Assessment of Photo Aesthetics with Efficiency

Kuo-Yen Lo[1], Keng-Hao Liu[2], and Chu-Song Chen[1,2]

[1]*Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan*
[2]*Institute of Information Science, Academia Sinica, Taipei, Taiwan*
kylo@citi.sinica.edu.tw, {keng3, song}@iis.sinica.edu.tw

Abstract

Photo quality assessment has been a popular research topic. Many previous works achieved high classification rates in photo aesthetics assessment by designing new aesthetic features. However, those hand-crafted features sometimes are not describable, or are very time-consuming and thus not applicable for real-time applications. In this paper, we propose aesthetic features with high efficiency to compute. The experimental results show that our proposed features reach considerable performance. The computation consumption for classifying an image is low so that it is possible to realize online assessment in photo capturing and provide instant feedback to users or fulfill photo rating system on portable devices.

1. Introduction

Photo quality assessment aims to classify the photographs into high or low quality automatically. Many achievements have been made before. Datta et al. [1] used a set of low-level features followed by a classifier to achieve photo quality assessment. Ke et al. [2] designed the semantic features based on human's perception to increase performance. These works are regarded as the earliest representatives in this topic.

Later, Luo et al. [3] developed subject region detection methods and regional features to improve assessment results. This work was refined by Luo et al. [4] by improving existing features via better subject detection algorithms. Dhar et al. [5] introduced a high-level attributes layer to make the subject-based framework more integrated. All those work used high-complexity and describable features to imitate the photography rules. The contribution is obvious but the computational overhead is also increased rapidly because salient subject-regions have to be segmented.

Other works [6,7] adopted bottom-up principle because many aesthetic factors cannot be simply defined by common photography rules. Despite these works set another benchmark in this topic, they also suffered from the problem of computation efficiency. Moreover, those bottom-up based rules are usually not describable so that their features cannot provide direct feedback to users.

(a) (b)

Figure 1. Five dominant colors generated by proposed color palette feature in flower scenes (a) High quality photo (b) Low quality photo.

In this paper, we move toward to a distinct direction. Our goal is to design a set of features that are describable, discriminative, and computationally efficient. With these advantages, the proposed method can be implemented and visualized in on-line aesthetic assessment systems such as live view screen on mobile camera or web-based photo rating system.

2. Methodology

The principle of our method primarily follows instance-based approach instead of conventional rule-specific approach because the former can preserve more complete information in training process. Moreover, we do not adopt any computation consuming techniques, such as subject detection or image segmentation in our scheme.

2.1. Proposed features

It is widely agreed that color and composition are the key factors to determine photo's quality, and highly related to human's perception. So our proposed features focus on color presentation and spatial composition, which includes color palette, layout/edge composition, and global texture features.

Color Palette (CP) A good combination of colors within an image is related to visual attractiveness. We

call such combination as *color palette*. To make our method efficient, we simply consider the color distribution of an image. The main issue is to find few dominant colors such that they occur frequently in the image and are dissimilar to each other.

In order to fulfill the goal, we divide each channel of the HSV color space into 16 bins, and so a total of 16^3=4096 bins are constructed. The center of each bin in the HSV space is called a candidate color, and so there are 4096 candidate colors. Our goal is to find, from the candidate colors, several key colors dominating the entire color distribution. First, we approximate the color distribution of the image by the histogram built on the candidate colors, $\mathbf{H}=\{h(i)|i=1...4096\}$, where $h(i)$ is the number of pixels associated with the i-th bin in the image. Denote $C_i \in \mathrm{R}^3$ to be the i-th candidate color, and we treat $h(i)$ as its weight. Let **D** be the dataset consisting of the weighted samples:

$$\mathbf{D}=\{(C_i, h(i)) \mid h(i)\geq 0,\ i=1...4096\}.$$

We then apply weighted k-means algorithm to **D** and obtain N cluster centers. Note that the clustering process is performed in only three-dimensional space and so it is very efficient to compute.

Despite the N colors associated to the cluster centers can be employed as the dominant colors, they could be suffered from the problem that these centers are not the colors appearing in the image since they are averages of candidate colors. In practice, we seek to find nearby candidate colors with high weights instead, which should be more representable. For each cluster j (j=1…N) we find the j-th dominant color by

$$Dom(j) = \underset{C_i \in \text{cluster } j}{\operatorname{argmax}} \ \alpha h(i) + (1-\alpha)||C_i - V_j||^{-1},$$

where V_j is the center of cluster j, and $\alpha \geq 0$ is a parameter balancing between the high-weight requirement and the closeness to the cluster center. The number of dominant colors is set as N=5 in our implementation. Fig.1 shows two examples of the dominant colors obtained by our method.

Once the dominant colors are obtained, an image is reduced to a 5×3(channels)=15-d vector. To conduct a feature for aesthetic-value assessment based on color information, we introduce an instance-based approach instead of using rule-based approaches such as color-harmony [9], since the former often performs better as more details are utilized. We employ a training dataset consisting of photos labeled as "high-quality" and "low-quality." Then, for the input image, we find its k nearest neighbors (kNN) among the training photos in the 15-d space. Let n_H and n_L be the numbers of high- and low-quality neighbors found by kNN, respectively, where $k=\mathrm{n}_H + \mathrm{n}_L$, we then construct the CP feature by their difference, $f_1 = \mathrm{n}_H - \mathrm{n}_L$. In our work, the training set typically contains hundreds of photos of each label. However, since kNN is only performed in 15-d space, it is still very efficient to compute.

Layout Composition (LC) We also utilize template-based principle instead of traditional rule-specific methods, such as rule of thirds and visual balance, to construct the LC features. They are obtained for the H, S, V, and H+S+V channels. We first average the high (low) quality training photos to build a high (low) quality template. Let the L1 distance between the input image and the high and low quality templates be d_H and d_L, respectively. The value $\mathrm{d}_L - \mathrm{d}_H$ for the four channels then serve as the LC features f_2 to f_5. They are proportional to the composition of high-quality photos.

Edge Composition (EC) These features are obtained in the same way of the LC features, but extracted from the edge-intensity images of the four channels. Features f_6 to f_9 are then obtained. Because edges in an image could reflect object boundaries, we assume that the spatial pattern of edges will benefit to the assessment of photos with salient objects.

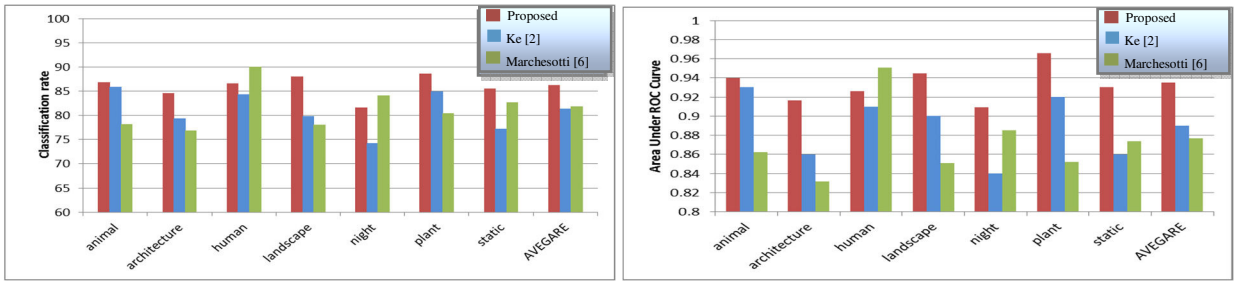
Global Texture (GT) We segment the image into 6 stripes uniformly in both of the vertical and horizontal directions, and compute the sum of differences of all the adjacent stripes for the four channels. Features f_{10} to f_{13} are thus generated. Similarly, features f_{14} to f_{17} are generated for the edge-intensity images.

2.2. General features

In addition to the features introduced above, we further use several common features in computational aesthetics: The feature **Blur** estimates the sharpness of a gray-level image by FFT (f_{18}). We also use **Dark channel** [4], but simply calculate sum of the minimum values of the RGB channels of pixels (instead of local patches) to reduce the computational complexity (f_{19}). **Contrasts** are regarded as another key factor of photo quality, and we follow [2] to compute them as the widths of 98% mass of both RGB and gray-level histograms (f_{20}, f_{21}). **HSV Counts** are the numbers of non-zero bins when quantizing each channel of the HSV into 16 bins (f_{22}, f_{23}, f_{24}). In general, high-quality photos have higher count values.

2.3. Validation

After feature extraction, a photo can be represented by a 24-d feature vector. We learn a binary classifier by using support vector machine (SVM) based on the dataset with high- and low-quality training photos that are also used for constructing the CP, LC, and EC features in Sec. 2.1. So for any testing image, it can be classified as "high" or "low" by the trained classifier.



(a) (b)

Figure 2. Classification performance in different categories with (a) ACC (b) AUC measures.

We choose the publicly available database provided by CUHK [4]. Each photo in this database has been assigned as "high-quality" or "low-quality" label. Total 7 categories of photos are included: Animal, Plant, Static, Human, Night, Architecture and Landscape. We define the former five categories as "subject photos" and the others as "scene photos". In our setting, for each category, we randomly select half of them as training samples and the rest as testing ones. The SVM classifiers of 7 categories are trained individually. The random partition repeats 10 times and averaged results are reported.

To evaluate classification performance, we use Classification Accuracy (ACC). Because the dataset contains different amount of high/low quality images, we further use Area Under the ROC Curve (AUC) since it is a better measure for unbalanced datasets.

3. Experimental results

We compare the methods of Ke et. al [2] and Marchesotti et. al [6]. Both methods have moderate computation times and good performance. The method in [2] is an early pioneer work not depending on complex techniques, and so it can be run very fast. The recent study [6] presents a bottom-up method by using bag-of-visual-words (BOV) features, which enhances the performance considerably. We implemented the 7 features of [2] and BOV feature of dense SIFT from [6] parameterized by 32x32 patch, 4 grid spacing, with 800 k-means centroids as thee visual words. It should be noted that BOV-SIFT feature is applied to gray level images only and so it does not use any color information.

Fig. 2(a-b) show the ACC and AUC measures of our method, the Ke et al. method [2], and the BOV-SIFT method [6]. It can be seen that our method consistently outperforms the Ke et al. method for both the ACC and AUC in all categories. We owe this to the reason that we have exploited approximate information in layout composition feature. The averaged ACC of our method reached 86% and AUC reached 0.93. It means the classification performance of our proposed method is well.

Compared to the BOV-SIFT method [6], our performance is still better for most categories, and only worse in the human category for both ACC and AUC. It can be seen that the BOV-SIFT method has better strength on human and night photos. The possible reason is that the BOV-SIFT feature has better ability to catch tiny details of image composition on photos with monotonous color. But it seems to lose dominance on assessing photos with plentiful colors. Therefore, for ideal photo aesthetic quality assessment, color presentation is still an indispensable factor which cannot be ignored.

Finally, we report the running time of processing a single testing photo. Each photo was rescaled such that either the width or height is no larger than 480 pixels. Our proposed method took averagely 0.26 seconds on a PC (Win7, Intel Core i5, 12GB RAM). The Ke et al. method [2] is also very efficient, which took averagely 0.16 seconds, and the BOV-SIFT method [6] took around 1.5 seconds. Ke et. al's and our methods have both acceptable computation times for nearly real-time applications. However, our classification rates are significantly better than Ke et al.'s as shown above. The BOV-SIFT method requires much more time because the SIFT image descriptor is more computationally demanding.

In addition to [2] and [6], we also investigate the running time for other high-complexity techniques [3,4]. We found that they took over 5 seconds to process one photo due to the fact that the subject detection is quite computationally expensive. There is no denying that the methods of focusing on foreground/background have their strength on assessing specific photos, but they also result in severe computational burden. By contrast, our method can achieve comparable performance with apparently lower computation time. For example, the AUC reported in [4] is 0.9044, while ours is 0.93 under similar settings of categories. Hence, our method has more potential to be carried out in instant manner on mobile or embedded devices for future applications.

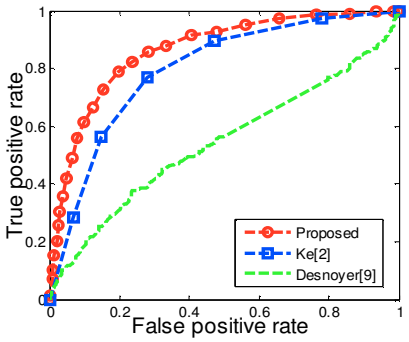


Figure 3. Comparison of ROC performance using three different kinds of color-based features.

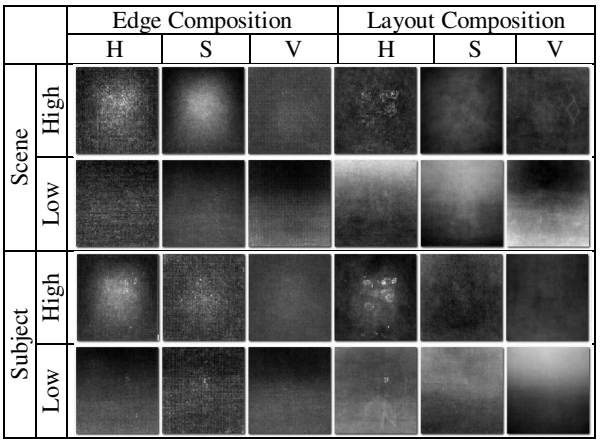


Figure 4. Edge/Layout templates of scene/subject photos.

4. Discussions

To evaluate the efficacy of the proposed color feature, Fig. 3 shows the ROC curves of our color palette feature, Ke's color distribution feature, and Desnoyer et. al's color harmony feature [9]. The feature in [9] is presented by 10 pair-wise harmony scores calculated by the 5 dominant colors generated by our proposed CP feature. Our feature outperforms them because that (1) directly using the information generated from training samples yields more reliability for classification, and (2) color harmony sometimes cannot reflect human's perception well because the pleasing colors of photos are often subjective and dependent to image contents.

Next, we look into the performance of two composition features: LC and EC. We apply them to subject and scene photos respectively, and then attempt to see the difference. We observed that EC feature has better classification accuracy in subject photos while the LC feature achieved better results for scene photos. Table.1 lists the corresponding AUC performance. This implies that assessing such "non-subject" photos seems to rely more on raw values instead of gradient information. The LC feature was lacking in the Ke et. al and Desnoyer et. al works. Fig.4 shows the layout/edge composition templates of both subject and scene photos. The difference of high/low quality templates in scene (non-subject) part is very distinctive.

Table 1. The AUC performance of layout and edge composition features operating on scene and subject photos respectively.

	Layout	Edge	Layout + Edge
Scene	**0.764**	0.609	0.758
Subject	0.644	**0.819**	0.837
Scene + Subject	0.673	0.758	0.814

5. Conclusions and Future Work

This paper demonstrates an efficient approach to assess photo quality. The proposed aesthetic features are not only efficient but also discriminative. The experimental results show that using simple techniques is sufficient to reach great classification performance on different variability of photos. In future works, we will analyze the advanced feature selection, photo categorization, and create an interface of instant photo quality rating system.

6. Acknowledgement

This work was supported in part by National Science Council, Taiwan, under the grants NSC 101-2631-H-001-007.

References

[1] R. Datta, D. Joshi, J. Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach," *ECCV* 2006.

[2] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," *CVPR* 2006.

[3] Y. Luo and X. Tang, "Photo and video quality evaluation:Focusing on the subject," *ECCV* 2008.

[4] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment," *ICCV* 2011.

[5] S. Dhar, V. Ordonez, T. Berg, "High level describable attributes for predicting aesthetics and interestingness," *CVPR* 2011.

[6] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," *ICCV* 2011.

[7] H. H. Su, T. W. Chen, C. C. Kao, W.H. Hsu, and S. Y. Chien, "Scenic photo quality assessment with bag of aesthetics-preserving features," *ACM Multimedia*, 2011.

[8] L. Yao, P. Suryannarayan. M. Qiao, et al. "OSCAR: On-site composition and aesthetics feedback through exemplars for photographers," *Int. J. Comp. Vis.*, 2011.

[9] M. Desnoyer and D. Wettergreen, "Aesthetics image classification for autonomous agents," *ICPR* 2010.