Multi-dimensional Aesthetics Mining for Social Photo Recommendation

Zhanwei Wu, ZhenYu Gu, and Zhanxun Dong

Department of Media and Design, Shanghai Jiaotong University
{Zhanwei_wu,zygu,dongzx}@sjtu.edu.cn

Abstract. The success of a social photo recommendation system mainly depends on its ability to provide high quality photos, which also means the recommended photos will have a greater chance to meet the interests of the users. We believe the quality of photos may originate from three dimensions. Two experiments was conducted to validate the relation of various features from these dimensions and the attractiveness of social photos. Result show, by integrated use of three dimensions, classifiers could be constructed effectively with fewer features.

Keywords: social photo, aesthetics, user study.

1 Introduction

Photo recommendation plays an important role in how people browse large photo collections on social network sites. To find visually appealing photos, many methods have been proposed from the perspective of visual aesthetics. The typical approach is to extract features and model it as a machine learning problem. Then, photos will be labeled or rated using classifiers and regression models constructed by those features. Most methods use low level characteristics [1][2][3][4], such as: HSV, contrast, level, histogram, etc. However, there exits too many low level features in photograph. It's hard to explain why some features should be selected, while others not. Recently, many high level attributes (which means more human centered) were proposed to address this problem [5][6][7]. However, most features are borrowed directly from photography theories without convincible reasons, rather than by studying user behaviors.

In this paper, three dimensions of aesthetic features are extracted and evaluated in a more intuitive way, by conducting user-centered studies. We believe, by integrating features that describe all the dimensions of the photography aesthetics, a more complete model could be constructed to achieve better performance. Moreover, not only image characteristics, but different types of data related to those photos should also be taken into consideration. For example, remarks may provide information that whether subject in the photo is attractive or not, mining EXIF data may reveal the quality of photography equipment.

C. Stephanidis (Ed.): Posters, Part II, HCII 2013, CCIS 374, pp. 407–412, 2013.

2 Experiment and Method

2.1 User Observation

First, a user observation experiment was conducted to learn what kind of key factors would affect user's social photo aesthetics evaluation. Take the fact into consideration that normal users lack of specific photography knowledge and could not describe their feeling precisely, we specially invited users with professional background as a complementation.

There are 10 normal users(non-design-majored students without professional photography training) and 10 professional users(design-majored students with professional photography training) taking part in our study. We requested the normal users, first, to randomly browse and then point out those they specifically interested in among all the social photos. By analyzing the observation result, we found that there were usually two stages of user behavior during the process of browsing pictures: the first stage was rapidly scanning the thumbnails in order to target the albums they may be interested in; while in the second stage, users would browse the pictures in sequence or view them in detail within the selected album.

The eye-tracking record showed that the average time cost in the first stage was relatively shorter than 0.7 second per picture. Moreover, most of the users' attentions were focused on the photo subjects. However, the average time spent on the second stage was significantly longer than on the first stage and the user behavior mode varied intensively according to user differences: some users scanned the whole album orderly, while the other skimmed or jumped over some pictures by "click and return". Still when viewing closely, users paid more attention to the subject.

Later we paired each nonprofessional user with one professional user and asked them to browse the social photos again. Under the assistance of professional users, nonprofessional users were supposed to pick up technical terms and describe the factors affecting their choices upon pictures in an accurate way. We concluded those factors into 3 dimensions: Attractiveness of subject (cute face, good body shape, graceful body curve and elegant pose), skill of photographer (subject size, angle, position, background simplicity, narrow DOF, accurate focus, light scheme) and quality of camera (aperture, shutter speed, noise, accurate color reproduction, sharpness, white balance accuracy).

2.2 Evaluation and Experiment Design

Though intuitive relationship could be found between visual quality and 3 dimensions mentioned above, to evaluate this relationship more objectively, experiments was proposed as follows:

Pre-processing of Photos

The main purpose of pre-processing was to filter out unsuitable photos, including three aspects:

- First, we found most pictures were screened out in the first phase of rapid scanning, without actual visiting, which consequently led to a lower click rate compared to the

average level of the whole album. The user interview suggested that by simulating this kind of user behavior we may be able to filter out low quality pictures effectively. Therefore we deleted all the photos under average click rate.

- Second, the object of our study was limited to portrait photography, so we deleted other photos by using face recognition.
- Last, since the object of our study was pictures taken and uploaded by users themselves, we simply deleted those photos without EXIF information.

The Evaluation about Attractiveness of Subject

According to survey, the attractiveness of subject included various aspects such as: face, body shape, body gestures, and etc. The aesthetic evaluation of these features is too complicated for computer vision to handle with. Considering the fact that social photos contain various social information which could be used for aesthetic evaluation, like: the click rate of photos, the total amount of replies, contents of replies and etc. We premised that:

H1.1 Good-looking subjects will induce more clicks in the rapid scanning stage.

H1.2 Good-looking subjects will make viewers more willing to comment and thus get more replies.

H1.3 Good-looking subjects will give viewers positive arousals and thus generate positive comments.

If the hypothesis above can be proved, we then be able to use these social features to construct a classifier. So we designed experiment 1-(1) to test and verify what we had proposed.

In Exp. 1-1, we asked a user to randomly browse social pictures and recorded his browsing history. All the browsed photos were sent to another 10 users for rating, and we used five point grade scale: -2(ugly), -1(bad looking), 0(so so), +1(good looking), +2(beautiful). The average rating from these 10 users was regarded as the final aesthetic rating and put into correlation test together with click rate, numbers of replies and positive comments. To notice that when we talk about positive replies, it meant that comments containing emotional remarks about affinity, admiration and praises. Instead of semantic analysis, we actually adopted the method of detecting the corresponding emoticons.

The Evaluation about Skill of Photographer and Quality of Camera

In recent studies, photography rules, for instance rule of thirds, have already been widely used in photo aesthetics evaluation. However, there is no evidence that using rules is always better. The fact is, many photographers believe that sometimes breaking the rule equals to breaking new ground. We argue that it's the photographer's experience and skills behind these rules makes photos more attractive, rather than a single rule. Thus, hypothesis were proposed as follows:

- H2.1 Photos meet the rule of thirds rank higher than those not.
- H2.2 Photos of which subject size can meet portrait photograph requirement rank higher than those not.
- H2.3 Photos with narrow DOF rank higher than those with wide DOF.
- H2.4 Photos with clear subjects rank higher than those with dim ones.
- H2.5 The more photography rules satisfied, the higher ranking photo will get.

Besides, we also took equipment evaluation into consideration at the same time because: First, some of photography rules request for equipment support. For instance, big aperture is usually necessary to narrow DOF, known as "selective focus". Technologically, it is unable to separate techniques from equipment. Second, veteran photographers tend to use professional cameras according to the survey. It is unnecessary to separate them apart either. As a result, we verified another hypothesis in the same experiment:

– H3 Social photos taken by professional camera are more aesthetic.

In order to verify hypothesis above, we designed Exp. 1-2:We collected 50 photos of one subject and asked 10 normal social network users and 10 professional users to make a rating. Still the average rating was regarded as the final result.
Then we tested photography rules on whether it was related to the result or not:

- Rule of thirds was measured using midpoint between two eyes. Eye recognition algorithm in OpenCV was used for targeting the eye positions. We used midpoint because it was the visual focus from most photographers' opinion. Eye midpoint ranging from 0.33±0.1 of the entire frame were regarded to satisfy the rule, otherwise not.
- Rule of subject size was measured using face size of subject, which decided by the face recognition algorithm in OpenCV. If the proportion of the face area in the whole frame reached the requirements listed below: 0.5%-5% as Full length photo; 5%-12.5% as Half length photo; 12.5%-50% as Close-up. Then we defined it to satisfy the rule. Otherwise not.
- Rule of DOF was measured by manual judgment (on whether background is blurred or not.), because image-sharpness calculating functions proposed in previous studies could not give a satisfactory accuracy.
- Rule of focus is measured by manual judgment for the same reason.

Quality of camera was measured using EXIF data and divided into 5 categories: 1. Cellphone Camera, 2. Normal Consumer Camera, 3. High-Quality Consumer Camera, 4. Professional Camera, 5. High-Quality Professional Camera. Camera model library was also built manually.

Photo Recommendation Using Multi-dimensional Features
Exp.1-1 and Exp.1-2 explored how the features from 3 dimensions affected viewers' aesthetic evaluations respectively. In order to combine all the three dimensions and evaluate its comprehensive effects upon practical photo recommendation, we chose part of the parameters in exp.1-1 to construct a classifier and used training set to train it in Experiment 2. Later, 10 users' social friends albums were randomly selected as testing set. We recommended photos that classifier evaluated as above "good looking" to another 10 users and collected actual feedbacks as a comparison.

3 Result and Discussion

H1.1: negative; H1.2: negative; H1.3: positive;
Both H1.1 and H1.2 were not supported by correlation test, though click rate and the total amount of replies are widely used in many photo recommendation system. After investigation, we found most social photo browse came from friends which suggest

browse and replies have significant relation with active friends. Moreover, users browse and comment a social photo for many reasons, including but not limited to aesthetics experience. However, H1.3 were supported. Feedback shows, emoticons are only used when users experienced strong emotional arousal, some are mostly related to aesthetics experience.

H2.1 negative; H2.2 negative; H2.3 negative; H2.4 negative; H2.5 negative; H2.1 to H2.4 are hypothesis regarding single photography rule. All four hypothesis are rejected by t-test with 95% confidence level, though photos satisfy certain rule usually have better score than photos satisfy none. Among them, the effect of subject sharpness (H2.4) is the least significant because users rarely upload photos that are seriously out of focus. The most significant one is DOF (H2.3). This hypothesis could be accepted if confidence level is 90%. Subject size (H2.2) is an interesting feature. Out of size photos and full length photos have almost the same score, but are significantly less than half length and close-up photos. The possible reason is, taking a good full length photo requires much more skills than half length or close-up, which is

H2.5 is rejected by correlation test with 95% confidence level. However, photos satisfy all rules have significantly higher mean rating than photos satisfy none. Since rules studied in this paper are just a small part of photography aesthetics, a more considerable effect should be observed if we take more rules into account.

H3 positive.H3 is accepted. Photos taken by professional camera (4, 5) have significant higher mean ratings than photos taken by consumer camera (1, 2, 3). Survey after experiment also shown photographer using professional camera have more photography knowledge and are more skillful.

After studying the results of experiment 1, several features were selected to construct a classifier, include: emoticon number and click rate (from dimension 1), subject size, DOF and usage of photograph rules (from dimension 2), camera type (from dimension 3). 100 photos were randomly chosen from 10 album and used as training set. Another 100 photos were used as test set. After classification, 14 photos from test set were labeled as good-looking or above. These 14 photos were recommended to 10 users for validation. Feedback shown, 10 of 14 were rated above good-looking by users.

4 Conclusion

Social photo recommendation is a challenging task. In this paper, a model consists of three dimensions was used to explain user's aesthetics preference. Integrated use of aesthetic features from those dimensions could help to address photo recommendation problem. However, additional studies should be carried out to verify the feasibility of this model. Future studies could determine whether this model hold true for different types of photo aesthetics and for various kind of users.

Moreover, there are considerable features from dimensions mentioned above. Some have not been discussed in this paper yet, especially many photography rules. Additional studies could focus on validation of these rules and take good use of photographer's experiences.

One classifier evaluated by participants in this study, using features mentioned above, proved to be effective for photo recommendation. However, several features

are decided by manual judgment instead of computer processing, though some calculating functions have already been proposed in previous studies. The main reason is to ensure accuracy. Thus, to improve the accuracy of these functions might be the basis for the future studies.

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
2. Luo, Y., Tang, X.: Photo and video quality evaluation: Focusing on the subject. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 386–399. Springer, Heidelberg (2008)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (1), pp. 886–893. IEEE Computer Society (2005)
4. Li, C., Chen, T.: Aesthetic visual quality assessment of paintings. IEEE Journal of Selected Topics in Signal Processing 3(2), 236–252 (2009)
5. Cerosaletti, C.D., Loui, A.C.: Measuring the perceived aesthetic quality of photographic images. In: First International Workshop on Quality of Multimedia Experience, QoMEX, pp. 47–52 (2009)
6. Jiang, W., Loui, A.C., Cerosaletti, C.D.: Automatic aesthetic value assessment in photographic images. In: ICME, pp. 920–925. IEEE (2010)
7. Dhar, S., Ordonez, V., Berg, T.L.: High level describable attributes for predicting aesthetics and interestingness. In: CVPR, pp. 1657–1664 (2011)