Summary Report

Hang Chu[1], Tsuhan Chen[1], Dong-Qing Zhang[2], Heather Yu[2]
[1]Advanced Multimedia Processing Lab, Cornell University
[2]Media Lab, FutureWei Technology

January 26, 2015

1 Summary

We have surveyed the state-of-the-art for photo capture guide and photo post-processing based on aesthetic models. We have proposed an improved aesthetic model, useful for both applications, and demonstrated the effectiveness of the proposed model. In addition, we have implemented algorithms for both photo capture guide and photo post-processing, based on the improved aesthetic model. This report summarizes our results.

2 Aesthetic models

To evaluate the aesthetic value of a photo, we first determine a set of bounding boxes indicating possible locations of objects in the photo. To do this, we construct a set of training images, with objects labeled by bounding boxes. We also generate randomly non-object bounding boxes as negative samples. For a box, we compute features such as SIFT, color contrastness, and main color components, to form a feature vector. Then, a support vector machine (SVM) based classifier between object boxes and non-object boxes is trained.

Given an input image, we can then apply the trained classifier using sliding windows of different sizes, and select windows with high object probability. This procedure produces a set of windows, each with a corresponding score specifying the likelihood of the window containing an object. This objectness measurement has advantages of being general and object category independent. Figure 1 shows some example outputs of objectness measurement. In most cases, our algorithm is able to produce image regions containing objects that attract most attention of a human viewer.

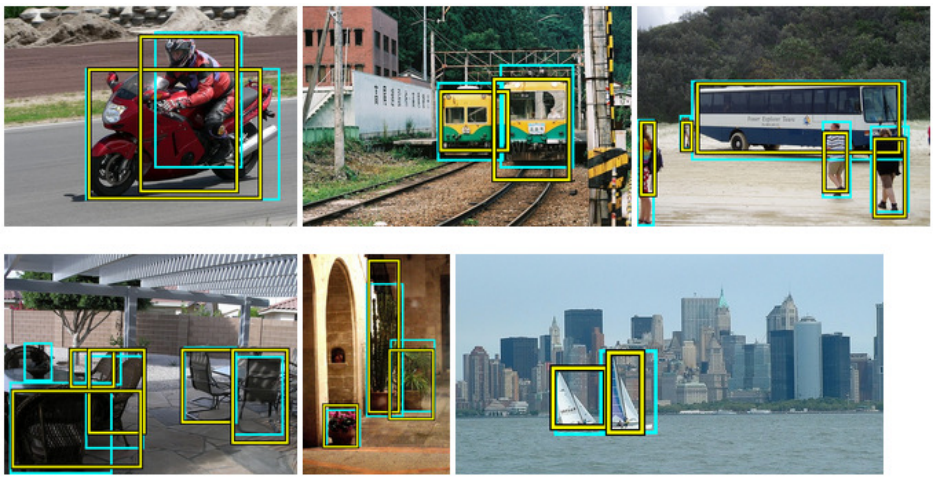

Figure 1: Examples of object measurement. Cyan and yellow boxes show detected object windows and ground truth object bounding boxes, respectively.

In the next step, we compute color, lighting, and composition features following a similar manner as the method in [6] for each detected bounding boxes. More specifically, we compute color and lighting statistics in the object region and lighting and color contrasts between the object region and background region. We also measure the spatial organization of detected objects and their relationship towards several popular photography rules, such as the rule of thirds, and the rule of visual balance.

Based on the extracted object-related features, we use a sparse regressor to compute coefficients for mapping features to an aesthetic score. As the concept of aesthetic is vague, it is difficult to tell which features actually affect human judgement. Thus we choose a sparsity-embedded method that

automatically determines a small subset of important features.

The proposed algorithm has two advantages. First, compared to the baseline algorithm, the proposed algorithm is able to generate an aesthetic score rather than just a binary label, i.e., *high quality* versus *low quality*. This gives more specific aesthetic information to the users and also enables wider range of potential applications. Second, compared to [6], the proposed algorithm can be applied to any type of images rather than just face images, which significantly expands the algorithm's usefulness.

3 Capture Guide and Post-Processing: for images

Once we have an improved aesthetic model, we can use it to serve as a capture guide. More specifically, we can post-process an input by cropping the input image into a sub-image that has better aesthetic value (or guide the photographer to move the camera and zoom in to the sub region, which has the same effect). Either to provide a capture guide or to post-process the image, we consider three types of features: the visual balance feature, the "rule of thirds" feature, and the relative size between object boxes and the cropping bounding box. In a visual balanced image, the visually salient objects are distributed evenly around the center and the center of the "visual mass" is nearby the center of the image. In the rule of thirds, the image is divided into nine equal parts by two equally spaced horizontal lines and two such vertical lines, the four intersections formed by these lines are referred to as the "power points," where the main subject should be placed at. Figure 2 shows examples of visual balance and the rule of thirds.

To find the best sub region, the search is performed in the space of the parameters that define the region. We keep the aspect ratio of the cropped image the same as that of the original image to limit the dimensionality of the search space. Therefore, only three parameters need to be determined: the position (x, y) of the cropped image, and its size. We apply the Particle Swarm Optimization technique to compute the final cropping box.

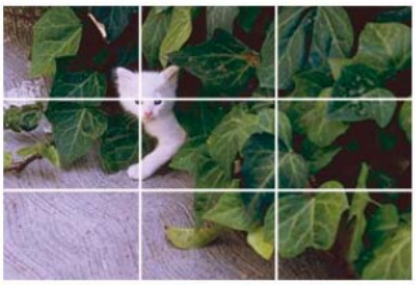

Figure 2: Examples the rule of thirds (left) and a visually balanced image (right).

3.1 Experimental results

We tested the baseline and proposed algorithms on the dataset from [6] and part of the AVA dataset [3]. Table 1 shows a comparison in average error in scores. It should be noted that as the baseline algorithm can only perform binary classification, we use 7.5 and 2.5 as scores produced by the baseline algorithm with good and bad labels.

Table 1: Quantitative comparison between baseline and proposed algorithms.

	baseline	proposed
Dataset from [6] $RMSE_{score}$	2.20	1.56
AVA dataset [3] $RMSE_{score}$	2.24	0.68

In addition, Figure 3 shows a comparison of results of the baseline and proposed algorithms.

Figure 3: Comparison of the baseline (top) and the proposed algorithms (bottom).

4 Capture Guide and Post-Processing: for videos

The usage of video recording devices has been increasing exponentially as smart phones are becoming more and more popular. However, we observed that most users lack the knowledge or awareness for positioning the camera to make aesthetically attractive videos. Based on our previous experience on aesthetic evaluation of photos, we investigate the problem of aesthetic-based video editing. The goal is to develop an algorithm that takes a raw video as input, and automatically produces an edited video with a higher aesthetic value.

4.1 Aesthetic Model

We started with a simple aesthetic model. In this model, two factors are considered to evaluate the aesthetic value of a single frame: the rule of thirds,

and the cropping size. In the definition of the rule of thirds, the image is equally divided into nine equal parts by two equally spaced horizontal lines and two equally spaced vertical lines, and the important compositional elements should be placed near the four intersecting points. We model the influence of the rule of thirds as a normal distribution centered at the nearest intersecting point, i.e.,

$$s_{rot} = \mathcal{N}(min\{x_{obj} - x_{inter}^{i=1,2,3,4}\}, \sigma_{rot})$$

where x_{obj} is the center of the main object. We model the influence of the cropping size simply as the ratio of the cropped size to the original frame size, i.e.

$$s_{cs} = \frac{Size_{cropped}}{Size_{original}}$$

Finally the aesthetic value of an image frame is computed as

$$s = s_{rot} \cdot s_{cs}$$

4.2 Video Editing

In the current version of our algorithm, we adopt an intuitive video editing approach. We select one frame after a constant time interval, and search the optimal cropping parameter for the selected frame. The cropping parameters for frames between two selected frames are then determined by interpolation.

Figure 4 shows a screen capture of our demo video, where the post-processed video demonstrates better aesthetic quality thanks to proper cropping. Note how the post-processed video follows nicely the rule of thirds by positioning the person one third to the left.

Figure 4: Our demo showing the videos before (left) and after (right) the proposed post-processing algorithm.

5 Future Work

Motivated by the success in our study, we would like to suggest the following directions for possible extensions:

- More photography compositional rules can be added to the aesthetic model, such as leading lines, symmetry, etc.
- More sophisticated keyframe selection methods can be used, such that the algorithm can intelligently select only the most important frames.
- Inter-keyframe smoothing can be added so that the produced video will not contain too much jerkiness due to sudden changes in the cropping parameters.

6 References

[1] K. Lo, K. Liu, and C. Chen, Assessment of Photo Aesthetics with Efficiency, *ICPR* 2012.

[2] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, Assessing the Aesthetic Quality of Photographs using Generic Image Descriptors, *ICCV* 2011.

[3] N. Murray, L. Marchesotti, and F. Perronnin, AVA: A Large-Scale Database for Aesthetic Visual Analysis, *CVPR* 2012.

[4] S. Dhar, V. Ordonez, and T. Berg, High Level Describable Attributes for Predicting Aesthetics and Interestingness, *CVPR* 2011.

[5] W. Luo, X. Wang, and X. Tang, Content-Based Photo Quality Assessment, *ICCV* 2011.

[6] C. Li, A. Loui, and T. Chen, Towards Aesthetics: a Photo Quality Assessment and Photo Selection System, *ACM MM* 2010.

[7] C. Li, and T. Chen, Aesthetic Visual Quality Assessment of Paintings, *IEEE Journal of Selected Topics in Signal Processing* 2009.

[8] B. Alexe, T. Deselares, and V. Ferrari, What is an Object?, *CVPR* 2010.

[9]H. Zhou, and K. Lange, A Path Algorithm fir Constrained Estimation, *Journal of Computational and Graphical Statistics*, 2013.

[10]J. Kennedy, and R. Eberhart, Particle Swarm Optimization, *IEEE Conf. on Neural Networks* 1995.