

Training Data Collection System for a Learning-based Photographic Aesthetic Quality Inference Engine

Razvan Orendovici^{1,2} James Z. Wang²

¹ Schreyer Honors College, ² College of Information Sciences and Technology
The Pennsylvania State University, University Park, Pennsylvania
{ruo3, jwang}@psu.edu

ABSTRACT

We present a novel data collection system deployed for the ACQUINE - Aesthetic Quality Inference Engine. The goal of the system is to collect online user opinions, both structured and unstructured, for training future generation learning-based aesthetic quality inference engines. The development of the system was based on an analysis of over 60,000 user comments of photographs. For photos processed and rated by our engine, all users are invited to provide manual ratings. The users can also choose up to three key photographic features that the user liked, from a list, or to add features not in the list. Within a few months that the system is available for public use, more than 20,000 photos have received manual ratings and key features for over 1,800 photos have been identified. We expect the data generated over time will be critical in the study of computational inferencing of visual aesthetics in photographs. The system is demonstrated at <http://acquine.alipr.com>.

Categories and Subject Descriptors

H.4.m [Information Systems Applications]: Miscellaneous; I.5.4 [Pattern Recognition]: Applications

General Terms

Design, Experimentation

Keywords

Aesthetics, User Data Collection, Photography

1. INTRODUCTION

In order to substantially improve the accuracy and usability of computational photographic aesthetics inferencing systems[1, 2, 3], we need to properly understand what combinations of features determine photos as aesthetically pleasing. It is important that computer scientists understand these various aesthetic qualities and their roles in photo rating

so that computational methods can be developed to model the rating process. Photo peer rating sites (e.g., photo.net) collect information about visual features in an unstructured manner, such as using free-text comments. The data is less useful because we cannot accurately parse these comments.

We have very limited knowledge of why we consider photos to be artistic, interesting, or visually appealing. We aim at collecting readily-usable data to further this knowledge. By understanding what qualities of a photo make it aesthetically pleasing we can correlate those qualities to lower-level computable visual features. With a large set of collected data researchers can potentially improve ACQUINE (Aesthetic Quality Inference Engine) or other photo rating systems. For instance, the data can help us to identify a set of photos with good lighting and color tones, which can be used to train a classification and rating system for such photos, based on extractable color, texture, shape, and other features.

The design of such a data collection system is non-trivial. There is a wealth of literature on designing effective human-computer interfaces [4]. Typical online users are impatient that they would not use the system if it is time consuming and without any rewards. Because the interaction with the system is entirely anonymous, no reward can be provided. Hence, the system needs to collect structured information without taking many mouse clicks for a typical user. Whereas the system needs to collect structured information primarily, it is important that unstructured information can be collected to supplement the structured information.

To achieve these objectives, we first analyzed common unstructured user comments about visual appearances of photographs. The analysis allowed us to design the system so that most frequently used visual features can be shown as clickable choices. Users can add features if the displayed choices are not sufficient for describing some photographs. In the next section, we show our analysis which has led to the current design.

2. THE DESIGN RATIONALE

The incredible database of photographs that has been acquired by flickr.com provides great insight into photographic patterns, and the way professional and amateur photographers judge photos. On February 11, 2010, we downloaded the comments for the top 1,500 most interesting photos on the site. These are photos that have receiving many views and comments, and are part of many photo pools. The Flickr API system gave us easy access to highly discussed photos that are appreciated for various qualities by the flickr

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10 ...\$10.00.

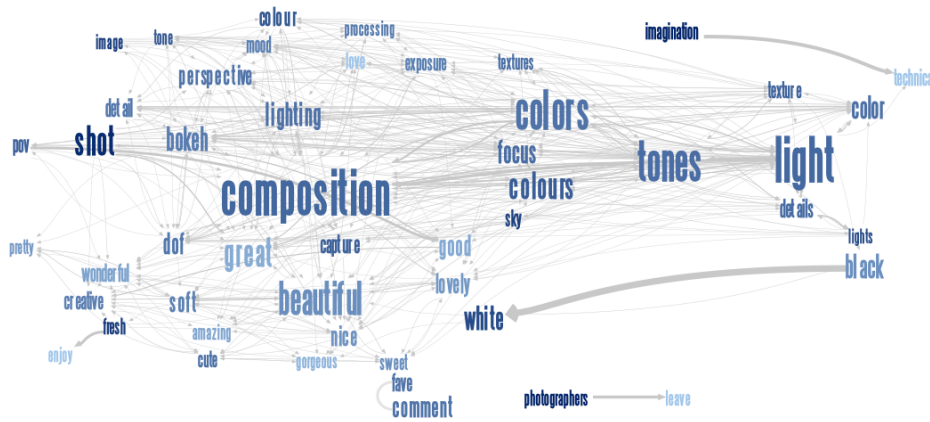
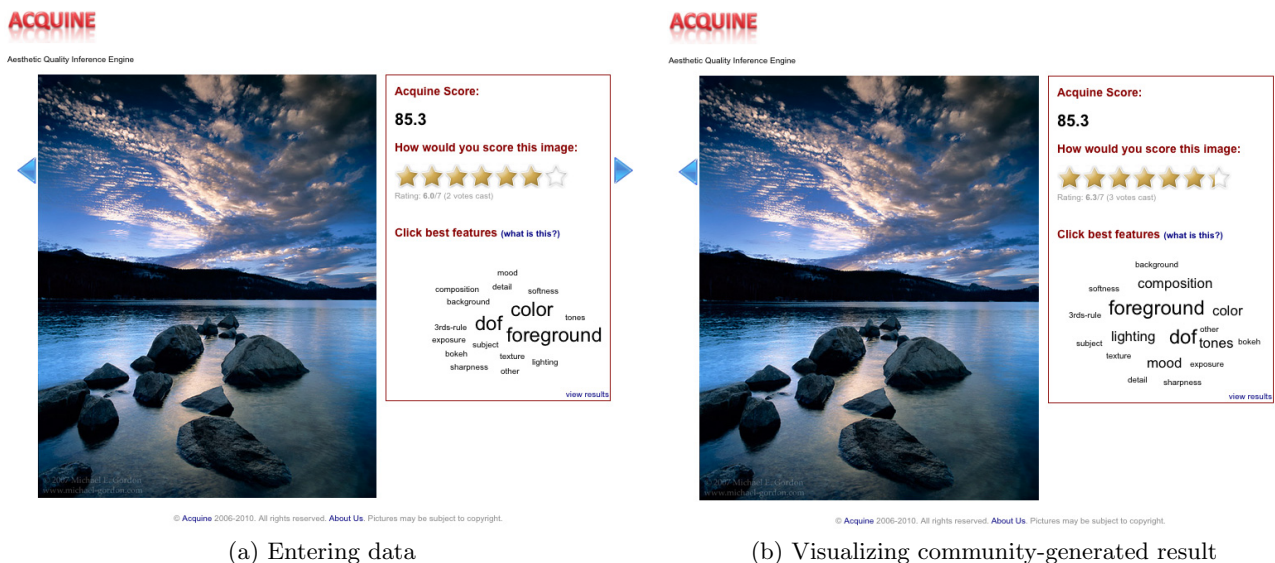


Figure 1: Visualization of phrases used in the user comments on 1500 most recent 'interesting' photos gathered on February 11, 2010, from Flickr. IBM Many Eyes is used for plotting. The arrows indicate an "and" association between the two terms. Top 50 of the 2225 terms are shown.



(a) Entering data (b) Visualizing community-generated result

Figure 2: The data collection process. Screenshots are shown. Copyright of the photo remains with the original owner(s).

community. We downloaded over 60,000 comments from these photos and imported them into the IBM Many Eyes word visualization system. This visualization provides us insights into the language used to describe photos. In Figure 1, we provide a screenshot of the visualization¹. The parameters of this visualization can be slightly altered to provide different insights into the data. From this visualization we chose 16 common terms to provide on the ACQUINE system for users to rate photos with. They are: *composition, tones, 3rds-rule, exposure, bokeh, background, softness, mood, detail, foreground, subject, texture, dof, color, sharpness, and lighting*.

3. THE DATA COLLECTION SYSTEM

As aforementioned, the large amount of word possibilities poses many problems for designing a proper rating system for a wide user community. Common option choices like drop down menus, radio buttons, and check boxes would

be visually displeasing to users and can detract them from using the system. We made all the choices clickable links and had to find a layout that would be more interesting and pleasing to use than a long list of terms. Due to popularity of tag-clouds in interactive Web pages, we chose to put the rating system as a randomly laid out tag cloud. This layout will prevent biases towards elements placed at the top of the list, and will be an interesting interface element which users should easily understand.

Each user is presented with the 16 terms and given the opportunity to select up to three of them as their favorite qualities of the photo. We provide interface feedback to the user by increasing the size of the vote whenever it is submitted. Once a user submits three votes towards the current photo we display the rating results from all the other users. We increase the size of terms linearly to display their prominence as an enjoyed quality of the photo. Additionally, we hide this final result from users to prevent biases from users voting the same way as previous users did. To make the system easy to use and interactive we use AJAX

¹http://manyeyes.alphaworks.ibm.com/manyeyes/visualizations/common-flickr-comment-phrases



Figure 3: Unstructured comments can be added by a user. Copyright of the photo remains with the original owner(s).

technologies to submit the votes and not distract users as the page refreshes after each vote.

We understand that the 16 terms from the Flickr visualization cannot accurately describe everyone’s impressions of a photo. For this reason we provide the ability for users to write free-text comments on the photo. A user can click on “other” to activate the comment box. These text-based comments can later be analyzed through language processing routines to provide more understanding of the ratings of that photo. Figures 2 and 3 demonstrate the interaction of the rating system.

The new data collection system in ACQUINE was launched in April 2010 for public use. In the last three months, it has resulted in the submission of over 5,300 visual feature selections (each click is counted once) for 1,800 photos in our ACQUINE database. As data collection increases we will analyze these with a seven star based overall aesthetic score as well as with our ACQUINE generated machine score.

We now show some preliminary analysis of the ratings as well as an analysis of the ACQUINE system. The distribution of the 7-star ratings provided by the users is similar to that shown in [3]. Many users value their own photos highly and often rate them with 7 stars. Our photo discovery system tends to bias the better rated photos by displaying highly ranked photos on the front page and ordering results by highly ranked photos. Additionally, we notice that users tend to spend more time on photos they find pleasing and are more willing to rate those photos as opposed to ones less pleasing.

Figure 4 shows the popularity of visual features across our entire database. Composition, color, subject, and lighting of photos are shown as the most important visual qualities. It is clear that the development of next generation aesthetic quality rating systems must focus on composition, despite the great difficulties in characterizing the quality

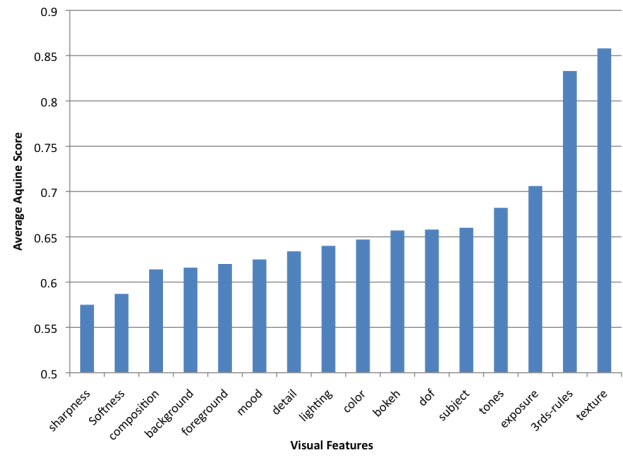


Figure 5: User-selected visual features and the average ACQUINE scores.

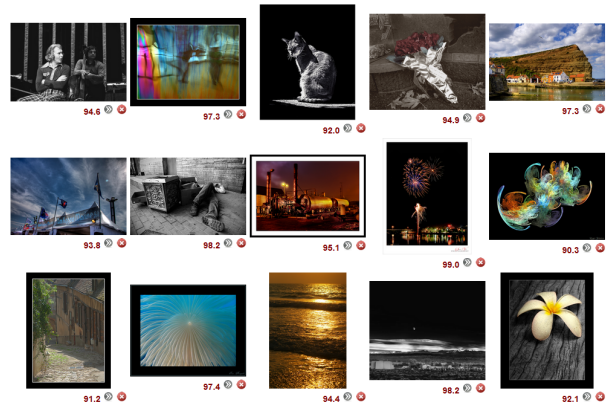


Figure 6: A random set of photos with high ACQUINE scores.



Figure 7: A random set of photos with high user-entered scores in a 7-star rating system.

of composition numerically. In Figure 5, we show the relationship between user-selected visual features and the average ACQUINE scores of images. Photos with good textures or following the rule of thirds have obtained



Figure 4: Aesthetically important visual qualities.

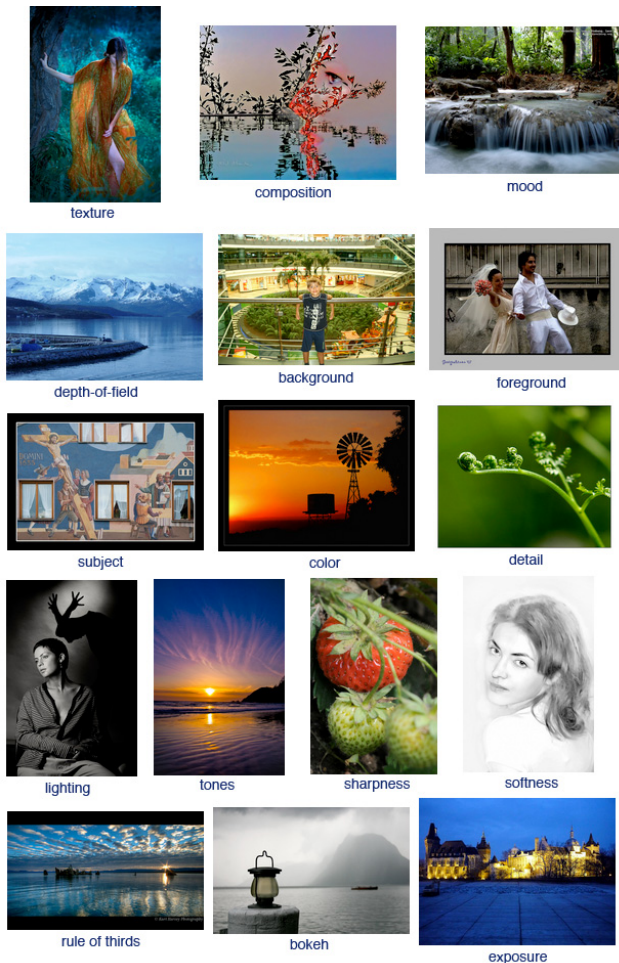


Figure 8: Example photos and the key visual features identified by the user.

higher ACQUINE scores, while sharpness, composition, the separation of foreground and background, and mood appear to be less important to ACQUINE.

Figures 6 and 7 show random sets of highly rated photos by ACQUINE and by users, respectively. Figure 8 shows 16 example photos and their key visual features identified by the users of the system.

4. CONCLUSIONS AND FUTURE WORK

Rating of visual aesthetics is a complex topic that provides challenges for any computational approach. Currently computer scientists have very limited understanding of aesthetic qualities and insufficient data to process them statistically. To address this, we developed a system that aims at collecting structured and unstructured data about the aesthetic qualities of photos. The data collected over time will be analyzed and used for developing and training next-generation systems.

We will collect data from the Website as we work on implementing more of these ideas to analyze textures, sharpness, softness, focus and composition of photos. We will add more functions and further improve the usability of the data collection system to better suit the research needs.

5. ACKNOWLEDGMENTS

The research is supported in part by the US National Science Foundation under Grant Nos. 0347148 and 0202007. We thank Ritendra Datta, Dhiraj Joshi, and Jia Li for valuable discussions and contributions to the ACQUINE project in its early stage. John Schleicher participated in the project. We collected research data from Flickr and Photo.net.

6. REFERENCES

- [1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," *Proc. of the European Conference on Computer Vision*, Part III, pp. 288-301, Graz, Austria, May 2006.
- [2] R. Datta, J. Li, and J. Z. Wang, "Algorithmic Inferencing of Aesthetics and Emotion in Natural Images: An Exposition," *Proc. of the IEEE International Conference on Image Processing (ICIP)*, pp. 105-108, San Diego, California, IEEE, October 2008.
- [3] R. Datta and J. Z. Wang, "ACQUINE: Aesthetic Quality Inference Engine - Real-time Automatic Rating of Photo Aesthetics," *Proceedings of the ACM International Conference on Multimedia Information Retrieval*, pp. 421-424, Philadelphia, Pennsylvania, ACM, March 2010.
- [4] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale, *Human-Computer Interaction*, 3rd Edition, Prentice Hall, 2003.