) Synthesizing the aged image I_t from the generative model
1.7 $I_t = J(G_t; \Delta_t)$
1.8 end

Algorithm 1: Inference Algorithm for Face Aging Sequence

B. Details of the algorithm

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

1) *Compute 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 11. This part of work is the integration and extension of Xu et al. [40] high resolution face model and Chen et al. [6] generative hair model.

(a) *Compute the hair representation*

Following the generative model of hair in Chen et al. [6], the geometric attributes $T_{\text{hair}}^{\text{geo}}$ of hair can be represented by its sketch, which includes a set of curves C_k and corresponding direction d_k . The sketch can be computed by a sketch pursuit algorithm in [6]. The photometric

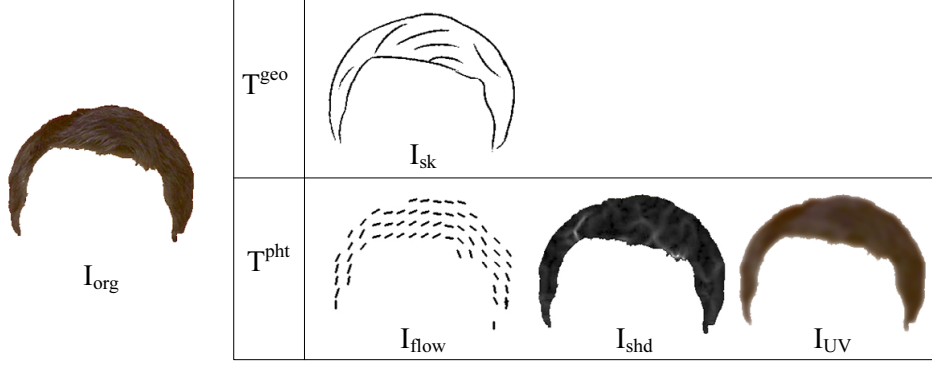


Fig. 5. For a hair image I_{org} , the geometric attributes are described by the directed curves in sketch image I_{sk} . Photometric attributes are described by three components: I_{flow} is the vector flow accounting for hair directions, I_{shd} represents the lighting and shading in the hair, and I_{UV} is the color channel of hair image.

attributes $T_{\text{hair}}^{\text{pht}}$ describing hair texture 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 the Gaussian basis simulating the lighting and shading of hair image. We refer to [6] for the details of the inference algorithms for these parameters. Based on $T_{\text{hair}}^{\text{geo}}$ and $T_{\text{hair}}^{\text{pht}}$, we classify hair into a number of styles indexed by l_{hair} .

(b) *Compute the face and facial components*

We represent the face and facial components with AAM [7] models. First we train a traditional AAM model for the first level with 90 feature points shown in Fig.6(a). For the second level, we build a number of AAM models for each facial component to refine the areas of eyebrows, eyes, nose, and mouth. Each component has several different prototypes with slightly different landmarks. For each component in the second level, l_{cmp} indexes the prototypes of the component, T_i^{geo} and T_i^{pht} are the coefficients of shape Eigenvectors and texture Eigenvectors. The variables are computed by minimizing reconstruction error.

(c) *Compute the wrinkles and pigments*

In the third level representation, we divide the face skin into 6 wrinkle zones as Fig.13(a)

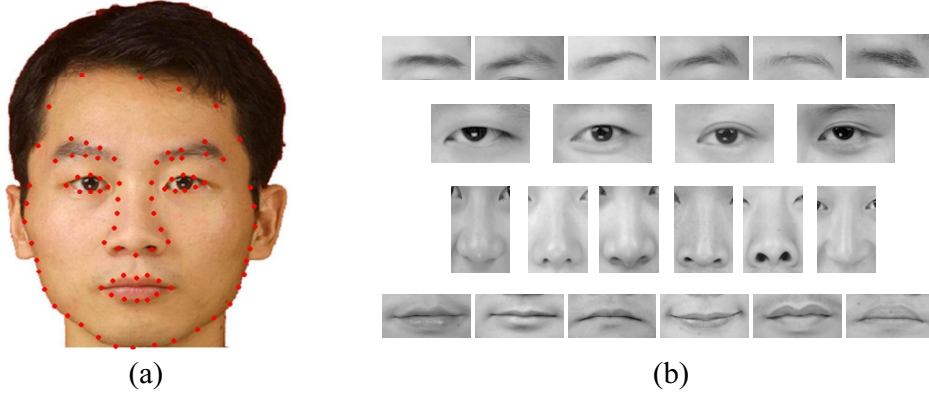


Fig. 6. (a) The 90 landmark points defined for the global AAM model. (b) Each face component has some sub-classes for different types of eyebrows, eyes, noses, and mouths. For each component type, a local AAM model is trained for precise feature points localization.

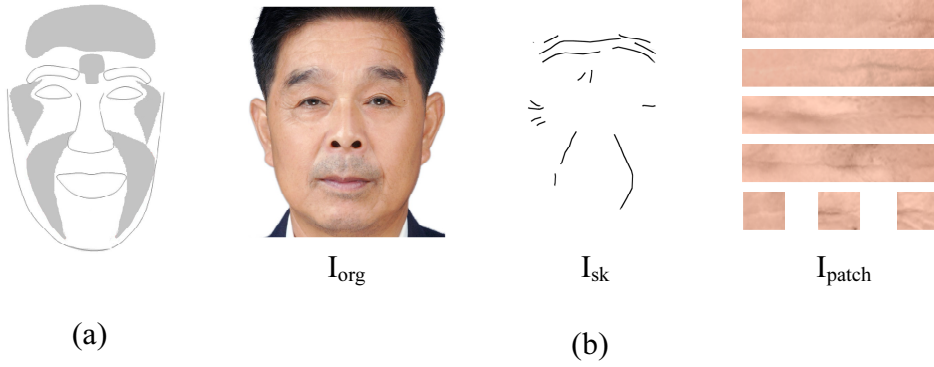


Fig. 7. (a) The skin is divided in 6 wrinkle zones. Wrinkles are added in each zone separately in our aging approach. (b) I_{org} is the input image. The curves and marks in I_{sk} and image patches I_{patch} account for the geometric and photometric attributes for wrinkles respectively.

shows. We locate the wrinkles (curve or sketch) in each zone with matching pursuit algorithm. The filters used include Gabor wavelets and blobs. The geometric variables T_{wk}^{geo} describe the position, length, orientation of the traced curves, and the position and scale of the marks. The photometric attributes T_{wk}^{pht} are represented directly by the wrinkle line images and the mark patches in Fig.13(c). Mostly there is no wrinkle for faces of age under 30, so the initial parse graph G_1 usually has only two levels.

2) *Simulating the evolution of parse graphs in aging: (a) learning the dynamic parameters*

To overcome the difficulty of collecting the photos of the same person at different age groups, our model decomposes face into parts, and learns the aging pattern for each part from similar images, which can be cropped from faces of different persons. Fig.8(a) gives a subset of the training data for eyes over three age groups.

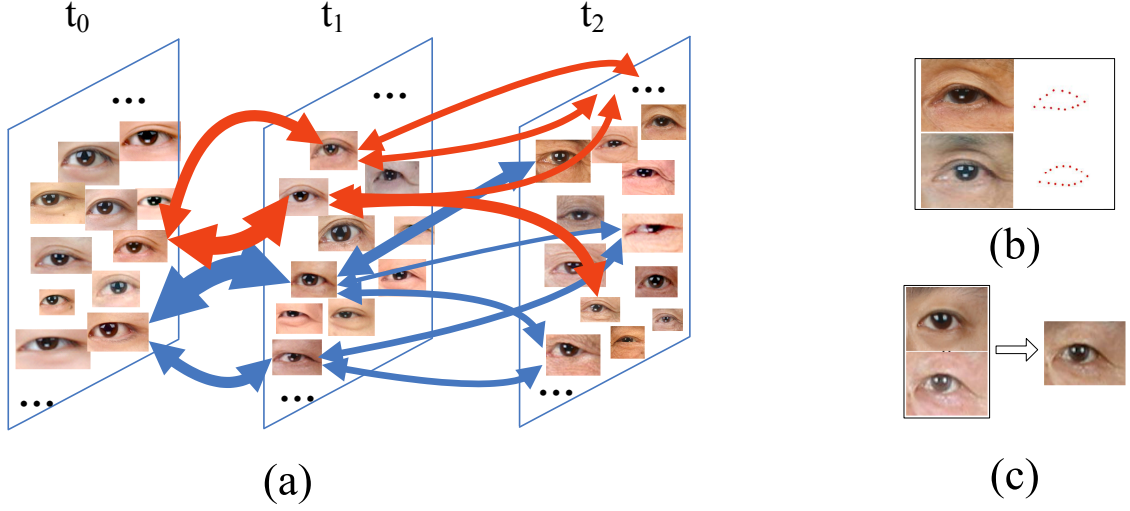


Fig. 8. Learning aging pattern from similar patches. (a) Eye examples in three age groups. Thickness of the arrows between two eye images indicates larger transition probabilities between the two images in consecutive age groups. (b) The labeled landmarks describing the contours of one pair of selected eyes from two adjacent age groups (c) One aging result of an eye.

As is described in Sec.II-B, our dynamic model is a first-order Markov Chain, since we model only the dynamics between adjacent groups. In the hierarchical representation, the dynamics is learned at three levels separately as shown in Eq.13. At each level we define energy functions measuring the similarity between an image pair, based on which the transition probabilities across two adjacent groups can be computed by using Eq.15 and Eq.16. In Fig.8(a) the thickness of the arrows reflects the transition probability.

The transition of a face component over age group is allowed only between prototype, i.e. the same landmarks. The similarity measures over the geometric and photometric attributes, i.e. $\psi()$ in eqns 15 and 16 follows the TPS model and AAM model respectively.

(b) probabilistic sampling to simulate parse graphs from the Markov chain model

For aging simulation, we use probabilistic sampling instead of maximizing the conditional

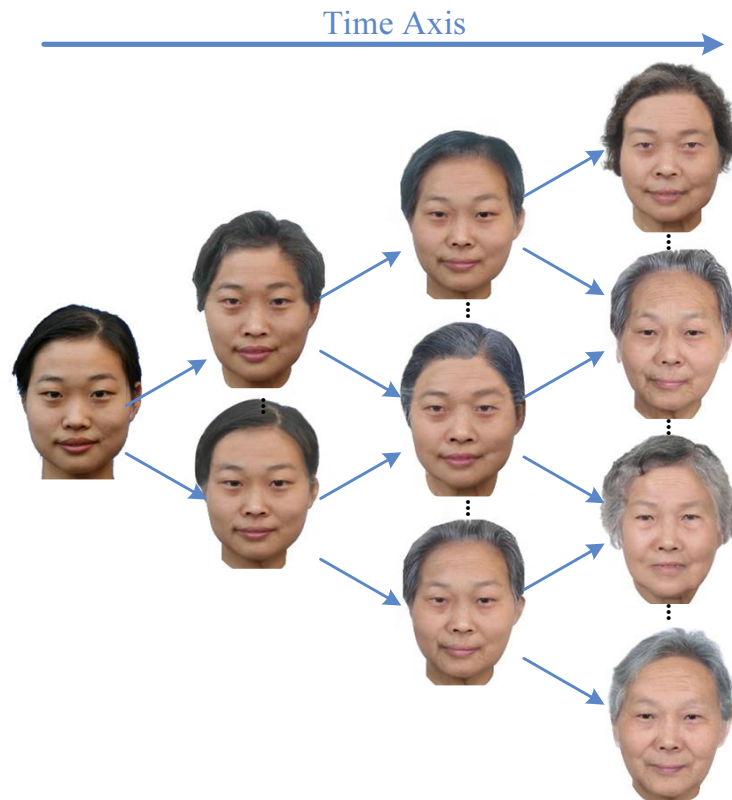


Fig. 9. The uncertainty of aging results increases with time. Give an input face image (leftmost), the algorithm simulate a series of aging results reflecting the uncertainty of face aging. The vertical column shows the plausible faces at the same age group.

probability $p(G_t|G_{t-1})$. This includes selecting the prototypes, emergence of new nodes, such as wrinkles, and their geometric and photometric attributes. Some details of these elements are referred to the next section.

The evolution of Markov chain is probabilistic. 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.9 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.

3) Synthesize image I_t from G_t : .

By the generative model, we synthesize the face image I_t from the 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 [41]. 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 [6]. 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 appearances. This step leads to higher resolution details and diverse appearance for these components. Thirdly, it generates the wrinkles and skin marks on the 6 skin zones based on $\omega_{3,t}$.

IV. IMPLEMENTATION ASPECTS

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

A. Level 1: Global Appearance Aging

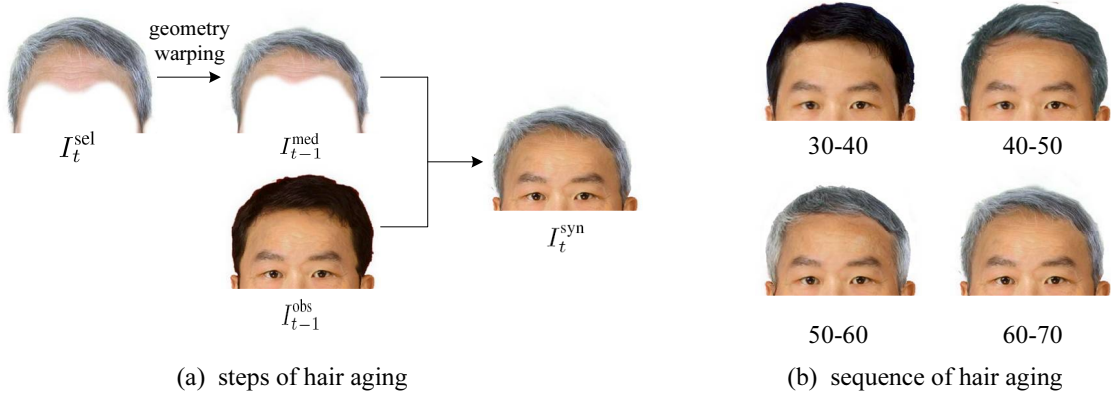
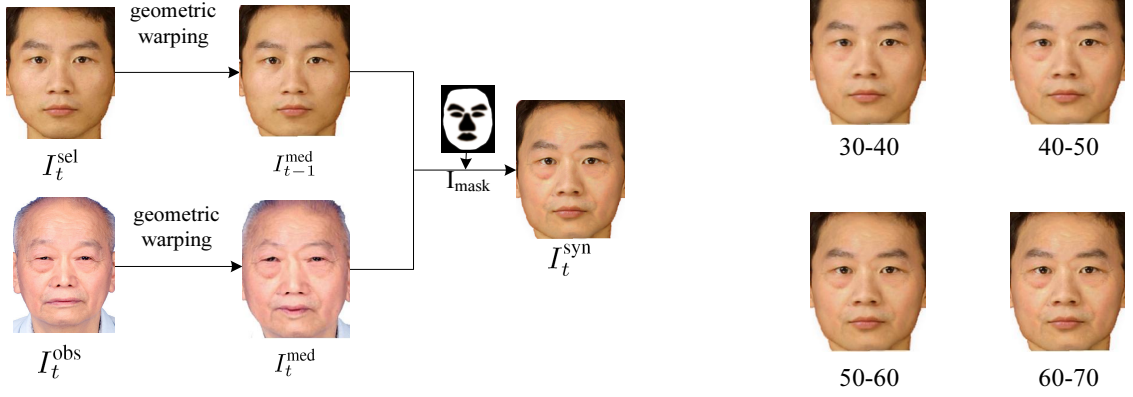


Fig. 10. Illustration of hair aging. (a) An illustration of hair aging procedure. I_t^{sel} is the selected hair image at age group t , and I_{t-1}^{med} is the intermediate results after geometric transformation under skull shape constraints. I_{t-1}^{obs} is the face image at age group $t - 1$. The final result of the hair aging is I_t^{syn} . (b) shows a final result of hair aging sequence.

1) *Hair Aging*: As we annotated 10,000 face images across the 5 age groups, a large set of hair images are collected for each group. Thus, when we need to select a hair image at group t based on the hair image at $t - 1$. The selection is based on two metrics, i.e., the geometric similarity and texture similarity. The geometric similarity between hair contours is computed



th]

(a) steps of face aging

(b) sequence of face aging

Fig. 11. Global face aging. (a) I_t^{sel} is the selected face image at age group t and I_{t-1}^{obs} is the young face. I_{t-1}^{med} and I_t^{med} are the intermediate results after applying geometric transformations with Anthropometric constraints. With a mask image excluding the facial components, we can synthesize an aged image as I_t^{syn} . (b) shows an aging sequence synthesized from the input young face I_t^{obs} in (a).

by a Thin Plate Spline model (TPS) as in shape context warping energy[1], while the texture similarity is computed by histogram of vector flow. After choosing the hair image at age group t , it is warped to fit the shape of the face. Some matting technique is used to remove artifacts along the boundary between the hair and face. Fig.10 shows an example of hair changing.

2) Face Aging:

At level 1, 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 $T_{fc,t}^{geo}$. Shape-context matching cost measuring the cost for aligning two face geometries is used as a natural shape distance. The appearance distance is computed as the Hanmin distances between histograms of corresponding filter responses(mean, variance, et al.) of two aligned faces. It is worth pointing out that the warping of face shape should be constrained by the prior knowledge of aging, e.g., some points should keep constant with age increases. Fig.11 shows the process of first level face aging.

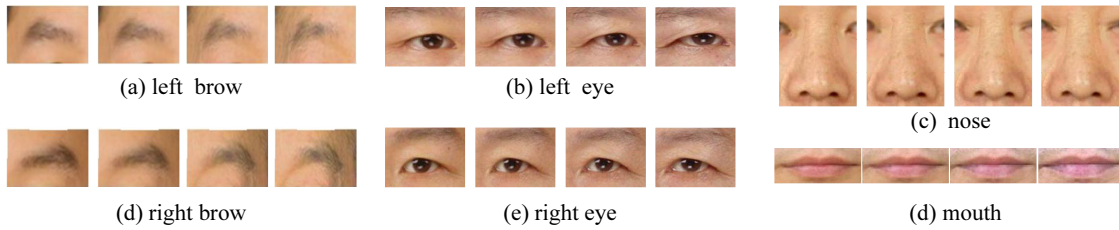


Fig. 12. The intermediate results of facial component aging.

B. 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 result, Therefore, we take the eye aging as an example to explain the component aging approach.

The resolution parameters for eye aging are learned from the dataset of eye patches across age groups, as is shown in Fig.8(a). By applying AAM searching with the local AAM models of eye, we can locate the landmarks of the components accurately as shown in Fig.8(b). Then the aging process is computed following Eq. 15,16. The geometric distance in Eq. 15 is measured by shape-context[1] bending energy between the images from adjacent age groups with the same topology, while the photometric distance in Eq. 16 is computed by summing over the squared intensity difference in the Gaussian window around the shape context matching points. Together we select eye image for aging pattern extraction of the component. After the selection of proper aging patterns, shape and texture transformations are applied to the young eye to generate aging results as shown in Fig.8 (c).

Another factor that needs to be considered in the component aging is the overall structural constraints. Symmetry of the two eyes is one of the most typical constraints, which is achieved by penalizing different aging patterns for the two eyes. The rule is also applied to the eyebrow aging.

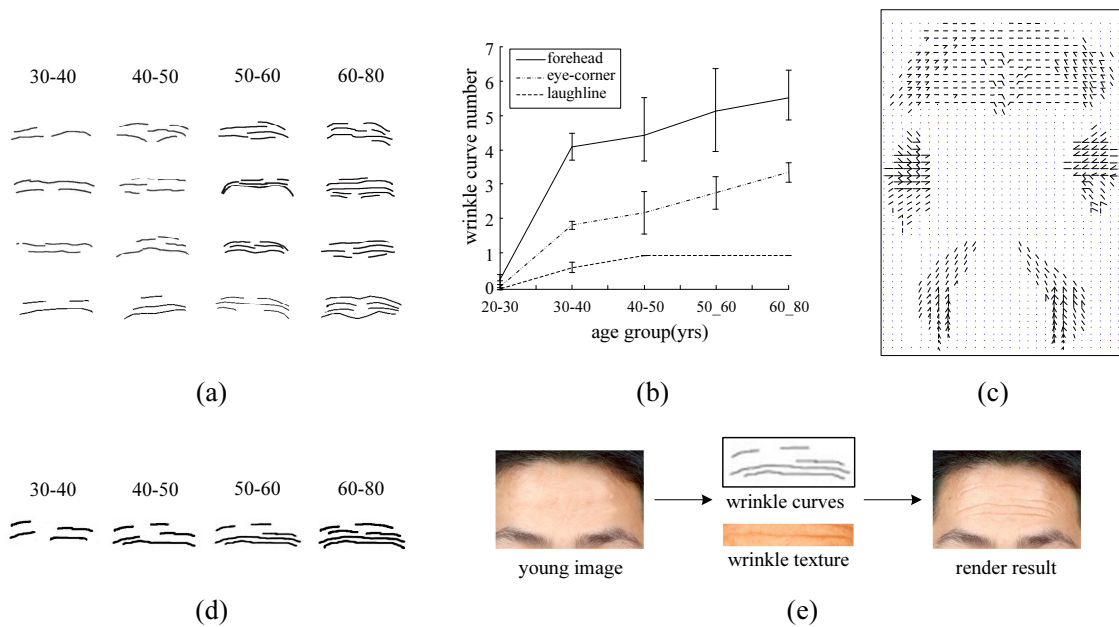


Fig. 13. Prior learning and wrinkle synthesis in different facial zones. (a) Some examples of wrinkle curves in the forehead zone. A large number of wrinkle curves are collected from the annotated dataset. (b) The frequency of wrinkle numbers in three wrinkle zones. (c) The prior distribution of wrinkle curve orientation in the six wrinkle zones, where the length of the arrow reflects the strength and the orientation describes the directions. (d) A sequence of synthesized wrinkle curves. (e) The process of rendering photo-realistic wrinkle images.

C. Wrinkle Addition

At level three, we model the aging patterns of wrinkles in different facial zones, as Fig.7 shows. Fig.13(a) shows some labeled forehead wrinkles collected from the dataset. According to the generative model, the procedure of wrinkle addition is completed in two steps: (i) Generation of curves in various wrinkle zones. The number of wrinkles and their positioning follow some prior probability densities shown in Fig.13(b) and (c). (ii) Transform the wrinkle curves to textured wrinkle image. From the annotated wrinkles, we collect the wrinkle intensity profiles perpendicular to the wrinkle curves for each age group. For a given wrinkle curve and intensity profile, the wrinkle image can be synthesized according to Eq.3. Fig.13(d) shows a series of wrinkle curves over the age groups and Fig.13(e) shows an example for generating the wrinkle image from the wrinkle curves.

1) *Learning the wrinkle statistics from labeled face data:* For wrinkle zone m , we model the number of wrinkles with Poisson distribution:

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

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.

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

in which M_t is the number of training images at age group t and $N_t^k(m)$ is the wrinkle number in zone m of the k th sample at age group t . $\lambda_t(m)$ equals to the mean value in Fig.13(b).

Similarly, we compute statistics about curve length, distance between two adjacent curves(if more than one curves are possible), et al. Prior distributions of curve position and the orientation are also inferred from the labeled data, as shown in Fig.13(c).

2) *Generating wrinkle curves:* Let $n_{3,t}$ denote the number of wrinkles at time t , the transition probability for the wrinkle number between two consecutive age groups are modeled by a bi-gram model.

$$P(n_{3,t} = k | n_{3,t-1} = j) = n_{3,t-1,t}(j, k) / N_{3,t-1}(j), \quad (19)$$

where $n_{3,t-1,t}(j, k)$ is the image number in the training dataset with $n_{3,t-1} = j$ and $n_{3,t} = k$, while $N_{3,t}(j)$ is the image number with $n_{3,t} = j$.

From the statistics of the wrinkle curves, we compute the geometric parameters of the wrinkle curves. Wrinkle number is computed by sampling from Eq.19. The other variables (length, position and orientation) can be sampled from the corresponding prior distribution. Fig.13(b)(c) show the prior information of the wrinkle geometry. With these geometric parameters we can generate a sequence of curve groups as shown in Fig.13(d).

3) *Generating realistic wrinkle images:* For the first emergence of wrinkles, we select the wrinkle line segment image randomly from the wrinkle dictionary for wrinkle addition. After warping the straight line to the shape of curves, we adopt Poisson image editing techniques to

render the curve group into realistic wrinkle image(as shown in Fig.13(e)). Because the wrinkle texture selected across age groups will not change much, we select similar wrinkle texture of the next age group based on the photometric distance. Fig.14 shows some intermediate results of wrinkle addition in different facial zones.



Fig. 14. Intermediate results of wrinkle addition and mark appearance in different facial zones at last age group.

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[28] techniques are adopted to obtain seamless fusion results. Some face aging results are displayed in Fig.16.

V. EXPERIMENTS AND EVALUATION

A. Dataset Collection

The most well-known dataset for face aging is the FG-Net aging database [43]. It includes 1,002 photos of 82 subjects in the age range between 0 and 69 and many images do not have very high resolution. We collected a database with about 50,000 images of Asian adults in the age range of [20, 80], including ID photos from a local Chinese police Bureau and ID photos at some photographers. All these face images have good resolutions with the between-eye distance being about 100 pixels. Besides this dataset, we also collected images from 20 people (friends and relatives) and each person has multiple images for the evaluation experiment. As the FG-Net dataset has not Asian faces, thus we did not use it in our study.

As life experiences affect face appearance, here we shall distinguish the appearance age from biologic age. A biologic age is the actual age of the subject while the appearance age is the perceived age. While the appearance age is often subjective and needs to be estimated through human experiments, the biologic age is not a sure thing either. In our dataset, we know the dates

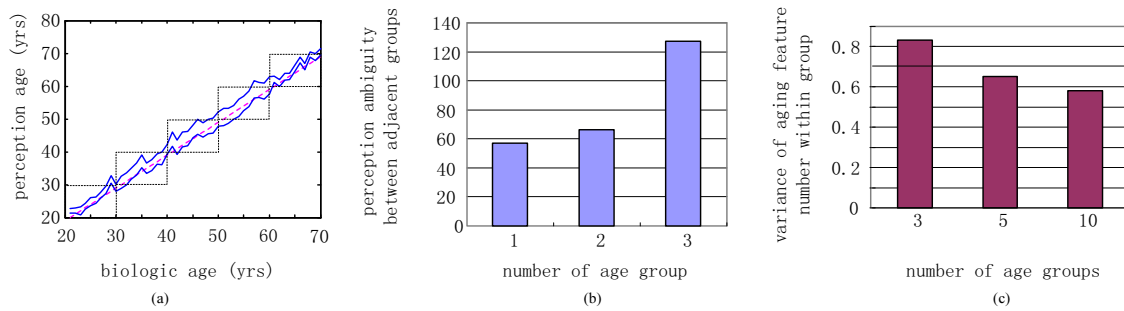


Fig. 15. Experiments for selecting the number of age groups. (a), Plot of the appearance age against the biologic age averaged over 500 faces in human experiment. The two blue lines show the upper and lower bounds of human age estimation results. (b) Vertical axis is total the age difference accumulated over 500 images in terms of total number of years between the biologic age and appearance age. The horizontal axis is the number of age group numbers. (c) The within group variance of the abrupt aging features for different age group numbers. The vertical axis is the statistical variance of certain face features.

of birth for the people in the ID photos images, and the time when a photo was taken. The latter is recorded in the time when the file was first created. We assume the mistakes of recording the date of birth and the time taking picture are rare.

In our first human experiment, we use 500 face images of different ages and asked 20 subjects (college students) to estimate the appearance age. Fig.15(a) plots the results. The two blue lines are the upper and lower-bound of the appearance 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: $[20, 30)$, $[30, 40)$, $[40, 50)$, $[50, 60)$, $[60, 80)$, based on the following reasons. (i) The difference between biologic age and appearance age is about 3 – 5 years for individuals. Thus the appearance ages between individuals in a certain age have an uncertainty interval of 6 – 10 years. (ii) As we increase the number of age groups, the perceptual errors between these groups increases (see Fig.15(b)), thus it is hard to evaluate the synthesis results. On the other hand, when the number of age groups increases, the feature variances within each group decreases, and make the model more accurate (see Fig.15(c)). As a tradeoff, we select 5 groups. (iii) The number of images withing the $[60, 80]$ group is relatively small, because less senior people took ID photos.

B. The face aging simulation experiment

We take 10,000 images from the 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 hair contour. Based on the annotation, our algorithm parses the face into parts and primitives, and then built the hierarchic dictionaries for each age group automatically. The learning of the dynamic model between age groups is discussed in Section IV.

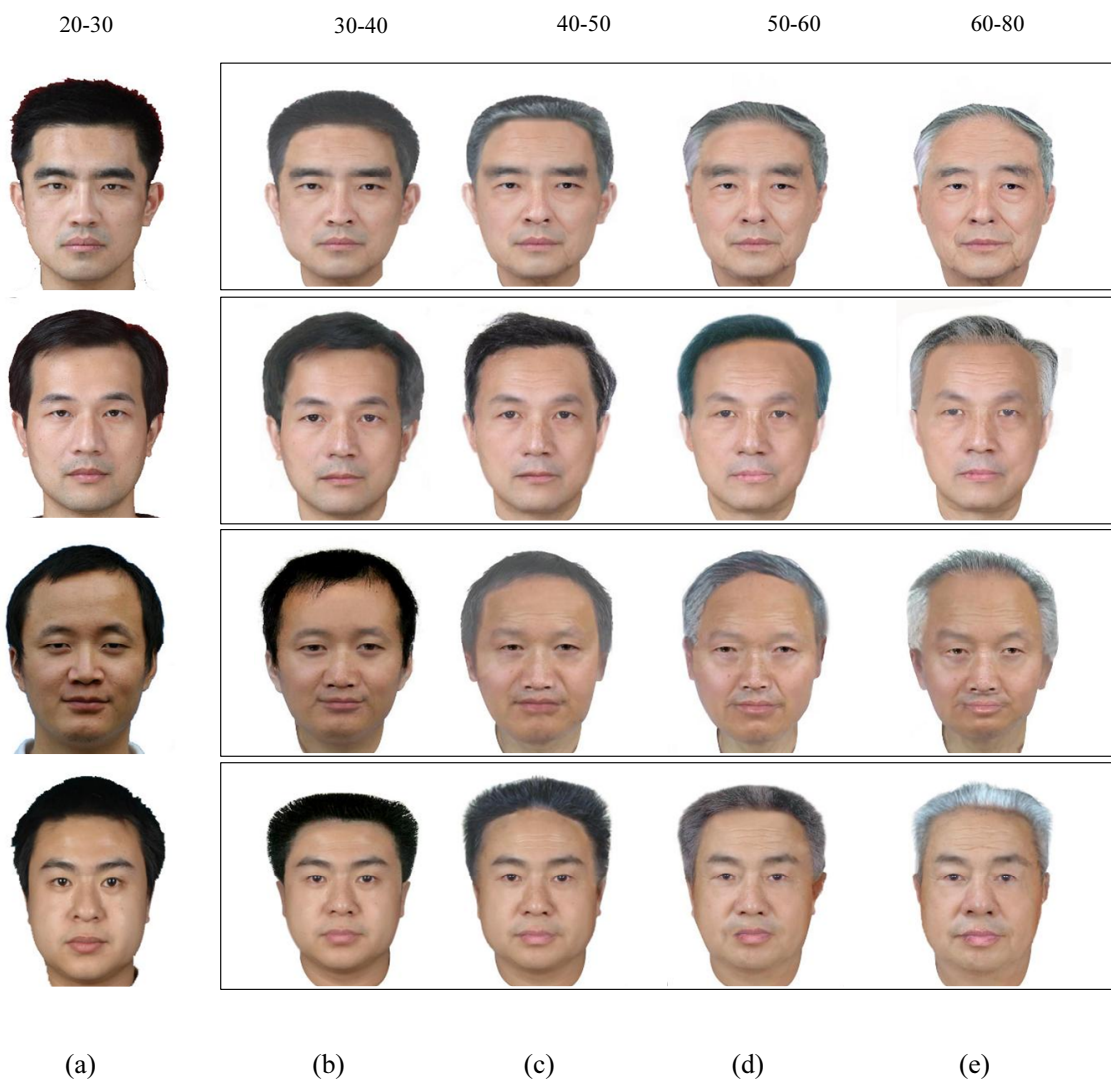


Fig. 16. Some aging simulation results. The leftmost column is the original input images of the subjects. The 2nd to 5th column are synthesized aged images at 4 consecutive age groups.

Based on the learned model, we test our inference and simulation algorithms a number of young faces in the $[20, 30)$ age range, and generate images for the other 4 age groups. Fig. 16 shows some of the aging results synthesized by our algorithms. Fig. 9 shows an example of simulating multiple plausible aging sequences for a person following the Markov chain model, as Fig.1 specifies.

Note that the people shown in Figures 16 and 1 are not in the training set, as we cannot show the ID photos for privacy reasons.

C. Background of human experiment for evaluation

In this subsection, we briefly introduce two statistical methods used in psychology experiment to quantify the human experiment results: ANalysis Of VAriance (ANOVA) and Multiple Regression Analysis (MRA). We shall use them in the next two subsections for evaluating the aging results.

i) *Analysis of variance (ANOVA)* is one of the most widely used statistical analysis approach in psychology experiment. It explores the effects of categorical factors on one or more measured quantitative variable(s) by analyzing the mean and variance across different levels of the factors. According to the number of factors included, we have One-Way ANOVA and Two-Way ANOVA. Two most important statistics for ANOVA is main effect *F-ratio* and power p .

F-ratio $F_{(a,b)}$ measures the main effect of one factor on the variable. Here a and b are the degrees of freedom of variables. with $a = A - 1$ is the number of different groups minus one and $b = A(N - 1)$, where N is the sample number in each group. F is the ratio between actual standard variation of each factor level and the expect standard variation.

Another important factor for ANOVA result is the p -value, which is the probability of accepting the hypothesis that the factor has great significance on variable by mistake. It can be used to compute the least significant difference between any of the means in terms of error probability. The typical value for p is 0.05.

ii) *Multiple Regression Analysis (MRA)* is a statistical technique that allows us to predict the score of one response variable on the basis of the scores of several other factors. The regression equation takes the following form:

$$Y = X\beta + E \quad (20)$$

In this equation, Y is the response variable, X is the data matrix containing the different factors responsible for Y . In our experiment, Y is the appearance age, and E is the prediction error.

Two most important factors are β and R . The value of β is a measure of how strongly each factor influences the response variable. R is a measure of the correlation between the observed value and the predicted value of the criterion variable, which indicates the proportion of the variance in the response variable accounted for by the model.

D. Experiment on the relative contribution of face components

Our aging approach adopts part based strategy and integrates aging features of different facial parts together. During the process, we notice that some features influence the age perception significantly while some others do not. This observation inspires us to study the relative contribution of each aging feature quantitatively to the age perception of face. The analysis results may provide further guidelines to improving the proposed method by, for instance, collecting more specialized data and refining the aging effects of the important features. The features considered in our experiment include both the internal factors (e.g., brow, eyes, nose, mouth, facial zone) 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. Eight mask images for face, hair, brow, eye, nose, mouth, forehead, and laughline respectively are used to extract these parts, as Fig.17 displays. Subjects are presented the masked images and asked to estimate the age of each part. Then we apply multivariate regression analysis to measure the contributions of each component to the perception age of the whole face. The results are shown in Table I.

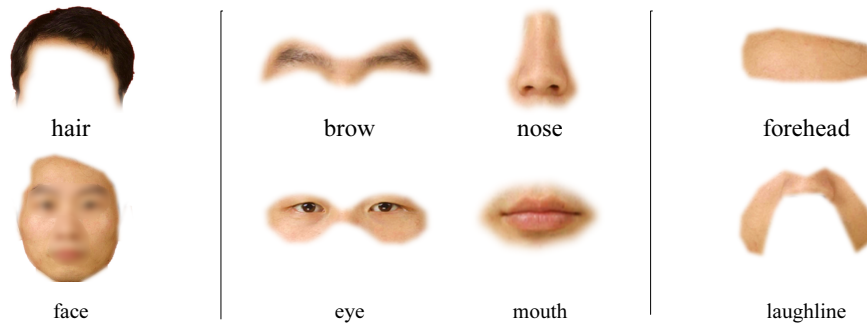


Fig. 17. Eight masks are designed to extract different parts for the experiment of relative contributions.

TABLE I

RELATIVE CONTRIBUTION OF EACH PART FOR AGE PERCEPTION

	face	eye	hair	forehead	laugh-line	mouth	nose	brow
β	0.357	0.205	0.179	0.09	0.083	0.058	0.041	0.013
R Square	0.907							

In the table, the large value of R square indicates that our model accounts for most of the age related changes. From the β value of each feature, we can see clearly that there are 5 features contribute much to the age perception of human face. The large contribution of the hair confirms the effectiveness of the hair feature in face age perception which was missing in previous studies.

E. Experiments on the aging results

To evaluate the goodness of the synthesized aging results, we use two criteria. (i) *The accuracy of age prediction*. We test whether the synthesized faces are indeed perceived to be the intended age group. (ii) *Preservation of the identity*, i.e. whether the synthesized faces are still recognized as the original person.

In this subsection, we conduct two human experiments with 20 subjects evaluating our aging results.

1. *Experiment on the accuracy of age prediction*. We compare two sets A and B in this experiment. For set A, we select 20 young faces in age group 1 [20 – 30) and synthesize one aging sequence for each person as Fig. 16 shows. Thus, set A has 80 synthesized images with 20

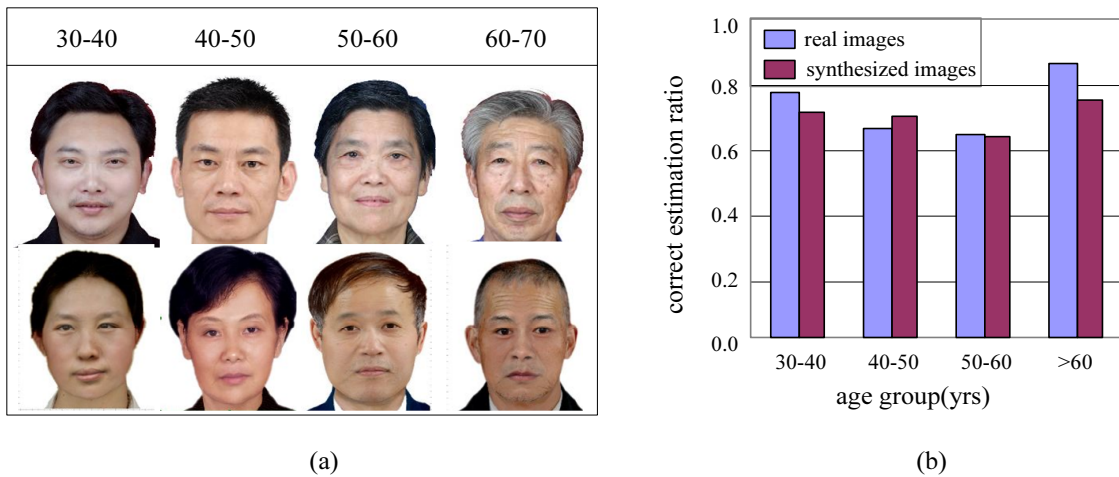


Fig. 18. The accuracy of age prediction. (a) The first row shows images from set B and the second row shows real images from set A. (b) The correct rate of age perception by human subjects.

TABLE II

ANOVA ANALYSIS OF AGE ESTIMATION RESULTS

		Sum of Squares	Degree of freedom	Mean Square	F-ratio	p
Main Effects	Combined	222.653 ^a	7	24.739	32.880	0.000
	Group	216.511	3	54.128	750.198	0.000
	Image Set	0.080	1	0.080	1.102	0.295
Interaction	Group*Image Set	1.131	3	0.283	3.918	0.004
Residule		13.709	190	0.072		
Total		236.362	200			

images in each of the age groups 2 – 5. For set B, we select randomly 20 real face images from the ID photo dataset for each of the age groups 2 – 5 respectively. We normalize the images in set B to the same resolution and grey level as the set A images.

Fig. 18(a) gives some example images for set A (2nd row) and set B (1st row). The subjects are asked to estimate the age of each face images in the two sets. Fig. 18(b) plots the human estimation error rates for sets A and B. This shows that the age estimation results of the synthesized images are consistent with those of real images.

We analyze the age estimation results of the synthesized aged face images by analysis of variance(ANOVA) and the analysis result is summarized Table. II. The large main effect of age group on age estimation($F_{4,156} = 216.511$) indicates that our model accounts for the aging related variations mostly and the small main effects($F_{1,158} = 0.080$) of image set on age estimation shows that the perception results of two sets do not differ significantly.

2. *Experiment on preservation of the identity.* We compare two data sets C and D. For set C, we use 20 real aging sequences from friends and relatives (this is different from the ID photo dataset) with images in some age groups missing. For each young face at the age group 1 in Set C, we synthesize one age sequence as Fig. 16. Thus we have 80 synthesized images and denote them as Set D. Fig. 19(a) shows some examples for set C (1st row) and set D (2nd row).

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

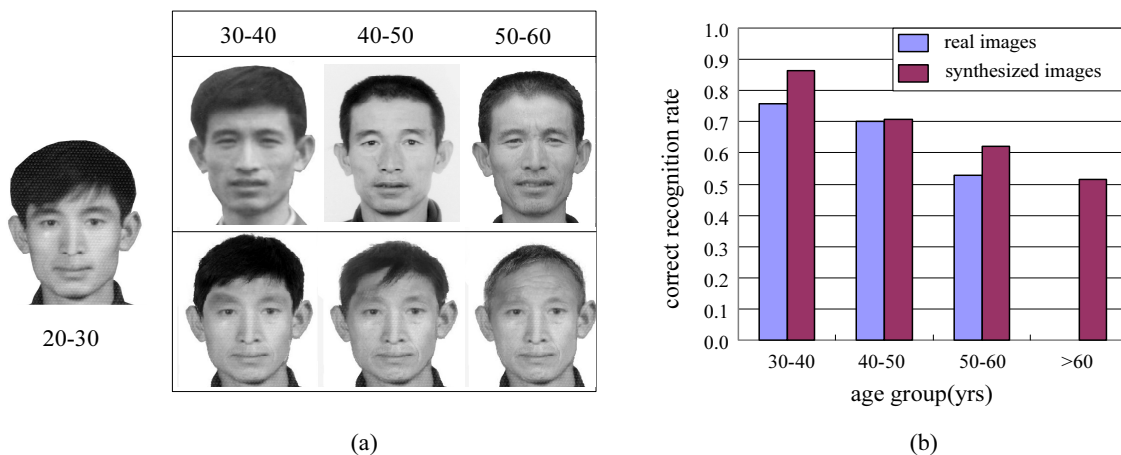


Fig. 19. Preservation of the identity. (a) The first row is real aging sequence in set C, and the second row is a sequence synthesized by our algorithm for set D. (b) The human recognition rate on sets C and D.

Fig.19(b) shows that the recognition rates of both sets in the four age groups. From the result,

TABLE III

ANOVA RESULT OF FACE RECOGNITION RESULT

		Sum of Squares	Degree of freedom	Mean Square	F-ratio	p
Main Effects	Combined	0.987 ^a	5	0.197	9.170	0.000
	Group	0.839	2	0.420	19.483	0.000
	Image Set	0.104	1	0.104	4.819	0.031
Interaction	Group*Image Set	0.044	2	0.022	1.032	0.361
Residule		1.809	84	0.022		
Total		46.469	90			

we can see that the recognition performance of two sets are quite similar, which means aging sequences synthesized by our aging approach retain the identity information of the input face very well. It has lower recognition rate for longer aging periods. The face recognition performance of the real images is a little lower than the synthesized images. This may be partially due to the other variations (such as illumination, expression, et al.) included in the real sequences.

In the same way as in Experiment 1, we apply ANOVA to the recognition results, and find that recognition rate is affected significantly by age group, while it does not differ significantly for the two image sets. Table.III shows the ANOVA analysis results.

VI. CONCLUSION

We study a compositional dynamic face model for face aging and develop algorithms for learning and simulations. Our results are evaluated through human experiments for the accuracy of age prediction and the preservation of identify. Our results are satisfactory and the good performance is attributed to two factors: (i). A large training set, and (ii) the expressive power of the compositional model, including some global appearance, such as hair color and hair style.

The current model shall be extended in the following directions. (i) Our annotated dataset is limited to Asian faces. Other ethnic groups need to be studied. (ii) The aging sequences are not

learned from real sequences due to lack of data, so the synthesized images are only plausible in visual perception. Thirdly, 3D head and muscle models should be important to modeling the deformation especially for the senior faces (60-80 years old).

ACKNOWLEDGMENT

The authors thank Zijian Xu and Feng Min for their assistance for using the hierarchic face model, and thank ChangHong Liu for his advice in designing the human experiments. The data used in this paper were provided by the Lotus Hill Annotation project [42].

REFERENCES

- [1] S. Belongie, J. Malik, J. Puzicha, "Shape matching and object recognition using shape contexts," *PAMI*, 2002, Vol. 24, No. 4, p.509-522.
- [2] A. C. Berg, "Aging of Orbicularis Muscle in Virtual Human Faces," *International Conference on Information Visualization*, 2003.
- [3] L. Boissieux et al., "Simulation of skin aging and wrinkles with cosmetics insight," *Computer Animation and Simulation*, 2000.
- [4] 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 Soc. London*, Vol. 259, 1995.
- [5] H. Chen, Z. J. Xu, et al., "Composite Templates for Cloth Modeling and Sketching," *CVPR*, 2005.
- [6] H. Chen and S. C. Zhu, "A Generative Sketch Model for Human Hair Analysis and Synthesis," *PAMI*, Vol. 28, No. 7, July 2006.
- [7] T. F. Cootes, G.J. Edwards, et al., "Active appearance models," *ECCV*, 1998.
- [8] D. DeCarlo, D. Metaxas, and M. Stone, "An anthropometric face model using variational techniques," *SIGGRAPH*, 1998.
- [9] L. G. Farkas, *Anthropometry of the Head and Face*, Raven Press, New York, 1994.
- [10] L. G. Farkas and I. R. Munro, *Anthropometric Facial Proportions in Medicine*, Charles C. Thomas, Springfield, Illinois, USA, 1987.
- [11] Y. Fu, Y. Xu, et al. "Estimating human age by manifold analysis of face pictures and regression on aging features", *ICME*, 2007.

- [12] Y. Fu and T. S. Huang "Human Age Estimation with Regression on Discriminative Aging Manifold", *Multimedia*, 2007.
- [13] M. Gandhi, "A method for automatic synthesis of aged human facial images," Master's thesis, McGill University, 2004.
- [14] A. Golovinskiy, et al., "A statistical model for synthesis of detailed facial geometry," *Graphics*, 2006.
- [15] X. Geng, et al. "Learning from Facial Aging Patterns for Automatic Age Estimation", *ICM*, 2006.
- [16] X. Geng, et al. "Automatic Age Estimation Based on Facial Aging Patterns", *PAMI*, Vol. 29, No. 12, Dec. 2007.
- [17] C. Hill, C. J. Solomon, S. J. Gibson, "Aging the Human Face-a statistically rigorous approach," *ICDP*, 2005.
- [18] W. B. Horng, C. P. Lee and C. W. Chen "Classification of Age Groups Based on Facial Features", *Tamkang Journal of Science and Engineering*, Vol. 4, No. 3, pp. 183-192, 2001.
- [19] Y. Kwon and N. Da Vitoria Lobo, "Age Classification from Facial Images", *CVIU*, 1999.
- [20] D. Kalamani and P. Balasubramanie "Age Classification using Fuzzy Lattice Neural Network", *Intelligent Systems Design and Applications*, 2006.
- [21] A. Lanitis, C. Draganova, C. Christodoulou "Comparing Different Classifiers for Automatic Age Estimation", *IEEE Trans. on Systems, man and cybernetics*, Vol. 34, No. 1, Feb. 2004.
- [22] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *PAMI*, Vol. 24, No. 4, p.442-455, 2002.
- [23] A. Lanitis, C. Dragonova and C. Christondoulou, "Comparing Different Classifiers for Automatic Age Estimation," *IEEE Trans. on Systems, Man and Cybernetics*, 2004.
- [24] F. R. Leta, et al., "Manipulating facial appearance through age parameters," *SIGGRAPH*, 1996.
- [25] L. S. Mark, J. T. Todd and R. E. Shaw, "Perception: a geometric analysis of how different styles of changes are distinguished," *Journal of Experimental Psychology Perception and Performance*, 7:855-868, 1981.
- [26] L. S. Mark, et al., "Wrinkling and head shape as coordinated sources of age level information," *Journal Perception Psychophysics*, 27(2):117-124, 1980.
- [27] E. Nelson, "Dynamical Theories of Brownian Motion," Princeton University Press, Princeton, 1967.
- [28] P. Perez, M. Gangnet, A. Blake, "Poisson image editing," *ACM Trans. on Graphics*, Vol. 22, Issue 3, 2003.
- [29] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events: implications for a theory of nonrigid shape perception," *Journal of Experimental Psychology:Human Perception and Performance*, 1975.
- [30] N. Ramanathan, R. Chellappa "Modeling age progression in young faces," *CVPR*, 2006.

- [31] N. Ramanathan and R. Chellappa, "Face verification across age progression," *CVPR*, 2005.
- [32] Y. Shan and Z. Y. Zhang, "Image based surface detail transfer," *CVPR*, 2001.
- [33] J. Suo, S.C. Zhu, S. Shan, X. Chen, "A Multi-Resolution Dynamic Model for Face Aging Simulation", *Proc. IEEE Conf. on CVPR*, 2007.
- [34] B. P. Tiddeman, D. M. Burt and D. I. Perrett, "Prototyping and transforming facial textures for perception research," *Computer Graphics and Applications*, 2001.
- [35] B. P. Tiddeman, M. Stirrat, and D. I. Perrett, "Towards realism in facial prototyping: results of a wavelet MRF method", *Proc. Theory and Practice of Computer Graphics*, 2006.
- [36] J. Y. Wang, Y. Shang, et al., "Age simulation for face recognition," *ICPR*, 2006.
- [37] Y. Z. Wang and S. C. Zhu, "Modeling complex motion by tracking and editing hidden Markov graphs," *CVPR*, 2004.
- [38] Y. Wu, et al. "A Dynamic Wrinkle Model in Facial Animation and Skin Aging," *Journal of Visualization and Computer Animation*, 6:195-205, 1995.
- [39] Y. Wu, et al. "Skin Aging Estimation by Facial Simulation," *Proc. of Computer Animation*, 1999.
- [40] Z. J. Xu, H. Chen, and S. C. Zhu, "A high resolution grammatical model for face representation and sketching," *CVPR*, 2005.
- [41] Z. J. Xu, et al. "A composite template for human face modeling and sketch," *IEEE Trans. on PAMI*, To appear in 2008.
- [42] Z. Y. Yao, X. Yang, and S. C. Zhu, "Introduction to a Large-Scale General Purpose Ground Truth Database: Methodology, Annotation Tool and Benchmarks," *EMMCVPR*, 2007.
- [43] "Face and Gesture Recognition," Network: FG-NET aging database. [online] Available:(<http://sting.cycollege.ac.cy/alanitis/fgnetaging/>).