1

Studying Aesthetics in Photographic Images Using a Computational Approach

Wei-Ta Chu

National Chung Cheng University

R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach" ECCV, 2006.

Introduction

- While the average individual may be interested in how soothing a picture is to the eyes, a photographic artist may be looking at the composition of the picture, the use of color and light, and etc.
- Aesthetic quality assessment is extremely subjective.
- However, there exist certain visual properties which make photographs more aesthetically beautiful.

Community-Based Photo Ratings

- Photo.net
- More than one million photographs
- These photos are peer-rated in terms of two qualities, namely *aesthetics* and *originality*, and given scores in the range of one to seven, with a higher number indicating better rating.
- Photos are rated by a relatively diverse group, ensuring generality in the ratings.

Computational Aesthetics Approach

- 1. Build a classifier to *qualitatively* distinguish between pictures of *high* and *low* aesthetic score.
- 2. Build a regression model to *quantitatively* predict the aesthetic score.

- 1. Measures are highly subjective, and there are no agreed standards for rating.
- 2. Lead to better understanding of the human vision.

Visual Feature Extraction

- Choice of features: (1) rules of thumb in photography, (2) common intuition, and (3) observed trends in ratings
- Using the HSV color space
- Image segmentation
- Totally 56 visual features are extracted.

Exposure of Light and Colorfulness

- Light: The average pixel intensity of a picture

$$f_1 = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_V(x, y)$$

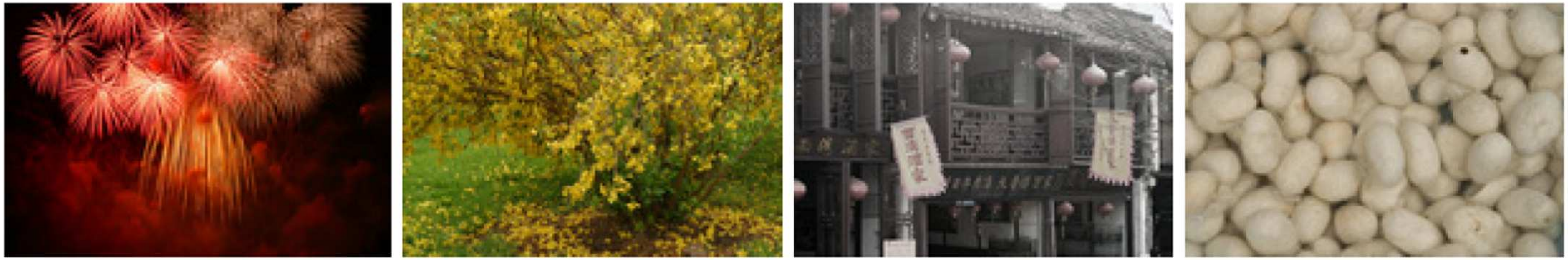
- Colorfulness:
  - Divide the RGB color space into 64 cubic blocks with four equal partitions along each dimension.
  - Distribution D_1: the color distribution of a hypothetical image such that for each of 64 sample points, the frequency is 1/64.
  - Distribution D_2: the color distribution of the given image

$$f_2 = EMD(D_1, D_2, \{d(a,b) | 0 \le a, b \le 63\})$$

$$d(a,b) = ||rgb2luv(c_a) - rgb2luv(c_b)||$$

Exposure of Light and Colorfulness

- The distribution D_1 can be interpreted as the ideal color distribution of a “colorful” image.
- How similar the color distribution of an arbitrary image is to this one is a rough measure of how colorful that image is.



High colorfulness

Low colorfulness

Saturation and Hue

- Saturation indicates chromatic purity. Pure colors in a photo tend to be more appealing than dull or impure ones.
- Average saturation:

$$f_3 = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_S(x, y)$$

- Average hue:

$$f_4 = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_H(x, y)$$

The Rule of Thirds

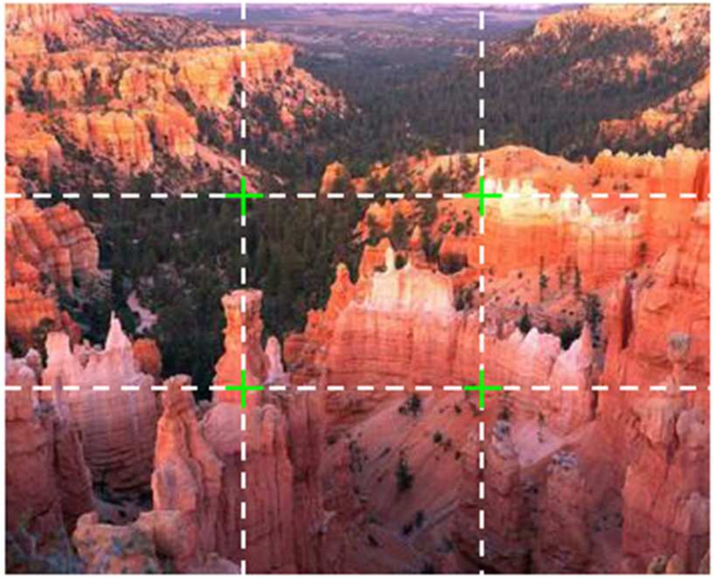
- The main element, or the center of interest, in a photograph should lie at one of the four intersections.
- The average hue:

$$f_5 = \frac{9}{XY} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_H(x,y)$$

- The average saturation
- The average intensity

$$f_6 = \frac{9}{XY} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_S(x,y)$$

$$f_7 = \frac{9}{XY} \sum_{x=X/3}^{2X/3} \sum_{y=Y/3}^{2Y/3} I_V(x,y)$$



Familiarity Measure

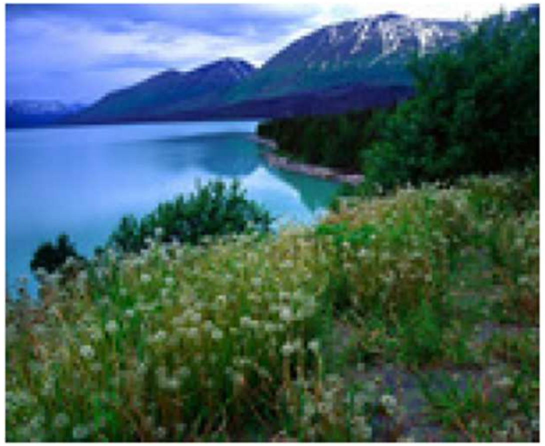
- Our opinions are often governed by what we have seen in the past.
- Integrated region matching (IRM) image distance
- Given a pre-determined anchor database of images with a well-spread distribution of aesthetic scores, we retrieve the top *K* closest matches in it with the candidate image as query.
- Let $\{q(i)|1 \leq i \leq K\}$ denote the IRM distances of the top matches

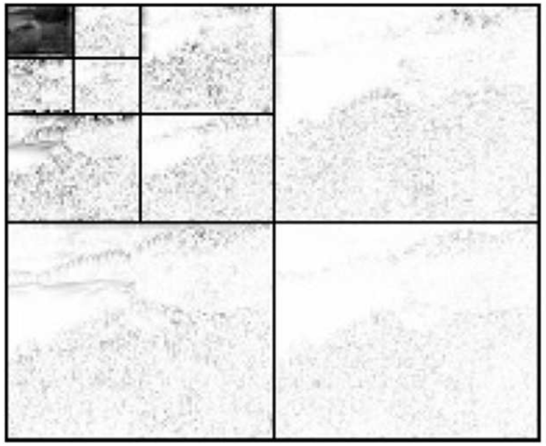
$$f_8 = \frac{1}{20}\sum_{i=1}^{20} q(i) \qquad f_9 = \frac{1}{100}\sum_{i=1}^{100} q(i)$$

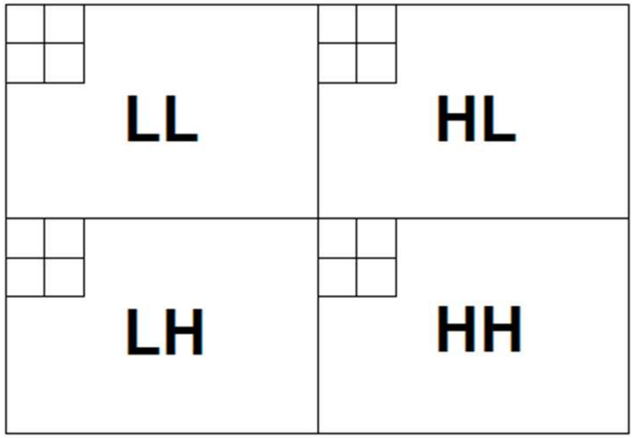
Wavelet-Based Texture

- Measure spatial smoothness. Perform three-level wavelet transform on all three color bands I_H, I_S, I_V
- Denoting the coefficients in level *i* for the wavelet transform on hue image I_H as $w_i^{hh}, w_i^{hl}, w_i^{lh}$

$$f_{i+9} = \frac{1}{S_i}\left\{\sum_x \sum_y w_i^{hh}(x,y) + \sum_x \sum_y w_i^{hl}(x,y) + \sum_x \sum_y w_i^{lh}(x,y)\right\}$$







Wavelet-Based Texture

- The corresponding wavelet features for saturation and intensity images are computed similarly to get f_{13} through f_{15} and f_{16} through f_{18} respectively.
- The sum of the average wavelet coefficients over all three frequency levels for each of *H*, *S*, and *V* are taken to form three addition features

$$f_{19} = \sum_{i=10}^{12} f_i \qquad f_{20} = \sum_{i=13}^{15} f_i \qquad f_{21} = \sum_{i=16}^{18} f_i$$

Size and Aspect Ratio

- Although scaling is possible in digital and print media, the size presented initially must be agreeable to the content of the photograph

$$f_{22} = X + Y$$

- 4:3 and 16:9 aspect ratios are well known, which are approximate the “golden ratio”

$$f_{23} = \frac{X}{Y}$$

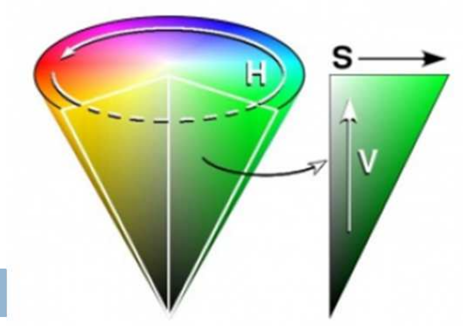
Region Composition

- Denote the set of pixels in the largest five connected components formed by the segmentation process as $\{s_1, \ldots, s_5\}$. The number of patches $t \leq 5$ which satisfy $|s_i| \geq \frac{XY}{100}$ denotes feature f_{24}
- The number of color-based clusters formed by K-means in the LUV space is feature f_{25}
- These two features combine to measure how many distinct color blobs and how many disconnected significantly large regions are present.

Region Composition

- The average H, S, and V values for each of the top 5 regions as features f_{26} through f_{30}, f_{31} through f_{35} and f_{36} through f_{40} respectively.
- Feature f_{41} through f_{45} store the relative size of each region with respect to the image.

Region Composition



- Average color spread around the color wheel and average complimentary colors among the top 5 region hues.

$$f_{46} = \sum_{i=1}^{5} \sum_{j=1}^{5} |h_i - h_j| \qquad h_i = \sum_{(x,y)\in s_i} I_H(x,y)$$
$$f_{47} = \sum_{i=1}^{5} \sum_{j=1}^{5} l(|h_i - h_j|)$$

where $l(k) = k \quad \text{if } k \le 180^\circ$, $l(k) = 360^\circ - k \quad \text{if } k > 180^\circ$

- The rough positions of each region. Divide the image into three equal parts along horizontal and vertical directions, locate the block containing the centroid of each region s_i , and set $f_{47+i} = 10r + c$, $(r,c) \in \{(1,1), ..., (3,3)\}$

Low Depth of Field Indicators

- Professional photographers often reduce the depth of field (DOF) for shooting single objects. DOF is the range of distance from a camera that is acceptably sharp in the photograph.
- Divide the image into 16 equal rectangular blocks $\{M_1, \ldots, M_{16}\}$, numbered in row-major order. Let $w_3 = \{w_3^{lh}, w_3^{hl}, w_3^{hh}\}$ denote the set of wavelet coefficients in the high-frequency of the hue image.

$$f_{53} = \frac{\sum_{(x,y)\in M_6 \cup M_7 \cup M_{10} \cup M_{11}} w_3(x,y)}{\sum_{i=1}^{16} \sum_{(x,y)\in M_i} w_3(x,y)}$$

Shape Convexity

- We hypothesize that convex shape like perfect moon, well-shaped fruits, boxes, or windows have an appeal, positive or negative, which is different from concave or highly irregular shapes.
- Find *R* patches $\{p_1, \ldots, p_R\}$ such that $|p_k| \geq \frac{XY}{200}$
- For each p_k , we compute its convex hull, denoted by $g(p_k)$. We define the shape convexity features as

$$f_{56} = \frac{1}{XY}\{\sum_{k=1}^{R} I(\frac{|p_k|}{|g(p_k)|} \geq 0.8)|p_k|\}$$

Shape Convexity

- This feature can be interpreted as the fraction of the image covered by approximately convex-shaped homogeneous regions, ignoring the insignificant image regions.

Fig. 6. Demonstrating the *shape convexity* feature. *Left*: Original photograph. *Middle*: Three largest non-background segments shown in original color. *Right*: Exclusive regions of the *convex hull* generated for each segment are shown in white. The proportion of white regions determine the convexity value.

Feature Selection, Classification, and Regression

- For the 3581 images downloaded, all 56 features were extracted and normalized to the [0,1] range to form the experimental data.
- Two classes of data are chosen, *high* containing samples with aesthetics scores greater than 5.8, and *low* with scores less than 4.2.
- For all experiments we ensure equal priors by replicating data to generate equal number of samples per class.

Feature Selection, Classification, and Regression

- Construct one-dimensional SVM classifiers.
- SVM is run 20 times per feature, randomly permuting the dataset each time, and using a 5-fold cross validation.
- The top 15 among the 56 features in terms of model accuracy are obtained.
- We proceed to build a classifier to separate *low* from *high* – SVM associated with the classification and regression trees (CART).

Feature Selection, Classification, and Regression

- Feature selection: combine filter-based method and wrapper-based method
  - (1) the top 30 features in terms of their one-dimensional SVM performance are retained
  - (2) Forward selection, a wrapper-based approach in which we start with an empty set of features and iteratively add one feature at a time that increases the 5-fold CV accuracy the most. We stop at 15 iterations and use this set to build the SVM-based classifier.

Feature Selection, Classification, and Regression

- We perform linear regression on polynomial terms of the features values to see if it is possible to directly predict the aesthetics scores.
- Quality of regression: residual sum-of-squares error

$$R_{res}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

  where $\hat{Y}_i$ is the predicted value of Y_i

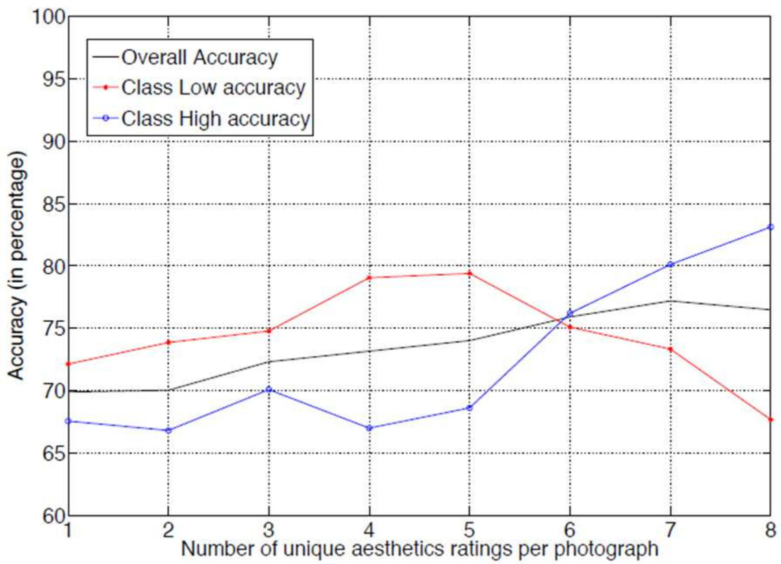
- In the worst case $\bar{Y}$ is chosen every time, yielding $R_{res}^2 = \sigma^2$.

Experimental Results

- For the one-dimensional SVM performed on individual features, the top-15 features are *f*31, 1, 6, 15, 9, 8, 32, 10, 55, 3, 36, 16, 54, 48, 22.
- The combined filter and wrapper method for feature selection yielded the 15 features: f31, 1, 54, 28, 43, 25, 22, 17, 15, 20, 2, 9, 21, 23, 6.
- The accuracy achieved with 15 features is 70.12%, with precision of detecting *high* class being 68.08%, and *low* class being 72.31%.

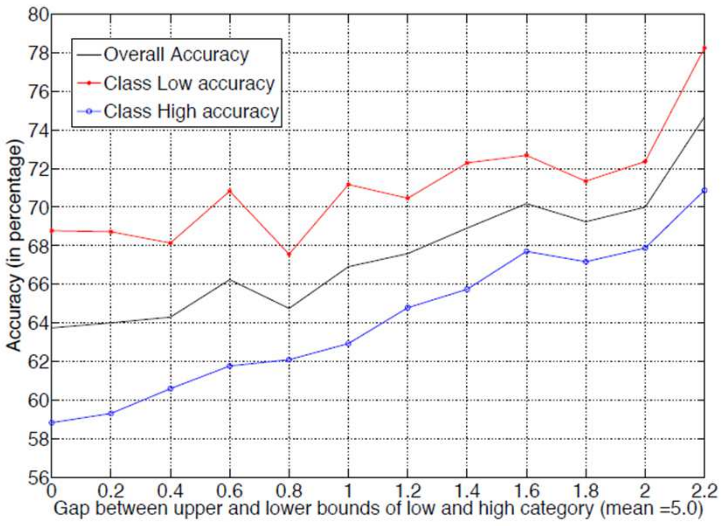
Experimental Results

- Stability of classification results
- Samples are chosen in such a way that each photo is rated by at least *K* unique users, *K* varying from 1 to 8
- Accuracy values show an upward trend with increasing number of unique ratings per sample, and stabilize somewhat when this value touches 5.



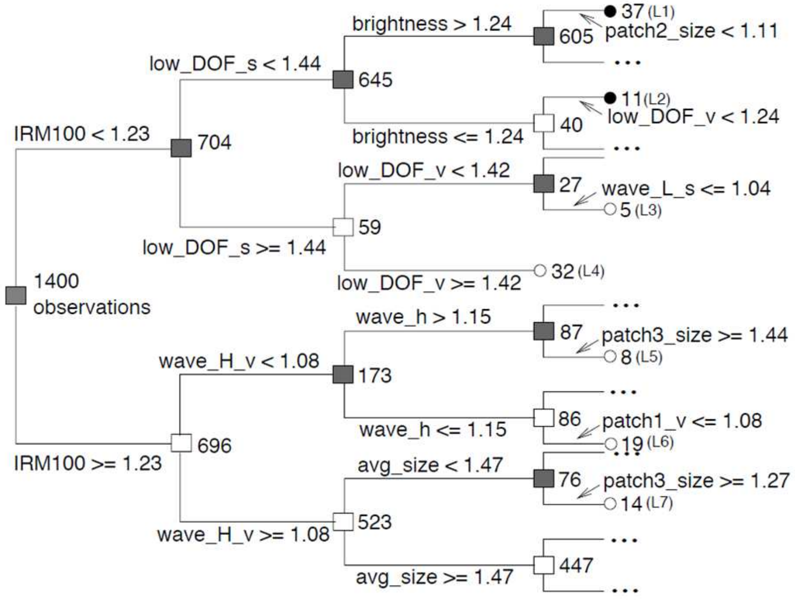
Experimental Results

- Experiment with how accuracy and precision varied with the gap in aesthetic ratings between the two classes *high* and *low*.
- So far we have considered ratings $\geq$ 5.8 as *high* and $\leq$ 4.2 as *low*. In general, considering that ratings $\geq 5.0 + \frac{\delta}{2}$ as *high* and ratings $\leq 5.0 - \frac{\delta}{2}$ as *low*.



Experimental Results

- The CART decision tree obtained using the 56 visual features.
- Shaded nodes have a higher percentage of *low* class pictures, while un-shaded nodes are those where the dominating class is *high*.



Experimental Results

- The variance of the aesthetics score over the 3581 samples is 0.69.
- We achieved a residual sum-of-squares $R^2_{res} = 0.5020$
- Visual features are able to predict human-rated aesthetics scores with some success.

Conclusion

- Certain visual properties tend to yield better discrimination of aesthetic quality than some others.
- SVM-based classifier is robust enough to produce good accuracy using only 15 visual features in separating *high* and *low* rated photographs.