ACQUINE: Aesthetic Quality Inference Engine – Real-time Automatic Rating of Photo Aesthetics

Ritendra Datta* James Z. Wang
The Pennsylvania State University, University Park, Pennsylvania
{datta, jwang}@psu.edu

ABSTRACT

We present ACQUINE - Aesthetic Quality Inference Engine, a publicly accessible system which allows users to upload their photographs and have them rated automatically for aesthetic quality. The system integrates a support vector machine based classifier which extracts visual features on the fly and performs real-time classification and prediction. As the first publicly available tool for automatically determining the aesthetic value of an image, this work is a significant first step in recognizing human emotional reaction to visual stimulus. In this paper, we discuss fundamentals behind this system, and some of the challenges faced while creating it. We report statistics generated from over 140,000 images uploaded by Web users. 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

Algorithm, Experimentation

Keywords

Aesthetics, Image Quality, Inference, Photography

1. INTRODUCTION

We as a society are continuing our efforts to improve automatic information access methods. One type of effort focuses on being able to isolate quality information from the rest, all else being equal. In particular, this paper deals with automatic photo quality assessment, a problem which if solved reasonably can be used to isolate high quality photos, which in turn can help with image acquisition, search, and organization. We present ACQUINE (Aesthetic Quality Inference Engine), a machine-learning based online system of computerized prediction of aesthetic quality for color natural photographic pictures. Because of the limited computational resources, much complexity reduction in this implementation is necessary, making it a demonstration intended mainly for assessing the potential of an automatic aesthetics assessment algorithm. We consider this to be an important step toward realization of the goal of emotion-based retrieval and filtering of image collections. Further, this system illustrates the potential for computers to exhibit emotional responses to visual stimulus.

Whereas the algorithmic aspects of the ACQUINE system was first presented in an earlier publication [1], we had at that point not implemented a publicly available system to help assess its real-world potential. When doing so, we did have to address questions of quality, scale, and usability. For example, in order to be able to generate quick responses, we had to compromise on feature extraction, limiting ourselves to only a subset of the features originally discussed [1]. We discuss some such issues in this paper. In the remainder of the paper, we devote one section each to the technical approach, user interface, and system statistics, before concluding with a brief discussion.

2. THE ACQUINE ENGINE

In this section, we briefly summarize the algorithmic components behind the ACQUINE system. The system is built upon the research presented in our earlier work [1].

The idea is to treat the problem of aesthetics inference as a standard two-class classification problem. Given that certain features are determined to have a good correlation with aesthetic quality, a Support Vector Machine (SVM) classifier is trained as follows. Suppose the labeled data consists of images rated by a number of users. Each image can be associated with the average over the user ratings received by it. As seen in the experiments reported in [1], and discussed more formally in a later publication [2], the more the number of ratings received by an image, the better the average is as a predictor of the true or intrinsic aesthetic quality of the image. The average values are bounded by the range of rating values allowed for the particular user feedback system involved.

With this data at hand, we train a two-class SVM classifier as follows. Let us say that in the rating scale, an average score above R_{high} typically indicates high-quality photos, and average score below R_{low} indicates low-quality ones. The set of photos falling between R_{low} and R_{high} (which

*R. Datta is currently affiliated with Google Inc., Pittsburgh, Pennsylvania.

MIR'10, March 29–31, 2010, Philadelphia, Pennsylvania, USA.

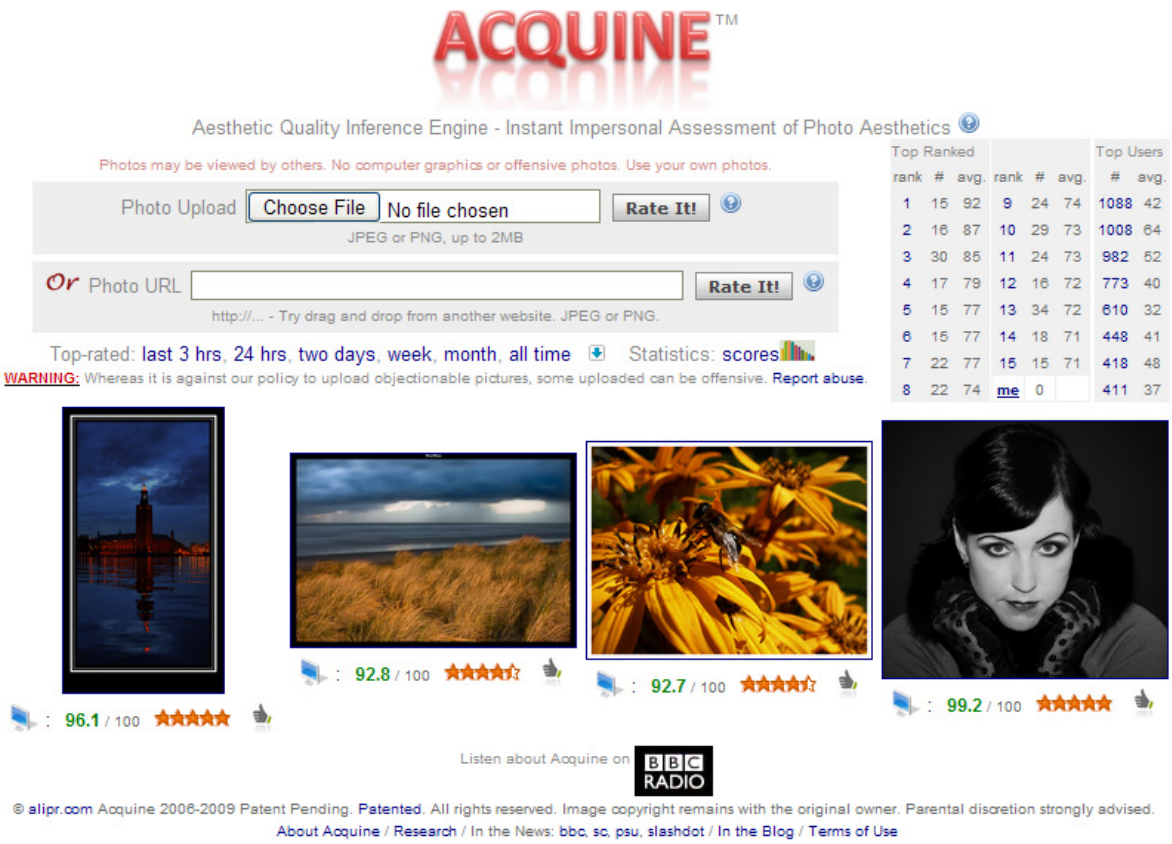


Figure 1: The front page of the ACQUINE system, launched in April 2009 at `http://acquine.alipr.com` . Users can upload photos from local storage, or paste an image URL. The lists of top users are shown on the right side. Photos with high ACQUINE aesthetics scores are randomly selected and shown on the page.

is termed as the 'band gap' [1]) are assumed too difficult to place into one of the two bins. Hence, the SVM classifier is trained with the extreme set of photos, one class of examples being the ones exceeding R_{high} and the other class being those below R_{low}. The SVM classifier training leads to the learning of an optimal hyperplane that hopefully isolates the appealing images from the plain ones.

Instead of just the class value, we wanted to be able to display a score in the 0 to 100 range for each image. The way we achieved this was by using a *sigmoid function* that maps the distance from the hyperplane to a 0 to 1 scale. The sigmoid function that takes in a distance *dist* and computes a *score* is as follows:

$$score(dist) = \frac{1}{1 + \exp^{-dist}}$$

As a point goes further away from the separating hyperplane in the negative side, *dist* tends to $-\infty$ and hence *score* tends to 0, while when it goes the other side of the hyperplane, *dist* tends to $+\infty$, as a result *score* tends to 1. Since it is more user friendly to show numbers between more typical score ranges like 0–10 or 0–100, we scale *score* to form the *ACQUINE score* between 0 and 100 by multiplying by 100.

More specific details on the SVM training are as follows. The data source for SVM training for the ACQUINE system is Photo.net [3]. Here, photos are rated on a 1 to 7 scale on their *aesthetics* and *originality*. A sizable fraction of the photos get rated by multiple users. We crawled photos from Photo.net with the condition that they were rated by at least 10 users (to ensure stability in scores, see [2]). Then, we split the set of photos into two groups: a 'LOW' group where the average scores were equal to or under 4.2, and a 'HIGH' group where the average scores equalled or exceeded 6.0. By sampling (without replacement) the two sets so as to balance the training data, we were left with about 20,000 training photos.

Features were extracted based on the results obtained in [1], but a number of the more computationally intensive features such as 'shape convexity' and 'familiarity measure' were not included. The latter feature is also dependent on the *anchor* database chosen, hence we felt that it may not generalize to arbitrary user uploads.

The SVM (with RBF kernel) was trained with this data, and then a held-out sample was used to tune the SVM parameters. Given that this was a larger training set than what was used in [1], the minimum number of ratings per photo was raised to 10, and due to better parameter tuning, we were able to get cross-validation accuracy in classifying HIGH vs. LOW of roughly 82%. Of course, the ACQUINE system setup is not identical, in that all photos must receive a score, not just those falling in the HIGH or LOW bands. The basic idea was to get the SVM to be able to distinguish between the extreme cases, and the ones in between get extrapolated correctly based on their distance to the hyperplane.

3. USER INTERFACE

The traffic into ACQUINE is very likely composed of a large fraction of otherwise non-technical photography enthusiasts. One reason for this assumption is that right after the public announcement of the system, a large number of blog posts mushroomed, and much of the discussion was spearheaded by photographers. To make the system accessible to the entire spectrum of people from domain experts to photography enthusiasts, we built a very simple interface which hides much of the technical underpinnings of the underlying classifier. A screenshot of the front page is shown in Fig. 1.

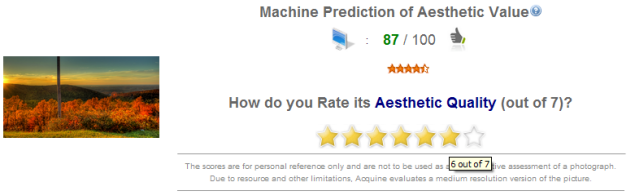


Figure 2: A screenshot of the ratings page that comes up subsequent to a photo upload. The interface also allows the user to rate her own photo, which can serve as a data point for future training.

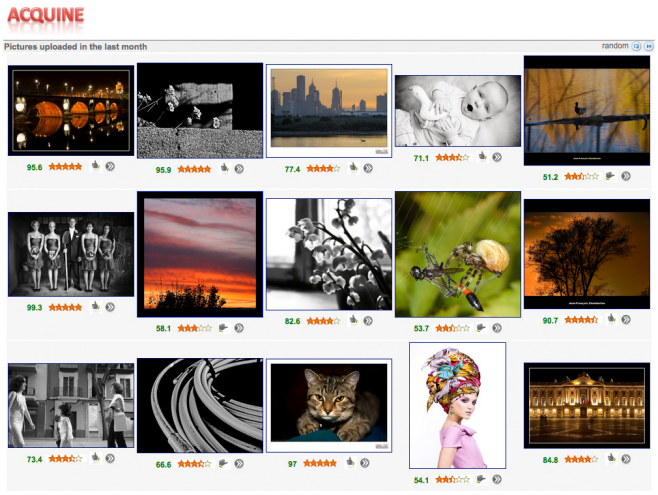
Images can be submitted for automatic rating either by uploading from local storage, or by pasting a URL which is then crawled in. From the time the file is obtained, i.e., factoring out the download time, ACQUINE typically takes under one second to rate a photo, varying slightly based on the image dimensions. Once an image is rated, results are presented as shown in Fig. 2. Alongside the computer prediction, there is also the option for the person uploading the photo to give it a 'human' rating on a 7-star scale (which corresponds with Photo.net's rating scale). This information is stored in our server for future validation and improvements to our classifier. It basically adds to our pool of labeled training data.

The user also has the option to browse through images, with high ACQUINE scores, that were uploaded either in the last 3 hours, last 24 hours, last two days, last week, last month, last three months, or from the very beginning (Fig. 3). The images uploaded within the specified time band are sorted in the descending order of their predicted aesthetics scores. As a policy, in these lists we do not show images receiving scores below 50.

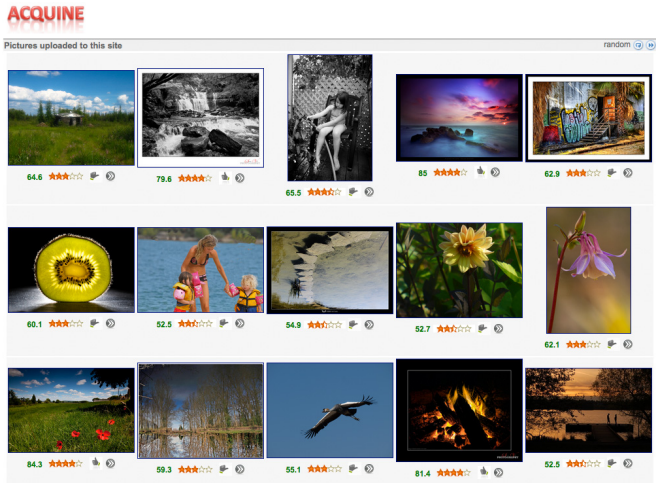
The system tracks user uploads using their IP addresses, which precludes the need for a login mechanism while still being able to collate per-user information (assuming that the same IP address is used by a given user). As can be seen in the right side of Fig. 1, we maintain two rank lists, one for the upload frequency and one for the average rating received. Upon clicking on a particular rank, the entire set of images uploaded by that user (with scores exceeding 50, as per policy) is available for browsing.

4. SYSTEM STATISTICS

The ACQUINE system was launched in April 2009 for public use, and has since resulted in the submission of over 140,000 photos from over 17,000 unique IP addresses (as of October 2009). Of these, over 18,000 have been rated by



(a) uploaded between Oct 15 and Nov 14, 2009



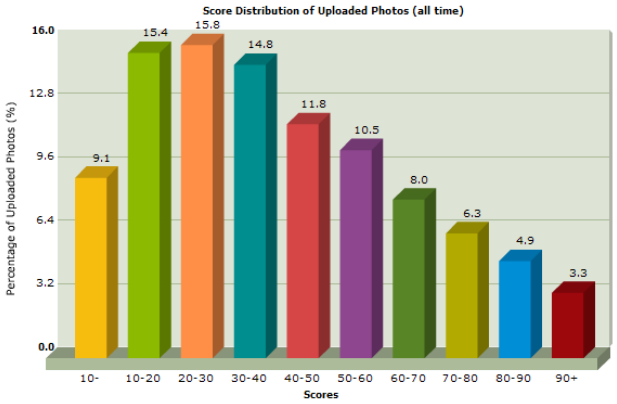
(b) uploaded after Apr 2009

Figure 3: Some random example images uploaded by Web users and rated by ACQUINE with scores higher than 50. Copyright of the photo remains with the original owner(s).

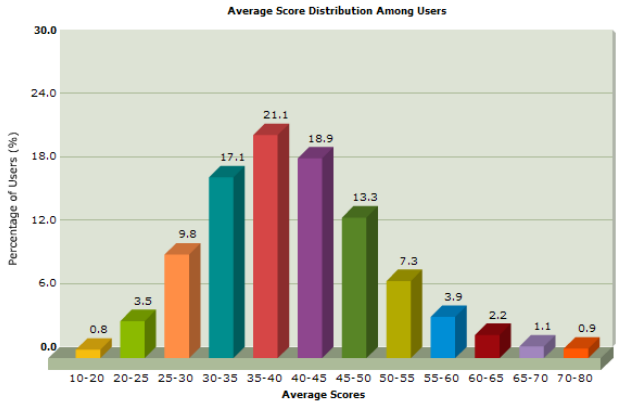
users (Fig. 2) subsequent to their upload. Top ten users each uploaded more than 400 photos.

We also maintain statistics about the kind of aesthetics scores the uploaded photos received from ACQUINE. Based on all photos uploaded from April to October of 2009, the distribution of scores can be seen in Fig. 4(a). Note that only about 3% of the uploads get scores in excess of 90, thereby making getting very high ACQUINE scores a rare occurrence. Most images tend to fall into the 10–60 bracket, with less than 10 being fairly rare as well. We make this graph available to the public via a link on the Website.

Another set of numbers that we maintain and make public is the distribution of average scores that the users get. Only those users who have uploaded at least 15 photos are included in the calculation. By collating the uploads by the IP address source, we compute average scores received over all the photos uploaded. The distribution of these average scores for photos uploaded between April and October of 2009 is shown in Fig. 4(b). It turns out that the per-user



(a) distribution of all ACQUINE scores



(b) distribution of average ACQUINE scores over all users

Figure 4: Statistics on the ACQUINE score distributions obtained with over 140,000 uploaded images from April to October 2009.

average distribution is even more spiked, with the mode being at the 35–40 bin. No user has been able to maintain an average exceeding 90, and only about one percent of all users have been able to maintain an average score above 70.

Through the user interface presented immediately after a photo upload (Fig. 2), users can submit what they think should have been a correct rating. Our hope was that this would add to our training pool of labeled photos. The distribution of the 18,000 user ratings over the 1–7 scale is shown in Fig. 5. When we look at this distribution and compare it with Fig. 4, or with the rating distribution of our training dataset, we find that they are very different shapes. This could be accounted for by various factors, but it is safe to say that users tend to over-rate their own photos, thus making it problematic to use these directly for training without any kind of adjustments. A more thorough analysis of this data is needed before we can make stronger assertions.

5. CONCLUSIONS

We have described our demo system ACQUINE. To the best of our knowledge, this is the first automatic aesthetics rating Website deployed for public use. While the core problem of aesthetics inference (and the general problem of reasoning about emotions computationally) continues to be

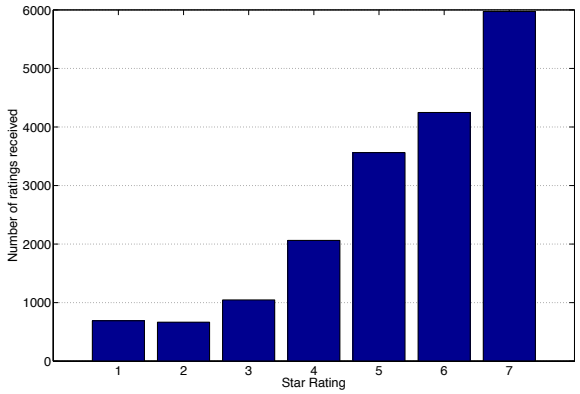


Figure 5: Distribution of ratings users gave to their own photos (From April to October 2009).

intriguing and highly challenging, we believe that a public demo system can go a long way toward generating interest and building better solutions.

In the process of building the real-time demo, we had to figure out ways to balance quality with response time, especially in the feature extraction part. While earlier research guided us in this pursuit, all features could not be computed in real-time with limited resources. We continue to have a steady trickle of users uploading photos every day. We also notice considerable attempts to game the system, e.g., uploading the same photos by making minor and major tweaks, such as cropping and de-colorizing, possibly to see if ACQUINE is consistent with its ratings. We consider this to be healthy skepticism toward the system, which in turn fuels our pursuit for better solutions to the core problem.

Finally, we hope that the ACQUINE system will stimulate public interest and support in developing next-generation computational or robotic systems that can exhibit aesthetics judgment, emotional responses, and other "intelligent" capabilities.

6. ACKNOWLEDGMENTS

The research is supported in part by the US National Science Foundation under Grant Nos. 0347148 and 0202007. We thank Jia Li for valuable discussions and Dhiraj Joshi for contributions to the project in its early stage.

7. 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)*, Special Session on Image Aesthetics, Mood and Emotion, pp. 105-108, San Diego, California, IEEE, October 2008.

[3] Photo.net, `http://photo.net`.