"Wow! You Are So Beautiful Today!"

LUOQI LIU, National University of Singapore
JUNLIANG XING, Chinese Academy of Sciences
SI LIU, National University of Singapore
HUI XU and XI ZHOU, Chinese Academy of Sciences
SHUICHENG YAN, National University of Singapore

Beauty e-Experts, a fully automatic system for makeover recommendation and synthesis, is developed in this work. The makeover recommendation and synthesis system simultaneously considers many kinds of makeover items on hairstyle and makeup. Given a user-provided frontal face image with short/bound hair and no/light makeup, the Beauty e-Experts system not only recommends the most suitable hairdo and makeup, but also synthesizes the virtual hairdo and makeup effects. To acquire enough knowledge for beauty modeling, we built the Beauty e-Experts Database, which contains 1,505 female photos with a variety of attributes annotated with different discrete values. We organize these attributes into two different categories, beauty attributes and beauty-related attributes. Beauty attributes refer to those values that are changeable during the makeover process and thus need to be recommended by the system. Beauty-related attributes are those values that cannot be changed during the makeup process but can help the system to perform recommendation. Based on this Beauty e-Experts Dataset, two problems are addressed for the Beauty e-Experts system: what to recommend and how to wear it, which describes a similar process of selecting hairstyle and cosmetics in daily life. For the what-to-recommend problem, we propose a multiple tree-structured supergraph model to explore the complex relationships among high-level beauty attributes, mid-level beauty-related attributes, and low-level image features. Based on this model, the most compatible beauty attributes for a given facial image can be efficiently inferred. For the how-to-wear-it problem, an effective and efficient facial image synthesis module is designed to seamlessly synthesize the recommended makeovers into the user facial image. We have conducted extensive experiments on testing images of various conditions to evaluate and analyze the proposed system. The experimental results well demonstrate the effectiveness and efficiency of the proposed system.

Categories and Subject Descriptors: H.3.3 [**Information and Storage Retrieval**]: Information Search and Retrieval—*Retrieval models*; I.2.6 [**Artificial Intelligence**]: Learning—*Knowledge acquisition*

General Terms: Algorithms, Experimentation, Performance

Additional Key Words and Phrases: Beauty recommendation, beauty synthesis, multiple tree-structured super-graphs model

ACM Reference Format:
Luoqi Liu, Junliang Xing, Si Liu, Hui Xu, Xi Zhou, and Shuicheng Yan. 2014. "Wow! You are so beautiful today!". ACM Trans. Multimedia Comput. Commun. Appl. 11, 1s, Article 20 (September 2014), 22 pages.
DOI: http://dx.doi.org/10.1145/2659234

L. Liu and J. Xing contributed equally to this work.
Authors' addresses: L. Liu (corresponding author), S. Liu, and S. Yan, National University of Singapore, Singapore; emails: {a0092770, dcslius, eleyans}@nus.edu.sg; J. Xing, Institute of Automation, Chinese Academy of Sciences, China; email: jlxing@nlpr.ia.ac.cn; H. Xu and X. Zhou, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, China; emails: {xuhui, zhouxi}@cigit.ac.cn.

DOI: http://dx.doi.org/10.1145/2659234

1. INTRODUCTION

Beauty is a language which enables people to express their personalities, gain self-confidence, and open up to others. Everybody, especially women, wants to be beautiful, and makeovers become indispensable for most modern women. Hairstyle and makeup are two main factors that influence one's judgment about whether someone looks beautiful or not. By choosing the proper hair and makeup style, one may enhance the whole aura of a person and thus look more attractive. However, people often encounter problems when they make choices. First of all, the effects of different makeup products and hairstyles vary for different individuals, and are highly correlated with one's facial traits (e.g., face shape and skin color). It is quite difficult, or even unlikely, for people to analyze their own facial features and make proper choices of care and products. Second, nowadays cosmetics has developed into a large industry and there exist an unimaginable variety of products. Making choices with so much to choose from could cost people a lot of time and money. Last but not least, the typical experimental procedure is tedious for both the shopping assistant and the customer. Therefore, choosing the proper hairstyle and makeup effectively and effectively becomes a challenge.

Many attempts have been made to alleviate this problem. Some virtual hairstyle and makeup techniques have been developed. Software like Virtual Haircut and Makeover[1], which allows people to change their hairstyle and apply virtual makeup on their photos, has been put into use. With software of this kind, users can input a facial image and then choose any hairstyle or makeup they prefer from the options provided by the system. Users can see the effects of applying the chosen hairstyles and makeup products on their faces and make decisions on whether to choose these products in reality. These functions make the makeover process easier but still require a lot of manual input. For example, users have to mark out special regions, such as corners of eyes, nose, mouth, or even pupil, etc., on their photos manually. Besides, these softwares do not have a recommendation function. Users have to make choices on their own, and adjust the synthetic effects of these choices manually. It is too complicated for people who are not cosmetic professionals to optimize this process.

For the issue of automatic beauty recommendations, the research is still quite limited, although some researchers have tried some approaches. Tong et al. [2007] extracted makeup from before-and-after training image pairs and transferred the makeup effect, defined as ratios, to a new testing image. Guo and Sim [2009] considered the makeup effect as existing in two layers of a three-layer facial decomposition result, and the makeup effect of a reference image was transferred to the target image. Some patents also try to address the hairstyle recommendation problem (e.g., Nagai et al. [2005]), which searches for hairstyles in a database from the plurality of the hairstyle parameters based on manually selected hair attributes by the user. These works all fail to provide a recommendation function, and the synthetic effects may not be suitable for every part of the face. Besides, to the best of our knowledge, most of these works need a lot of user interaction, and the final results are highly dependent on the efforts of users.

The aim of this work is to develop a novel Beauty e-Experts system which helps users to select hairstyle and makeup automatically and produces the synthesized visual effects [Liu et al. 2013]. The main challenge is modeling the complex relationships among different beauty and beauty-related attributes for reliable recommendation and natural synthesis. To address this challenge, we built a large dataset, the Beauty e-Experts Dataset, which contains 1,505 images of female figures selected from professional fashion websites. Based on this Beauty e-Experts Dataset, we first annotate all the beauty

[1]http://www.goodhousekeeping.com/beauty.

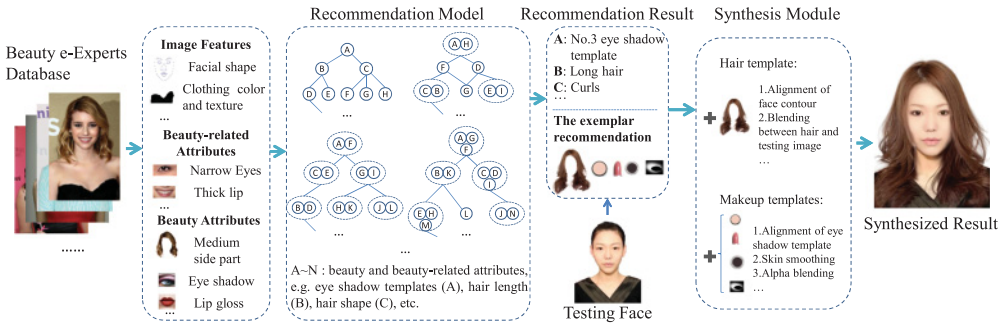

Fig. 1. System processing flowchart. We first compiled the Beauty e-Experts Database of 1,505 facial images with different hairstyles and makeup effects. With the extracted facial and clothing features, we propose a multiple tree-structured supergraph model to express the complex relationships among beauty and beauty-related attributes. Here, the results from multiple individual supergraphs are fused based on voting strategy. In the testing stage, the recommended hair and makeup templates for the testing face are then applied to synthesize the final visual effects.

attributes with discrete values for each image in the whole dataset. These beauty attributes, including different hairstyles and makeup applications, are all adjustable in daily life. Their specific combination are considered the recommendation objective of the Beauty e-Experts System. To narrow the gap between high-level beauty attributes and low-level image features, a set of mid-level beauty-related attributes, such as facial traits and clothing properties, are also annotated for the dataset to help the system perform recommendation.

Based on all these attributes, we propose learning a multiple tree-structured supergraph model to explore the complex relationships among these attributes. As a generalization of a graph, a supergraph could theoretically characterize any type of relationship among different attributes and thus provide more effective and powerful recommendations. We propose using its multiple tree-structured approximations to reserve the most important relationships and make the inference procedure tractable. Based on the recommended results, an effective and efficient facial image synthesis module is designed to seamlessly synthesize the recommended results into the user facial image and show it back to the user. Experimental results on 100 testing images show that our system can obtain very reasonable recommendation and appealing synthesis results. The whole system processing flowchart is illustrated in Figure 1.

The contributions of this work are summarized as follows.

—A comprehensive system considering both hairstyle and makeup recommendation and synthesis is explored for the first time.
—A large database called the Beauty e-Experts Database is constructed, including 1,505 facial images with various makeover effects, and fully annotated with different types of attributes.
—A multiple tree-structured supergraph model is proposed for hairstyle and makeup recommendation.

The remaining sections are organized as follows. Section 2 introduces the Beauty e-Experts Database and features as well as attributes design. Section 3 illustrates our recommendation model in detail, including model formulation, parameter learning, and inference. Section 4 explains the synthesis process of hairstyle and makeup based on the recommendation. Experiments are presented in Section 5. At last, Section 6 concludes our work and discusses future work.

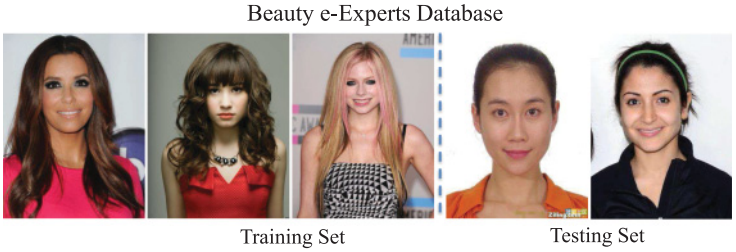

Fig. 2. Some exemplar images from the Beauty e-Experts Dataset and the additional testing set. The left three images are from the Beauty e-Experts Dataset used for training, while the right two images are from the testing set.

2. DATASET, ATTRIBUTES, AND FEATURES

Hairstyle and makeup products capture a lucrative market among female customers, but no public datasets for academic research exist. Most previous research [Scherbaum et al. 2011; Guo and Sim 2009; Tong et al. 2007] only work for a few samples. Chen and Zhang [2010] released a benchmark for facial beauty study, but their focus is geometric facial beauty, not facial makeup and hairstyle. Wang et al. [2012] sampled 1,021 images with regular hairstyles from the Labeled Faces in the Wild (LFW) Database [Huang et al. 2007], which is designed for hair segmentation but is not suitable for hairstyle recommendation. The sampled LFW database contains only a few hairstyles, since it is designed only for hair segmentation. In order to obtain enough knowledge to perform beauty modeling, we need a large dataset specifically designed for this task. In the following, we will describe the construction of the Beauty e-Experts Dataset, as well as its attribute annotation and feature extraction process.

2.1. The Beauty e-Experts Dataset

To build our Beauty e-Experts Dataset, we downloaded $\sim$ 800K images of female figures from professional hairstyle and makeup websites (e.g., `www.stylebistro.com`). We searched for these photos with key words such as *makeup, cosmetics, hairstyle, haircut,* and *celebrity*. The initial downloaded images are screened by a commercial face analyzer[2] to remove images with no face detected, and then 87 key points are located for each image. Images with high resolution and confident landmark detection results are retained. The retained images are further manually selected, and only those containing female figures who are considered "attractive" and with complete hairstyle and obvious makeup effects are retained. The final 1,505 retained images contain female figures in distinct fashions and with clear frontal faces, and are thus very good representatives for beauty modeling. We also collected 100 face images with short/bound hair and no/light makeup, which better demonstrate the synthesized results, for experimental evaluation purpose. Figure 2 shows some exemplar images in the dataset.

2.2. Beauty Attributes

To acquire beauty knowledge from the dataset, we comprehensively explore different beauty attributes on these images, including various kinds of hairstyles and facial makeups. All these beauty attributes can be easily adjusted and modified in daily life and thus have practical meaning for our beauty recommendation and synthesis system. We carefully organize these beauty attributes and set their attribute values based on some basic observations or preprocessing on the whole dataset. Table I lists the names and values of all the beauty attributes considered in the work. The

[2]OMRON OKAO Vision: `http://www.omron.com/r_d/coretech/vision/okao.html`.

Table I. List of the High-Level Beauty Attributes

Name	Values
hair length	long, medium, short
hair shape	straight, curled, wavy
hair bangs	full, slanting, center part, side part
hair volume	dense, normal
hair color	20 classes
foundation	15 classes
lip gloss	15 classes
eye shadow color	15 classes
eye shadow template	20 classes

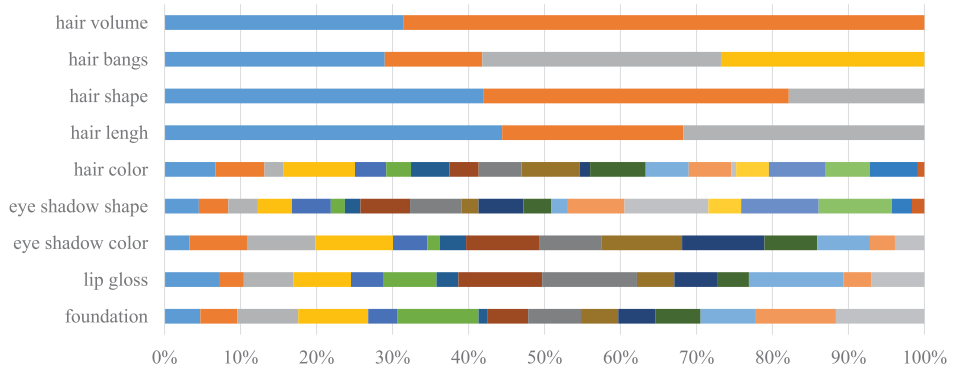

Fig. 3. Statistics of beauty attribute values labeled in the training set.

values of the first four beauty attributes in Table I, are set intuitively. The last five include the shape attributes of an eye shadow template and the color attributes of a hair template, foundation, lip gloss, and eye shadow. The cluster numbers are determined empirically to get the best partition of colors in practice. We show the distribution of values according to each kind of beauty attribute in Figure 3. The vertical axis is the type of beauty attributes, and the horizontal axis is the percentage of each value occurring in the database. The beauty attribute values are roughly uniformly distributed in the database, which indicates that bias was minimized when collecting the data. Next we will introduce the process of extracting the shape and color attributes for respective regions.

2.2.1. Shape Attribute Extraction. Eyes are the main channel by which to convey intrinsic human beauty and mood. Eye makeup is used to emphasize and change attractiveness and moods shown on the face. Similar eye shadow products with different shape patterns may present quite distinct moods [Aucoin 2000]. Therefore it is crucial to extract and analyze the shape attribute values of eye shadow. Eye shadow is not a hard segment on the face, however. It lies on background skin with opacity. The opacity value is called an alpha value in image-matting technology. In this work, we utilize spectral matting [Levin et al. 2007] to extract eye shadow from the enlarged eye shadow region based on eye contour landmarks.

Spectral matting generalizes matting to K layers $\sum_{k=1}^{K} A^k \odot F^k$, where $A^k \in [0, 1]^{w \times h}$ is the alpha channel of the kth image layer $F^k \in \mathbb{R}^{w \times h}$ (w and h are the width and height of the enlarged eye shadow region), and $\odot$ means element-wise product. The summation of each pixel in A^k across all the K layers should be equal to 1, $\sum_{k=1}^{K} A^k = 1_{w \times h}$. The

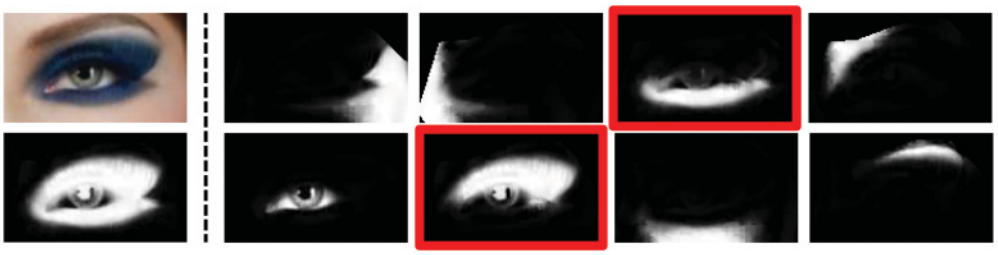

Fig. 4. An example of the extracted eye shadow and its labeling process: the original image, the extracted eye shadow, and the eye shadow component. The selected eye shadow components are marked in red rectangles. Note that eyes may not be correctly segmented away from eye shadow, but they can be easily cropped out by using the detected eye landmarks.

matting component A^k can be solved by finding the smallest eigenvectors of the matting Laplacian matrix [Levin et al. 2007]. The advantage of this method is that the matting process is unsupervised, and foreground and background prior knowledge can be added in the next step.

After the K matting components are extracted, we need to determine whether each matting component belongs to foreground eye shadow or background facial skin, which can be further formulated as a binary labeling problem and solved by graph cut [Boykov and Veksler 2001]. Each matting component A^k is labeled as either foreground eye shadow $f_k = 1$ or background skin $f_k = 0$. An energy function is defined on matting components and their labels, and the optimal labeling is found by minimizing

$$E(f) = \sum_{k=1}^{K} D_k(f_k) + \sum_{m=1}^{K}\sum_{n=1}^{K} V_{mn}(f_m, f_n), \tag{1}$$

where D_k and V_{mn} are the unary and pairwise potentials. For ease of computation, we threshold the matting component $A^k \in [0,1]^{w\times h}$ to obtain binary mask $\hat{A}^k \in \{0,1\}^{w\times h}$. The unary potential models the cost of a matting component A^k belonging to foreground eye shadow $f_k = 1$ or background skin $f_k = 0$ based on location and the skin probability,

$$D_k(f_k) = -\log G_k(f_k) - \log S_k(f_k). \tag{2}$$

Here, $G_k(f_k)$ models the location prior, and $G_k(f_k = 1) = \lambda_1 e^{-dist_g^2(\hat{A}^k, eye)}$, where λ_1 is set as 0.6 empirically and $dist_g(\cdot,\cdot)$ is the smallest Euclidean distance between the center of the binary component mask $\hat{A}^k$ and eye contour. A smaller distance indicates a higher probability of A^k being part of the eye shadow. Similarly, $S_k(f_k)$ is the skin prior, and $S_k(f_k = 0)$ equals the likelihood of the matting component belonging to the skin area, which is the direct output of a skin detector [Jones and Rehg 1999].

The pairwise potential models the similarity of two matting components A^m and A^n,

$$V_{mn}(f_m, f_n) = [f_m \neq f_n] e^{-dist_h^2(F^m, F^n)}, \tag{3}$$

where $[\cdot]$ is Iverson bracket notation, that is, $[\cdot]$ equals 1 if the expression is true, and 0 otherwise, and $dist_h(\cdot,\cdot)$ is the χ^2 distance between the color histograms extracted from F^m and F^n in HSV space, which is more perceptually relevant to human vision than RGB space. One example of the extracted eye shadow and the labeling process is illustrated in Figure 4. The components in the red bounding box are labeled as eye shadow.

We then cluster the extracted eye shadows into several shape templates. Due to facial symmetry, we only consider the left eye in the clustering step. Procrustes analysis [Cootes et al. 1995] is utilized to align the eye landmarks and the mean eye shape

Fig. 5. Visual examples of the specific values for some beauty attributes.

is calculated. After warping all the eye images to the mean eye by thin plate spline method [Bookstein 1989], the cluster centers from k-means are used to represent the eye shadow shape templates, namely, values of the eye shadow shape attribute. To remove noise, Robust Principal Component Analysis [Wright et al. 2009] is applied before clustering.

2.2.2. Color Attributes Extraction. The values of color attributes are obtained by running the k-means clustering algorithm on the training dataset from the corresponding face regions (the cluster number is determined empirically according to each specific attribute). The pixel values within the specific facial regions on each training image are clustered by Gaussian mixture models (GMM) in RGB color space. The centers of the largest GMM components are used as the representative colors. Then the colors are clustered by k-means to obtain the color attributes of makeup templates. We show visual examples of specific attribute values for some beauty attributes in Figure 5.

2.3. Beauty-Related Attributes and Features

A straightforward way of predicting the values of these high-level beauty attributes is using some low-level features to train some classifiers. However, since there is a relatively huge gap between high-level beauty attributes and low-level image features, and because the beauty attributes are intuitively related to some mid-level attributes like eye shape and mouth width, we further explore a set of mid-level beauty-related attributes to narrow the gap between high-level beauty attributes and low-level image features. Table II lists all the mid-level beauty-related attributes annotated for the dataset. These mid-level attributes mainly focus on facial shapes and clothing properties, which are kept fixed during the recommendation and synthesis process.[3]

After the annotation of high-level beauty attributes and mid-level beauty-related attributes, we further extract various types of low-level image features on the clothing and facial regions for each image in the Beauty e-Experts Dataset to facilitate further beauty modeling. The clothing region of an image is automatically determined based on its geometrical relationship with the face region. Specifically, the following features are extracted for image representation:

—RGB color histogram and color moments on clothing region;
—Histograms of oriented gradients (HOG) [Dalal and Triggs 2005] and local binary patterns (LBP) [Ahonen et al. 2006] features on clothing region;
—Active shape model [Cootes et al. 1995] based shape parameters;
—Shape context [Belongie et al. 2002] features extracted at facial points.

All of these features are concatenated to form a feature vector of 7,109 dimensions, principal component analysis (PCA) [Jolliffe 2002] is then performed to reserve 90% of

[3]Although the clothes of a user can be changed for optimal appearance, they are fixed in our current Beauty e-Experts system.

Table II. List of Mid-Level Beauty-Related Attributes Considered in This Work

Names	Values	Names	Values
forehead	high, normal, low	smiling	smiling, neutral
eyebrow	thick, thin	lip thickness	thick, normal
eyebrow length	long, short	fatness	fat, normal
eye corner	upcurved, downcurved, normal	jaw shape	round, flat, pointed
eye shape	narrow, normal	face shape	long, oval, round
nose tip	wide, narrow	mouth opened	yes, no
cheek bone	high, normal	mouth width	wide, normal
nose bridge	prominent, flat	race	Asian, Western
ocular distance	hypertelorism, normal, hypotelorism		
collar shape	strapless, v-shape, one-shoulder, high-necked, round, shirt collar		
clothing pattern	vertical, plaid, horizontal, drawing, plain, floral print		
clothing material	cotton, chiffon, silk, woolen, denim, leather, lace		
clothing color	red, orange, brown, purple, yellow, green, gray, black, blue, white, pink, multi-color		

the energy. The compacted feature vector with 173 dimensions and the annotated attribute values are then fed into an SVM classifier to train a classifier for each attribute.

3. THE RECOMMENDATION MODEL

A training beauty image is denoted as a tuple $(\mathbf{x}, \mathbf{a}^r, \mathbf{a}^b)$. Here, $\mathbf{x}$ is the image features extracted from the raw image data; $\mathbf{a}^r$ and $\mathbf{a}^b$ denote the set of beauty-related attributes and beauty attributes, respectively. Each attribute may have multiple different values, that is, $a_i \in \{1, \ldots, n_i\}$, where n_i is the number of attribute values for the ith attribute. Beauty-related attributes $\mathbf{a}^r$ act as the mid-level cues to bridge between low-level image features $\mathbf{x}$ and high-level beauty attributes $\mathbf{a}^b$. The recommendation model needs to uncover the complex relationships among the low-level image features, mid-level beauty-related attributes, and high-level beauty attributes, to make the final recommendation for a given image.

3.1. Model Formulation

As we can see, there is no ranking information available for different beauty attribute combinations in the training data, therefore a generative model is more appropriate than a discriminative model in this setting. Based on this consideration, we model the relationships among the low-level image features, the mid-level beauty-related attributes, and the high-level beauty attributes from a probabilistic perspective. The aim of the recommendation system is to estimate the probability of beauty attributes, together with beauty-related attributes, given the image features, that is, $p(\mathbf{a}^b, \mathbf{a}^r|\mathbf{x})$, which can be modeled using Gibbs distribution [Wainwright and Jordan 2008],

$$p(\mathbf{a}^b, \mathbf{a}^r|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(-E(\mathbf{a}^b, \mathbf{a}^r, \mathbf{x})), \tag{4}$$

where $Z(\mathbf{x}) = \sum_{\mathbf{a}^b, \mathbf{a}^r} \exp(-E(\mathbf{a}^b, \mathbf{a}^r, \mathbf{x}))$, known as the partition function, is a normalizing term dependent on the image features, and $E(\mathbf{a}^b, \mathbf{a}^r, \mathbf{x})$ is an energy function measuring the compatibility among the beauty attributes, beauty-related attributes, and image features. The beauty recommendation results therefore can be obtained by finding the most likely joint state of the beauty attributes $\hat{\mathbf{a}}^b = \arg\max_{\mathbf{a}^b} \max_{\mathbf{a}^r} p(\mathbf{a}^b, \mathbf{a}^r|\mathbf{x})$.

The capacity of this probabilistic model is entirely dependent on the structure of the energy function $E(\mathbf{a}^b, \mathbf{a}^r, \mathbf{x})$. Here we propose learning a general supergraph structure to build the energy function which could theoretically be used to model any relationships among the low-level image features, mid-level beauty-related attributes,

and high-level beauty attributes. To begin with, we give the definition of the supergraph.

Definition 1 (*Supergraph*). A supergraph $\mathcal{G}$ is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is called supervertexes, consisting of a set of nonempty subsets of a basic node set, and $\mathcal{E}$ is called superedges, consisting of a set of two-tuples, each of which contains two different elements in $\mathcal{V}$.

It can be seen that a supergraph is actually a generalization of a graph in which a vertex can have multiple basic nodes and an edge can connect any number of basic nodes. When all the supervertexes only contain one basic node, and each superedge is forced to connect to only two basic nodes, the supergraph then becomes a traditional graph. A supergraph can be naturally used to model the complex relationships among multiple factors, where the factors are denoted by the vertexes and the relationships are represented by the superedges.

Definition 2 (*k-Order Supergraph*). For a supergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, if the maximal number of vertexes involved with the superedges in $\mathcal{E}$ is k, $\mathcal{G}$ is said to be a k-order supergraph.

Based on these definitions, we propose using the supergraph to characterize the complex relationships among the low-level image features, mid-level beauty-related attributes, and high-level beauty attributes in our problem. For example, the aforementioned pairwise correlations can be sufficiently represented by a two-order supergraph (traditional graph), while other more complex relationships, such as one-to-two and two-to-two relationships, can be represented by other higher-order supergraphs. Denote the basic node set A as the union of the beauty attributes and beauty-related attributes, that is, $A = \{a_i | a_i \in \mathbf{a}^r \cup \mathbf{a}^b\}$. The underlying relationships among all the attributes are represented by a supergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{\mathbf{a}_i | \mathbf{a}_i \subset A\}$[4] is a set of nonempty subsets of A, $\mathcal{E}$ is the superedge set that models their relationships, and the energy function can then be defined as

$$E(\mathbf{a}^b, \mathbf{a}^r, \mathbf{x}) = \sum_{\mathbf{a}_i \in \mathcal{V}} \phi_i(\mathbf{a}_i, \mathbf{x}) + \sum_{(\mathbf{a}_i, \mathbf{a}_j) \in \mathcal{E}} \phi_{ij}(\mathbf{a}_i, \mathbf{a}_j). \tag{5}$$

The first summation term is called FA (feature to attribute) potential, which is used to model the relationships between the attributes and low-level image features; the second is called AA (attribute to attribute) potential and is used to model the complex relationships among different attributes represented by the superedges. $\phi_i(\mathbf{a}_i, \mathbf{x})$ and $\phi_{ij}(\mathbf{a}_i, \mathbf{a}_j)$ are the potential functions of the corresponding inputs, which can be learned in different ways. Generally, the FA potential $\phi_i(\mathbf{a}_i, \mathbf{x})$ is usually modeled as a generalized linear function in the form

$$\phi_i(\mathbf{a}_i = \mathbf{s}_i, \mathbf{x}) = \psi_{\mathbf{a}_i}(\mathbf{x})^\top \mathbf{w}_i^{\mathbf{s}_i}, \tag{6}$$

where $\mathbf{s}_i$ is the set of values for attribute subset $\mathbf{a}_i$, $\psi_{\mathbf{a}_i}(\mathbf{x})$ is a set of feature mapping functions for the attributes in $\mathbf{a}_i$ by using SVM on the extracted features (see Section 2.2), and $\mathbf{w}_i$ is the FA weight parameters to be learned for the model. The AA potential function $\phi_i(\mathbf{a}_i, \mathbf{a}_j)$ is defined by a scalar parameter for each joint state of the corresponding superedge,

$$\phi_{ij}(\mathbf{a}_i = \mathbf{s}_i, \mathbf{a}_j = \mathbf{s}_j) = w_{i,j}^{\mathbf{s}_i \mathbf{s}_j}, \tag{7}$$

[4]Note that in this article, we use $\mathbf{a}_i$ to denote a nonempty attribute set and a_i to denote a single attribute.

where $w_{i,j}^{\mathbf{s}_i\mathbf{s}_j}$ is a scalar parameter for the corresponding joint state of $\mathbf{a}_i$ and $\mathbf{a}_j$ with the specific value $\mathbf{s}_i$ and $\mathbf{s}_j$.

3.2. Model Learning

The learning of the supergraph-based energy function includes learning the structure of the underlying supergraph and the parameters in the potential functions.

3.2.1. Structure Learning. Learning a fully connected supergraph structure is generally an NP-complete problem, which makes finding the optimal solution technically intractable [Koller and Friedman 2009]. However, we can still find many good approximations which can model a very large proportion of all the possible relationships. Among all the possible approximations, tree structure provides a very good choice which can be learned efficiently using many algorithms [Koller and Friedman 2009]. Another merit of tree structure is that the inference on a tree can be efficiently performed using methods such as dynamic programming. Based on these considerations, we therefore use the tree-structured supergraph to model the underlying relationships. To remedy the information loss during the approximation procedure, we further propose simultaneously learning multiple different tree-structured supergraphs to collaboratively model the objective relationships. Learning multiple tree-structured supergraphs should also produce more useful recommendation results, since it is intuitively similar to the daily recommendation scenario. These tree-structured supergraphs could be different recommendation experts, each of which is good at modeling some kind of relationship. The recommendation results generated by these experts are fused to form the final recommendation result.

For a two-order supergraph, learning a tree-structured approximation can be efficiently solved using the maximum spanning tree algorithm [Chow and Liu 1968]. The edge weights in the graph are given by the mutual information between the attributes, which can be estimated from the empirical distribution from the annotations in the training data. For higher-order supergraphs, however, learning their tree-structured approximations will not be a trivial task, since the choices of vertex subsets for each superedge are combinational.

Suppose for a supergraph built on basic node set $A = \{a_1, \ldots, a_M\}$ with M elements, we need to find a k-order tree-structured supergraph for these vertexes. We first calculate the mutual information between each pair of vertexes and denote the results in the adjacency matrix form, that is, $W = \{w_{ij}\}_{1\le i,j\le M}$. Then we propose a two-stage algorithm for finding the k-order tree-structured supergraph.

In the first stage, we aim to find the candidate set of basic node subsets $\mathcal{V} = \{\mathbf{a}_i | \mathbf{a}_i \in \mathcal{V}\}$, which will be used to form the superedges. The objective here is to find the set of subsets that has the largest amount of total mutual information in the resulting k-order supergraph. Here we first define a function that calculates the mutual information of a subset set with a specified mutual information matrix,

$$f(\mathcal{V}, W) = \sum_{|\mathbf{a}_i|\ge 2} \sum_{a_j, a_k \in \mathbf{a}_i} w_{jk}. \tag{8}$$

Based on this definition, we formulate the candidate set generation problem as the following optimization problem

$$\begin{aligned} \underset{\mathcal{V}}{\text{argmax}} \quad & f(\mathcal{V}, W), \\ \text{s.t. } & |\mathbf{a}_i| \le \left\lfloor \frac{k+1}{2} \right\rfloor, \forall i, \\ & |\mathcal{V}| \le m, \end{aligned} \tag{9}$$

ALGORITHM 1: Candidate Set of Subset Generation for Supergraph Structure Learning

Input: basic node set $A=\{a_1, \ldots, a_M\}$, adjacency matrix $W = \{w_{ij}\}_{1\le i,j\le M}$, desired order of the supergraph k.
Output: candidate set of subsets $\mathcal{V} = \{\mathbf{a}_i | \mathbf{a}_i \in A\}$.

Initialization: set $\mathcal{V}$ with m unique subsets randomly collected from A, each of which has no more than $\lfloor (k+1)/2 \rfloor$ elements. Set $w_{\max} = f(\mathcal{V}, W)$.
while not converged **do**
  for $i = 1 \to M$ **do**
    for $j = 1 \to m$ **do**
      $w_j = f(\mathcal{V}, W)$ if move a_i to $\mathbf{a}_j$.
    end for
    $w_l = \text{argmax}_j(\{w_j\})$.
    if $l > w_{\max}$ **then**
      Move a_i to $\mathbf{a}_l$, $w_{\max} = w_l$.
      if $|\mathbf{a}_l| > \lfloor (k+1)/2 \rfloor$ **then**
        Split $|\mathbf{a}_l|$ into two subsets.
        $m \leftarrow m + 1$.
      end if
      if $m > \lceil 2 \times M/(k-1) \rceil$ **then**
        Merge two smallest subsets.
        $m \leftarrow m - 1$.
      end if
    end if
  end for
end while
Generate candidate vertex subsets.

where the first inequation is from the k-order constraint from the resulting supergraph, $\lfloor \cdot \rfloor$ is the floor operator, and parameter m in the second inequation is used to ensure that the generated subsets have a reasonable size to cover all the vertexes and make the inference on the resulting supergraph more efficient. Specifically, its value can be set as

$$m = \begin{cases} M, & k = 2, \\ 2\lceil M/(k-1) \rceil, & \text{otherwise}, \end{cases} \tag{10}$$

where $\lceil \cdot \rceil$ is the ceil operator. To solve this optimization problem, we design a k-means-like iterative optimization algorithm to find the solution. The algorithm first initializes some random vertex subsets and then reassigns each vertex to the subsets that result in maximal mutual information increment; if one vertex subset has more than $\lfloor (k+1)/2 \rfloor$ elements, it will be split into two subsets; if the total cardinality of the vertex subset set is larger than $2\lceil M/(k-1) \rceil$, two subsets with the smallest cardinalities will be merged into one subset. This procedure is repeated until convergence. Algorithm 1 gives the pseudocode description of this procedure.

Based on the candidate vertex subsets, the second stage of the two-stage algorithm first calculates the mutual information between the element pair that satisfies the order restrictions in each vertex subset. Then it builds a graph by using the calculated mutual information as an adjacency matrix, and the maximum spanning tree algorithm [Chow and Liu 1968] is adopted to find its tree-structured approximation.

This two-stage algorithm is run many times by setting different k values and initializations of subsets, which then generates multiple tree-structured supergraphs with different orders and structures. In order to make parameter learning in the following tractable, the maximal k-value K is set to equal 5. The detailed description of this process is summarized in Algorithm 2.

ALGORITHM 2: Learning Multiple Tree-Structured Supergraphs

Input: basic node set $A = \{a_1, \ldots, a_M\}$, adjacency matrix $W = \{w_{ij}\}_{1 \le i,j \le M}$, number of desired supergraphs T.
Output: T tree-structured supergraphs $\mathbf{G} = \{\mathcal{G}_t\}_{t=1}^{T}$.
1: **Initialization**: set $\mathbf{G} = \emptyset$, $K = 5$.
2: **for** $t = 1 \rightarrow T$ **do**
3: Generate a random variable $k \in \{2, \ldots K\}$.
4: Obtain a candidate vertex subsets $\mathcal{V}$ using Alg. 1.
5: Calculate the mutual information between the elements pair with no more than k vertexes in $\mathcal{V}$.
6: Make a graph using the calculated mutual information as adjacency matrix.
7: Find its maximal spanning tree using the algorithm in Chow and Liu [1968].
8: Form the k-order tree-structured supergraph $\mathcal{G}_t$.
9: $\mathbf{G} \leftarrow \mathbf{G} \cup \{\mathcal{G}_t\}$.
10: **end for**
11: Generate tree-structured supergraph set $\mathbf{G}$.

3.2.2. Parameter Learning and Inference. For each particular tree-structured supergraph, its parameter set, including the parameters in the FA potentials and the AA potentials, can be denoted as a whole as $\Theta = \{\mathbf{w}_i^{\mathbf{s}_i}, w_{ij}^{\mathbf{s}_i \mathbf{s}_j}\}$. We adopt the maximal likelihood criterion to learn these parameters. Given N independent and identically distributed training samples $\mathbf{X} = (\mathbf{x}_n, \mathbf{a}_n^r, \mathbf{a}_n^b)$, we need to minimize the loss function

$$\begin{aligned} \mathcal{L}(\Theta|\mathbf{X}) &= \sum_{n=1}^{N} \mathcal{L}_n(\Theta|\mathbf{x}_n) + \frac{1}{2}\lambda \sum_{i,\mathbf{s}_i} \left\| \mathbf{w}_i^{\mathbf{s}_i} \right\|_2^2 \\ &= \sum_{n=1}^{N} \left\{ -\ln p\left(\mathbf{a}_n^b, \mathbf{a}_n^r | \mathbf{x}_n\right) \right\} + \frac{1}{2}\lambda \sum_{i,\mathbf{s}_i} \left\| \mathbf{w}_i^{\mathbf{s}_i} \right\|_2^2, \end{aligned} \tag{11}$$

where λ is the trade-off parameter between the regularization term and log-likelihood. Since the energy function is linear with respect to the parameters, the log-likelihood function is concave, and the parameters can be optimized using gradient-based methods. The gradient of the parameters can be computed by calculating their marginal distributions [Mensink et al. 2013]. Denoting the value of attribute $\mathbf{a}_i$ for training image n as $\hat{\mathbf{a}}_i$, we have

$$\frac{\partial \mathcal{L}_n}{\partial \mathbf{w}_i^{\mathbf{s}_i}} = ([\hat{\mathbf{a}}_i = \mathbf{s}_i] - p(\mathbf{a}_i = \mathbf{s}_i | \mathbf{x}_n)) \psi_{\mathbf{a}_i}(\mathbf{x}_n), \tag{12}$$

$$\frac{\partial \mathcal{L}_n}{\partial w_{ij}^{\mathbf{s}_i \mathbf{s}_j}} = [\hat{\mathbf{a}}_i = \mathbf{s}_i, \hat{\mathbf{a}}_j = \mathbf{s}_j] - p(\mathbf{a}_i = \mathbf{s}_i, \mathbf{a}_j = \mathbf{s}_j | \mathbf{x}_n). \tag{13}$$

Based on the calculation of the gradients, the parameters can be learned from different gradient-based optimization algorithms [Koller and Friedman 2009]. In the experiments, we use the implementation by Schmidt[5] to learn these parameters. The learned parameters, together with the corresponding supergraph structures, form the final recommendation model.

Here each learned tree-structured supergraph model can be seen as a beauty expert. Given an input testing image, the system first extracts the feature vector $\mathbf{x}$, and then each beauty expert makes its recommendation by performing inference on the tree

[5] http://www.di.ens.fr/~mschmidt/Software/UGM.html.

structure to find the maximum posteriori probability of $p(\mathbf{a}^b, \mathbf{a}^r|\mathbf{x})$. The recommendation results output by all the Beauty e-Experts are then fused by majority voting to make the final recommendation to the user.

3.3. Relations with Other Models

The proposed multiple tree-structured supergraph model characterizes the complex relationships among different attributes from a probabilistic prospective. When the maximal order value K is set to 2, our model degenerates into the classical graphical model used by most previous works [Boykov and Veksler 2001; Wang et al. 2012; Mensink et al. 2013], where only the one-to-one pairwise correlations between two attributes are considered when modeling complex relationships. Our model can generally model any order of relationships. When the maximal order value K of the supergraph is set to 5, many other types of relationships, for example, one-to-two, two-to-two, and two-to-three, can be simultaneously modeled.

The pairwise correlations are also extensively modeled using the latent SVM model [Wang and Mori 2010] from a deterministic perspective, which has been successfully applied into the problem, such as object detection [Felzenszwalb et al. 2010], pose estimation [Yang and Ramanan 2011], image classification [Wang and Mori 2010], as well as clothing recommendation [Liu et al. 2012]. Compared with the latent SVM model, our tree-structured supergraph model can not only consider much more complex relationships among the attributes, but also is more efficient, since tree structures make both the learning and the inference process much faster. For a tree-structured model with n nodes and k different values for each node, the time complexity of the inference process is only $O(k^2n)$, while a fully connected model (e.g., latent SVM) has the complexity of $O(k^n)$. Actually, during the training of the latent SVM model, some intuitively small correlations have to be removed manually to accelerate the training process. Our tree-structured supergraph model, on the contrary, can automatically remove the small relationships during the structure learning process in a principled way. By extending to multiple tree-structured supergraphs, our model can produce much more reliable and useful recommendations, as verified in the experimental part, since it can well simulate the common recommendation scenario in our daily life, where one usually asks many people for recommendations and takes the majority or the most suitable one as the final choice.

4. THE SYNTHESIS MODULE

With the beauty attributes recommended by the multiple tree-structured supergraph model, we further synthesize the final visual effect of hairstyle and makeup for the testing image. To this end, the system first uses beauty attributes to search for hair and makeup templates. A hair template is a combination of hairstyle attributes, such as long curls with bangs. We use the recommended hairstyle attributes to search the Beauty e-Experts Database for suitable hair templates. As mentioned in Section 2.2, each hair template is extracted from a training image. If more than one template is obtained, we randomly select one from them. If we cannot find the hair template with exactly the same hairstyle attribute values, we use the one which has the values that most approximate the recommended hairstyle attribute values. Each makeup attribute forms a template which can be directly obtained from the dataset. These obtained hair and makeup templates are then fed into the synthesis process, which mainly has two steps: alignment and alpha blending, as shown in Figure 6.

In the alignment step, both the hairstyle and makeup templates need to be aligned with the testing image. For hair template alignment, a dual linear transformation procedure is proposed to put the hair template on the target face in the testing image.

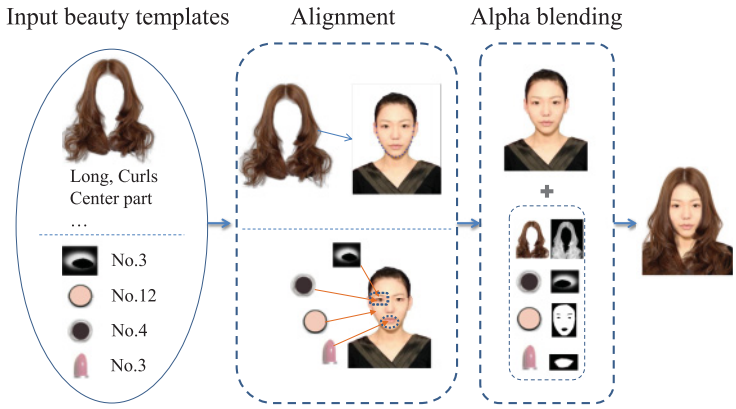

Fig. 6. The flowchat of the synthesis module.

The dual linear transformation process first uses a linear affine transformation to perform rough alignment and then adopts a piecewise-linear affine transformation [Goshtasby 1986] to perform precise alignment. In the linear affine transformation, the 21 face contour points generated by the face analyzer are adopted to calculate an affine transformation matrix between the hair template and the testing face. The hair template then can be roughly aligned to the testing face using the transformation matrix. In piecewise-linear affine transformation, three subsets of keypoints, namely, the inner subset, the middle subset, and the outer subset, are sampled based on the result of rough alignment. Eleven points are sampled interlacedly from 21 face contour points to consist the inner subset. The points in the inner subset are extended on the horizontal direction with half of the distance between two eye centers. They form the middle subset of 8 points. Ten points in the outer subset are fixed on the edge of the image. Their coordinates are determined by the image corners or the horizontal lines of eye centers and mouth corners. Note that points of the middle and outer subsets are at the same position in both the hair template and the testing image, which aims to keep the overall shape of the hairstyle. The total 29 points in the three subsets are then used to construct a Delaunay triangulation [Berg et al. 2008] to obtain 46 triangles. Then affine transformations are applied within the corresponding triangles between the testing face and the hair template. After that, these points on the hair template are precisely aligned with the testing face.

For the makeup template alignment, only the eye shadow template needs to be aligned to the eye region in the testing image. Other makeup templates can be directly applied to the corresponding regions based on the face keypoint detection results. To align the eye shadow template to contour the eye on the face, we use the thin plate spline method [Bookstein 1989] to warp the eye shadow template by using the eye contour points. Because the eye shadow template attributes are learned by clustering from the left eye, the left template is mirrored to the right to obtain the right eye shadow template.

In the alpha-blending step, the final result R is synthesized with the hair template, makeup, and the testing face I. The synthesis process is performed in CIELAB color space. L^* channel is considered as lightness because of its similarity to human visual perception. a^* and b^* are the color channels. We first use the edge-preserving operator on image lightness channel L^* to imitate the smoothing effect of foundation. We choose the guided filter [He et al. 2010], which is more efficient and has better performance near the edges among all the edge-preserving filters. It is applied to the L^* channel

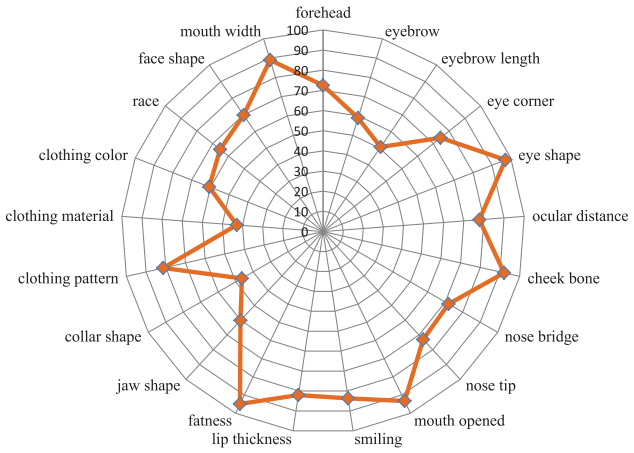

Fig. 7. Accuracies of the predicted beauty attributes from the SVM classifier.

of the facial region determined by facial contour points. Note that since we do not have contour points on the forehead, the forehead region is segmented out by GrabCut [Rother et al. 2004]. The final synthesis result is generated by alpha-blending the testing image I and hair and makeup template T in the L^*, a^*, and b^* channels, respectively,

$$R = \alpha I + (1 - \alpha)T, \tag{14}$$

where α is a weight to balance I and T. For hair and eye shadow templates, the value of α is obtained from the templates themselves. For foundation and lip gloss, the α value is set to 0.5 for L^* channel, and 0.6 for a^* and b^* channels.

5. EXPERIMENTS

In this section, we design experiments to evaluate the performance of the proposed Beauty e-Experts system from different aspects. We first visualize and analyze the intermediate result of model learning processing. Then the recommendation result is evaluated by comparison with several baselines, such as latent SVM [Wang and Mori 2010], multiclass SVM [Chang and Lin 2011], and neural network [Haykin 1999]. The synthesis effects are finally presented and compared with some commercial systems related to hairstyle and makeup recommendation and synthesis.

5.1. Model Learning and Analysis

We look into some intermediate aspects for deep insight into the recommendation model structure and learning process. In Figure 7, we present the accuracy of predicted beauty-related attributes from the SVM classifier, which is used to build the FA potential function in the energy function to model the recommendation distribution (see Eq. (5)). Note that we do not present the accuracy of beauty attributes, for we cannot obtain their ground truth label in this stage. It can be seen that most classifiers have accuracies of more than 70%. It is sufficient to provide enough information to predict beauty attributes. Clothing-related attributes have accuracy of 40%∼50%, which is a little bit low. This is mainly caused by the large number of categories of clothing-related attributes. By analyzing the errors of clothing-related attributes, we have found that most of the errors are classifying an attribute value to other quite similar values. This is unlikely to have a serious effect on the final recommendation results.

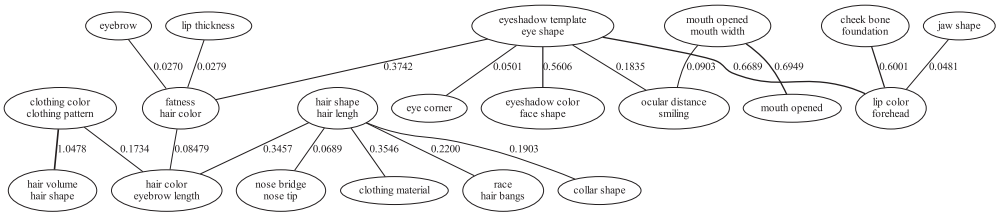

Fig. 8. Visualization of one learned tree-structured supergraph model.

We also visualize one example of the learned tree structure in the recommendation model in Figure 8. This tree structure is of order 4, and each supervertex can only include two attributes at most (see Algorithm 1). The weight of the superedge represents the mutual information between two related supervertexes. From the results, we make some observations. First, meaningful relationships are learned as shown in the tree structure. The superedges between supervertex "hair shape, hair length" and the other five supervertexes are retained, while the superxsvertex "eye corner" only retains one superedge with other supervertexes. This means that "hair shape, hair length" is more important and has broader relationships with other nodes than "eye corner" in this structure. Second, some highly correlated attributes are clustered into one supervertex, such as "hair shape" with "hair length" and "eye shadow template" with "face shape." This fits well with the intuitive perception of humans. Long hair may match well with curled hair, and certain shapes of eye shadow templates may also fit to some face shapes. Third, the correlation between supervertexes is represented on the superedges. The super-vertex "eye shadow template, eye shape" has weight 0.5606 with "eye shadow color, face shape," which is higher than the weight 0.0501 with "eye corner." This means that "eye shadow template, eye shape" has a stronger correlation with "eye shadow color, face shape."

5.2. Recommendation Results Evaluation

For the recommendation model in the Beauty e-Experts system, we also implement some alternatives using multiclass SVM, neutral network, and latent SVM. The first two baselines only use the low-level image features to train classifiers for high-level beauty attributes. The latent SVM baseline considers not only the low-level image features but also the pairwise correlations between beauty and beauty-related attributes. We use the 100 testing images to evaluate the performance of the three baseline methods and our algorithm. To evaluate the recommendation result of the Beauty e-Experts system quantitatively, human perception of suitable beauty makeup is considered as the ground truth measured on 50 random combinations of the attributes for all 100 testing images. We asked 20 participants (staff and students in our group) to label the ground truth of ranking results of the 50 types of beauty makeup effects for each testing image. Instead of labeling absolute ranks from 1 to 50, we use a k-wise strategy similar to that of Liu et al. [2012]: labelers are shown k images as a group each time, where k is set to 10. They only need to rank satisfying levels within each group. $C(k, 2)$ pairwise preferences can be obtained from the k ranks, and then the final rank is calculated across groups by ranking SVM [Joachims 2002].

In Figure 9, we plot the comparison results of our model and other baselines. The comparison between a multiple tree-structured supergraph model and single tree-structured supergraph model (denoted as "Single-tree") is also included in the figure. The single tree-structured supergraph model we select is the one with the highest recommendation performance. The performance is measured by normalized discounted cumulative gain (NDCG) [Siddiquie et al. 2011], which is widely used to evaluate

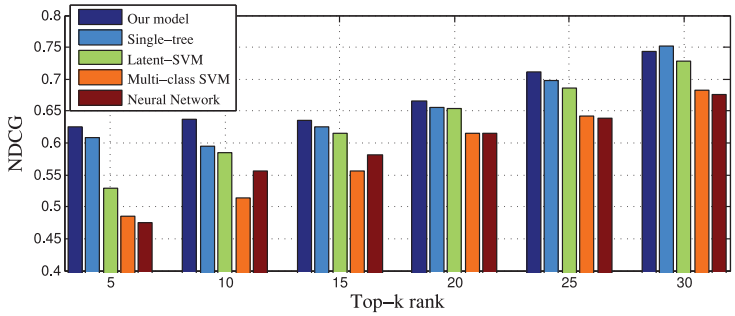

Fig. 9. NDCG values of the multiple tree-structured supergraphs model and other baselines. The horizontal axis is the rank of top-k results, while the vertical axis is the corresponding NDCG value. Our proposed method achieves better performance than the single tree-structured supergraph model, the latent SVM model, and other baselines.

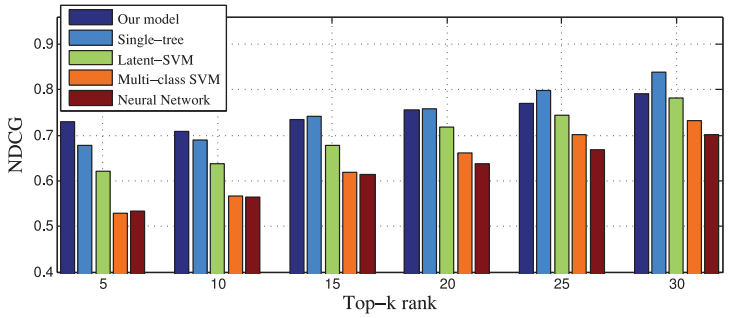

Fig. 10. Self-evaluation performance of the multiple tree-structured supergraphs model and other baselines. The horizontal axis is the rank of top-k results, while the vertical axis is the corresponding NDCG value.

ranking systems. From the results, we can observe that the multiple tree-structured supergraph model, single tree-structured supergraph model, and latent SVM significantly outperform multiclass SVM and neural network. This is mainly because our models and the latent SVM method are equipped with mid-level beauty-related attributes to narrow the semantic gap between low-level image features and high-level beauty attributes. These models are able to characterize the co-occurrence information to mine the pairwise correlations between every two factors. From Figure 9 we can further find that our model has overall better performance than the latent SVM method, especially in the top 5 recommendations. With higher-order relationships embedded, our model can express more complex relationships among different attributes. What is more, multiple tree-structured supergraphs shows better performance than single tree-structured supergraphs. By integrating multiple supergraph information, our model tends to obtain more robust recommendation results.

We further do the self-evaluation experiment to verify whether female users are satisfied with the recommendation results. We asked ten females (including staff and students in our university) to provide high-quality photos of themselves, and tested on these photos using our system. We then labeled these photos as in the preparation of the training set. The evaluation criterion is similar with those previously mentioned. The comparison results are shown in Figure 10. From the results, we can observe that our proposed tree-structured supergraph model consistently outperforms the latent SVM, multiclass SVM, and neural network.

Fig. 11. Comparison results between Guo and Sim [2009] and ours. The first column and the second column show the original testing image and reference image used by Guo and Sim [2009]. The other columns show the results from Guo and Sim [2009] and our proposed method, respectively. It can be seen that our method has much better visual effects than those from the method of Guo and Sim [2009]. (Best viewed in $\times 3$ size of original color PDF file.)

5.3. Synthesis Results Evaluation

5.3.1. Synthesis Results Evaluation with Other Research Works. Since there is no previous work focusing on hair synthesis in the multimedia area, we only compare the makeup effects between our method and that of Guo and Sim [2009]. The methods of Tong et al. [2007] and Scherbaum et al. [2011] require before-and-after makeup image pairs or 3D information, which is more restricted than the method of Guo and Sim [2009], and thus we do not compare with these two methods. The first and second columns of Figure 11 show the original testing image and the reference image which is only applied for the method of Guo and Sim [2009]. The reference image is selected based on the similarity of facial shapes between the testing image and the reference image. The other columns show the results from Guo and Sim [2009] and our algorithm, respectively.

The results show that our method has better visual effects. First, it can be seen that our method precisely synthesizes the makeup effects on the lips and eyes, but the results from Guo and Sim [2009] have obvious artifacts due to its usage of thin plate spline to holistically warp the facial points between the reference image and testing image. However, in real cases, the reference image and testing image may have quite different appearances caused by facial expressions and other factors. For the given reference face image, the appearance of a closed mouth is transferred to an opened mouth, which causes the unexpected artifacts.

Second, the eye shadow from the method of Guo and Sim [2009] is unclear. The reason being that Guo and Sim [2009] assume makeup mainly exists in the decomposed detail layer. It only transfers the detail layer between the testing and reference image, leaving the structure layer untouched. However, this assumption may not always be correct. For example, smoky eye shadow is significant in the L^* channel, and its energy is mainly decomposed into the structure layer. If the structure layer is not transferred, the synthesized eye shadow will be very weak.

We perform a user study to compare the synthesized visual effects generated by our method with those of Guo and Sim [2009] on the testing set. The comparison mainly focused on the quality of visual effects and the degree of harmony. To generate the results of Guo and Sim [2009], we use each testing face to search the training set for the nearest neighbor based on raw features $\mathbf{x}$, which is used as the reference image. For our method, we apply the rank-1 recommended makeup attributes for each testing face. The same 20 participants as for our previous experiment were invited to compare each pair of results from the two methods, which were shown to them in random order. The comparison results of our method and Guo and Sim [2009] were divided into five degrees: "much better," "better," "same," "worse," and "much worse." The percentage of respective chosen degree is calculated, as shown in Figure 12. About 66% of results generated by our method are better or much better than those of Guo and Sim [2009], which well demonstrates the effectiveness of our method.

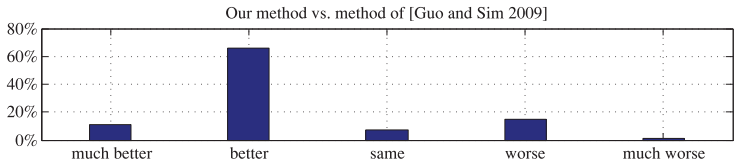

Fig. 12. Synthesis effects comparison of our method and Guo and Sim [2009] on the testing set. The horizontal axis is the degree of comparison, and the vertical axis is the percentage of the testing images for each degree.

Table III. Comparisons of Several Popular Hairstyle and Makeup Synthesis Systems

	VH	IHM	DM	VMT	TAAZ	Ours
hairstyle	✓	✓	✓	×	✓	✓
makeup	×	✓	✓	✓	✓	✓
face detection	×	✓	✓	×	✓	✓
easy of use	×	×	✓	×	×	✓
500+ templates	×	×	×	×	✓	✓
composition freedom	×	✓	×	✓	✓	✓
recommendation	×	×	×	×	×	✓

5.3.2. Synthesis Results Evaluation with Commercial Systems. We compare our Beauty e-Experts system with some commercial virtual hairstyle and makeup systems, including Virtual Hairstyle (VH)[6], Instant Hair Makeover (IHM)[7], Daily Makeover (DM)[8], Virtual Makeup Tool (VMT)[9], and the virtual try-on website TAAZ.[10] They are all very popular among female customers on the Internet.

We first compare these systems in an overview manner, which means that we focus on the comparison of the main functionalities among these systems. The comparison results are summarized in Table III. It can be seen that IHM, DM, and TAAZ systems can provide both hairstyle and makeup synthesis functions. They also provide face detection, which can largely reduce the manual workload. IHM, VMT, and TAAZ ask users to choose makeup and hair products to perform composition, while VH and DM cannot support this, since their methods are mainly based on holistic transformation between the testing face and the example template. However, all these systems cannot support a large dataset with more than 500 templates and do not provide hairstyle and makeup recommendation functions. In contrast, our Beauty e-Experts system can support all the functions just mentioned. What is more, it is fully automatic and can work in more general cases. The recommendation function of our system is really useful for helping female users choose suitable hairstyle and makeup products.

We then compare the hairstyle and makeup synthesis results with these commercial systems. As shown in Figure 13, the first column is the testing images, and the other four columns are the results generated by DM, IHM, TAAZ, and our system, respectively. The reason why we selected these three systems is that only these three can synthesize both the hairstyle and makeup effects. It is seen that the results from these three websites have obvious artifacts. The selected hair templates cannot cover the original hair area. IHM cannot even handle the mouth-opened cases. We further present several testing faces with different hairstyle and makeup effects in Figure 14.

[6]http://www.hairstyles.knowage.info.
[7]http://www.realbeauty.com/hair/virtual/hairstyles.
[8]http://www.dailymakeover.com/games-apps/games.
[9]http://www.hairstyles.knowage.info.
[10]http://www.taaz.com.

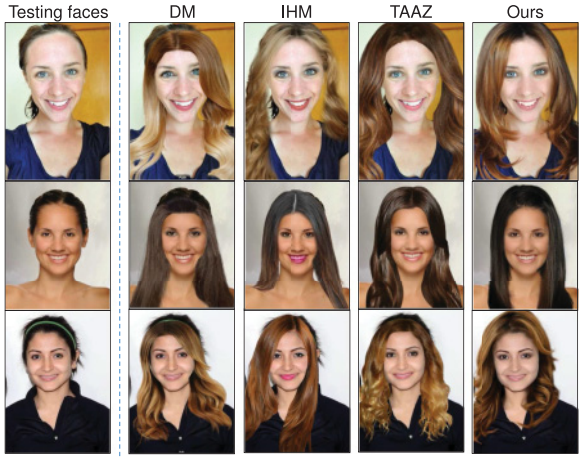

Fig. 13. Contrast results of synthesized effect among websites and our system.

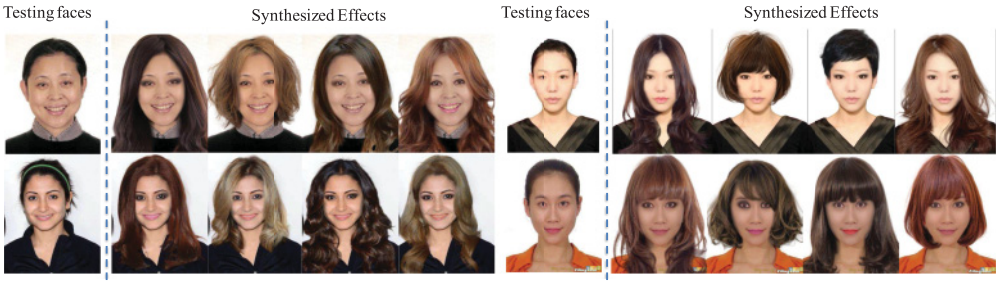

Fig. 14. More synthesized results of the proposed Beauty e-Experts system.

These results still look natural with a variety of style changes, which demonstrates the robustness of our system. Since our supergraph model is tree-structured, the learning and testing process is very efficient. In current implementation, the multiple tree-structured supergraphs are learned in about two hours for each. The test of one face image takes only 2∼3 seconds on a PC with 3GHz CPU and 8G memory.

6. CONCLUSIONS AND FUTURE WORK

In this work, we have developed the Beauty e-Experts system for automatic makeover recommendation and synthesis. To the best of our knowledge, it is the first study to investigate into a fully-automatic system that simultaneously deals with hairstyle and makeup recommendation and synthesis. Based on the proposed multiple tree-structured supergraph model, our system can capture the complex relationships among the different attributes and produce reliable and explainable recommendation results. The synthesis model in our system also produces nature and appealing results. Extensive experiments on a newly built dataset have verified the effectiveness of our recommendation and synthesis models.

The current system is definitely not perfect yet. We are planning to further improve the system in the following directions. First, we may further consider segmenting the hair region and filling in the uncovered region with example-based inpainting techniques [Criminisi et al. 2004]. Second, we may extend the current system for male users by constructing a male dataset. Finally, we also plan to extend the current

system to perform occasion-aware and personalized recommendation by introducing more occasion-related attributes [Liu et al. 2012] and employing information collected from personal photo albums in a user's social network.

REFERENCES

Timo Ahonen, Abdenour Hadid, and Matti. Pietikainen. 2006. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*

K. Aucoin. 2000. *Face Forward*. Little, Brown Company.

Serge Belongie, Jitendra Malik, and Jan Puzicha. 2002. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*

Mark Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. 2008. *Computational Geometry: Algorithms and Applications* (3rd ed.). Springer-Verlag.

Fred Bookstein. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*

Yuri Boykov and Olga Veksler. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*

Fangmei Chen and David Zhang. 2010. A benchmark for geometric facial beauty study. In *Proceedings of the International Conference on Medical Biometrics*.

C. Chow and C. Liu. 1968. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory.*

Timothy Cootes, Christopher Taylor, David Cooper, and Jim Graham. 1995. Active shape models-their training and application. *Comput. Vision Image Understand.*

Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.*

Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. 2010. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*

Ardeshir Goshtasby. 1986. Piecewise linear mapping functions for image registration. *Pattern Recogn.*

Dong Guo and Terence Sim. 2009. Digital face makeup by example. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Simon Haykin. 1999. *Neural Networks*. Prentice Hall.

Kaiming He, Jian Sun, and Xiaoou Tang. 2010. Guided image filtering. In *Proceedings of the European Conference on Computer Vision*.

Gary Huang, Manu Ramesh, Tamara Berg, and Erik Learned. 2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report. University of Massachusetts.

Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Ian Jolliffe. 2002. *Principal Component Analysis*. In *Encyclopedia of Statistics in Behavioral Science*, Springer.

Michael J. Jones and James M. Rehg. 1999. Statistical color models with application to skin detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Daphne Koller and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Anat Levin, Alex Rav-Acha, and Dani Lischinski. 2007. Spectral matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Luoqi Liu, Hui Xu, Junliang Xing, Si Liu, Xi Zhou, and Shuicheng. Yan. 2013. "Wow! You are so beautiful today!" In *Proceedings of the 21st ACM International Conference on ACM Multimedia*.

Si Liu, Jiashi Feng, Zheng Song, Tianzhu Zhang, Hanqing Lu, Changsheng Xu, and Shuicheng Yan. 2012. "Hi, magic closet, tell me what to wear". In *Proceedings of the 20th ACM International Conference on ACM Multimedia*.

T. Mensink, J. Verbeek, and G. Csurka. 2013. Tree-structured CRF Models for interactive image labeling. *IEEE Trans. Pattern Anal. Mach. Intell.*

Yuki Nagai, Kuniko Ushiro, Yoshiro Matsunami, Tsuyoshi Hashimoto, and Yuusuke Kojima. 2005. Hairstyle suggesting system, hairstyle suggesting method, and computer program product. U.S. Patent US20050251463 A1. Filed May 5, 2005; Accepted November 10, 2005.

Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*

Kristina Scherbaum, Tobias Ritschel, Matthias Hullin, Thorsten Thormählen, Volker Blanz, and Hans Seidel. 2011. Computer-suggested facial makeup. *Comput. Graph. Forum* 30, 2.

Behjat Siddiquie, Rogério Schmidt Feris, and Larry Davis. 2011. Image ranking and retrieval based on multi-attribute queries. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Wai Tong, Chi Tang, Michael Brown, and Ying Xu. 2007. Example-based cosmetic transfer. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*.

Martin J. Wainwright and Michael I. Jordan. 2008. *Graphical Models, Exponential Families, and Variational Inference*. Foundations and Trends® in Machine Learning.

Nan Wang, Haizhou Ai, and Feng Tang. 2012. What are good parts for hair shape modeling? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Yang Wang and Greg Mori. 2010. A discriminative latent model of object classes and attributes. In *Proceedings of the European Conference on Computer Vision*.

John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. 2009. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Proceedings of the Neural Information Processing Systems Conference*.

Yi Yang and Deva Ramanan. 2011. Articulated pose estimation with flexible mixtures-of-parts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Received February 2014; revised June 2014; accepted June 2014