

Intelligent Photographing Interface with On-Device Aesthetic Quality Assessment

Kuo-Yen Lo¹, Keng-Hao Liu², and Chu-Song Chen^{1,2}

¹ Research Center for Information Technology Innovation, and

² Institute of Information Science, Academia Sinica, Taipei, Taiwan
{kylo,keng3,song}@iis.sinica.edu.tw

Abstract. This paper proposes a efficient method for instant photo aesthetics quality assessment that can be implemented on general portable devices. The classification performance is guaranteed to 0.89 on benchmark photo database. We also port our method onto a middle-level tablet computer to execute instantly and we find it reaches good acceptable efficiency. Moreover, an aesthetic information display to present the aesthetics evaluation results to users is introduced.

1 Introduction

Photo aesthetic quality assessment aims to classify the photographs into high or low quality automatically. Tong et al. [1] attempted to classify photographs into those taken by professionals or home users using low-level features derived from computer vision techniques. Datta et al. [2] also employed a set of low-level features then followed by a classifier to achieve photo quality assessment. Ke et al. [3] designed more semantic features based on the perceptual factors that present the difference between high and low quality photos to increase the performance. These works are the earliest representatives in this topic.

Later, Luo et al. [4] proposed regional features to improve assessment results by utilizing subject region detection methods. This work was refined by Luo et al. [5] by improving existing features with a variety of subject detection algorithms, such as super-pixel segmentation, layout and human detection. Dhar et al. [6] introduced a high-level attributes layer to make the subject-based framework more integrated. Those works tended to use more high-complexity and describable features to imitate the photography rules. The contribution is indeed significant but the computational overhead is increased rapidly.

Other works [7,8] adopted bottom-up strategies to acquire more improvement than conventional rule-specific method. They extracted the hidden composition relations of image by using bag-of-words since many aesthetic factors cannot be simply described by common photography rules. Despite those works set another benchmark in this topic, they also suffered from the issue of computation efficiency. Furthermore, those bottom-up features are usually not describable so that they cannot provide direct feedback to users. Recently, a few web-based applications using the techniques of photo aesthetics were proposed.

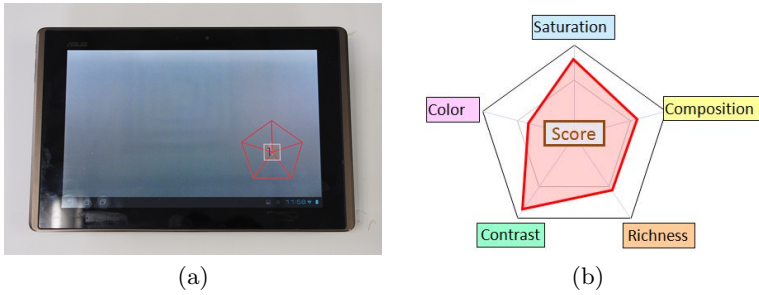


Fig. 1. The overview of our proposed mobile application. (a) Tablet device with our instant photo aesthetics system (b) The proposed five aesthetic perspectives of photography.

Datta et al. [9] created a web-based aesthetic quality inference system (ACQUINE) which automatically rates the quality of photos that uploaded from web users. Yao et al. [10] proposed an on-site composition and aesthetics feedback system that can provide aesthetic score and retrieved exemplars for mobile consumers. Both works are considered as off-line applications since they require supports of network transitivity, computational ability and storage of server.

As the rapid growth of mobile commodity shown in the market, more and more people use them to take life photos. Compared to existing off-line applications, creating an autonomous photo aesthetics quality assessment system on mobile devices is urgently demanded. With the advantage of programming flexibility and considerable computational ability of mobile devices, it is possible to design a near instant aesthetics evaluation system which can assist users to take photos. However, those well-crafted works [10][5] cannot be run on such devices because of the huge computational complexity that portable devices cannot afford. Besides, using a bunch of aesthetics features without appropriate manipulation can not give describable feedback to users. Therefore, the key to reach this goal is to adopt more efficient and describable techniques for implementation on mobile devices. For instance, Vazquez et al. [11] proposed an assisted photography method for people with visual impairments to improve picture composition by using efficient methods.

In this paper, we design an efficient way of running instant photo quality assessment on popular portable devices. We first design a set of aesthetics features that are describable, discriminative, and computationally efficient. Then those features are clustered hierarchically into five groups that are semantically independent to display describable aesthetic information to users. Finally, the proposed scheme is implemented and visualized in on-line (near real-time, or instant) aesthetic assessment system through the live view screen of mobile device as a strong aid for creation of photos.

This paper is organized as follows: Section 2 describes the feature extraction for our assessment work. Section 3 evaluates the feature performance and corresponding runtime on two different platforms. Section 4 introduces aesthetic information display method. Finally, the conclusion is drawn in Section 5.



Fig. 2. Five dominant colors generated by proposed CP feature in flower scenes (a) High quality photo (b) Low quality photo.

2 Aesthetic Feature Extraction

This section introduces several efficient aesthetic features for our assessment system. They follow the rules of photography. Our principle of designing features mainly follows instance-based (data driven) rule, where the features are created by analyzing the training database which contains large amount of high/low quality-labeled photos. The rule-specific approach is also adopted to make assessment work more effective. In addition, we do not utilize any computation consuming techniques, such as subject detection or image segmentation. Only low-cost features are considered in our system to mitigate the computational burden of mobile devices.

2.1 Color Combination

A good combination of colors within an image is directly related to visual appearance and attractiveness. We call such combination as Color Palette (CP). 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.

We divide each channel of the HSV color space into 16 bins to construct a $16^3 = 4096$ bins color histogram. The center of each bin in the HSV space is called a candidate color. Our first goal is to find several key colors dominating the entire color distribution from 4096 candidate colors. First, we approximate the color distribution of the image by the histogram built on the candidate colors, $\mathbf{H} = \{h(i) \mid i = 1, \dots, 4096\}$, where $h(i)$ is the number of pixels associated with the i -th bin in the image. Denote $C_i \in R^3$ to be the i -th candidate color, and we treat $h(i)$ as its weight. Let $\mathbf{D} = \{C_i, h(i) \mid h(i) > 0, i = 1, \dots, 4096\}$ be the dataset consisting of the weighted samples. We then apply weighted k-means algorithm to \mathbf{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 suffer 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 ($j=1,\dots,N$) we find the j -th dominant color by

$$\arg \max_{C_i \in \text{cluster } j} \alpha h(i) + (1 - \alpha) \|C_i - V_j\|^{-1}. \quad (1)$$

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 $N=5$ and α is set as $10/\max(h(i))$ 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 utilize an instance-based approach instead of using rule-based approaches such as color-harmony [12]. Our empirical study finds that the former often performs better as more details can be utilized. For any input image, we find its k -Nearest Neighbors (k NN) among the training photos in the 15-d space via Hungarian Algorithm [13]. Let n_H and n_L be the numbers of high and low-quality neighbors found by k NN with $k=25$, respectively, where $k=n_H + n_L$, we then construct the CP feature by their difference, $f_1 = (n_H - n_L)/2k + 0.5$. In our work, the training set typically contains thousands of photos for each label. Since k NN is only performed in 15-d space, it is still very efficient to compute.

2.2 Composition

A proposed efficient composition feature is called Edge Composition (EC). We follow instance-based learning principle to measure image composition measured from training images, instead of using traditional rule-specific methods (e.g., rule of thirds or visual balance). Such a design not only directly reflect the composition properties of image but also requires less complexity. This feature is operated on H, S, and V channels individually. We first calculate the average of edge-intensity maps of thousands of high (low) quality training photos to build a high (low) quality edge template. The edge-intensity is measured by laplacian filter. Since the edges in an image could reflect object boundaries, it assumes that the spatial pattern of edges will benefit to the assessment of photos with salient objects. Let the L1 distance between the input image and the high/low quality edge templates be d_H and d_L respectively. The value $d_H - d_L$ for the three channels then serve as the EC features f_2 to f_4 . Those values are proportional to the composition of high-quality photos.

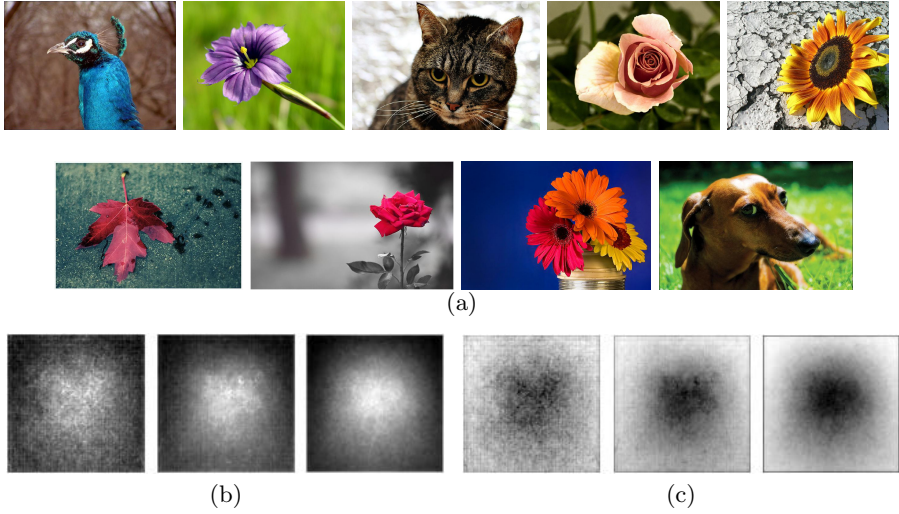


Fig. 3. Examples of photos containing subjects and normalized subject/background-prior maps obtained by EC feature (H, S, and V channels). (a) Photo examples. (b) Subject-prior maps. (c) Background-prior maps.

2.3 Contrast

Contrast is considered as important aesthetic factor in rules of photography. It measures the dynamic range of photos. There are two types of contrast feature we use: Histogram Contrast and Spatial Contrast.

Histogram Contrast. Histogram Contrast (HC) calculated the width of dominant range in color histograms of image. We follow [3] to compute them as the widths of 98% mass of both RGB-mixed and gray-level histograms (f_5 , f_6). In general, high-quality photos have higher contrast values in common.

Spatial Contrast. The general contrast feature merely measures the range of color histograms, where the information of spatial contents is ignored. We propose a new feature called Spatial Contrast (SC) which measures the related contrast between subject and background regions in the spatial domain.

To obtain SC, we first need prior information of subject region and background region. We directly employ the edge template obtained by EC feature as a subject-prior map (\mathbf{MAP}_{sub}) to save computation power. The background-prior map (\mathbf{MAP}_{bkg}) can be further calculated by the subtraction from subject-prior map. Both maps have fixed size and are normalized to sum-to-one. Fig. 3(b-c) show the examples of prior maps. For any query image \mathbf{I} with the same size of prior maps, we create three 16-bin histograms for H, S, and V channels to present color distribution of subject regions, denoted as \mathbf{H}_{sub}^H , \mathbf{H}_{sub}^S , \mathbf{H}_{sub}^V , and other three present color distribution of background regions, denoted as \mathbf{H}_{bkg}^H ,

$\mathbf{H}_{bkg}^S, \mathbf{H}_{bkg}^V$, by using prior maps. To create those histograms, the basic unit of count (vote) for each color bin is deployed by the weighting values on corresponding spatial locations of \mathbf{MAP}_{sub} and \mathbf{MAP}_{bkg} , instead of by one. More specifically, let $\mathbf{h}_A(k)$ presents the weight of k -th bin of histogram \mathbf{H}_A in a certain channel, where $A \in \{sub, bkg\}$ and $k=1,2,\dots,16$. Thus, the $\mathbf{h}_A(k)$ can be obtained by:

$$\mathbf{h}_A(k) = \sum_{i,j} \mathbf{MAP}_A(i, j), \quad (2)$$

where (i, j) denotes the spatial locations of the pixels contributing to k -th bin. Therefore, the histogram \mathbf{H}_A not only retains statistics of color distribution but also takes the spatial properties into account. Once six histograms are obtained, the SC feature is defined by $f_7 = \mathbf{dist.}(\mathbf{H}_{sub}^H - \mathbf{H}_{bkg}^H)$, $f_8 = \mathbf{dist.}(\mathbf{H}_{sub}^S - \mathbf{H}_{bkg}^S)$, and $f_9 = \mathbf{dist.}(\mathbf{H}_{sub}^V - \mathbf{H}_{bkg}^V)$ respectively, where $\mathbf{dist.}$ denotes L2 norm between any two 16-d histograms. The SC feature measures the relative difference between subject region and background region. In general, high quality photos usually possess higher SC.

2.4 Richness

People usually feel more pleasant to those images containing richness of spatial contents. A feature that can measure the variability of image content is further required. Two features are proposed to measure the richness in both spatial and color aspects.

Spatial Richness. Since the human eye used to scan image in horizontal or vertical way, the image with severe changes in both directions usually attracts more humans attention. We propose a simple feature, called Spatial Richness (SR), to measure such properties. We segment the image into 6 stripes uniformly in both vertical and horizontal directions, and compute the sum of differences of edge-intensity maps of all the adjacent stripes for H, S, and V channels. Features f_{10} to f_{12} are thus generated.

Color Richness. A good low-level statistic can also contribute to aesthetics prediction. We design a feature called Color Richness (CR) that assumes that high-quality photos always have more colors with higher hue counts than low-quality ones [3]. To obtain CR, we directly make use of the 16-bin HSV histograms obtained from CP feature. Then we calculate the number of bins with corresponding counts higher than a given threshold for each histogram. Features f_{13} to f_{15} are presented as the HSV counts for H, S, and V channels respectively.

2.5 Average Saturation

Average Saturation (AS) is considered as another indispensable statistic in computational aesthetics for high-quality photos [2]. It calculates the average value

of saturation channel in HSV color space for an image (f_{16}). Basically, photos with higher saturation usually produce more aesthetic feeling.

It should be noted that we do not employ the averages of H and V channels as features due to the reasons: The average of Hue is generally not related to aesthetic quality because it just corresponds to human's preference. The average of Value is also not a factor to aesthetics since the exposure status is automatically balanced by camera devices in most of cases. Therefore, only the average of Saturation can be useful statistic to express aesthetic emotion of image.

3 Feature Evaluations and Experiments

After developing the describable features for photo aesthetics assessment, it is necessary to evaluate the classification performance of those features and corresponding on-device computation time. In this section we introduce the photo database and validation methods used in this paper. The performance of proposed aesthetics features will be demonstrated in detail, and the corresponding efficiency report of on-device computation will be also presented.

3.1 Database and Setting

We choose the publicly available photo database provided by CUHK [5] for experiments. It consists 7 categories of photos and each photo in the database has been assigned as high-quality or low-quality label. We only use those photos with obvious subjects inside the scene, referred to as Animal, Plant, Static categories of photos, which totally contain 2078 high-quality photos and 7573 low-quality photos, for our implementation. Half of them are selected as training photos and the rest as testing ones.

Once feature extraction is done, a binary classifier can be learned by using Support Vector Machine (SVM) based on high/low-quality training photos to evaluate the classification performance for testing photos. In our setting, the random partition repeats 10 times and the averaged results are reported. The performance index we use is Area Under the ROC Curve (AUC) since it is a better measure for unbalanced datasets. In general, higher AUC presents higher classification ability.

3.2 Feature Performance

To balance between performance and computational efficiency, we assign a small scale 240x180 as working resolution in feature extraction for each image. Table. 1 tabulates the classification performance and computational consumption of our proposed features. First row "AUC" shows the AUC value of each individual aesthetics feature, referred to as Edge Composition (EC), Color Palette CP), Histogram Contrast (HC), Spatial Contrast (SC), Spatial Richness (SR), Color Richness (CR), and Average Saturation (AS) respectively. To evaluate the performance of single feature, we trained a individual SVM classifier for each feature

Table 1. The classification performance (AUC) of proposed aesthetic features

	Edge Comp.	Color Palette	Spatial Contrast	Hist. Contrast	Spatial Richness	Color Richness	Avg. Saturation	Overall (16-d)
AUC	0.80	0.77	0.67	0.67	0.79	0.70	0.64	0.89

separately to obtain its own AUC result. From Table. 1, it can be observed that those "well-designed" features, such as EC, CP, and SR, have better classification abilities while other low-statistical features, such as HC and AS, just give fair results. We owe this to the reason that those well-designed features adopted instance-based principle to extract statistical information from training samples so that the values they generated are much accordance with the characteristics of datasets. Nevertheless, other simple features designed by traditional rule-specific approach have fair performance but could help to promote overall classification rate. The final column in row AUC shows the classification result using all 7 features (16-d) simultaneously. It reached considerable classification accuracy 0.89. Furthermore, it is noticed that the CP feature performed the best among all features we used. It suggests that the color combination is indeed one of the crucial factor to determine photo's quality. However, such a superior feature usually requires relatively higher computational resource to implement. It is obviously a trade-off between effectiveness and efficiency.

3.3 On-Device Computation Time

Table. 2 tabulates the required computation time for each feature implemented on PC (Win7 64-bit, Intel Core i5 CPU at 3.4MHz, 16GB RAM, Matlab) and middle-level mobile device (i.e. Tablet PC) (ASUS Transformer TF101, Android 3.2, NV Tegra2 CPU at 1.0MHz, 1GB RAM) respectively. There are two observations: 1. Those features with higher computational complexity require much computation time. For example, the CP feature contains complex techniques such as histogram construction, weighted k-means, and k NN algorithms, and Hungarian matching so that it requires much more time to process than other features. 2. PC runs faster than mobile device does for all features. The PC requires averaged 82ms to run whole 16-d features per image while the mobile needs 288ms. This result matches our expectation because PC has more advantages on computation power in terms of higher CPU clock and larger capacity of memories. Even though the mobile device has such inherent disadvantages, it can run whole features instantly in near 2.3fps including all system loading factors.

Besides, we found both systems own different properties on computing the same features. For instance, CP feature needs close processing time on both platforms, but EC feature requires greatly different processing time. This implies that to make the aesthetics evaluation system working more efficiently, appropriate optimization on programming is further required.

Table 2. The average computation time (millisecond) of each proposed aesthetics feature on different platforms.

	Edge Comp.	Color Palette	Spatial Contrast	Hist. Contrast	Spatial Richness	Color Richness	Avg. Saturation	Overall (16-d)
On PC	11	40	12	3	2	14	0.1	82
On Mobile	160	54	22	10	25	16	1	288

4 Aesthetic Information Display

For building a good aesthetics system, it is necessary to create an interface to display aesthetic information once the aesthetic prediction is calculated. Opposite to most previous works where the assessment results are simply obtained by feature extraction followed by a trained classifier, additional describable, referable and useful-feedback information that can suggest users to re-frame current scene is indispensable for an instant photo aesthetic assessment system.

We simply utilize hierarchical structure to insert an additional layer that constitutes five dominant aesthetics indices for such a purpose. Finally, we design a simple interface to display such information on portable devices. The details are going to be described as follows.

4.1 Hierarchical Structure for Aesthetics Evaluation

In order to reveal the "hidden aesthetic issues" when evaluating an image, the diagram of our aesthetics assessment system is designed as a three-layer hierarchical structure. Fig. 4 shows the diagram of hierarchical structure of proposed system. From Layer 1 (named as Feature layer) to Later 2 (named as Aesthetic index layer), we first group our 16-d aesthetic feature set to five dominant groups according to their attributes. They are Composition, Saturation, Color, Contrast, Richness

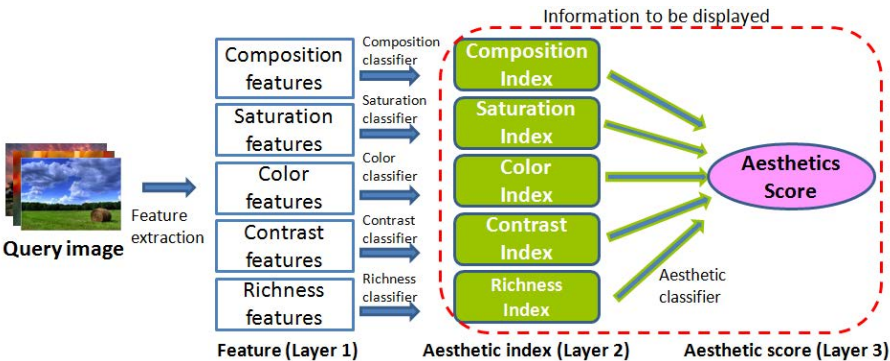


Fig. 4. The hierarchical structure of proposed aesthetics system

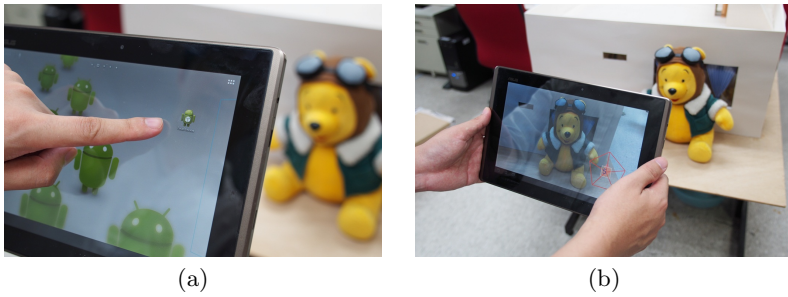


Fig. 5. Demonstration of proposed instant photo aesthetic quality assessment system. (a) Mobile program (c) Real-life scenario.

Contrast, and Richness respectively, which are considered as five common aesthetic indices in photography. The Composition index presents the degree of good composition by using EC feature (f_2 to f_4). The Saturation index measures saturation degree of image by simply using Avg. saturation feature (f_{16}). The Color index presents the degree of color combination of image using CP feature (f_1). The Contrast index evaluates both intensity contrast (histogram-based) and spatial contrast by considering HC and SC features simultaneously (f_5, f_6, f_7 to f_9). Finally, The Richness index detects the richness of image content in both spatial and color aspects by taking SR and CR features (f_{10} to f_{12} , and f_{13} to f_{15}).

For each group, we then train an independent SVM classifier (via probability model) with training database to measure the degree of each index then it is normalized to $[0,1]$ according to corresponding index range. (The index range are the lowest and largest values of the distribution of a certain aesthetics index in the database. Before acquiring this range, we removed top and bottom 3% samples to avoid the influence of extremely good or bad samples.) These main purpose of aesthetics indices is to inform users the current sub aesthetic degree of five general perspective in photography. Users can improve overall aesthetics quality of current photo scene by means of adjusting those ones which are lower than a certain threshold.

Since each image has been represented as a 5-d feature vector in Layer 2, from Layer 2 to Layer 3 (named as Aesthetic score layer), another SVM classifier can be trained to evaluate the overall aesthetics quality (aesthetics score) for a query image. The aesthetics score indicates the quality of image for a simple reference. This score is generally high if those indices in Layer 2 are high enough. It is worthy to notice that the evaluation performance of photo quality will not be significant improved if Layer 2 is removed from our system. The classification rate of using hierarchical maintains 0.88 according to our extended experiment. It implies that the components in Layer 2, referred to as five aesthetics indices, could preserve integrated statistical information for photo aesthetics assessment.

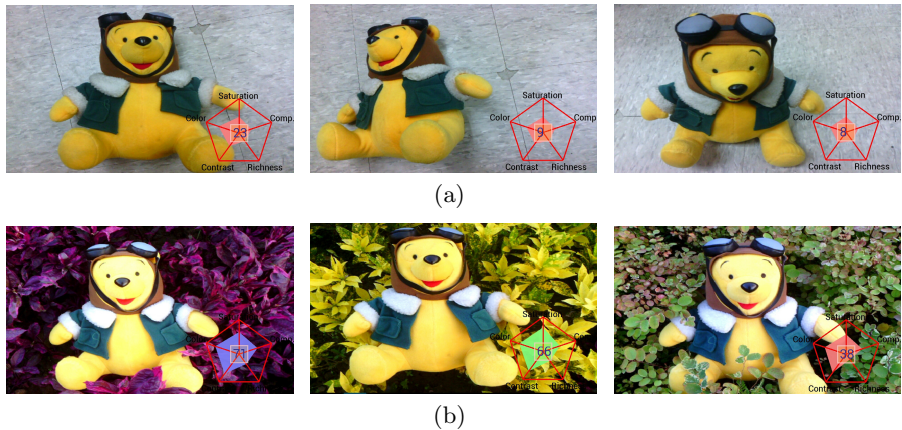


Fig. 6. Assessment examples of proposed instant photo aesthetic quality assessment system. (a) Scenes with different poses (b) Scenes with different background.

4.2 Interface Design on Display

An ideal photo quality assessment system not only needs precise evaluation ability, but also requires a good presentation to users. To display the aesthetics information on the screen of mobile device, we design a pentagon graph to display the information of aesthetics indices and overall score introduced in Section 4.1. The illustration of pentagon graph is shown in Fig. 1(a). The outside pentagon defines five aesthetics indices while the inner pentagon indicates corresponding index values. Figs. 5(a-b) demonstrate the display interface of our aesthetics system where the pentagon graph is set in the lower right corner on the screen. Fig. 6 further demonstrates the assessment results of proposed system implementing on the scenes with different scenarios. By means of such layout design for aesthetics presentation, user can easily understand the current status of photo aesthetics when taking photos, and would have intention to improve those indices with low values to obtain higher aesthetics score.

5 Conclusions and Future Work

This paper introduces an on-device photographing interface with aesthetic quality feedback on mobile devices. We first experimentally show that our proposed aesthetic features could reach great classification performance. Later, the implementation on a modern tablet computer with considerable response speed is achieved. The experiments demonstrate that the computational ability of middle-level mobile device is sufficient to implement image processing techniques in instant manner. In future works, we will investigate more discriminative features with better evaluation accuracy, explore the ways to optimize the system overall efficiency, and create an interactive system coupled with more useful feedback for advanced photographing interface.

Acknowledgement. This work was supported in part by National Science Council, Taiwan, under the grants NSC 101-2221-E-001-015-MY2.

References

1. Tong, H., Li, M., Zhang, H.-J., He, J., Zhang, C.: Classification of Digital Photos Taken by Photographers or Home Users. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3331, pp. 198–205. Springer, Heidelberg (2004)
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying Aesthetics in Photographic Images Using a Computational Approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
3. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: CVPR (2006)
4. Luo, Y., Tang, X.: Photo and Video Quality Evaluation: Focusing on the Subject. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 386–399. Springer, Heidelberg (2008)
5. Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: ICCV (2011)
6. Dhar, S., Ordonez, V., Berg, T.: High level describable attributes for predicting aesthetics and interestingness. In: CVPR (2011)
7. Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G.: Assessing the aesthetic quality of photographs using generic image descriptors. In: ICCV (2011)
8. Su, H.H., Chen, T.W., Kao, C.C., Hsu, W.H., Chien, S.Y.: Scenic photo quality assessment with bag of aesthetics-preserving features. In: ACM Multimedia (2011)
9. Datta, R., Wang, J.Z.: Acquine: Aesthetic quality inference engine real-time automatic rating of photo aesthetics. In: ACM MIR (2010)
10. Yao, L., Suryannarayan, P., Qiao, M., Wang, J.Z., Li, J.: Oscar: On-site composition and aesthetics feedback through exemplars for photographers. IJCV (2011)
11. Vazquez, M., Steinfeld, A.: An assisted photography method for street scenes. In: WACV (2011)
12. Desnoyer, M., Wettergreen, D.: Aesthetics image classification for autonomous agents. In: ICPR (2010)
13. Burkard, R., Dell’Amico, M., Martello, S.: Assignment Problems. SIAM (2009)