

# High-Resolution Face Fusion for Gender Conversion

Jinli Suo, Liang Lin, Shiguang Shan, *Member, IEEE*, Xilin Chen, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

**Abstract**—This paper presents an integrated face image fusion framework, which combines a hierarchical compositional paradigm with seamless image-editing techniques, for gender conversion. In our framework a high-resolution face is represented by a probabilistic graphical model that decomposes a human face into several parts (facial components) constrained by explicit spatial configurations (relationships). Benefiting from this representation, the proposed fusion strategy is able to largely preserve the face identity of each facial component while applying gender transformation. Given a face image, the basic idea is to select reference facial components from the opposite-gender group as templates and transform the appearance of the given image toward the selected facial components. Our fusion approach decomposes a face image into two parts—sketchable and nonsketchable ones. For the sketchable regions (e.g., the contours of facial components and wrinkle lines, etc.), we use a graph-matching algorithm to find the best templates and transform the structure (shape), while for the nonsketchable regions (e.g., the texture area of facial components, skin, etc.), we learn active appearance models and transform the texture attributes in the corresponding principal component analysis space. Both objective and subjective quantitative evaluation results on 200 Asian frontal-face images selected from the public Lotus Hill Image database show that the proposed approach is able to give plausible gender conversion results.

**Index Terms**—And-Or graph, face fusion, gender conversion.

## I. INTRODUCTION

FACE image fusion is attracting increasing attention from both computer vision and graphics due to its many interesting applications, such as psychological experiment, forensics, digital makeup, face image editing, etc. [29]. The central objective of face image fusion is to integrate information from multiple face images to achieve task-oriented visual results.

In this paper, we propose an automatic gender conversion approach that is able to convert any given face to the opposite gender visibly and preserve its face identity subjectively. Fig. 1



Fig. 1. Two examples of gender conversion generated by our approach.

shows two examples. This approach can be used in many interesting applications, e.g., creating transsexual makeup effects in filmmaking, looking for lost opposite-gender siblings, etc.

### A. Related Work

To the best of our knowledge, the only work on gender transformation is the prototyping-based approach [36], which computes the average faces (named prototypes) of two gender groups to describe the typical characteristics of males and females, respectively, and defines the difference between them as the gender transformation axis. Considering that some details that are crucial for gender perception are smoothed out during prototype computation, Tiddeman *et al.* [43] modify the amplitude of multiscale edges (via wavelets) in the prototype and produce some improvements. Later, they make further improvements in [44], where an input image is transformed first into a wavelet domain as in [43], the conditional probability densities of facial parameters in high-frequency bands are then altered toward that of the target gender group. In spite of these efforts, the prototyping approach is still not sufficient for high-resolution gender conversion. Another disadvantage of the prototyping-based method is that the complex intergender discrepancies are simplified as the difference between prototypes in coarse scale and divergence between wavelet coefficient distributions in fine-scale facial texture. To compensate for these problems, the proposed framework integrates analysis approaches for gender classification in computer vision and image fusion techniques in computer graphics. Here, we briefly review the related work in the two fields.

1) *Gender Perception and Classification*: As one of the most important issues of face perception, gender classification is widely studied in computer vision and has a wide application foreground, such as preclassification in face recognition [49] on large databases (e.g., CAS-PEAL database [10] and others), development of gender-specific human-computer interfaces [30], etc.

As described in empirical surveys on gender classification [25], [26], a large number of classification approaches vary in two aspects: the proposed features and the adopted classifiers. The various features used by researchers include intensity,

Manuscript received November 14, 2007; revised January 25, 2009 and July 17, 2009; accepted December 19, 2009. Date of publication August 30, 2010; date of current version January 19, 2011. This work was supported by the National Natural Science Foundation of China under Grants 60970156 and 60728203, by the National High-Technology Research and Development Program of China (863 Program) under Grant 2007AA01Z340, and by the National Program on Key Basic Research Projects (973 Program) under Grant 2009CB320902. This paper was recommended by Associate Editor L. Gunderson.

J. Suo is with the Graduate University of the Chinese Academy of Sciences, Beijing 100190, China (e-mail: jlsuo@jdl.ac.cn).

L. Lin is with the School of Software, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: linliang@ieee.org).

S. Shan and X. Chen are with the Key Laboratory of Intelligent Information Processing and the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: sgshan@ict.ac.cn; xlchen@ict.ac.cn).

W. Gao is with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCA.2010.2064304

TABLE I  
SUMMARY OF PREVIOUS WORK ON GENDER CLASSIFICATION

Reference	Features	Classifier
Wiskott95 [48]	graph and Gabor wavelet	Bayesian classifier
Fellous97 [9]	geometric metrics	rules from metric difference between males and females
Gutta98 [14]	image intensities	RBF and deductive tree
Gutta00 [13]	image intensities	integrating RBF, deductive tree and SVM
Graf02 [11]	PCA or LLE	SVM
Moghaddam02 [28]	image intensities	SVM
Lian05 [21]		
Leng08 [18]	Gabor wavelet	SVM
Ueki04 [45]	external metrics	Bayesian classifier
Costen04 [8]	AAM	exploratory basis pursuit classification and sparse kernel classifier
Kim06 [17]	image intensities	Gaussian process classifiers
Samal06 [38]	measurements from human faces	stepwise discriminant analysis and PCA
Baluja07 [1]	binary features computed by pixel comparison	Adaboost
Lu08 [24]	pixel-pattern-based textural feature	SVM

active appearance model (AAM) [7] parameters, nonfacial features [16], iris [42], etc. Sun *et al.* [40] attempt to select distinctive features for gender classification. Some researchers also focus on exploring different classifiers for gender classification, such as support vector machines (SVMs), neural networks, adaboost, etc. Table I summarizes the large variety of approaches proposed for gender classification.

The aforementioned methods mostly utilize a global or a flat-layer face model. Differently, Wiskott and Fellous [48] represent a human face with a graph that encodes the face shape and local texture (named jet) separately. The final classification result is obtained by combining the proposals from the subclassifiers built on different jets.

2) *Image Fusion Techniques in Graphics*: Image fusion is a well-studied field in computer graphics. It seeks to integrate information from multiple images into one single image and has been widely used in remote sensing images, medical imagery, entertainment, and other applications. A lot of fusion approaches have been proposed for seamless image editing. Principal component analysis (PCA) methods [3], [4] and intensity–hue–saturation transformation methods are two simple single-resolution approaches. Motivated by the human visual system’s property of being sensitive to local contrast, pyramid-based schemes are proposed, such as wavelet method [6], [15], [20], [31], gradient method [33], etc. Some others extend the pixel method to regions [19], [27], [34]. Poisson image editing [32] is a popular method due to its good editing results and easy implementation. Wen *et al.* [47] extend the Poisson image-editing approach to fuse a low-spatial-resolution multispectral image with a high-spatial-resolution panchromatic image together.

In sum, the statistical approaches and large data sets in computer vision provide a basis for studying gender discrepancies, while the fusion methods in graphics promise seamless face image fusion results. Thus, we take the advantages from both

fields to build a gender conversion framework. The diagram of the proposed framework is shown in Fig. 2, in which the upper part displays the process of learning the probabilistic transition model from a large data set, and the lower part illustrates the flow of generating fusion results using the learned model and graphic techniques.

## B. Method Overview

Among the literature on gender classification, the AAM model is a popular face representation due to its simplicity and effectiveness. Whereas according to the viewpoint of Xu *et al.* [50], the original AAM model is not able to capture the large variance of high-resolution human faces, so we adopt the stochastic graph model first proposed by Zhu and Mumford [52] in our framework. In the proposed hierarchical representation, a face is represented by a multilayer graph, and multiple facial parts are described with graph nodes at different layers. Simultaneously, a dictionary of typical templates is built for each node. From a large database including face images of both genders, we learn a transition matrix to perform gender conversion for the given faces.

Following the primal sketch model [12], we decompose a face image into two parts: sketchable and nonsketchable ones. The former encodes the facial structure, such as the contours of facial components, wrinkles, etc., and the latter describes the facial texture, such as the inner regions of facial components, skin, etc. For the sketchable part, we use the graph-matching algorithm to select a proper template from the opposite gender and transform the profile along the graph edges using the Poisson image-editing approach [32]. For the nonsketchable part, we model first the texture of each part with AAM models, learn the distribution of AAM parameters in two gender groups, and then transform image parameters toward the distribution of the opposite gender.

The evaluation of gender conversion results is a nontrivial task for the following two reasons. First, there is no ground truth for gender conversion (since they do not exist). Second, the previously proposed evaluation measurement for fusion methods [35], [37], [46] evaluates mainly the visual results of synthetic images, and is insufficient for task-specific fusion tasks. In this paper, we propose three evaluation criteria for the generated gender conversion results and then conduct both subjective and objective evaluation experiments to validate our fusion framework.

The key contributions of this paper are summarized as follows.

- 1) The statistical methods in computer vision and the seamless image-editing methods in computer graphics are integrated for high-resolution face fusion, which is applied to gender conversion.
- 2) A hierarchical fusion framework is proposed. The algorithm developed under this framework selects gender-discriminative features in different facial regions and performs gender transformation on them separately. In addition, global constraints are imposed on these local transformations to ensure the consistency among facial regions.

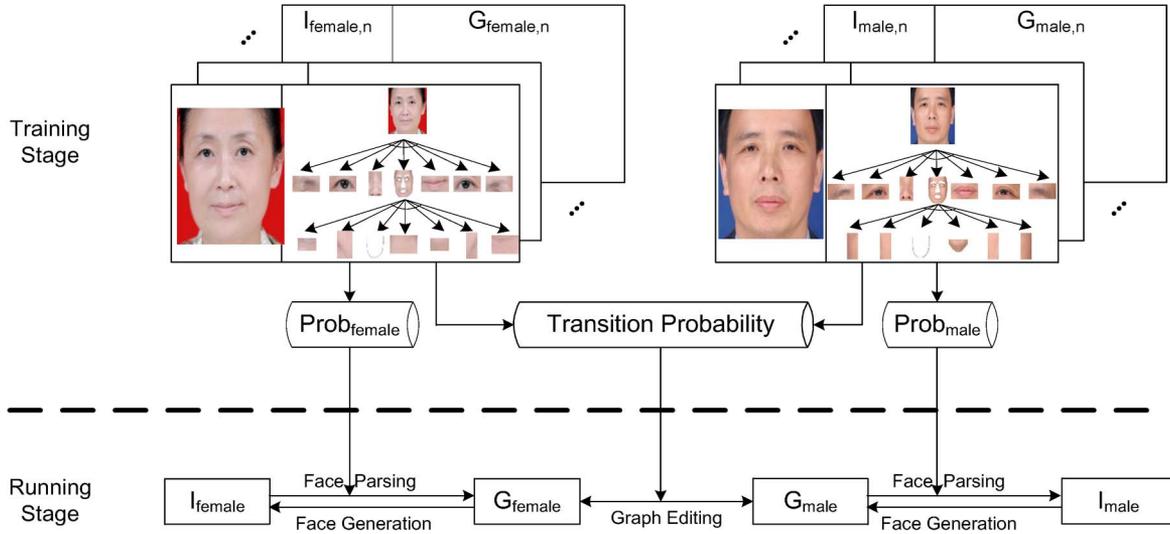


Fig. 2. Proposed gender conversion framework. During the training stage, a large set of face images  $I$ 's of each gender group are parsed into their graph representation  $G$ 's, and then, a union probability  $Prob$  of graph parameters is learned for each gender, including the probability of each facial part and the relationships among them. Based on the predefined distances, we learn a transition probability between two gender groups from a large set of opposite-gender pairs. In the running stage, for an input image  $I_m$ , we compute first its graph representation  $G_m$  and then compute its corresponding graph in the opposite-gender group  $G_n$ , which generates the final gender converted image  $I_n$ .

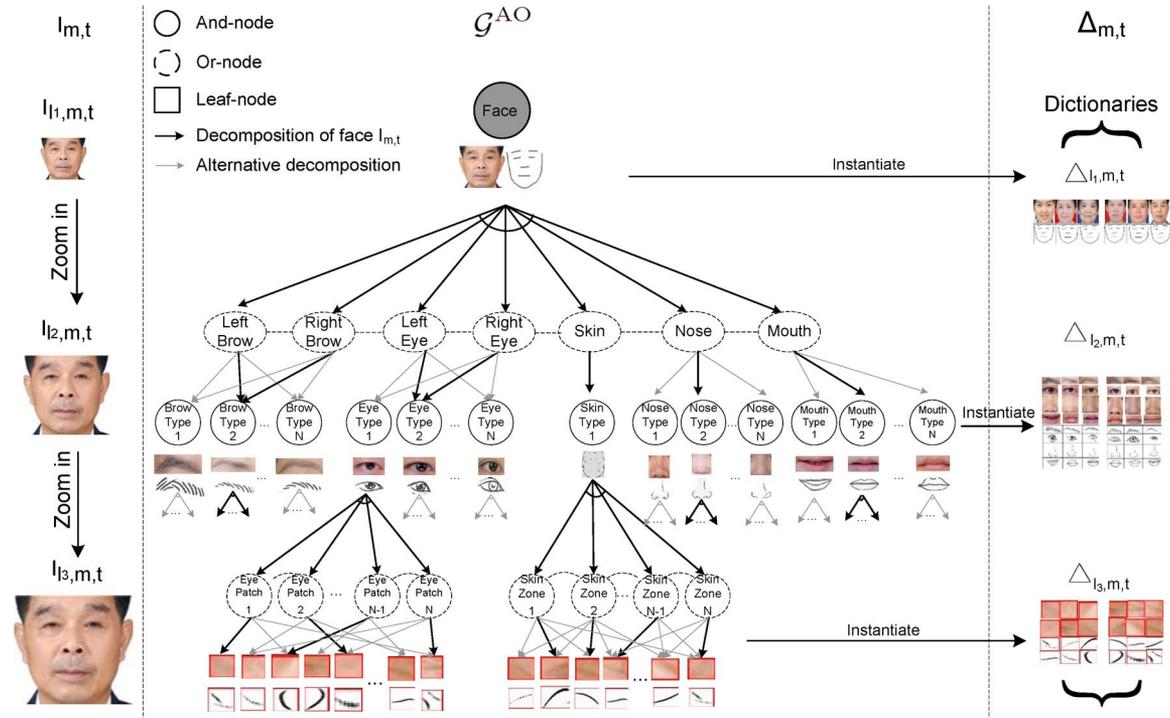


Fig. 3. Adopted hierarchical face representation. (Leftmost column) A high-resolution face image of gender  $m$  at age group  $t$   $I_{m,t}$  is represented in three resolutions— $I_{1,m,t}$ ,  $I_{2,m,t}$ , and  $I_{3,m,t}$ . (Second column) All the face images are represented collectively by a hierarchic And-Or graph  $G^{AO}$ , which generates specific face instances together with a dictionary (third column)  $\Delta_{m,t}$  of corresponding gender and age group.

3) Three novel criteria are proposed to evaluate the gender conversion results, and a series of quantitative experiments is designed to validate the proposed framework.

Section V, we conduct a series of experiments to validate the proposed approach. This paper concludes in Section VI with some discussions on the proposed approach and future work.

The rest of this paper is organized as follows. In Section II-A, we explain the adopted face model—the And-Or graph. Section II-B gives the fusion strategy based on our graph representation and the large number of templates. The detailed computation aspects are described in Sections III and IV. In

## II. REPRESENTATION

In this section, we explain the adopted face representation and the probabilistic transition model for gender conversion.

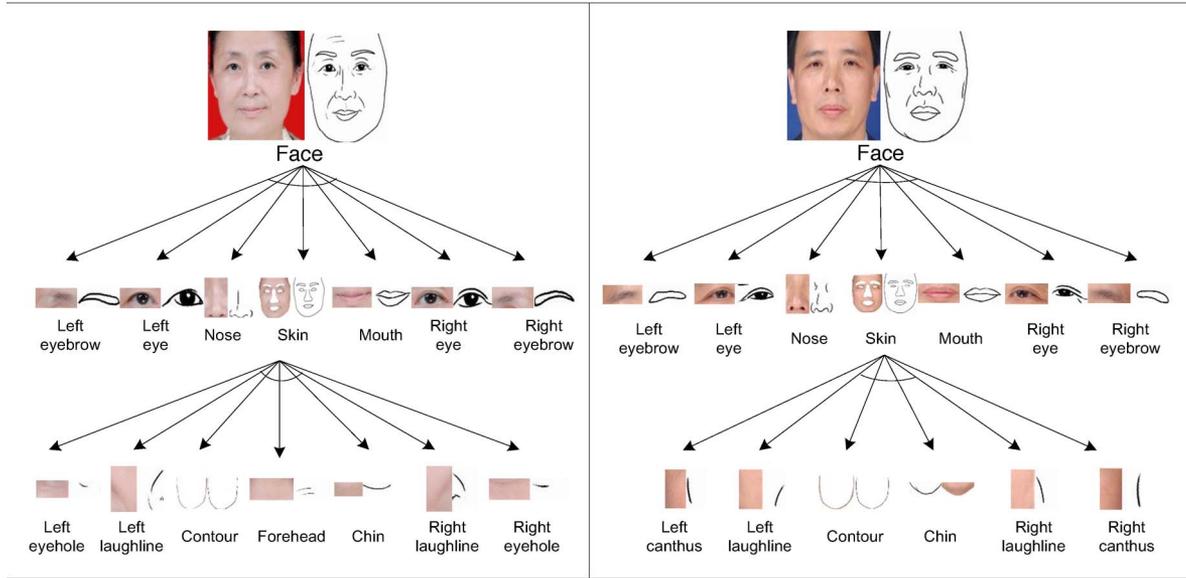


Fig. 4. Three-layer parse graphs of two individuals of opposite genders.

### A. Image Model

Our image model follows the face representation of Xu *et al.* [50] and the decomposition rule of age range in the work of Suo *et al.* on face aging simulation [41]. Supposing that  $I_{m,t}$  denotes a face image of gender  $m$  and at age  $t$ , as shown in the leftmost column of Fig. 3, it is decomposed into parts at three coarse-to-fine layers and is formally formulated as follows:

$$I_{m,t} = (I_{l_1,m,t}, I_{l_2,m,t}, I_{l_3,m,t}). \quad (1)$$

Here,  $I_{l_1,m,t}$  is the low-resolution face image accounting for the general face shape, skin texture, etc.  $I_{l_2,m,t}$  describes the appearances of facial components (eyes, eyebrows, nose, mouth, etc.), whose details are not well preserved in the global representation layer  $I_{l_1,m,t}$ .  $I_{l_3,m,t}$  further refines the details in different skin zones, including the wrinkles, skin marks, and pigments.

In our high-resolution face model, all human faces are collectively represented with an And-Or graph  $\mathcal{G}^{\text{AO}}$  [5], [22] (see the second column of Fig. 3), where an And node (in solid ellipse) represents decomposition and an Or node (in dashed ellipse) represents the candidate alternatives. After the selection of all the Or nodes, the And-Or graph turns into a *parse graph*  $G_{m,t}$  representing a specific face image  $I_{m,t}$ . Fig. 4 shows two exemplar parse graphs.

The And-Or graph is supported by a dictionary  $\Delta_{m,t} = \{\Delta_{l_1,m,t}, \Delta_{l_2,m,t}, \Delta_{l_3,m,t}\}$ , which is composed of typical templates of graph nodes at different layers, as shown in the rightmost column of Fig. 3.  $G_{m,t}$  generates an image  $I_{m,t}$  together with the dictionary  $\Delta_{m,t}$

$$G_{m,t} \xrightarrow{\Delta_{m,t}} I_{m,t}. \quad (2)$$

The generation is dominated by three hidden variables

$$G_{m,t} = \{w_{l_1,m,t}, w_{l_2,m,t}, w_{l_3,m,t}\}. \quad (3)$$

Each graph node contains different attributes, including topological information  $T^{\text{top}}$ , geometric information  $T^{\text{geo}}$ , and photometric information  $T^{\text{pht}}$ . The topological attributes account for the large variations in facial component structures; the geometric attributes include a set of affine transformations  $A = (Sx; Sy; \theta; k)$ , where  $Sx$ ,  $Sy$ ,  $\theta$ , and  $k$  denote horizontal scaling, vertical scaling, rotation, and shearing, respectively; and the photometric attributes  $T^{\text{pht}}$ , a vector including the types of intensity/color profiles and their contrasts, describe the texture information of the nodes in the parse graph. These variables produce together a rich set of human faces. All these attributes are described uniformly by the hidden variable  $w$

$$w_{l_i,m,t} = \{T_{l_i,m,t}^{\text{top}}, T_{l_i,m,t}^{\text{geo}}, T_{l_i,m,t}^{\text{pht}}\}, \quad i = 1, 2, 3. \quad (4)$$

Overall, the probabilistic face model is as follows:

$$p(I_{m,t}|G_{m,t}) = \prod_{i=1}^3 p(I_{l_i,m,t}|w_{l_i,m,t}, \Delta_{l_i,m,t}) \quad (5)$$

Here, we use  $i$  as the index of image resolution.

### B. Transition Model

The proposed framework transforms the gender attributes of a given face by altering the node attributes of its parse graph. To minimize the alteration and preserve the face identity effectively, we transform the attributes toward a similar template of the opposite gender but in the same age group.

In order to select a proper template, we define topological distance  $D^{\text{top}}$ , geometric distance  $D^{\text{geo}}$ , and photometric distance  $D^{\text{pht}}$  between  $I_{m,t_1}$  and  $I_{n,t_2}$ . For the nonsketchable regions, we use the Euclidean distance in PCA space. The detailed form of distances for the sketchable regions is explained in Section III-B.

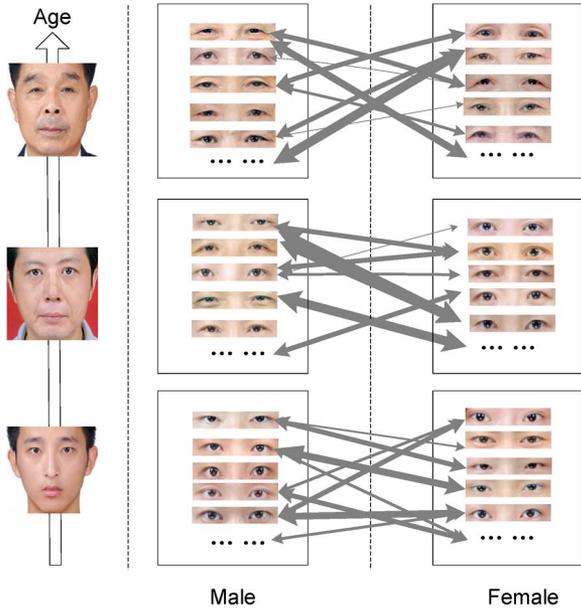


Fig. 5. Parameter learning scheme for the proposed gender conversion algorithm. The second and third columns give a subset of eye dictionaries of male and female groups, respectively. Our gender conversion algorithm transforms the attributes toward that of the opposite gender but the same age group. The arrows between two images reflect the transition probability, which is denoted by the thickness of the arrows.

For identity preservation, we favor similar image pairs and penalize large alternations in transformation. The probabilistic transition model is defined as follows:

$$p(T_{m,t_1}^{\text{top}} | T_{n,t_2}^{\text{top}}) \propto \exp(-\mathcal{D}^{\text{top}}(T_{m,t_1}^{\text{top}}, T_{n,t_2}^{\text{top}})) \quad (6)$$

$$p(T_{m,t_1}^{\text{geo}} | T_{n,t_2}^{\text{geo}}) \propto \exp(-\mathcal{D}^{\text{geo}}(T_{m,t_1}^{\text{geo}}, T_{n,t_2}^{\text{geo}})) \quad (7)$$

$$p(T_{m,t_1}^{\text{pht}} | T_{n,t_2}^{\text{pht}}) \propto \exp(-\mathcal{D}^{\text{pht}}(T_{m,t_1}^{\text{pht}}, T_{n,t_2}^{\text{pht}})). \quad (8)$$

Here,  $m$  and  $n$  are the indices of gender class, and  $t_1$  and  $t_2$  are the indices of age group. It is worth noting that the transition occurs only between image pairs of the same age group and from the opposite gender.

The transition probability is defined as follows:

$$p(G_{m,t} | G_{n,t}) = \prod_{i=1}^3 p(T_{i,m,t_1}^{\text{top}} | T_{i,n,t_2}^{\text{top}}) \cdot \prod_{i=1}^3 p(T_{i,m,t_1}^{\text{geo}} | T_{i,n,t_2}^{\text{geo}}) \cdot \prod_{i=1}^3 p(T_{i,m,t_1}^{\text{pht}} | T_{i,n,t_2}^{\text{pht}}) \quad (9)$$

In Fig. 5, we denote the transition probability with the thickness of the arrows.

### III. COMPUTATIONAL ASPECTS

Following the primal sketch model [12], the image lattice  $\Lambda$  is divided into sketchable part  $\Lambda_{\text{sk}}$  and nonsketchable part  $\Lambda_{\text{nsk}}$ , accounting for the structural and textural regions, respectively

$$\Lambda = \Lambda_{\text{sk}} \cup \Lambda_{\text{nsk}}, \quad \Lambda_{\text{sk}} \cap \Lambda_{\text{nsk}} = \phi. \quad (10)$$

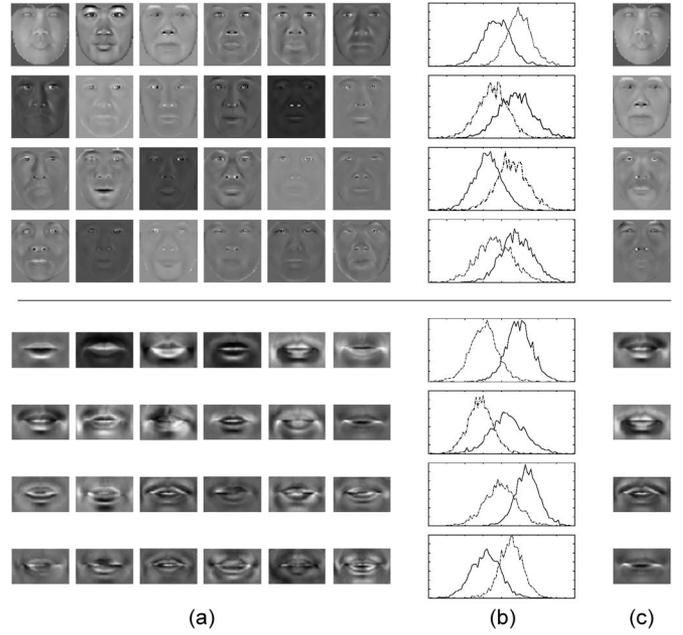


Fig. 6. Statistical analysis of nonsketchable part of multiple facial parts. (a) First 24 eigenvectors. (b) Histograms of projection coefficients on four most discriminative eigenvectors (solid curves for female and dash-dotted curves for male). (c) Four most discriminative principal components selected according to the coefficient distribution in (b).

In a face image, the nonsketchable part describes the general skin textural appearances (e.g., smoothness, color, etc.), and the sketchable part refines the facial details (e.g., component contour, wrinkles, etc.). We adopt different fusion methods for these two types of regions according to their intensity characteristics.

#### A. Nonsketchable Part

For the nonsketchable part, we model its appearance with AAM models. Adopting the coarse-to-fine strategy [50], we first train a global AAM model from 8000 labeled faces for the low-resolution face description. The first 24 eigenfaces are shown in the upper part of Fig. 6(a). For each facial component and skin region, we run a cluster algorithm to classify the images coarsely into subclasses and learn one local AAM model for each subclass. The lower part of Fig. 6(a) shows the first 24 eigenvectors of one specific type of mouth.

Letting  $pt$  denote a specific facial part and  $l$  denote the index of subclass, the corresponding AAM model is as follows:

$$I_{pt,l} = \overline{I}_{pt,l} + \alpha * \mathbf{V}_{pt,l}. \quad (11)$$

Here,  $\overline{I}_{pt,l}$  is the mean vector, and  $\mathbf{V}_{pt,l}$  and  $\alpha$  are the eigenvectors and corresponding coefficients, respectively.

Being similar to feature selection in recognition tasks, we select the principal components that are most discriminative for gender classification based on the histograms of PCA parameters [see Fig. 6(b)] and perform alternations on these selected components, as shown in Fig. 6(c).

The selected vectors form a low-dimensional space, in which face images cluster into two gender groups, as shown in Fig. 7.

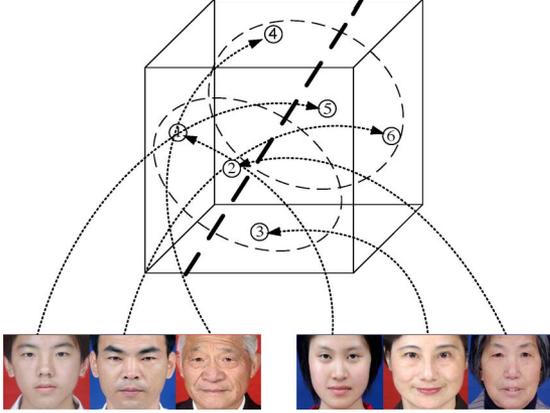


Fig. 7. Gender conversion scheme of the nonsketchable parts. (Circled numbers) Each sample is projected as a high-dimensional point in PCA space, and (dashed ellipses) the facial parameters of two genders cluster into two groups with some overlaps. We compute a gender transformation axis across the center of two clusters, as displayed by the thick dashed line.

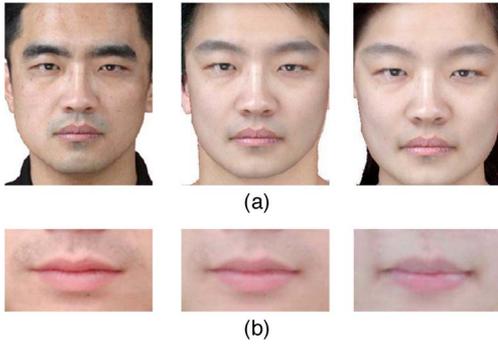


Fig. 8. Intermediate results of gender conversion. From the three images reconstructed with the coefficients along the transformation axis in Fig. 7, one can see that masculinity decreases as the coefficients shift from the male cluster to the female one. Conversion examples of (a) global skin and (b) mouth.

We define the direction perpendicular to the gender classification surface as the gender transformation axis, shifting the parameters along which the gender attributes of the subject will be changed. Fig. 8 shows intermediate results of gender conversion, and one can see that masculinity decreases from left to right.

### B. Sketchable Part

The detailed information that is crucial for face perception is lost in the AAM models, so we try to refine the details in skin area (e.g., wrinkles, marks, etc.) and contours of facial components in this section. Similar to the structure part in the primal sketch model, these details can be described with explicit graphs.

As aforementioned, each facial part can be of different topologies, so we build various graph templates for each part [see Fig. 9(a)]. During the conversion process, we use a stochastic graph-matching algorithm [23], [53] to select similar templates from the opposite gender group and apply transformation toward them. The adopted graph-matching algorithm include topological editing operations to tolerate the geometric defor-

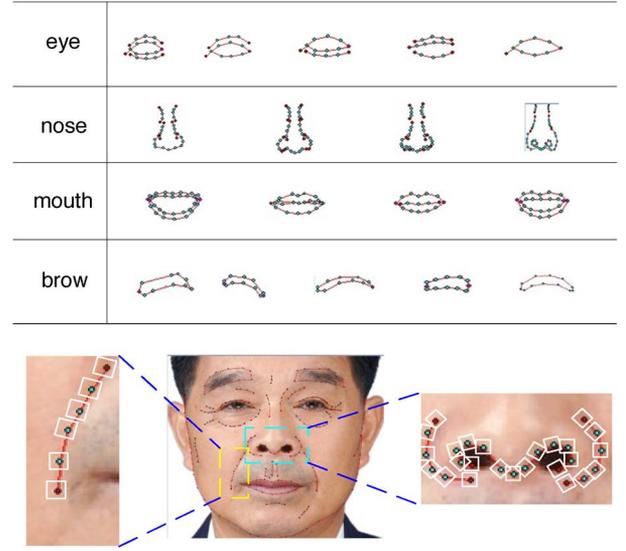


Fig. 9. Representation for the sketchable facial parts. For each facial component, we build different types of templates in (a) to account for the large variations. A template is described by a graph, as shown in (b).

mations and some structural differences between two groups. A graph is denoted as

$$G_{m,t} = \{V_{m,t}, E_{m,t}, A_{m,t}\}. \quad (12)$$

Here,  $V$ ,  $E$ , and  $A$  denote the set of vertices, edges, and relationships among vertices (e.g., collinearity, perpendicularity, etc.) in the layered graph, respectively.

Given two graphs  $G_{m,t}$  and  $G'_{m,t}$ , let  $g_{i,m,t}$  and  $g'_{i,m,t}$  be their subgraphs at layer  $i$ , and we define matching functions  $\psi_{i,m,t}$  and editing operators  $\phi_{i,m,t}$  between them

$$\psi_{i,m,t} : V_{i,m,t} \longrightarrow V'_{i,m,t} \cup \{\phi\}. \quad (13)$$

For each node  $v \in V$ , we have  $v' \in V' \cup \phi$ , which specifies the matching relationships between two graphs

$$\phi_{i,m,t} = \left\{ \phi_{i,m,t}^{\text{top}}, \phi_{i,m,t}^{\text{geo}}, \phi_{i,m,t}^{\text{pht}} \right\} \quad (14)$$

in which  $\phi_{i,m,t}^{\text{top}}$ ,  $\phi_{i,m,t}^{\text{geo}}$ , and  $\phi_{i,m,t}^{\text{pht}}$  are the topological, geometric, and photometric transformations between two subgraphs, respectively. Correspondingly, we define the following three energy terms to quantify them.

- 1) Our graph-matching approach defines the neighbors of vertex  $v \in V_{i,m,t}$  in graph  $g_{i,m,t} = \{V_{i,m,t}, E_{i,m,t}, A_{i,m,t}\}$  as

$$N_v = \{u : \langle u, v \rangle \in E_{i,m,t}, u \in V_{i,m,t}\} \quad (15)$$

Supposing that  $v' = \phi_i(v) \in V'_i$  is a matched node of  $v$ , if  $N_v = N_{v'}$ , then  $E^{\text{top}}$  is zero; else, some graph-editing operators [53] are needed for geometry alignment. We define all the costs of the necessary operators as the energy of topology transformation

$$E^{\text{top}} = \sum_{k=0}^K \text{cost}(\text{op}_k). \quad (16)$$

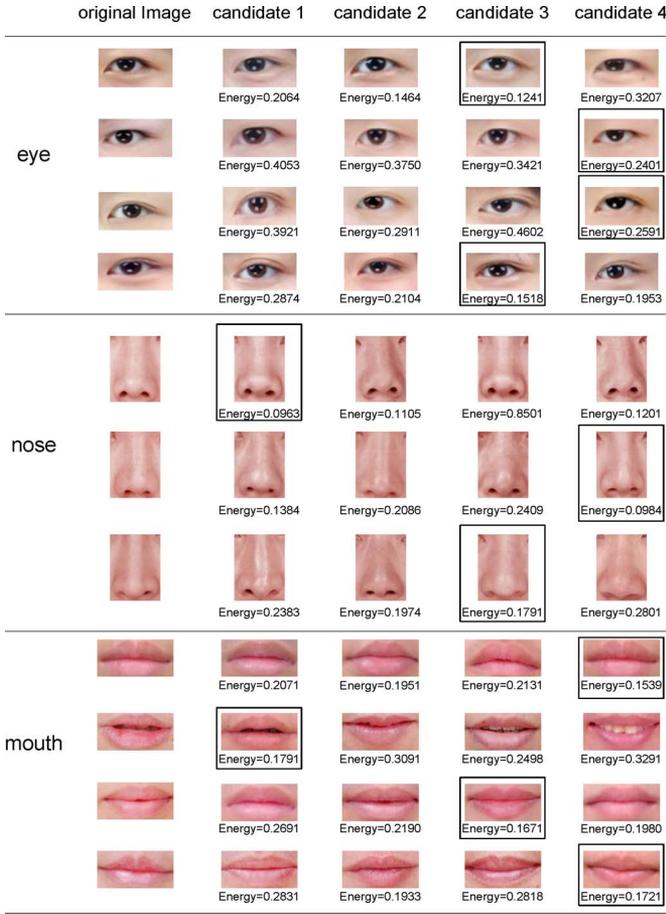


Fig. 10. Examples of template selection based on the matching energy defined in the graph-matching algorithm (the one with minimum energy is emphasized with thick boundary). Here, we show the top four candidates for each input example.

Here,  $K$  is the number of necessary graph-editing operators.

- 2) The geometric transformation  $\phi_{i,m,t}^{\text{geo}}$  from  $g_{i,m,t}$  to  $g'_{i,m,t}$  applies a global affine transformation  $A_{i,m,t}$  and a thin-plate spline warping  $F_{i,m,t}$

$$E^{\text{geo}}(\phi_{i,m,t}^{\text{geo}}) = E^{\text{geo}}\{A_{i,m,t}\} + E^{\text{geo}}\{F_{i,m,t}\}. \quad (17)$$

- 3) After topology transformation and geometric alignment, the distance between two graphs is measured by the dissimilarity of photometry around each vertex, which is described by the intensity profiles perpendicular to the edges or bars. We denote  $\mu_v$  as the intensity vectors around vertex  $v$

$$E^{\text{pht}} = \|\mu_v - \mu'_v\|. \quad (18)$$

Some final matching candidates of given eye, nose, and mouth images are shown in Fig. 10. After the target template is selected, we transfer the profiles along the graph edges to the selected image.

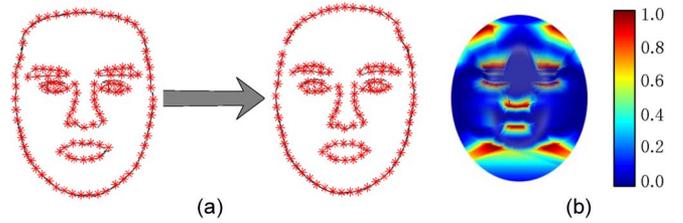


Fig. 11. Configuration difference between two opposite genders. (a) Shape-context matching between (left) average male face shape and (right) average female face shape. (b) Difference between two average face shapes.

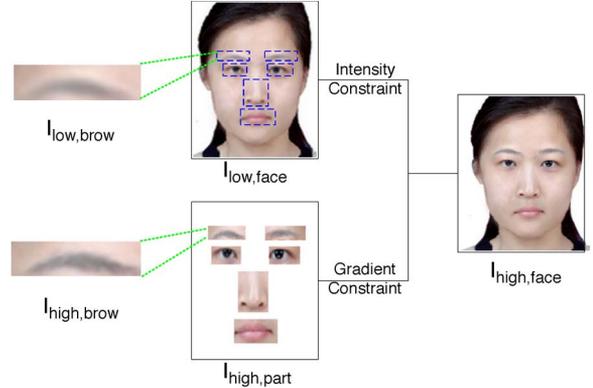


Fig. 12. Flowchart of high-resolution image fusion. First, we generate low-resolution face image  $I_{\text{low,face}}$  by nonsketchable part fusion and high-resolution part image  $I_{\text{high,part}}$  by sketchable part fusion. Then, an improved Poisson image-editing method is adopted to obtain high-resolution face image  $I_{\text{high,face}}$ , with  $I_{\text{low,face}}$  being the intensity constraint and  $I_{\text{high,part}}$  being the gradient constraint.

### C. Configurations

We compute the warping energy between average face shape of 2000 males and that of 2000 females to measure the difference between the facial configurations of two gender groups. Since we have fixed number of feature points for two faces after graph matching, only bending energy is used to measure the energy necessary for aligning two average face shapes [2]. The visualization of the bending energy is shown in Fig. 11. In our approach, we compute the target shape by simply superimposing the difference onto the original face.

## IV. IMPROVED POISSON IMAGE EDITING

As explained in Sections II-A and B, we conduct image fusion in two aspects separately. Fusion in the nonsketchable region generates low-resolution result, while fusion in the sketchable region synthesizes high-resolution details. We introduce image-editing techniques to merge the synthetic low-resolution face image  $I_l$  and high-resolution face image  $I_h$  seamlessly to generate the final fusion result.

The target is similar to that in the work of Wen *et al.* [47] on remote sensing image fusion, i.e., to enhance the low-resolution image  $I_l$  by fusing it with the high-resolution one  $I_h$  to produce a new image  $I$  (see Fig. 12). Our editing problem can be formulated as follows.

Let  $S$  denote the fusion result image domain, and let  $\Omega$  denote the uniformly sampled point set from  $I_l$ . Our objective is to find the interpolation of  $I$  over  $S \setminus \Omega$  to make its gradient field close to  $I_l$  and satisfy the boundary condition in  $\Omega$

$$\min \iint_{S \setminus \Omega} |\nabla I - \nabla I_h| dx dy, \quad \text{with } K(I)|_{\Omega} = I_l|_{\Omega}. \quad (19)$$

Here,  $\nabla$  is the gradient operator, and  $K$  is the Gaussian degradation operator. The first term requires the gradient of the resultant image to be consistent with high-resolution image  $I_h$ . The second term is the added soft constraint as the boundary condition, which forces the resultant image to be equal to  $I_l$  when degraded.

For a discrete image, (19) can be further simplified as

$$\min \sum_p \sum_{q \in N^p} (I^p - I^q - I_h^p + I_h^q)^2 + \alpha^2 \sum_p \left( \sum_{r \leq R} I_h^r K_M(p, r) - I_l^p \right)^2. \quad (20)$$

Here,  $N^p$  is the neighborhood of pixel  $p$ ,  $R$  is the radius of the degradation operator, and  $K_M(p, r) = k(|p - r|)$ .

This minimization problem can be computed by simple iterations

$$\begin{aligned} & \left( |N^p| + \frac{\alpha^2}{2} K(p, p) \right) * I^p - \sum_{q \neq p} I^q \\ & = \frac{\alpha^2}{2} \sum_{q \in N^p} K_M(p, q) I_l^q + \sum_{q \in N^p} (I_h^p - I_h^q). \end{aligned} \quad (21)$$

Here,  $K_{p,q} = \delta_{q \in N^p} - (\alpha^2/2)K_W(p, q)$  and  $K_W(p, q) = \sum_r K_M(p, r)K_M(q, r)$ . Setting the initial guess as the enlarged low-resolution fusion result  $I_l$ , we are able to obtain fusion result  $I$  with details in facial parts and skin area refined after several iterations.

## V. EXPERIMENTAL RESULTS

In this paper, we collect 8000 high-resolution Asian face images, among which 4000 are males and 4000 are females. The between-eye distance is around 100 pixels, and all the images are of frontal faces taken under controlled light conditions. For each image in this database, 90 landmarks are labeled manually. Based on these labels, we build the graphical face model and learn the transition probabilities between two gender groups. In this paper, we discretize the age range into four age groups and synthesize fusion results in two opposite directions for each age group. Fig. 13 shows eight exemplar results. We also display the results enhanced with external features (hair and clothes) in Fig. 13. Note that, in the following experiments, only experiment five is conducted on the enhanced faces, while the results of other experiments are from images without external features.

### A. Criteria for Quantitative Evaluation

Being different from the other face synthesis tasks (e.g., pose rectification, delighting, etc.), there is no ground truth for gender conversion. Therefore, we propose the following three criteria for quantitative evaluation of the synthetic results in a series of subjective and objective experiments. (In this paper, 20 volunteers are recruited for subjective evaluation.)

- 1) *Realism of the synthesis.* Obtaining realistic synthetic images is the most important criterion for image fusion algorithms, particularly in subjective judgment. This criterion requires the synthesized face images to be realistic enough, e.g., without apparent artifacts, and consistency among regions, among others.
- 2) *Face identity preservation.* After changing the gender of a human face, the identity information should be preserved effectively, i.e., the resulting face images should bear sufficient similarity to the input face in terms of identity.
- 3) *Appearance similarity to the target gender.* Since the target is to transform the input face to the opposite gender group, the synthetic results should bear the typical characteristics of the intended gender group.

### B. Experiment 1: Studies on Relative Contributions of Facial Parts to Gender Perception

We observe that, for some facial parts, males are significantly different from females, while for the other parts, there exists no large difference. We introduce multiple regression analysis (MRA), a statistical analysis approach widely used in psychological experiments, to study the relative contribution of each facial part to gender perception. In this experiment, we randomly select 100 males and 100 females from our training database. After landmark localization, the images are decomposed into facial parts, which are then given to the volunteers for gender determination.

The result is shown in Table II. In the table,  $R^2 \in [0, 1]$  is a quantitative measurement reflecting the validness of our compositional face model. The large  $R$ -square value shows that our conversion model includes most of the features that are crucial for gender classification. For each facial part, MRA obtains a  $\beta$  value measuring its importance to gender perception. From the table, we can see that three features, i.e., skin, chin, and brow, are with the largest  $\beta$  values, which indicates that males are significantly different from females in these attributes. We should thus give high priority to the features with high  $\beta$  values in both fusion algorithm and gender classification approaches.

### C. Experiment 2: Tests on the Realism of the Fusion Results

First, we perform gender conversion on the 200 face images selected in Experiment 1 (Fig. 13 gives one subset of the results) and mix these 200 synthetic images with other 200 real images from the training database. Then, the 20 volunteers are presented with the 400 images at three increasing resolutions ( $90 \times 120$ ,  $150 \times 200$ , and  $240 \times 320$  pixels) and asked to pick out as many synthetic ones as possible. The result is shown in Table III.

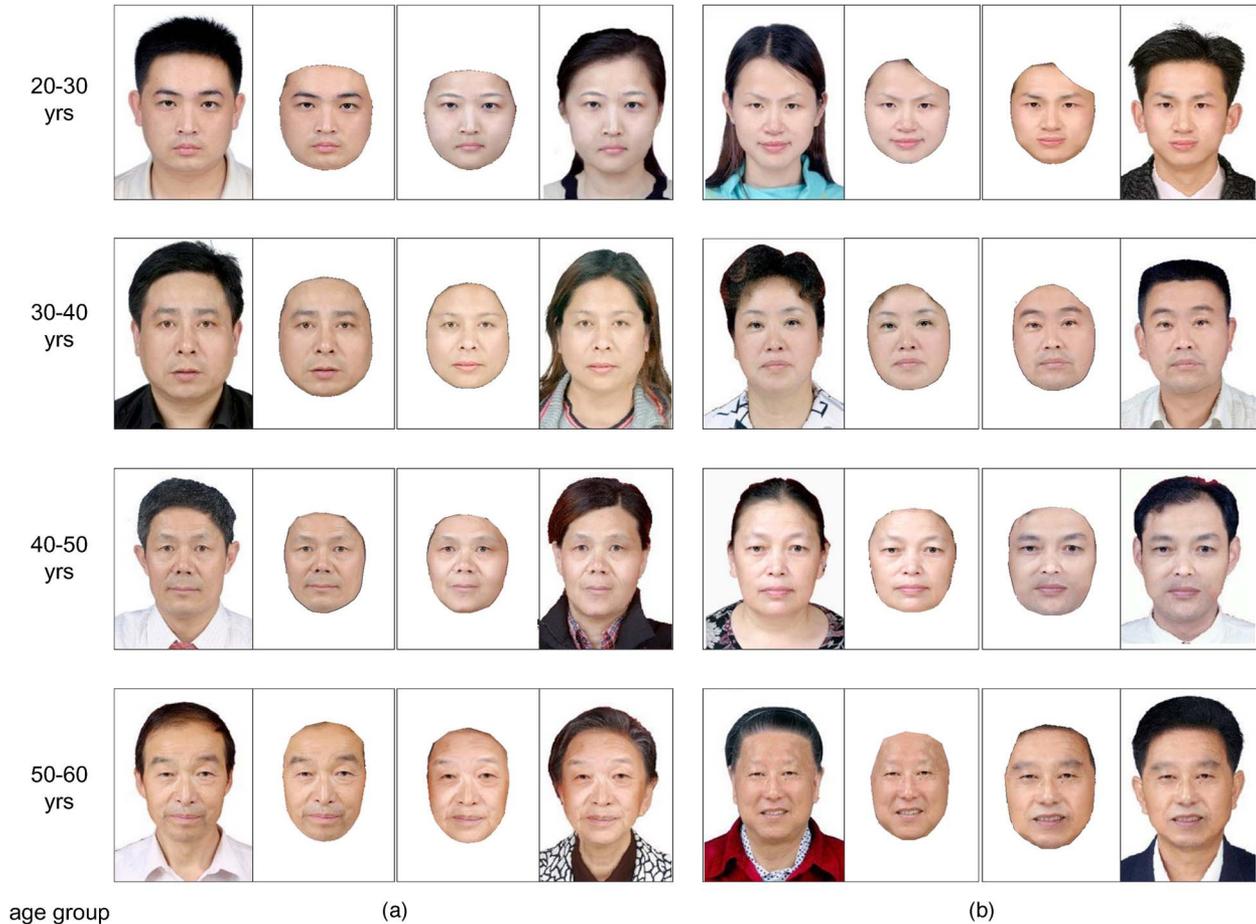


Fig. 13. Examples of gender conversion results produced by our algorithm. For each age group, we show two examples in two opposite directions. In each conversion example, the images from left to right are as follows: input image, the input face with background cut out, the conversion result, and the result enhanced with external factors. (a) From male to female. (b) From female to male.

TABLE II  
RELATIVE CONTRIBUTION OF GENDER CLASSIFICATION

	skin	chin	brow	forehead	contour	mouth	eye	eyecorner	nose	laughline	eyehole
$\beta$	0.392	0.214	0.161	0.090	0.074	0.068	0.061	0.024	0.023	0.021	0.019
$R^2$	0.907										

TABLE III  
PERCENTAGE OF SYNTHETIC FACES WITH NOTICEABLE ARTIFACTS

Image resolution (pixels)	$90 \times 120$	$150 \times 200$	$240 \times 320$
Percentage of pick-out images (%)	1.5	6.625	18.75

The small proportion of “picked-out” images indicates that our fusion algorithm is able to generate results with few artifacts, and shows a great potential in real applications. In addition, the percentage of picked-out images increases with image resolution. This consists with the intuition that slight fusion artifacts become invisible at lower resolutions. However, even for a resolution of  $240 \times 320$  pixels, the 18.75% picked-outs are still an encouraging result.

#### D. Experiment 3: Tests on the Appearance Similarity to the Target Gender

In this experiment, we conduct subjective and objective experiments to compare the performances of gender classification

on the 200 synthetic face images and 200 real images used in Experiment 2. In the subjective experiment, we present the 400 selected images to the 20 volunteers and ask them to label their gender. For the objective evaluation, considering the effectiveness of the SVM algorithm on gender classification, we train an SVM classifier that adopts the graph parameters at three resolutions as features. Here, the classifier is trained on 500 images in our database, and none of the selected 400 images is included.

The classification performance is shown in Fig. 14(b), from which we can see that the accuracy of the objective gender classification is comparable to that of the subjective classification. What is more important is that, regardless of classification type, i.e., objective or subjective, the accuracy of synthetic images is quite comparable to that of real face images, which can strongly indicate that our fusion results indeed bear sufficient appearance characteristics of the intended gender group.

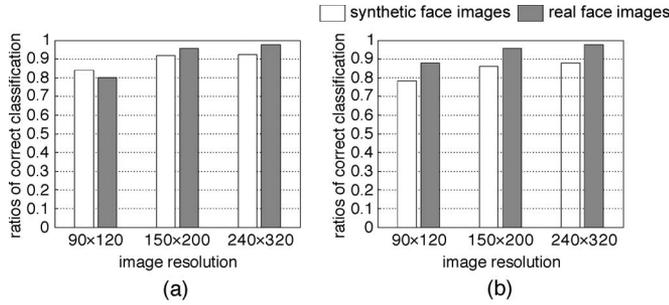


Fig. 14. Gender classification performance comparison on synthetic faces and real images.

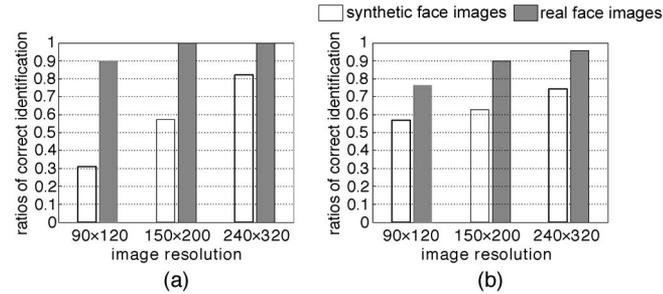


Fig. 15. Performance comparison between face identification performances on synthetic and real face images. (a) Subjective identification. (b) Objective identification.

#### E. Experiment 4: Tests on the Preservation of Identity

While we denote the 200 synthetic images in previous experiments as set A, here, the input images of these 200 synthetic faces that we mix with 200 images randomly selected from our training data set are denoted as set B. In the subjective experiment, the 20 volunteers are asked to identify the images in set A from set B. Aside from the subjective evaluations, we also conduct objective face recognition for comparison. In the objective experiment, a face recognition algorithm developed on Gabor features and linear discriminant analysis [39] is used to recognize the images in set A, with the whole data set being the gallery set. Fig. 15 shows both the subjective and objective recognition results. From the high recognition rates, one can see that the identity is preserved effectively in the process of gender conversion.

#### F. Experiment 5: Benefits of External Features to Gender Perception

Life experience tells us that some external features (hair and clothes) are informative for gender classification in real-life communications, so we enhance the synthetic faces with exemplar hairstyles and clothes from the intended gender group (see Fig. 13), based on which a series of subjective experiments is conducted to test the effects of external features quantitatively. To study the impacts of external features on gender perception, we perform a comparison between gender classification rates on synthetic images with and without external features. In the same way, we study the impacts of external features on subjective face recognition by comparing the face recognition

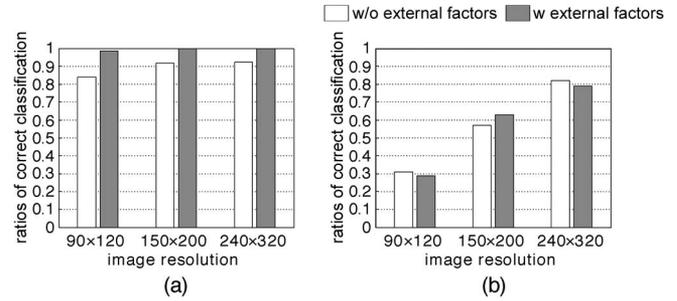


Fig. 16. Quantitative analysis on the impact from external features. (a) Gender classification. (b) Face identification.

performances on the same data. The results are shown in Fig. 16.

From the gender classification results in Fig. 16, one can see that the external features (hair and clothes) help gender classification largely. With external features included, subjective gender classification of middle- and high-resolution images achieves nearly perfect results. However, in the task of face identification, we can see no significant help from external features. This may be due to the fact that human beings are good at recognizing faces after millions of years of evolution, so subjective face recognition is robust to variations in external features.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a fusion strategy for gender conversion. With the support of a large face database and advances in gender classification, we tried to alter the gender attributes of an input face and preserve its face identity simultaneously. In addition, we obtained photorealistic visual results by adopting image-editing techniques in computer graphics.

Due to the nonexistence of ground truth in gender conversion, three task-oriented criteria have been proposed for result evaluation, based on which both subjective and objective experiments have been conducted to validate the proposed strategy.

The visually photorealistic and statistically reasonable results would potentially benefit some real-world applications: 1) providing some reference templates to help look for the lost opposite-sex siblings of the given subjects; 2) generating transsexual makeup results that can be applied in entertainments such as filmmaking and computer games; 3) producing stimuli for gender-related psychological experiments; and 4) extending to fusion between other groups and introducing some interesting applications, e.g., fusion between two age groups, between film stars and ordinary people, etc.

In future work, we will study more meaningful similarity metrics for facial component matching and extend the strategy to more extensive data sets (e.g., faces from different races, images with expression variations, etc.).

#### ACKNOWLEDGMENT

This work was partially done at the Lotus Hill Institute (LHI), and the data used in this paper are provided by the LHI annotation project [51].

## REFERENCES

- [1] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, Jan. 2007.
- [2] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
- [3] W. Cao, B. Li, and Y. Zhang, "A remote sensing image fusion method based on PCA transform and wavelet packet transform," in *Proc. Int. Conf. Neural Netw. Signal Process.*, 2003, pp. 978–981.
- [4] H. Chen, "A multiresolution image fusion based on principal component analysis," in *Proc. Int. Conf. Image Graph.*, 2007, pp. 737–741.
- [5] H. Chen, Z. Xu, Z. Liu, and S.C. Zhu, "Composite templates for cloth modeling and sketching," in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 943–950.
- [6] L. J. Chipman, T. M. Orr, and L. N. Graham, "Wavelets and image fusion," in *Proc. Int. Conf. Image Process.*, 1995, pp. 248–251.
- [7] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
- [8] N. P. Costen, M. Brown, and S. Akamatsu, "Sparse models for gender classification," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 201–206.
- [9] J. M. Fellous, "Gender discrimination and prediction on the basis of facial metric information," *Vis. Res.*, vol. 37, no. 14, pp. 1961–1973, Jul. 1997.
- [10] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
- [11] A. B. Graf and F. A. Wichmann, "Gender classification of human faces," in *Proc. Int. Workshop Biologically Motivated Comput. Vis.*, 2002, pp. 491–500.
- [12] C. Guo, S. C. Zhu, and Y. Wu, "Primal sketch: Integrating texture and structure," *Comput. Vis. Image Understanding*, vol. 106, no. 1, pp. 5–19, Apr. 2007.
- [13] S. Gutta, J. Huang, P. Jonathon, and H. Wechsler, "Mixture of experts for classification of gender, ethnic origin, and pose of human faces," *IEEE Trans. Neural Netw.*, vol. 11, no. 4, pp. 948–960, Jul. 2000.
- [14] S. Gutta, H. Wechsler, and P. J. Phillips, "Gender and ethnic classification of face images," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 1998, pp. 194–199.
- [15] P. Hill, N. Canagarajah, and D. Bull, "Image fusion using complex wavelets," in *Proc. 7th Int. Conf. Inf. Fusion*, 2002, pp. 487–496.
- [16] Z. Ji, X. Lian, and B. Lu, "Gender classification by information fusion of hair and face," in *State of the Art in Face Recognition*. Vienna, Austria: IN-TECH, 2009.
- [17] H. Kim, D. Kim, Z. Ghahramani, and S. Bang, "Appearance-based gender classification with Gaussian processes," *Pattern Recognit. Lett.*, vol. 27, no. 6, pp. 618–626, Apr. 2006.
- [18] X. Leng and Y. Wang, "Improving generalization for gender classification," in *Proc. Int. Conf. Image Process.*, 2008, pp. 1656–1659.
- [19] J. Lewis, R. J. Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah, "Region-based image fusion using complex wavelets," in *Proc. Int. Conf. Image Fusion*, 2004, pp. 555–562.
- [20] H. Li, B. S. Manjunath, and S. K. Mitra, "Multi-sensor image fusion using the wavelet transform," in *Proc. Int. Conf. Image Process.*, 1994, pp. 51–55.
- [21] H. Lian, B. Lu, E. Takikawa, and S. Hosoi, "Gender recognition using a min-max modular support vector machine," in *Proc. 1st Int. Conf. Natural Comput.*, 2005, pp. 438–441.
- [22] L. Lin, S. Peng, J. Porway, S. C. Zhu, and Y. Wang, "An empirical study of object category recognition: Sequential testing with generalized samples," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [23] L. Lin, S. C. Zhu, and Y. Wang, "Layered graph match with graph editing," in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.
- [24] H. Lu, Y. Huang, Y. Chen, and D. Yang, "Automatic gender recognition based on pixel-pattern-based texture feature," *J. Real-Time Image Process.*, vol. 3, no. 1/2, pp. 109–116, Mar. 2008.
- [25] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, Mar. 2008.
- [26] E. Makinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1544–1556, Jul. 2008.
- [27] N. Mitianoudis and P. T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," in *Proc. Int. Conf. Inf. Fusion*, 2007, pp. 131–142.
- [28] B. Moghaddam and M. Yang, "Learning gender with support faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 707–711, May 2002.
- [29] U. Mohammed, S. J. D. Prince, and J. Kautz, "Visio-lization: Generating novel facial images," *ACM Trans. Graphics*, vol. 28, no. 3, article 57, Aug. 2009.
- [30] A. F. Norcio and J. Stanley, "Adaptive human-computer interfaces: A literature survey and perspective," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 2, pp. 399–408, Mar./Apr. 1989.
- [31] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [32] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, Jul. 2003.
- [33] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.
- [34] G. Piella, "A general framework for multiresolution image fusion," *Inf. Fusion*, vol. 4, no. 4, pp. 258–280, Dec. 2003.
- [35] C. Ramesh and T. Ranjith, "Fusion performance measures and a lifting wavelet transform based algorithm for image fusion," in *Proc. Int. Conf. Inf. Fusion*, 2002, pp. 317–320.
- [36] D. A. Rowland and D. I. Perrett, "Manipulating facial appearance through shape and color," *IEEE Comput. Graph. Appl.*, vol. 15, no. 5, pp. 70–76, Sep. 1995.
- [37] F. Sadjadi, "Comparative image fusion analysis," in *Proc. 2nd Int. Workshop Object Tracking Classification*, 2005, pp. 8–15.
- [38] A. Samal, V. Subramani, and D. Marx, "Analysis of sexual dimorphism in human face," *J. Vis. Commun. Image Represent.*, vol. 18, no. 6, pp. 453–463, Dec. 2007.
- [39] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [40] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proc. 6th IEEE Workshop Appl. Comput. Vis.*, 2002, pp. 165–170.
- [41] J. Suo, F. Min, S. C. Zhu, S. Shan, and X. Chen, "A multi-resolution dynamic model for face aging simulation," in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.
- [42] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," in *Proc. Int. Conf. Biometrics: Theory, Appl. Syst.*, 2007, pp. 1–5.
- [43] B. Tiddeman, M. Burt, and D. Perrett, "Prototyping and transforming facial textures for perception research," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 42–50, Sep./Oct. 2001.
- [44] B. P. Tiddeman, M. Stirrat, and D. Perrett, "Towards realism in facial transformation: Results of a wavelet MRF method," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 449–456, Sep. 2005.
- [45] K. Ueki, H. Komatsu, S. Imaizumi, K. Kaneko, N. Sekine, J. Katto, and T. Kobayashi, "A method of gender classification by integrating facial, hairstyle, and clothing images," in *Proc. Int. Conf. Pattern Recog.*, 2004, pp. 446–449.
- [46] M. E. Ulug, "A quantitative metric for comparison of night vision fusion algorithms," *Proc. SPIE*, vol. 4051, pp. 80–88, 2000.
- [47] J. Wen, Y. Li, and H. Gong, "Remote sensing image fusion on gradient field," in *Proc. Int. Conf. Pattern Recog.*, 2006, pp. 643–646.
- [48] L. Wiskott and J. M. Fellous, "Face recognition and gender determination," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 1995, pp. 92–97.
- [49] D. B. Wright and B. Sladden, "An own gender bias and the importance of hair in face recognition," *Acta Psychologica*, vol. 114, no. 1, pp. 101–114, Sep. 2003.
- [50] Z. Xu, H. Chen, S. C. Zhu, and J. Luo, "A hierarchical compositional model for face representation and sketching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 955–969, Jun. 2008.
- [51] B. Yao, X. Yang, and S. C. Zhu, "Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks," in *Proc. 6th Int. Workshop Energy Minimization Methods Comput. Vis. Pattern Recog.*, 2007, pp. 169–183.
- [52] S. C. Zhu and D. Mumford, "A stochastic grammar of images," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 4, pp. 259–362, Jul. 2006.
- [53] S. C. Zhu and A. L. Yuille, "A flexible object recognition and modeling system," *Int. J. Comput. Vis.*, vol. 20, no. 3, pp. 187–212, Dec. 1996.



**Jinli Suo** received the B.S. degree from Shandong University, Jinan, China, in 2004. She is currently working toward the Ph.D. degree at the Graduate University of the Chinese Academy of Sciences, Beijing, China.

Her research interests mainly include face aging modeling, gender classification, face recognition, facial sketching, and caricature generation.



**Liang Lin** was born in 1981. He received the B.S. and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1999 and 2008, respectively. He was a joint Ph.D. student in the Department of Statistics, University of California, Los Angeles (UCLA), during 2006–2007.

He was a Postdoctoral Research Fellow in the Center for Image and Vision Science, UCLA, and a Senior Research Scientist with the Lotus Hill Research Institute, Wuhan, China, during 2007–2009.

He is currently an Associate Professor with the School of Software, Sun Yat-Sen University, Guangzhou, China. His research interests include but not limited to computer vision, statistical modeling and computing, and pattern recognition.



**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, China, in 2004.

He has been with ICT since 2002, where he has been an Associate Professor with the Key Laboratory of Intelligent Information Processing since 2005 and is also the Vice Director of the ICT-ISVISION Joint Research and Development Laboratory for Face

Recognition. His research interests include image analysis, pattern recognition, and computer vision. He is particularly focusing on face-recognition-related research topics and has published more than 120 papers on related research topics.

Dr. Shan received the State Scientific and Technological Progress Awards in 2005 in China for his work on face recognition technologies. One of his coauthored CVPR 2008 papers won the Best Student Poster Award Runner-up. He also won the Silver Medal of the Scopus' Future Star of Science Award in 2009.



**Xilin Chen** (M'00–SM'09) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1988, 1991, and 1994, respectively.

He was a Professor with the Harbin Institute of Technology from 1999 to 2005. He was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. Since August 2004, he has been with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, where he is also with the Key Laboratory of Intelligent Information Processing and the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition. His research interests include image processing, pattern recognition, computer vision, and multimodal interfaces.

Dr. Chen has served as a program committee member for more than 20 international and national conferences. He has received several awards, including the State Scientific and Technological Progress Award in 2000, 2003, and 2005 in China for his research work.



**Wen Gao** (M'92–SM'05–F'09) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1985 and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was a Professor in computer science with the Harbin Institute of Technology from 1991 to 1995 and with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, from 1996 to 2005. He is currently a Professor with the School of Electronics Engineering

and Computer Science, Peking University, Beijing. He is also an Adjoint Professor of ICT. He has been leading research efforts to develop systems and technologies for video coding, face recognition, sign language recognition and synthesis, and multimedia retrieval. He has published four books and over 500 technical articles in refereed journals and proceedings in the areas of signal processing, image and video communication, computer vision, multimodal interfaces, pattern recognition, and bioinformatics.

Dr. Gao served the academic society as the General Cochair of the 2007 IEEE International Conference on Multimedia and Expo and has been the Head of the Chinese delegation to the Moving Picture Expert Group of the International Standard Organization since 1997. He is also the Chairman of the working group responsible for setting a national audio–video coding standard for China. He has received many awards, including five national awards for his research achievements and activities.