Automatic Aesthetic Photo-Rating System

Chen-Tai Kao
Stanford University
chentai@stanford.edu

Hsin-Fang Wu
Stanford University
hfwu@stanford.edu

Yen-Ting Liu
Stanford University
eggegg@stanford.edu

ABSTRACT

Growing prevalence of smartphone makes photography easier than ever. However, the quality of photos varies widely. Because judging the aesthetic of photos is based on several "rule-of-thumb", it remains difficult for computers to rate photos without manual intervention.

In this work, we utilize aesthetic features of photos and machine learning techniques to automatically distinguish good photos from bad ones. Our system is able to achieve 10-fold cross-validation rate of 82.38%. We believe this technique forms the basis of various novel applications, including real time view-finding suggestion, automatic photo quality enhancement, and massive photo rating.

INTRODUCTION

Rating image aesthetic, as observed in [3] [4], is a very challenging problem. The difficulties are manifold. First, determining image quality remains very subjective. Abundant experience is necessary for being a professional photographer, and there is no effective way to digitalize those rules-of-thumb. Second, the same photo, if viewed by different people with different aesthetic accomplishment, might receive contradicting scores. There lacks consistent principles to classify photos based on their quality. To solve this problem, we need a universal representation of those photography rules, and teach computers to discern good photos from bad ones.

Automatic rating is important because it forms the ground stone of various novel applications useful in multiple stages of digital imaging. Applications spanning from creation, post-processing, and social sharing, are all based on this technique. For example, intelligent camera could have real-time suggestions built into the view-finder, letting the user know where to point and shoot. It would be far greater than simply showing a 3-by-3 grid without any active suggestion, as shown in Figure 1. Also, post-processing software can automatically determine the best way to enhance photos without any manual intervention. Furthermore, if equipped with this technology, social websites like Facebook and Flickr would be able to recommend great photos more frequently than photos with poor-quality. In short, we see a high demand in automatic photo-rating that has the potential to make photography friendlier and more intelligent.

Figure 1. An example of passive suggestion, showing a 3-by-3 grid on an iPhone when user takes a photo.

In this work, we picked multiple aesthetic features and modeled them as simple and intuitive features. These features were trained using automatic classifiers such as random forest, SVM and Bayes network. Finally, a model is generated to predict the aesthetics class of any photo. Figure 2 shows the framework of the overall system.

We collect a dataset of 1942 images from DPChallenge, a photograph contest website [1], where people submitted photos to be rated by the public. One advantage of adopting this photo database is that these photos have been quantitatively scored from 1 to 10 by a large set of users. We collect 1000 top rated images with average rate between 7.4 to 8.6 points as the high quality photos, and 942 lowest rated images that are scored between 1.8 to 3.2 points as low quality photos.

The rest of this paper is organized as follows. First, we introduce aesthetic features used in the model. The training methods are presented thereafter. Finally, experimental results are illustrated and discussed.

AESTHETIC FEATURES

To design features representing photo quality, we determine the perceptual criteria that people used to judge photos. We reference principles of photography and select several important criteria used by professional photographers to improve photo quality. In our system, we need saliency map as a way to segment object and determine area of interest. We adopt the saliency map proposed by [12], which is fast and robust.

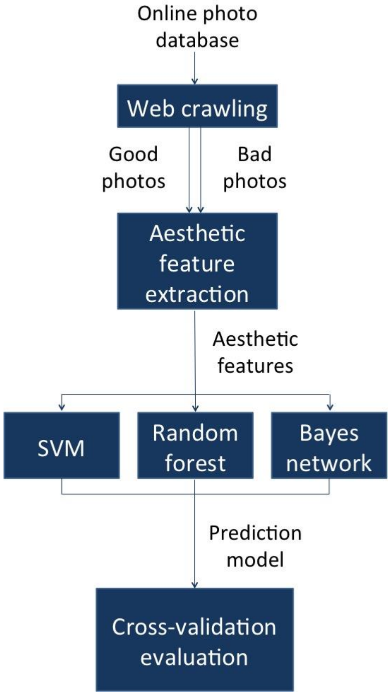

Figure 2. Framework of our automatic photo-rating system.

Background Complexity

Attractive photos usually contain simple background as a way to highlight the object in the foreground. In an image, rejoin with high saliency is considered as foreground and the rest are considered as background. We use ratio of edges in background to indicate the background complexity. The intuition behind this feature is that complex background is very likely to contain large amount of edges. A set of background complexity features with 10 dimensions is extracted per photo.

$$f_{\text{Background\_Complexity}} = \frac{\text{number of background pixels that are edge}}{\text{number of total pixels}}$$

Blurriness

A blurry photo is usually considered low quality. To model this effect, we calculate the Laplacian pyramid of the image with three stacks. For each stack of the pyramid, the ratio of pixels that are edge is used as a feature. This is because blur photo tends to have wider edges, which are more likely to be detected at higher stack of the pyramid.

$$f_{\text{Blurriness},k} = \frac{\text{number of edge pixels at stack k of the pyramid}}{\text{number of total pixels at stack k of the pyramid}},$$

where $k = 1, 2, 3$, since we used three layers.

A set of blurriness features with 3 dimensions is extracted per photo.

Centroid of saliency and color

Good photos have good composition, meaning that all objects are balanced around center. That is, if there is an object on the left side, then there should be another object on the right side to balance it, preventing the photo from tilting toward one side. Therefore, we hypothesize that good photos should have their saliency centroid in some certain position. To obtain the centroid, we consider all pixels with high saliency, and calculate the centroid regarding the coordinate of those pixels. Note that we use percentage to denote the centroid, e.g. $(x, y) = (50\%, 50\%)$, which is more general regardless of the image's resolution.

$$f_{\text{Saliency\_Centroid}} = \frac{\sum_i r_i s_i}{\sum_i s_i},$$

where $r_i = [x_i, y_i]$ is the coordinate of pixel i and s_i is the saliency value of pixel i.

We also consider centroid of color value. A well-known phenomenon is that different color has different weight, e.g. red is often considered heavy, and yellow is often regarded light. In this work, we adopt Ou's model [10] to assign weight to each pixel based on its color. We then compute the color centroid of all pixels.

$$f_{\text{Saliency\_Centroid}} = \frac{\sum_i r_i w_i}{\sum_i w_i},$$

where $r_i = [x_i, y_i]$ is the coordinate of pixel i and w_i is the weight of pixel i given by

$$w_i = -1.8 + 0.04(100 - L) + 0.45\cos(h - 100),$$

where L and h are HSL space value of pixel i.

Both saliency centroid and color centroid are used as feature. See Figure 3 for an example of color centroid. A set of centroid features with 4 dimensions (2 for saliency centroid and 2 for color centroid) is extracted per photo.

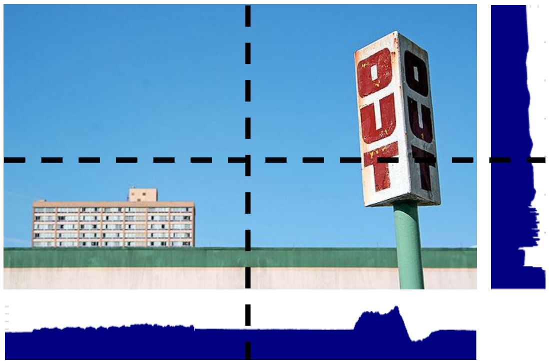

Figure 3. An example showing color centroid of an image. Histogram on the side roughly illustrates the distribution of pixel weight. The centroid of each coordinate is shown on the image, where the intersection is the final centroid.

Contrast

Human visual system is more sensitive to contrast than color or luminance. Figure 4 illustrates the effect of varying contrast of an image. There are many ways to calculate the contrast of an image. Here, we use root-mean-square contrast, the standard deviation of RGB value, to evaluate the contrast of a photo.

$$f_{\text{Contrast}} = \sqrt{\frac{1}{MN}\sum_i \sum_j \left(I_{ij} - \bar{I}\right)^2},$$

where M and N are the width and height of the image, respectively. A set of contrast features with 3 dimensions (RGB) is extracted per photo.

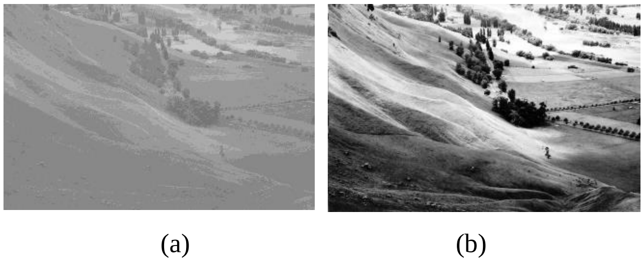

Figure 4. (a) The image with low contrast. (b) The same image with higher contrast. In general, (b) is considered better than (a).

Color Histogram

We hypothesize that the color distribution of an image encodes some information of photo quality. For example, pixels with warm color tend to dominate sunset photos. We use color histogram of RGB, YUV, and HSV, to model this effect. A set of color histogram features with 256x9 dimensions is extracted from each photo.

Noise

Noisy photos are often considered low quality. To calculate the amount of noise, we perform non-local means denoising [11] to obtain the denoised photo, which is then subtracted from the original photo to obtain the noise amount. Root-mean-square of the noise is used as feature.

$$f_{\text{Noise}} = RMS(I - I_{\text{Denoised}}),$$

where I and $I_{Denoised}$ are the original image and the denoised one, respectively. A set of noise features with 3 dimensions is extracted per photo.

Rule of Thirds

Rule of thirds is a popular aesthetic rule in photography. Consider dividing the image into 3-by-3 grids. It is preferred that objects being placed near the intersection of the grid. Figure 5 demonstrates the comparison of two photos where the subject is aligned with the grid in one photo but not the other. We model this feature by applying four 2-D Gaussian distributions as weighting function on the grid such that the center of each Gaussian is placed at each intersection of the grid. Therefore, pixels near the grid are multiplied by higher weight, and pixels farther from the grid are weighted less. We then use the weighted sum of saliency values to represent this feature.

By varying the parameter of Gaussian distribution and the saliency threshold, we have a set of rule of thirds features with 50 dimensions for each photo.

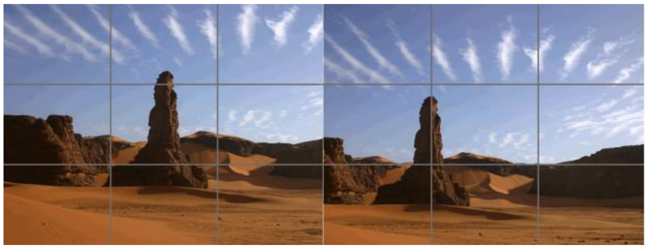

Figure 5. Demonstration of rule of thirds. Photos are regarded as better if objects are placed around the 3-by-3 grid, especially on the intersections.

Symmetry

Sometimes, symmetry implies a sense of beauty. In this work, top-down and left-right symmetry are calculated by convolving saliency value of pixels on the two halves. The result of convolution is used as this feature. To tolerate small amount of inexact symmetry, we compute the convolution ten times, each time shifting one half a little bit (2% of the width). At last, we pick the largest convolution among all iterations to denote this feature.

$$f_{\text{Symmetry}} = \sum_i \sum_j I_1(i,j)\, I_2(i,j),$$

where I_1 is the left (or upper) half of the image, and I_2 is the other half that is shifted. A set of symmetry features with 2 dimensions (top/down and left/right) is extracted per photo.

Gray scale

Noticing that a great amount of good photos are gray scale image, we add this feature to distinguish whether the photo is gray scale.

$$f_{\text{Grayscale}} = \begin{cases} 1, & \text{if image is grayscale} \\ 0, & \text{otherwise} \end{cases}.$$

Mean and variance of color

The mean and variance descriptors are utilized to describe statistic properties of an image. We calculate mean and variance as a pair of the nine layers extracted from RGB, HSV, and YUV, color space of the image. A set of blurriness features with 18 (2x9) dimensions is extracted per photo.

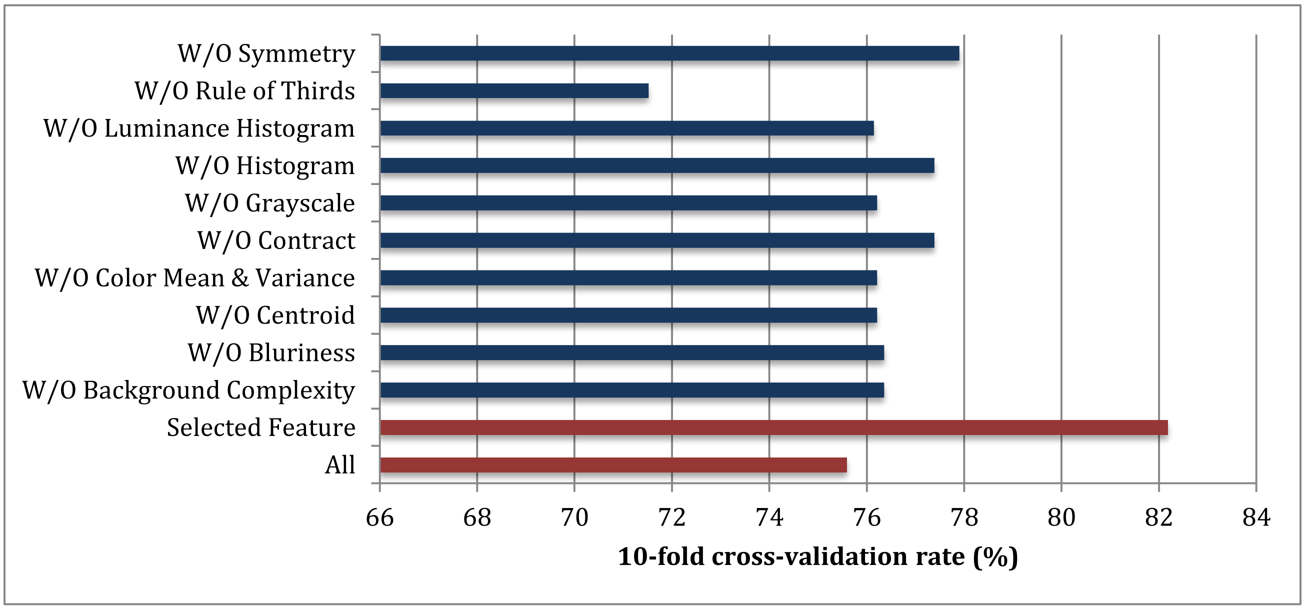

Figure 6. Photo quality classification accuracy with different combination of aesthetic features

TRAINING METHODS

We trained the feature data with 3 different learning methods: SVM, random forest and Bayes network. We selected the parameters of SVM by performing a grid search on the C and γ. For random forest, we constructed a forest of 300 random trees in training phase. The Bayes network was constructed by K2 algorithm.

The original data contains 10 sets of features with 2634-dimension. We performed forward feature selection to remove potentially ineffective dimensions. Correlation-based feature subset selection method was utilized to reduce the feature data to 27-dimension.

Table 1 compares the performance of 3 learning methods, SVM, random forest (RF), and Bayes network (BN). Random forest outperforms the other two methods both in all feature case and selected feature case. By selecting effective feature, random forest achieves 82.38% of 10-fold cross-validation accuracy.

	SVM	RF	BN
All	76.44%	80.33%	70.23%
Selected features	80.48%	82.38%	80.99%

Table 1. Learning method comparison.

EXPERIMENTAL RESULTS

To evaluate the effectiveness of each aesthetic feature, we performed a single iteration of backward feature selection process. That is, we remove one set of feature each time, and then we train and calculate the 10-fold cross-validation rate using random forest consisted of 100 random trees. Figure 6 shows the accuracy with different combination of aesthetic feature sets. Rule of thirds plays an important rule due to the 4.07% decrease in accuracy without rule of thirds feature. It's also obvious that after selecting effective features, the accuracy increased by 7%.

Table 2 presents the overall performance measurement of random forest, including true positive (TP), false positive (FP), precision, and recall. While the rate of true positive is high, the false positive rate remains low enough so that precision rate is in a reasonable range.

	TP	FP	Precision	Recall
Good	86.4%	21.9%	80.7%	83.5%
Bad	78.1%	13.6%	84.4%	81.1%
Avg	82.4%	17.9%	82.5%	82.3%

Table 2: Performance Measurement of Random Forest

DISCUSSION

Collecting bad photo into our dataset is one of the biggest challenges we face. Existing photo databases often contain good photos; therefore, it's hard to obtain massive bad quality photos online.

In this work, we design several aesthetic features based on principles of photography. However, there exists ways to extend the feature set, such as dividing images into patches and local binary patterns (LBP), which is popular in many classification problem of computer vision. Moreover, some experiments can be conducted with the aesthetic feature set not only on general photographs, but also on different topics, such as scenic photos or portrait photo with human faces.

The proposed automatic photo rating system can be further used in many applications, as illustrated in the first chapter.

Some examples include automatically remove low-quality photo and real-time recommendation of view-finding.

REFERENCES

1. http://www.dpchallenge.com/
2. H.H Su, T.W.Chen, C.-C. Kao, W.H.Hsu, and S.-Y. Chien, "Preference-aware view recommendation system for scenic photos based on bag of aesthetics-preserving features," *IEEE ToMM*, Vol. 14, No. 33, 2012.
3. R.Datta, D.Joshi, J.Li and J.Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," *Proc. ECCV*, 2006.
4. Y. Ke, X. Tang, and F. Jing, "The Design of High-Level Features for Photo Quality Assessment," *Proc. CVPR*, 2006.
5. Lok, S., Feiner, S., and Ngai, G. "Evaluation of visual balance for automated layout," *Proceedings of the 9th international conference on Intelligent user interface*, 2004.
6. Y. Luo and X. Tang. "Photo and video quality evaluation: Focusing on the subject," *ECCV*, 2008.
7. Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol*. 2, 3, Article 27, 2011.
8. Leo Breiman. "Random Forests," *Machine Learning*. 45(1):5-32, 2001.
9. G. Cooper, E. Herskovits. "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, 9(4):309-347, 1992.
10. Li-Chen Ou, M. Ronnier Luo, Andree Woodcock, and Angela Wright. "A study of colour emotion and colour preference," *Color Research and Application*, 29(3):232-240, 2004.
11. Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "Non-Local Means Denoising," *Image Processing On Line*, vol. 2011.
12. Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency Detection via Graph-based Manifold Ranking," *CVPR*, 2013.