People Search and Activity Mining in Large-Scale Community-Contributed Photos

Yan-Ying Chen
National Taiwan University, Taipei, Taiwan
yanying@cmlab.csie.ntu.edu.tw
Advised by
Winston H. Hsu, Hong-Yuan Mark Liao

ABSTRACT

A growing number of users are contributing a huge amount of photos containing people (e.g., family, classmates, colleagues, etc.) to social media for the purpose of photo sharing and social communication. There arises a strong need for automatically analyzing the people shown in large-scale photos because these visual data comprise abundant consumer activities which greatly benefit demographic analysis and enhance marketing research. In this work, we aim at learning facial attributes (gender, race, age, etc.) by these publicly available photos and exploiting the detected facial attributes for locating designated persons, profiling user preferences and predicting social group types. In addition, community-contributed data possess rich contexts such as tags, geo-locations and time stamps, which strongly correlate with user intentions and preferences. The knowledge would be informative to actively refine the recognition models and promising towards improvement of photo management, personalized recommendation and social networking. Most importantly, this framework effectively relieves costly annotation efforts and ensures scalability for large-scale media.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.4 [**Image Processing and Computer Vision**]: Feature Measurement; H.2.8 [**Database Applications**]: Data mining, Spatial databases and GIS

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Facial attributes, image retrieval, personalized recommendation, social subgraphs

MM'12, October 29–November 2, 2012, Nara, Japan.

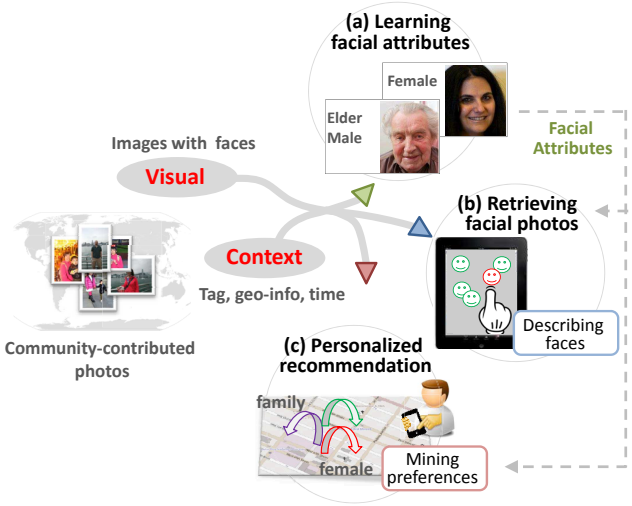

Figure 1: we propose to manipulate community-contributed photos and the associated contexts for learning facial attributes with less human intervention (as (a)) and exploiting the automatically detected facial attributes for retrieving photos containing designated persons (as (b)) and mining user preferences for personalized/group recommendation (as (c)). Most importantly, the framework is operative without intensive annotation labor and thus ensures the scalability for a growing amount of consumer photos. (Best seen in color. Photo courtesy of Flicker users [1] under Creative Commons license)

1. INTRODUCTION AND RELATED WORKS

With the prevalence of social media and the success of many photo-sharing websites, like Flickr and Picasa, the volume of community-contributed photos has increased drastically. Isola et al. [11] indicates that the most memorable photos for people in general usually contain faces, especially recognizable persons like family members or friends. Consequently, it motivates users to share those photos via social media to maintain close relationships within their social circles. In our study, more than 17 million photos were retrieved from Flickr using the search keyword "family." We found that around 60% of them contain at least one face. Such publicly available media provide a cost-effective way to obtain demographic information, e.g., the statistics for

user preferences regarding certain events or locations such as restaurants, hotels, landmarks, etc.

Moreover, these plentiful and publicly available photos contain rich metadata such as tags, time, and geo-locations (or geo-tags). These overwhelming amounts of context data, though noisy, are tremendously essential for many multimedia applications including annotation, searching, marketing, advertising and recommendation [18]. To deal with the *big data*, many studies are devoted to automatic image content analysis to ease the pains of manual annotations. Facial image analysis such as face detection and facial attribute detection (e.g., gender, age, race, etc.), is quite a challenging and attractive problem because it is very practical for consumer products and is also one of the enabling technologies for human behavior analysis and effective manipulation regarding large-scale images and videos.

Although this research area has made remarkable progress through decades of related works [13], most of them highly relied upon supervised learning with manually collected training photos from limited sources that induce intensive human intervention. On the other hand, we aim to acquire effective training images from community-contributed photos for learning facial attributes [4]. It is promising since social media are full of user activities via the photos along with tags, comments, locations, etc. However, simply acquiring training images by utilizing keywords (e.g., "beard") brings a significant amount of false positives due to an uncontrolled annotation quality; learning with such noisy data degrades the accuracy of facial attribute detectors. We propose to measure the quality of training images by discriminative visual features, which are automatically selected according to the relative discrimination in unlabeled images. We further exploit the rich context cues (e.g., tags, geo-locations, etc.) associated with these publicly available photos for mining more semantically consistent but visually diverse training images around the world.

Once facial attributes are detected from photo contents, those automatic annotations would be effective to locate specific persons in large-scale photos and beneficial for personal photo management. Users may forget where or when they took the photos but they would remember the basic traits of their friends and family. Therefore, it is imaginable to exploit facial attributes to formulate their search intention in regard to the people they are looking for. Furthermore, reviewing the retrieved images allows users to recall more scenes from their memory. For example, a little boy seems to be standing to the left of his mother. The scenes of the photo in mind can be organized intuitively by graphically arranging people on a query "canvas" and refined by designating more facial attributes. Rather than laboriously sketching detailed appearances [3] or typing text [23], our work allows users to formulate a query canvas by placing "icons" of desired facial attributes at desired positions and in desired sizes [15, 14]. Although consumer photos are naturally lacking annotations, automatic facial attribute detection would make the scenario more economical and scalable.

Seeing the power of describing people in photos by facial attributes, it is promising to acquire the statistics of demographic data from community-contributed media. Those media, as millions of human-sensors [20], carry rich spatio-temporal information such as geo-tags or time stamps which benefit the mining of travel-related data automatically. A focus of recent interests in the use of user-contributed resources involves the textual travelogues (i.e., blogs or logs) [12, 10, 9] and photos taken during such trips [2, 17]. However, previous studies solely consider the travel logs and ignore the rich facial attributes which provide another important aspect regarding travel demographics (e.g., gender, race, age) and are promising for personalized travel recommendations. Through information-theoretic measurements, we did observe that such (detected) facial attributes do correlate with traveling between locations. In this sense, we conduct facial attribute detection for the sequential photos according to their geo-tags and then provide a probabilistic recommendation model based on the user's profiles and travel logs [5].

In fact, consumer activities and user intentions are not limited to individuals. Group recommendations, which recommend to a group of people instead of only individuals, are vital for daily life. In Li et al.'s work [16], they analyzed specific transaction logs and found that different types of consumer groups (e.g. family, friends, couple) have quite different preferences when searching for travel accommodations. However, transaction logs are not easily accessible due to privacy and commercial issues. As a substitute for transaction logs, group activities can be observed from growing and freely available sources – social media. In addition, mining from that rich media overcomes the huge language gaps and culture differences that may exist. It has been evidenced that certain social interactions and relationships can be observed from the social contexts found within a photo [21, 8, 22]; for example, a mother usually stands close to her child(ren) and they naturally form a social subgroup. Therefore, we propose a novel framework to discover informative subgraphs, which resembles social subgroups in communities. Those automatically mined subgraphs would be the informative features to categorize specific group types (e.g., nuclear family) or events (e.g., a party for friends of similar ages).

To sum up, we aim to utilize community-contributed photos and the associated contexts to achieve large-scale photo management and extensive demographic study which are still very costly and challenging at present. In this proposal, we aim to (1) learn facial attributes with less human intervention (as Fig. 1 (a)), (2) retrieve images containing designated persons by their facial attributes (as Fig. 1 (b)) and (3) improve personalized and group recommendations by mining facial attributes and social relationships from community-contributed photos (as Fig. 1 (c)).

2. PROPOSED APPROACH

Witnessing the sheer amount of "people images" from the crowds on the web, we propose to conduct people retrieval and knowledge mining from the growing amount of user-contributed data; meanwhile, the framework is operative without intensive annotation labor for ensuring the scalability.

2.1 Problem Statement

Generally, we would concentrate on the four tasks, (1) facial attribute learning with less human intervention, (2) photo retrieval by facial attributes and (3) knowledge mining from faces in social media for personalized recommendations and (4) social subgraph discovery for predicting social group types. All the inputs are the freely available photos from social media as well as the associated metadata.

Consequently, we might confront the following major challenges: (1) the noisy information naturally existing in user-contributed data, (2) the multiple sorts of contexts associated with these photos and (3) how to enhance personalized and group recommendations by the mined demographic information.

2.2 Methodology

2.2.1 Weakly supervised facial attribute learning

The Internet images inherently provide better generalization capability for learning facial attributes because they cover more photos of diverse communities. The associated metadata of these images are able to alleviate the pains caused by manual annotation, for example, collecting photos tagged by "woman" for learning female attribute [4]. However, the training data of this sort are very easy to introduce unexpected noisy labels. Previous works have presented the effectiveness in rejecting noisy labels by visual relevance [19]; however, the strategy may decrease the visual diversity in training images and require certain human intervention (e.g., manually selecting discriminative visual features). For ensuring less supervision and more generality, two critical issues remain unsolved: (1) automatically selecting discriminative features as visual relevance cues; (2) preventing training images from being dominated by the majority of visual appearances and by those obtained from few specific locations.

Conventionally, it is possible to use a verification strategy based on potential visual features such as regional textures, edges or color to filter out incorrectly labeled images. This move can considerably resist the interference of noisy labels. Another factor that may mislead the facial attribute training process is diverse appearance of the same facial attribute. That is, the intra-class variance of the same facial attribute is large; for example, the appearances of female faces from the world are very diverse. In order to balance the unfavorable effects caused by noisy labels and the diversity problem of facial attributes, we propose to introduce context cues, tags and geo-locations, to ease the problem of solely relying on visual relevance. Augmented by feature selection for verifying discriminability of the potential features, the framework would be generalized for adaptively learning numerous facial attributes. Consequently, our approach requires no tedious manual annotations and thus ensures the scalability.

2.2.2 Photo retrieval by facial attributes

Facial attributes are important characters for describing a human face. Nowadays, because of the prevalence of camera devices, people are growing accustomed to preserving important moments in life by photos. With an increasing number of personal photos, it is difficult and inefficient for users to indicate the exact file location in the storage even though they are well categorized by time stamps or geo-locations. Some photo sharing websites employ crowd-sourcing to obtain free tags semantically associated to images, but the mechanism cannot be duplicated to personal photo management because users are not expected to actively annotate their photos. Recently, certain commercial software began to exploit technologies of face recognition and face clustering; such solutions still lack the capability of searching for scenes with faces deployed in a specific layout. In light of this observation, we attempt to make consumer photo retrieval faster and easier by facial attributes, face similarity and overall layout. We are (1) to analyze "wild photos" with no tag information at all by automatic facial attribute detection and face similarity estimation [13], (2) to advance search pattern from query by single face instance to query by multiple attributed faces allocated on a canvas and (3) to enable rapid search response by block-based indexing approach. The framework has been realized in a touch-based user interface which allows interactively refining the query canvas [15, 14].

2.2.3 Personalized travel recommendation

By intuition, we know that some landmarks are female-favored, and some are male-favored. So are by other attributes such as race, age. To examine the correlation between travel behavior and facial attributes, we measure the entropy and the mutual information [6] in predicting next travel location by facial attributes. Taking the correlation between gender attribute and the travel route from Madison Square in Manhattan as an example, the mutual information gained from the facial attribute is 0.5329 (bits), about 25% reduction of the entropy. The result can be illustrated like this, if there are 4 random choices for the next destination, after knowing the facial attribute (e.g., for the male only), the number of choice is down to 3. We can see that the preferences can be partially observed by facial attributes; therefore, the proposed approach involves facial attributes for improving the recommendation performance. At first, in order to mine the travel information within each city, we crawl the photos from the on-line photo-sharing websites (i.e., Flickr). We then use a mean-shift based method on geo-locations of these photos to generate the important locations in each city for the following user trip mining process. We can further identify the demographic information (via automatically detected facial attributes) within travel paths by analyzing the associated photos. By mining the travel patterns users' day trips, we further propose two personalized travel recommendation applications – mobile travel recommendation and route planning, which are entailed by a probability Bayesian model and dynamic programming technology [5].

2.2.4 Travel group type prediction

In fact, consumer activities and user intentions are not limited to only individuals. Group recommendations are essential for daily life, for example, recommending a family-friendly travel path for family group. For group analysis, it has been evidenced that the cohesive subgroups represent an important construct to study a group and individuals [7]. The social links in a group photo resemble a graph parameterized by facial attributes and topological information. Therefore, we propose a novel graph representation to model the potential social subgroups among a group of people. The inputs of our approach are consumer photos containing faces with estimated gender and age attributes (extendable to other attributes as well). The faces in a photo are modeled as a face graph. From the face graphs, we propose to automatically discover the informative subgraphs which resemble the social subgroups commonly appearing in communities. We further represent a group photo by the occurrence patterns of social subgraphs which act as effective features for classifying social group types by supervised learning.

3. EXPECTED CONTRIBUTION

The main idea of this proposal is to leverage the plentiful knowledge in social media to augment photo management and demographic study in large-scale media with as less human intervention as possible. In summary, four contributions are expected and listed as follows,

- Proposing a framework for learning facial attributes with less annotation efforts.
- Devising an efficient manner to retrieve photos by locating faces with specific facial attributes at desired positions and in desired scales
- Improving personalized recommendation by mining people attributes from community-contributed photos
- Discovering informative social subgraphs in communities to predict social group types and conduct group recommendations

Currently, we have figured out several potential scenarios such as consumer photo retrieval and travel recommendation that highly coincide with the hybrid people-related information. Meanwhile, we have demonstrated the feasibility of certain concepts and the preliminary results showed an impressive improvement in cost-effective personalized services (e.g., 20% relative improvement in destination prediction gained by minded travel demographic information [5]). The achievements persuade us to continually extend those ideas to more influential and realistic solutions.

4. CONCLUSION

We saw a sheer amount of consumer photos and most of them involve rich consumer activities which greatly profit people-oriented applications. In this proposal, we propose the motivations and the concepts for learning facial attributes, retrieving photos comprising people, and mining demographic information from community-contributed photos. The proposed approaches consider the noisy labels inherently existed in crowdsourced media, the user intentions in searching for designated photos and the potential knowledge for improving personalized services. Meanwhile, the mined knowledge is informative feedback to actively refine the recognition and prediction models. Most importantly, the proposed framework operates without intensive human intervention and is thus scalable to a growing number of consumer photos.

5. REFERENCES

[1] Flickr user rogerblackwell, http://www.flickr.com/photos/rogerblackwell/3083886325; trevino, http://www.flickr.com/photos/trevino/3756126025.

[2] Y. Arase, X. Xie, T. Hara, and S. Nishio. Mining people's trips from large scale geo-tagged photos. In *ACM international conference on Multimedia*, 2010.

[3] Y. Cao, C. Wang, L. Zhang, and L. Zhang. Edgel index for large-scale sketch-based image search. In *CVPR*, 2011.

[4] Y.-Y. Chen, W. H. Hsu, and H.-Y. M. Liao. Learning facial attributes by crowdsourcing in social media. In *International Conference on World Wide Web*, 2011.

[5] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W. H. Hsu, and H.-Y. M. Liao. Personalized travel recommendation by mining people attributes from community-contributed photos. In *ACM International Conference on Multimedia*, 2011.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[7] K. A. Frank. Identifying cohesive subgroups. In *Social Networks*, 1995.

[8] A. C. Gallagher and T. Chen. Understanding images of groups of people. In *CVPR*, 2009.

[9] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua, and R. Jain. W2go: a travel guidance system by automatic landmark ranking. In *ACM international conference on Multimedia*, 2010.

[10] Q. Hao, R. Cai, C. Wang, R. Xiao, J.-M. Yang, Y. Pang, and L. Zhang. Equip tourists with knowledge mined from travelogues. In *International Conference on World Wide Web*, 2010.

[11] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[12] R. Ji, X. Xie, H. Yao, and W.-Y. Ma. Mining city landmarks from blogs by graph modeling. In *ACM international conference on Multimedia*, 2009.

[13] N. Kumar, P. Belhumeur, and S. Nayar. Facetracer: A search engine for large collections of images with faces. In *European Conference on Computer Vision*, 2008.

[14] Y.-H. Lei, Y.-Y. Chen, L. Iida, B.-C. Chen, and W. H. Hsu. Where is who: Large-scale photo retrieval by facial attributes and canvas layout. In *ACM SIGIR*, 2012.

[15] Y.-H. Lei, Y.-Y. Chen, L. Iida, B.-C. Chen, H.-H. Su, and W. H. Hsu. Photo search by face positions and facial attributes on touch devices. In *ACM International Conference on Multimedia*, 2011.

[16] B. Li, A. Ghose, and P. G. Ipeirotis. Towards a theory model for product search. In *Proceedings of the 20th International Conference on World Wide Web*, 2011.

[17] X. Lu, C. Wang, J.-M. Yang, Y. Pang, and L. Zhang. Photo2trip: generating travel routes from geo-tagged photos for trip planning. In *ACM international conference on Multimedia*, 2010.

[18] T. Mei, W. H. Hsu, and J. Luo. Knowledge discovery from community-contributed multimedia. *IEEE MultiMedia*, 17, October 2010.

[19] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *ACM Multimedia*, 2009.

[20] V. K. Singh, M. Gao, and R. Jain. Social pixels: genesis and evaluation. In *ACM international conference on Multimedia*, 2010.

[21] P. Singla, H. Kautz, A. Gallagher, and J. Luo. Discovery of social relationships in consumer photo collections using markov logic. In *CVPRW*, 2008.

[22] G. Wang, A. Gallagher, J. Luo, and D. Forsyth. Seeing people in social context: Recognizing people and social relationships. In *ECCV*, 2010.

[23] H. Xu, J. Wang, X.-S. Hua, and S. Li. Image search by concept map. In *SIGIR*, 2010.