iGlasses: A Novel Recommendation System for Best-fit Glasses

Xiaoling Gu Lidan Shou Pai Peng Ke Chen Sai Wu Gang Chen
College of Computer Science and Technology, Zhejiang University, China
{belizabeth, should, pengpai_sh, ck, wusai, cg}@zju.edu.cn

ABSTRACT

We demonstrate *iGlasses*, a novel recommendation system that accepts a frontal face photo as the input and returns the best-fit eyeglasses as the output. As conventional recommendation techniques such as collaborative filtering become inapplicable in the problem, we propose a new recommendation method which exploits the implicit matching rules between human faces and eyeglasses. We first define fine-grained attributes for human faces and frames of glasses respectively. Then, we develop a recommendation framework based on a probabilistic graphical model, which effectively captures the correlation among these fine-grained attributes. Ranking of the frames (glasses) is done by their similarity to the query facial attributes. Finally, we produce a synthesized image for the input face to demonstrate the visual effect when wearing the recommended glasses.

Keywords

Eyeglasses Recommendation; Probabilistic Graphical Model

1. INTRODUCTION

People will not stop pursuing beauty and fashion. Everyone wants to be attractive and gains self-confidence in the presence of other people. Among all fashionable accessories, eyeglasses are probably one of the most important tools to show one's temperament. However, people often face the problem of selecting a right pair of eyeglasses that fit his/her face shape and hairstyle the most. The main difficulty is that everyone is unique and has distinctive appearance which makes it hard to refer to others' recommendations. Moreover, making the right choice requires prior knowledge of understanding style and color matching between glasses and one's face. For instance, rectangular shape of eyeglasses can make a round face appear thinner and longer. To address the above problem, we develop a novel mobile APP named *iGlasses*, which recommends the best-fit eyeglasses based on one's facial traits like a domain expert.

SIGIR '16, July 17-21, 2016, Pisa, Italy

DOI: http://dx.doi.org/10.1145/2911451.2911453

Most state-of-the-art recommender systems rely on collaborative filtering or content-based recommendation techniques [1], which either learn users' preferences from their ratings or identify user's interests based on item descriptions and historical data. For example, Netflix can recommend movies similar to those previously viewed by a user, while Amazon can recommend goods for one based on ratings given by similar users. Although these systems work well for items such as movies and books, their underlying techniques are not applicable in the context of eyeglasses recommendation, due to two reasons: First, unlike movies or books, the eyeglasses rating data and historical purchase data are hard to acquire. Second, learning user's preferences or interests is not suitable for eyeglasses recommendation, as people choose eyeglasses by considering their own appearences (face shape, hairstyle, skin color, etc.).

In this demonstration, we present *iGlasses*, a novel prototype which is able to automatically recommend personalized eyeglasses for users. *iGlasses* takes a face photo from a smartphone as input and returns a ranked list of eyeglasses. When the user selects a particular recommendation result, the prototype performs image synthesis to demonstrate the visual effect of the eyeglasses on the users's face. Specifically, we propose a novel recommendation method, which relies on a probabilistic graphical model capturing the correlation among facial attributes and eyeglasses attributes. Compared with conventional recommendation methods, our proposed method (1) does not require collecting user rating data or historical data as the training set; (2) does not suffer from cold start or sparsity problem; (3) provides personalized recommendation by mining the matching patterns between human faces and eyeglasses. Preliminary experiments show that our approach outperforms baseline methods in most cases.

2. SYSTEM STRUCTURE

The *iGlasses* recommendation framework is illustrated in Figure 1. Given a run-time query, namely a frontal face image, its *facial attributes* are predicted from a set of classifiers, which are pre-trained in the offline phase. The predicted facial attributes are then fed into the *recommendation model* to produce a set of matched attributes of frames (namely *frame attributes*). Such frame attributes are used to generate a ranked list of candidate glasses in the *retrieval module*. Finally, the *synthesis module* crafts synthesized images from the face image and candidate glasses images as the representation of final recommendation results. In what follows, we provide a brief overview of the major system components.

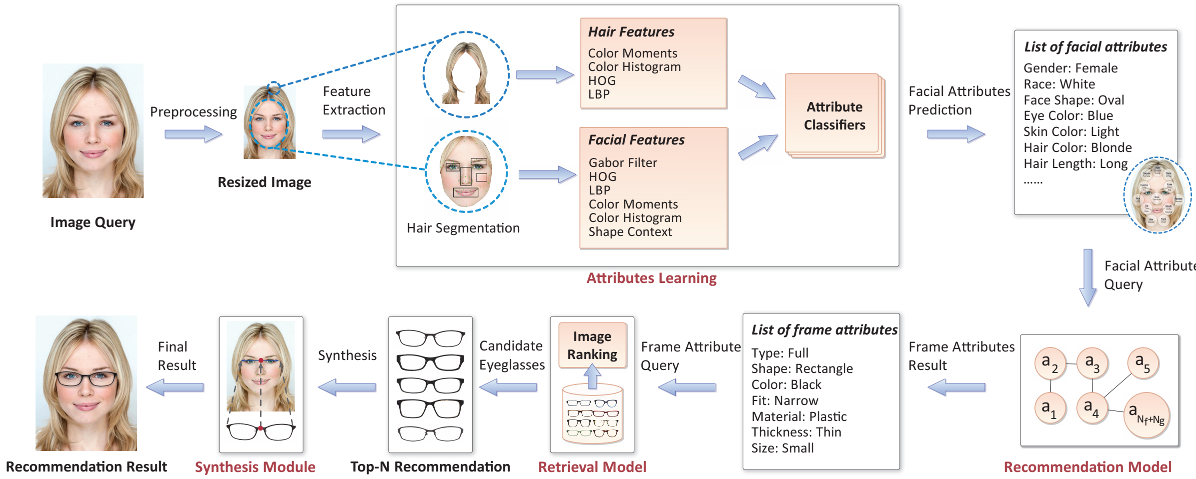


Figure 1: Framework Structure of *iGlasses*

2.1 Facial Attributes Learning

We utilize 17 fine-grained facial attributes to construct user profiles, including *gender*, *race*, *eyebrow thickness*, *eyebrow length*, *eyes shape*, *eyes color*, *nose bridge*, *mouth width*, *smiling*, *lip thickness*, *lip color*, *skin color*, *jaw shape*, *face shape*, *fatness*, *hair color* and *hair length*. We focus on the facial attributes as they include rich information of people and act as the main influence factors in the context of eyeglasses recommendation.

For facial attribute extraction and learning, we make use of the adaptive framework proposed by Kumar et al. [3], where SVM and Adaboost are combined to automatically select on screen space the face region and feature type for each facial attribute. Firstly, all face images are preprocessed before feature extraction. We align faces to a canonical pose by choosing the corners of both eyes as two fiducial points. Then, all face images are cropped and resized to a fixed width by running Viola Jones face detector [6]. To extract hair features, hair segmentation is done by using the GrabCut algorithm [5]. Once the feature extraction is completed, a set of SVM classifiers with an RBF kernel are trained for each attribute by extracting visual features (e.g., Gabor Filter, HOG, Color Moments, Color Histogram, LBP, Shape Context) from different face components (e.g., whole face, eyes, nose, mouth). Experiments on our face photo dataset achieve an average accuracy higher than 80% in detecting the facial attributes.

2.2 Recommendation Model

Our recommendation model is based on a probabilistic graphical model where complex relationships among facial attributes and frame attributes are explored. Similar to the facial attributes, we define 7 discriminative frame attributes, namely *type*, *shape*, *color*, *fit*, *material*, *thickness* and *size*. Specifically, we construct a tree-structured conditional random fields model[4] where each node represents an attribute, and each edge represents the mutual dependencies between two nodes. The reason for using tree model is that inference is easily tractable and can be learned efficiently.

A training image is denoted by a vector of attributes $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g\}$, where $\boldsymbol{\alpha}^f = \{a_1^f, ..., a_{N_f}^f\}$, $\boldsymbol{\alpha}^g = \{a_1^g, ..., a_{N_g}^g\}$, N_f is the number of facial attributes (in our case is 17) and N_g is the number of frame attributes (in our case is 7). Each attribute a_i has multiple possible values, i.e. $a_i \in \{1, ..., n_i\}, i \geqslant 1$. Given an image $\boldsymbol{x}$, the joint probability of a specific configuration $\boldsymbol{\alpha}$ can be represented as:

$$p(\boldsymbol{\alpha}|\boldsymbol{x}) = p(\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} exp(-E(\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g, \boldsymbol{x}))) \quad (1)$$

where $Z(x) = \sum_{\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g} exp(-E(\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g, \boldsymbol{x}))$ is the partition function, and $E(\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g, \boldsymbol{x})$ is an energy function scoring the compatibility among an image $\boldsymbol{x}$, facial attributes $\boldsymbol{\alpha}^f$, and frame attributes $\boldsymbol{\alpha}^g$. The recommendation results can be obtained by the most likely joint attributes state $\hat{\boldsymbol{\alpha}}^g = \arg\max_{\boldsymbol{\alpha}^g} \max_{\boldsymbol{\alpha}^f} p(\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^g|\boldsymbol{x})$.

2.3 Image Ranking

While the recommendation model generates a query vector of frame attributes $q = (q_1, q_2, ..., q_{N_g})$ (N_g=7), the image ranking module ranks the relevant eyeglasses images from the product database D, where each product image is labeled with frame attributes that annotated as $d^{(k)} = (d_1^{(k)}, d_2^{(k)}, ..., d_{N_g}^{(k)})$, $1 \leq k \leq |D|$. We define the similarity function between q and $d^{(k)}$ as follows:

$$f(q, d^{(k)}) = \sum_{i=1}^{N_g} w_i[q_i = d_i^{(k)}] \quad (2)$$

where $[\cdot]$ is the Iverson bracket notation, i.e. $[\cdot]$ equals 1 if the expression is true and 0 otherwise. We use Ridge Regression to learn a model w that represents the weight vector of frame attributes.

2.4 Synthesis

This module displays the recommended frame images synthesized with the input face image. To generate quality synthesis results, the frame should be placed on the face reasonably. We observe a variety of people wearing eyeglasses in real life and summarize some important rules. For example, the width of the frame should fit the face width, the height of the frame should be an appropriate proportion of the size of eyes, and the center of frame should be aligned at the nose bridge. There are mainly three key steps during the synthesis process. Firstly, active shape model [2] is used to

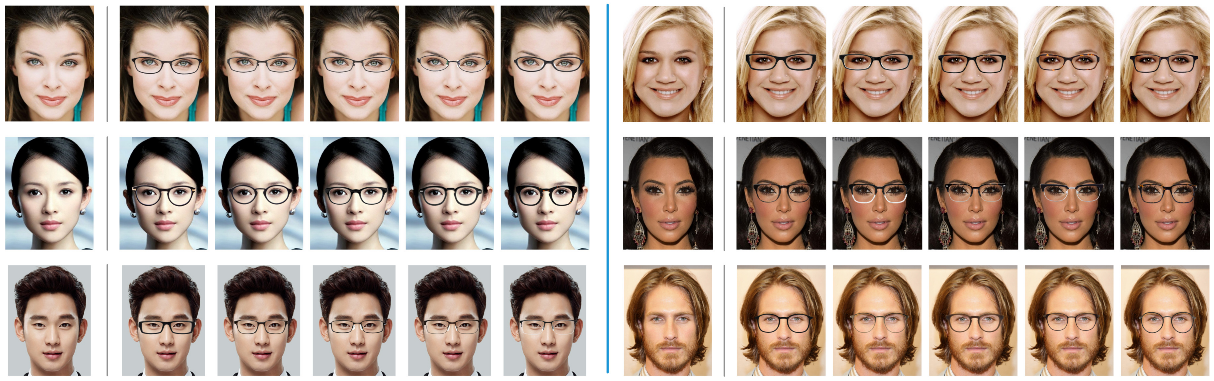


Figure 2: The final top 5 recommendation results on several testing faces

locate 77 feature points on the input face image to gain the size of eyes, the width of face and etc. Secondly, the frame image is scaled according to the human face width. Finally, the face photo is overlaid with the frame image by aligning the center of the face against the center of the frame.

3. EVALUATION

3.1 Dataset

We construct two datasets, including a face photo (FP) dataset and an eyeglasses product (GP) dataset, for evaluating the *iGlasses* system.

FP Dataset: We collect face photos, in which people wearing eyeglasses, from search engines and photo sharing web sites (e.g. Flicker, Pinterst) by using queries such as *celebrities with eyeglasses* and *stars with eyeglasses*. We use the images of movie stars and celebrities who wear eyeglasses as our training images, since those people always have a great sense of style. Viola Jones face detector [6] is utilized to remove those crawled images with no face detected. Then we retain 3039 face images where the eyeglasses is considered to match with the face. Each retained face image is annotated with both facial attributes and frame attributes. We leverage FP dataset to learn facial attributes and build our recommendation model.

GP Dataset: We collect eyeglasses product images from a variety of popular online eyewear stores (e.g. Warby Parker, Coastal, LensCrafters) for the recommendation results. In total, 2035 eyeglasses images are collected and each product image is labeled with frame attributes.

3.2 Baseline Methods

Our recommendation approach bridges the gap between low-level facial features and frame attributes by introducing facial attributes. To evaluate the effectiveness of intermediate facial attributes, we implement two alternative recommendation methods which directly predict the frame attributes with low-level features extracted from face images. More specifically, a set of classifiers are trained for each frame attribute with FP Dataset by using multi-class SVM and neutral network methods. Here different types of features (Gabor Filter, HOG, LBP, Color Moments, Color Histogram, Shape Context, etc.) are extracted from the whole face and concatenated to form a feature vector. Then the concatenated feature vectors are utilized to train a classifier

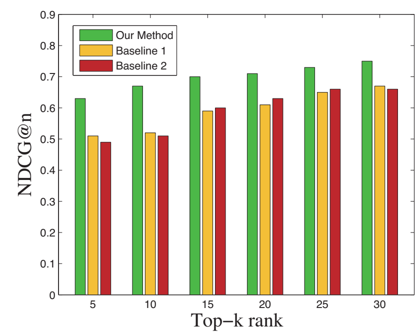


Figure 3: Performance of three recommendation methods measured with NDCG values

for each frame attribute. It is worthwhile to note that our recommendation approach and other two baseline methods share the same image ranking module and synthesis module.

3.3 Recommendation Results Evaluation

In order to qualitatively evaluate the final recommendation result, we conduct a user study with a crowdsourcing platform. We use 150 testing images to evaluate the performance of the two baseline methods and our approach. Given a testing face image, three top-k (k=30) ranked lists are generated by our method and other two baselines respectively. Every participant is asked to score the matching degree (1-5) between the recommended frame and query face for each ranked list. We collect scores for 150×3 ranked lists from 500 participants. Finally, the average scores are calculated for each ranked list as the ground truth.

The performance of our recommendation method is measured by Normalized Discounted Cumulative Gain (NDCG), which is widely used in ranking systems. Figure 3 reports the NDCG values of our recommendation method and other two baselines. From the results, we can observe that our model outperforms multi-class SVM and neural network, especially in the top 10 recommendations. This is mainly because that our model reserves the intrinsic relationship among frame attributes and captures the dependencies between frame attributes and facial attributes, while other two baseline methods neglect these correlations. Figure 2 presents the final top 5 recommendation results of our method on several testing faces.

4. DEMONSTRATION

4.1 Demonstration Setup

The *iGlasses* prototype includes an Android APP and a server backend. The mobile client is implemented on an Android smartphone and the web server is deployed with Apache Tomcat. The client is mainly in charge of displaying images, such as face images, eyeglasses images and synthesized images. While the server side is responsible for computing and communication with the client side. In the offline phase, the server builds facial attribute classifiers, constructs three recommendation models, and learns image ranking model. In the online phase, the server recognizes facial attributes and returns the synthesized images to the client.

4.2 Walkthrough Example

The *iGlasses* demo consists of the following steps, as illustrated in Figure 4:

Step 1: The user can either use *iGlasses* APP to take a photo or select an existing photo from his/her album as the input. Once the user chooses a photo, the query photo will be displayed at the client.

Step 2: As soon as the user submits the facial attribute query on the client, the above input photo is forwarded to the server. The server utilizes active shape model [2] to locate 77 feature points on the received face image and returns 17 fine-grained facial attributes by using the pre-trained SVM classifiers. Once the client obtains the facial attribute values, these facial attributes will be displayed along with face photo at the client.

Step 3: With the recognized/predicted facial attributes, *iGlasses* can recommend proper eyeglasses based on one's facial traits once the user submits a recommendation request. The server uses our probabilistic graphical model and image ranking model to recommend the best-fit eyeglasses. For comparison purpose, we also implement two alternative recommendation models on server using multi-class SVM and neural network methods, which can be configured in the APP.

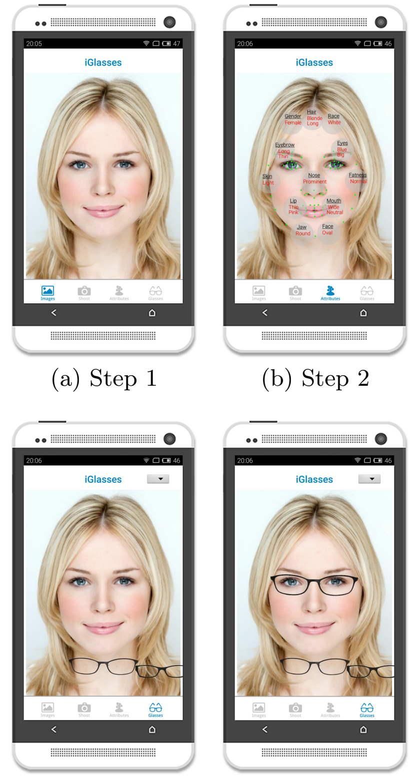
Step 4: Once the ranked list of recommended eyeglasses are returned to the user, the *iGlasses* system can generate the synthesis result efficiently. Users can have a preview of how the recommended eyeglasses look on their faces. In most cases, our approach outperforms the baseline methods and produces high quality recommendation results.

5. CONCLUSION

In this demonstration, we presented *iGlasses*, a novel recommendation system for recommending the best-looking eyeglasses. *iGlasses* used a new recommendation approach based on a probabilistic graphical model, which exploited implicit matching rules between faces and eyeglasses. To qualitatively evaluate our recommendation method, we also implemented two baseline methods for comparison and conducted a user study. Experiments showed promising results from our approach.

6. ACKNOWLEDGEMENT

This research is supported by the National Key Basic Research Program of China (GrantNo. 2015CB352400). This research is also partially supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centre in Singapore Funding Initiative.



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Figure 4: The Screenshots of *iGlasses* System

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995.

[3] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011.

[4] T. Mensink, J. J. Verbeek, and G. Csurka. Tree-structured crf models for interactive image labeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013.

[5] C. Rother, V. Kolmogorov, and A. B. 0001. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[6] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.