Content-Based Photo Quality Assessment

Xiaoou Tang, *Fellow, IEEE,* Wei Luo, and Xiaogang Wang, *Member, IEEE*

***Abstract*—Automatically assessing photo quality from the perspective of visual aesthetics is of great interest in high-level vision research and has drawn much attention in recent years. In this paper, we propose content-based photo quality assessment using both regional and global features. Under this framework, subject areas, which draw the most attentions of human eyes, are first extracted. Then regional features extracted from both subject areas and background regions are combined with global features to assess photo quality. Since professional photographers adopt different photographic techniques and have different aesthetic criteria in mind when taking different types of photos (e.g. landscape versus portrait), we propose to segment subject areas and extract visual features in different ways according to the variety of photo content. We divide the photos into seven categories based on their visual content and develop a set of new subject area extraction methods and new visual features specially designed for different categories. The effectiveness of this framework is supported by extensive experimental comparisons of existing photo quality assessment approaches as well as our new features on different categories of photos. In addition, we propose an approach of online training an adaptive classifier to combine the proposed features according to the visual content of a test photo without knowing its category. Another contribution of this work is to construct a large and diversified benchmark dataset for the research of photo quality assessment. It includes $17,673$ photos with manually labeled ground truth. This new benchmark dataset can be down loaded at http://mmlab.ie.cuhk.edu.hk/CUHKPQ/Dataset.htm.**

***Index Terms*—Photo Quality Assessment, Content-based, Hue Composition, Scene Composition, Dark Channel, Clarity Contrast, Composition Geometry**

I. Introduction

AUTOMATIC assessment of photo quality based on aesthetic perception gains increasing interest in the computer vision community in recent years. With the rapid development of the Internet and the popularization of digital cameras, the number of photos accessible to end users is growing quickly. This demands fast and effective computing algorithms which can automatically conduct aesthetic assessment of large scale photo datasets. Such techniques enable high-quality photos to be harvested from massive volume of online sources and have many potential applications. For example, they help professionals, such as newspaper editors, to select high-quality photos from a large collection. They can help home users automatically select and manage appealing photos from a large amount of photos taken with their digital cameras. With automatic photo quality assessment, web image search engines can provide users search results which are both relevant to the queries and aesthetically pleasing. Finally, studying whether a computer can perform what was perceived as a human-only task is an interesting problem itself for brain cognitive study.

A. Photo Quality Assessment by Professionals

Defining aesthetics in photography is no easy task, since quality assessment is subjective. Normal people may simply regard the quality of a photo as the degree to which it appeals to human eyes, or whether it is attractive. Professional photographers, however, take into consideration various criteria such as sharpness, composition, lighting balance, topics of photos, and even the usage of special photographic skills. On the other hand, it is widely agreed that there are many rules of thumbs regarding what *generally* makes a photo appealing. For example, people would prefer a sharp and clear photo to a photo that is blurred; balanced lighting with proper contrast is considered better than dim lighting that makes most details unclear; a photo with a clear topic is seen as more pleasing than one with unnecessarily distracting background. These generally accepted rules make it possible to develop computing algorithms to automatically select photos that are more likely to be aesthetically pleasing. These rules can be categorized into *composition*, *lighting*, *color arrangement*, *camera settings* and *topic emphasis* as described below. Some examples of these rules are shown in Figure 1.

Composition. Photographic composition refers to the arrangement of visual elements in a photo. Photographers compose lines, shapes, patterns, texture, balance, symmetry, depth, perspective, and scale in a single artwork to communicate their message to viewers. Many rules of thumbs have been developed on how to make a good composition. First, it is vital that one keeps a photo *simple*. Instead of squeezing many subjects into a single frame, aiming for simplicity is often a good strategy to make a decent composition and to emphasize the topic. Meanwhile, the arrangement of lines helps to create a memorable image: diagonal lines, leading lines, and curved lines should be carefully placed to keep good visual balance; vertical and horizontal lines often frame the scene or serve as visual boundaries. Finally, the geometrical locations of objects are important when compositing a photo. Guidelines, such as the rule of thirds, are often used to place important elements to make a photo interesting and pleasing. According to the rule

X. Tang is with the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong and Key Lab of Computer Vision and Pattern Recognition, the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. Email: xtang@ie.cuhk.edu.hk; phone: +852-39438379; fax: +852-26035032.

W. Luo is with the Department of Information Engineering, the Chinese University of Hong Kong. Email: awesomekeane@gmail.com; phone: +852-39438206; fax: +852-26035032.

X. Wang is with the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong. Email: xgwang@ee.cuhk.edu.hk; phone: +852-39438382; fax: +852-26035558.

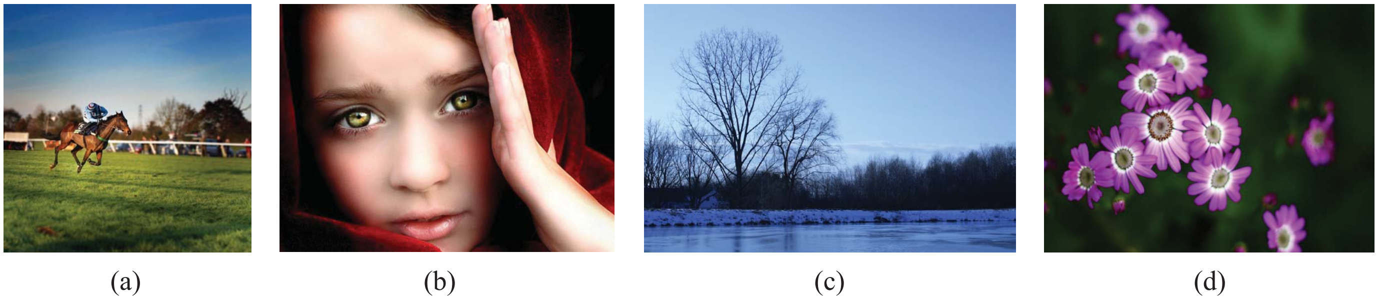


Fig. 1. Examples of applying the rules of photography. (a) Example of a wildlife photo with good composition. Semantic lines are properly arranged and the placement of visual elements satisfies the rule of thirds. The subject also has a good color contrast with the background. (b) Professional studio lighting on a portrait. (c) Cold color scheme with a calming effect. (d) Shallow depth of field photography. See details in the text of Section I-A.

of thirds, if a photo is divided into nine parts by two equally-spaced horizontal lines and two equally-placed vertical lines, the most important compositional elements should be placed along these lines or their intersections. An example is shown in Figure 1(a).

Lighting. A badly lit scene ruins a photo as much as poor composition. Photographers try hard to get a perfect lighting in both natural and studio settings to express certain moods and to create artistic effects. Weather is one of the most important factors in wildlife photography to provide preferred natural lightings. In studio settings, various lighting sources are carefully adjusted to create stunning effects in a portrait shot or a static close-up (Figure 1(b)). Lighting not only makes details in a photo more vivid, but also enhances the 3D impression of objects using highlights and shadows. The lighting contrast between the subject and the background helps to emphasize the area of interest in a photo.

Color Arrangement. Much of what viewers perceive and feel about a photo is through colors. Although their color perception depends on the context and is culture-related, recent color science studies show that the influence on human emotions or feelings from a certain color or a certain color combination is usually stable under different culture backgrounds [1], [2]. Professional photographers use various methods to control the color palette in a photo, and use specific color combination to raise specific emotions of viewers. For example, a color scheme dominated by warm colors create excitement, joy, and dynamism. Cold colors, on the other hand, tend to have a calming effect (Figure 1 (c)). Photographers enforce contrast using the combination of complementary colors and make a photo more smooth using analogous colors [3]. Color arrangement can also make a photo look “surreal”.

Camera Settings. Most non-professional photographers use point-and-shoot cameras in the “auto” mode, while professional photographers carefully adjust aperture, shutter speed, and ISO of cameras to create satisfying pictures. In many cases, special equipments, such as a fish-eye lens, a macro lens, or a LOMO camera, are used to achieve special effects. Such efforts lead to different degrees of sharpness, lighting, smoothness, and motion blur for different elements in a photo: a changing degree of sharpness in a photo leads to a feeling of depth and puts strong emphasis on clear objects (Figure 1 (d)); motion blur is considered as a powerful effect to stress the dynamics of a moving object; noise and grain in a photo are usually unwanted, while sometimes they describe certain textures exceptionally well. Note that in many cases, camera settings can be read from meta-data that comes with photos. People can also infer such settings from images alone in case such data is not available.

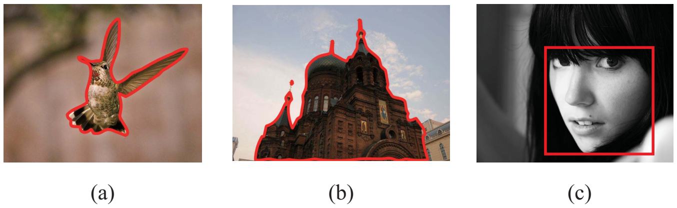


Fig. 2. Subject areas of three different types of photos. They can be extracted in different ways. (a) Close-up for a bird. (b) Architecture. (c) Human Portrait.

Topic Emphasis. High-quality photos generally satisfy three principles: *a clear topic, gathering most attention on the subject, and removing objects that distract attention from the subject*. Impressive photos usually treat the foreground subject and the background differently (Figure 2) to highlight the topic of a photo [4]–[6]. Professionals use various ways to isolate the subject from the background, such as *background out of focus* (Figure 1 (d)), *color contrast* (Figure 1 (a)), or *lighting contrast* (Figure 1 (b)).

B. Automatic Quality Assessment

Various methods of high-level photo quality assessment were proposed in recent years [7]–[16]. In early works [7], [8], only *global* visual features, such as global edge distributions, color histograms, and exposure, were used. However, our previous study [9] showed that regional features could be more effective, since in many cases human beings perceive subject areas differently from the background (see examples in Figure 2). After extracting the subject areas, which draw the most attentions of human eyes, regional features are extracted from the subject areas and the background separately and are used for assessing photo quality. In this paper, both global

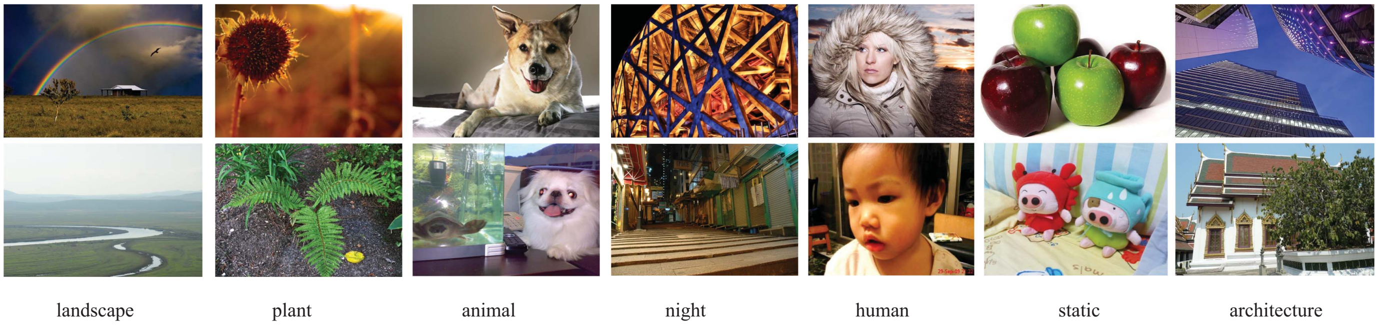


Fig. 3. Examples of photos belonging to seven categories according to visual content. First row: high quality photos; Second row: low quality photos.

and regional features are used and compared. Experimental evaluation shows that they complement each other.

Existing methods treat all photos equally without considering the diversity of their content. It is known that professional photographers adopt different photographic techniques and have different aesthetic criteria in mind when taking different types of photos [17], [18]. For example, for close-up photographs (e.g. Figure 2 (a)), viewers appreciate the high contrast between the foreground and the background regions. In human portrait photography (e.g. Figure 2 (c)), professional photographers use special lighting settings [19] to create aesthetically pleasing patterns on human faces. For landscape photos, well balanced spatial structure, professional hue composition, and proper lighting are considered as traits of professional photography. Also, subject areas of different types of photos should be extracted in different ways. In a close-up photo, the subject area is emphasized using the shallow depth of field technique, which leads to blurred background and clear foreground. However, in human portrait photos, the background does not have to be blurred since the attentions of viewers are automatically attracted by the presence of human faces. Their subject areas can be better detected by a face detector. In landscape photos, it is usually the case that the entire scene is clear and tidy. Their subject areas, such as mountains, houses, and plants, are often vertical standing objects. This can be used as a cue to extract subject areas.

C. Overview of Our Approach

Building upon these considerations and our previous works, we propose a content-based photo quality assessment framework [7], [9], [16]. Using high-level computer vision techniques such as classification, segmentation, detection, and feature extraction, the framework is composed of three levels. First, we classify the photo collection into several categories based on visual content. Secondly, in each category, based on the content of the category, a new subject area detection and segmentation method is developed to separate the subject from the background. Thirdly, high-level computer vision features based on modeling of human perception are computed to extract features from both the subject region and the background region. Both regional and global features are used to assess photo quality based on photo content.

Photos are manually divided into seven categories based on their visual content: “animal”, “plant”, “static”, “architecture”, “landscape”, “human”, and “night”. See examples in Figure 3. Regional and global features are selected and combined in different ways when assessing photos in different categories. More specifically, we propose three methods of extracting subject areas.

- ***Clarity-based subject area detection*** combines blur kernel estimation with image segmentation to accurately extract the clear region as the subject area. It is suitable to extract subject areas of photos belonging to the categories of “animal”, “plan”, and “static”, where the foreground objects are often highlighted using the technique of shallow depth of field.
- ***Layout-based subject area detection*** analyzes the layout structure of a photo and extracts vertically standing objects as the subjects. It is applicable to the categories of “architecture” and “landscape”, where both background and subjects have high clarity.
- ***Human based detection*** locates faces in a photo with a face detector or a human detector. Apparently, it is for the category of “human”.

Based on the extracted subject areas, multiple types of regional features are proposed.

- ***Dark channel feature*** measures the sharpness and the colorfulness of the subject area.
- ***Clarity contrast feature*** captures the clarity contrast between the subject area and the background.
- ***Lighting contrast feature*** quantifies the lighting contrast between the subject area and the background.
- ***Composition geometry feature*** evaluates geometry composition of photos by considering the location of the subject area.
- ***Complexity features*** measure the spatial complexity of the subject area and the background.
- ***Human based features*** capture the clarity, brightness, and lighting effects on human faces.

In addition, two types of global features are proposed.

- ***Hue composition feature*** fits photos with color composition schemes to evaluate their color arrangement.
- ***Scene composition features*** capture spatial structures of photos by detecting semantic lines, such as horizons and surfaces of water.

The design of these new methods of extracting subject areas as well as the new regional and global features well considers the criteria for photo quality assessment introduced in Section I-A. Their details are described in Section III-V.

Through extensive experiments on a large and diverse benchmark dataset, which includes $17,673$ photos with manually labeled ground truth, different subject area extraction methods and different features are evaluated and compared on different photo categories. Experimental results show that their effectiveness highly depends on the visual content. To the best of our knowledge, it is the first systematic study of photo quality features on different photo categories. Then these features are combined with a SVM classifier trained on each of the categories separately. Experimental comparisons show that the new features significantly outperform existing features. This large scale dataset is released to the public[1].

The knowledge of photo categories is assumed in these experiments. There are various ways of automatically classifying photos into different categories based on their visual content, such as scene classification [20], object recognition [21], and image categorization [22], [23]. Besides visual cues, additional information is available in text format, such as tags and surrounding texts, and could be extremely helpful for predicting photo categories. Some websites already categorize their photos, but not in all the cases. Therefore, the difficulty level of photo categorization depends on application scenarios. However, if only visual information is available, the problem of photo categorization is still challenging. Instead, we propose an approach of automatically online learning an adaptive classifier only using training samples which are similar to the photo to be assessed in visual content. Experimental results show that this automatic approach gives comparable performance to the case of knowing the information of photo category. It is also efficient enough for real-time application, since the number of features in use is small enough for fast online training.

II. Related Work

Automatic photo quality assessment is challenging because the two classes, high- and low-quality photos, are subjectively defined with high variations on ratings and it is not obvious what kind of features are suitable to differentiate the two classes. Most of the works on image quality assessment are not based on aesthetic perception. Some of them [24]–[27] require the original undistorted image to assess the quality of a degraded images. A low-quality image is typically degraded by compression or certain noise models. Subsequently, there have been works on directly estimating the quality of a single image without the undistorted one [28]–[30]. Unlike our work, they focused on the psychovisual study of photo quality or quality degradation caused by JPEG compression artifacts. Tong *et al.* [31] used boosting to combine 846 global low-level features for the classification of professional and amateurish photos. These features, such as color histograms, wavelets, and DCT moments, were widely used in image retrieval applications and were not specially designed for photo quality assessment. Such black box approaches of using low-level features give little insight on the reasons why particular features are chosen and how to design better features for classification.

As the first attempt to utilize high-level human aesthetic perception, we [7] designed a set of high-level semantic features based on perceptual criteria that people used for rating photos, and measured the global distributions of edges, color distributions, hue counts, blurriness, contrast, and brightness. The method outperforms low-level features with a much smaller number of features. Datta *et al.* [8] designed 56 features based on analysis of aesthetic perception of photos, including low-level features such as saturation, and high-level features such as the shallow depth-of-field indicator. They selected 15 best performing ones from them through SVM training to assess photo qualities. Nishiyama *et al.* [32] assessed the aesthetic quality by evaluating the color harmony of photos and proposed “bags-of-color-patterns” to characterize color variations in local regions. The performance of this feature was improved by combining with blur, edges, and saliency features.

Some approaches employed regional features. Datta *et al.* [8] divided a photo into 3×3 blocks and assumed the central block to be the subject area. They subsequently extracted regional features, such as lighting, saturation, and wavelet features based on this subject region. However, such assumption is not valid in many high quality photos. In fact, it is generally known that professionals deliberately avoid putting subjects right at the center of image. Wong *et al.* [12] and Nishiyama *et al.* [33] used the saliency map to extract the subject areas, which were assumed to have higher brightness and contrast than other regions. However, if a certain part of the subject area has very high brightness and contrast, other parts will be ignored by this method. See an example in Figure 6. Lo and Chen [34] proposed an approach to assess photo quality based on spatial relations of image patches.

Dhar *et al.* [35] used a set of high-level describable attributes, such as “presence of a salient object”, “Rule of Third” and “clear skies”, to predict the perceived aesthetic quality of photos. These attributes characterize the layout, content, and illumination of photos. The presence of an attribute is predicted by a binary classifier based on low-level features. All the attributes need to be labeled on each training photo and the problem of predicting various attributes of test photo is also challenging. Teh and Cheng [36] proposed relative features. In order to evaluate the quality of a photo, information of multiple photos from the same physical scene is utilized.

All the works discussed above universally apply global and regional features to photos without considering the variety of their content. Our experimental results show that they all badly perform on certain types of photos.

III. Global Features

Professionals follow certain rules of color composition and scene composition to produce aesthetically pleasing photographs. For example, photographers focus on artistic color combination and properly put color accents to create unique composition solutions and to invoke certain feelings among the viewers of their artworks. They also try to arrange objects

[1] http://mmlab.ie.cuhk.edu.hk/CUHKPQ/Dataset.htm

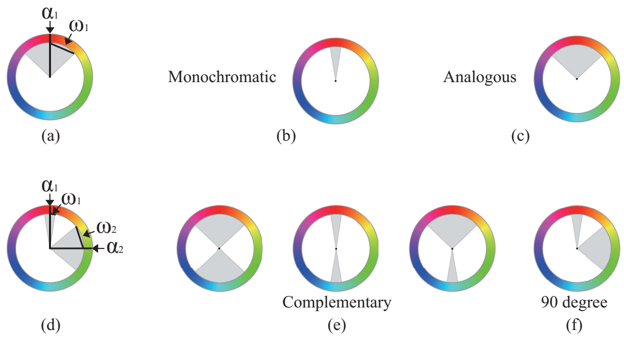


Fig. 4. Harmonic templates on the hue wheel used in [38]. An image is considered as harmonic if most of its hue fall within the gray sectors on the template. The shapes of templates are fixed. Templates may be rotated by an arbitrary angle. The templates correspond to different color schemes.

in the scene according to such empirical guidelines like "rule of thirds". Based on these techniques of photography composition, we propose two global features to measure the quality of hue composition and scene composition.

A. Hue Composition Feature

Proper arrangement of colors engages viewers and creates inner sense of order and balance. Major color templates [3], [37] can be classified as *subordination* and *coordination*. Subordination requires photographers to set a dominant color spot and to arrange the rest of colors to correlate with it in harmony or contrast. It includes certain color schemes, such as the 90^o color scheme and the Complementary color scheme, which leads to aesthetically pleasing images. With *coordination*, the color composition is created with the help of different gradation of one single color. It includes the Monochromatic color scheme and the Analogous color scheme. See examples in Figure 4.

Color templates can be mathematically approximated on the color wheel as shown in Figure 4. A coordination color scheme can be approximated by a single sector with the center (α_1) and the width (w_1) (Figure 4 (a)). A subordination color scheme can be approximated by two sectors with centers (α_1, α_2) and widths (w_1, w_2) (Figure 4 (d)). Although it is possible to assess photo quality by fitting the color distribution of a photo to some manually defined color templates, our experimental results show that such an approach is suboptimal. It cannot automatically adapt to different types of photos either. We choose to learn the models of hue composition from training data. The models of hue composition for high- and low-quality photos will be learned separately. The learning steps are described below.

Given an image I, we first decide whether it should be fitted by a color template with a single sector (T_1) or two sectors (T_2) by computing the following metric,

$$E_k(I) = \min_{T_k} \sum_{i \in I} D(H(i), T_k) \cdot S(i) + \lambda A(T_k)$$

where $k = 1, 2$. i is a pixel on I. $H(i)$ and $S(i)$ are the hue and saturation of pixel i. $D(H(i), T_k)$ is zero if $H(i)$ falls in the sector of the template; otherwise it is calculated as the arc-length distance of $H(i)$ to the closest sector border. $A(T_k)$ is the width of the sectors ($A(T_1) = w_1$ and $A(T_2) = w_1 + w_2$). λ is empirically set as 0.03. $E_k(I)$ is calculated by fitting the template T_k, which has adjustable parameters, to image I. T_1 is controlled by parameters (α_1, w_1) and T_2 is controlled by parameters $(\alpha_1, w_1, \alpha_2, w_2)$. This metric is inspired by the color harmony function [38]. However, we assume that the width of the sector is changeable and add a penalty on it. The single sector is chosen if $E_1(I) < E_2(I)$ and vice versa.

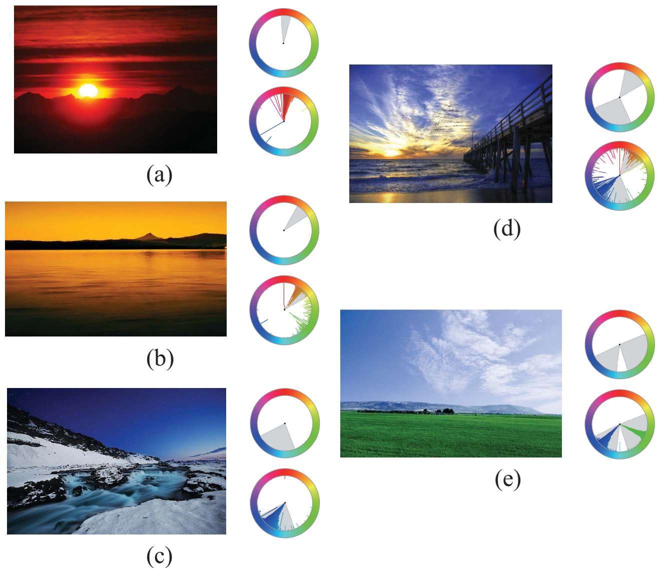
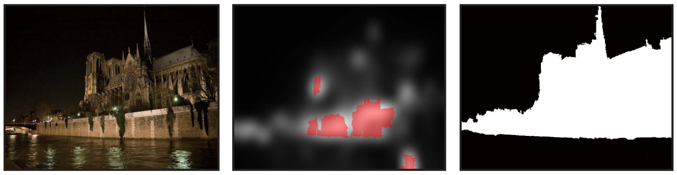


Fig. 5. (a),(b),(c): Mixture components for images best fitted with single sector templates. Color wheels on the top right side show the mixture components. The center and width of each gray sector are set as the mean and the standard deviation of each mixture component. Color wheels on the down right side show the hue histograms of images. (d),(e): Mixture components for images best fitted with double sector templates.

If I is fitted with a single-sector template, the average saturation s_1 of pixels inside this sector is computed. s_1 and α_1, the hue center of the fitting sector, are used as the hue composition features of this photo. If I is fitted with a two-sector template, a four dimensional feature vector $(\alpha_1, w_1, \alpha_2, w_2)$, which includes average saturations and hue centers, are extracted from the two sectors. Based on the extracted hue composition features, two Gaussian mixture models are separately trained for the two types of templates.

Examples of training results of high-quality photos in the category "landscape" are shown in Figure 5. Among 410 training photos, 83 are fitted with single-sector templates and 327 are fitted with two-sector templates. Three Gaussian mixture components are used to model hue composition features of photos belonging to single-sector templates [2]. Two Gaussian mixtures components are used to model the hue composition features of photos belonging to two-sector templates. One photo best fitting each of the mixture components is shown in Figure 5. We find some interesting correlations between the learned components and the color schemes. For examples,

[2] If we choose more than three components, the results degenerate to three components.



(a) (b) (c)

Fig. 6. (a) Input image (b) Saliency map with the subject area (red regions) extracted by the method in [12]. Because of the very high brightness in the red regions, other subject area is ignored. (c) Subject area (white regions) extracted by our clarity-based region detection method described in Section IV-A.

the components in Figure 5(a) and (b) correlates more with the monochromatic schemes centered at red and yellow. The components in Figure 5(c) and (e) more correlate with the analogous color scheme and the complementary color scheme.

The likelihood ratio $P(I|high)/P(I|low)$ of a photo being high-quality or low-quality can be computed from the Gaussian mixture models and is used for classification.

B. Scene Composition Feature

High-quality photos show well-arranged spatial composition to hold the attention of a viewer. Long continuous lines often bear semantic meanings, such as horizons and surfaces of water, in those photos. They can be used to compute scene composition features. For example, the location of the horizon in an outdoor photo was used by Bhattacharya *et al.* [14] to assess visual balance. We characterize scene composition by analyzing the locations and orientations of semantic lines. Our scene composition features include the average orientations of horizontal lines and vertical lines, the average vertical position of horizontal lines, and the average horizontal position of vertical lines.

We take the Hough transform of a given photo and extract top prominent lines present in the scene. Those lines are then classified into horizontal lines and vertical lines based on orientations. We define the orientation features as:

$$f_1 = \frac{1}{\|H\|} \sum_{l_k \in H} \theta_k, \; f_2 = \frac{1}{\|V\|} \sum_{l_k \in V} \theta_k$$

where H and V are the sets of horizontal and vertical lines, and θ_k is the orientation of line l_k.

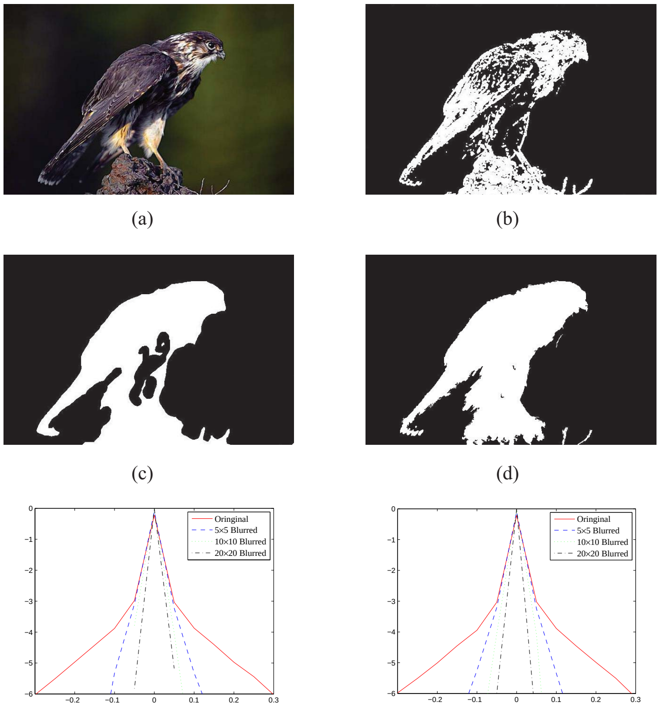
The location features are defined as:

$$f_3 = \frac{1}{\|H\|} \sum_{l_k \in H} \frac{y_{k1} + y_{k2}}{2}, \; f_4 = \frac{1}{\|V\|} \sum_{l_k \in V} \frac{x_{k1} + x_{k2}}{2}$$

where (x_{k1}, y_{k1}) and (x_{k2}, y_{k2}) are the two endpoints of line l_k.

IV. Subject Area Extraction Methods

The subject area of a photo is defined as the region that viewers pay attention to. Wong *et al.* [12] and Nishiyama *et al.* [33] used the saliency map [39] to extract the subject areas, which were assumed to have higher brightness and contrast than other regions. A saliency detector may seem a natural choice to extract the region of interest. However, computing the saliency map may fail to correctly segment the subject region in many cases. For instance, if a certain part of the subject area has very high brightness and contrast, other parts will be ignored by this method. See examples in Figure 6.



(a) (b)

(c) (d)

(e)

Fig. 7. Clarity-based subject area extraction. (a) Input image. (b) Clear region generated by mask U_0 (white area). (c) Mask refined by local convex hull. (d) Mask further improved by superpixels. It is used as the subject area. (e) Log histograms of the horizontal (left) and vertical (right) derivatives of the original image and images blurred by kernels f_5, f_{10}, and f_{20}, respectively.

The way to detect subject areas in photos depends on photo content. When taking close-up photos of animals, plants, and statics, photographers often use a macro lens to focus on the main subjects, such that photos are clear on the main subjects and blurred in other areas. For human portraits, viewers' attentions are often automatically attracted by human faces. In outdoor photography, architectures, mountains, and trees are often the main subjects.

We thus propose a clarity-based method to find clear regions in shallow depth of field photos, which take the majority of high-quality photographs in the categories of "animal", "plant", and "static". We adopt a layout-based method [40] to segment vertical standing objects, which are treated as subject areas by us, in photos from the categories of "landscape" and "architecture". For photos in the category of "human", we use a human detector and a face detector to locate human faces.

A. Clarity-Based Subject Area Extraction

We extend the 1D motion blur detection scheme proposed by Levin [41] to identify 2D blurred regions in an image [9]. We use a kernel of size $k \times k$ with all coefficients equal to $1/k^2$ to blur a photo. As shown in Figure 7 (e), blurring significantly changes the shapes of the derivative histograms of the photo. Therefore, the statistics of the responses of derivative filters can be used to tell the difference between clear and blurred regions.

The input photo I is convoluted with mentioned blurring kernels f_k $(k = 1, 2, \cdots, 50)$ of size $k \times k$. We then compute the horizontal and vertical derivatives from $I * f_k$ to obtain the distributions of the horizontal and vertical derivatives:

$$p_{xk} \propto hist(I * f_k * d_x), \qquad p_{yk} \propto hist(I * f_k * d_y),$$

where $d_x = [1, -1]$, and $d_y = [1, -1]^T$ are spatial derivative operators.

For a pixel i in I, we define a log-likelihood of the derivatives in its rectangular neighboring window $\Omega(i)$ with respect to each of the blurring model as:

$$L_k(i) = \sum_{i' \in \Omega(i)} (\ln p_{xk}(I_x(i')) + \ln p_{yk}(I_y(i')))$$

where $I_x(i')$ and $I_y(i')$ are the horizontal and vertical derivatives at pixel i' respectively. $L_k(i)$ measures how well pixel i's neighboring window is explained by a $k \times k$ blurring kernel. Defining $k^*(i) = \arg\max_k L_k(i)$, we generate a mask U_0, which labels pixel i as clear when $k^*(i) = 1$, as blurred when $k^*(i) > 1$. This clarity mask is then improved by an iterative procedure (see Figure 7). A pixel is labeled as clear if it falls in the convex hull of its neighboring pixels labeled as clear. The step repeats until convergence. Then a photo is oversegmented into super-pixels [42]. A super-pixel is labeled as clear if more than half of its pixels are labeled as clear.

B. Layout-Based Subject Area Extraction

Hoiem *et al.* [40] proposed a method to recover the surface layout from an outdoor image by learning the appearance-based models for each geometric class. The scene is segmented into three classes: sky regions, ground regions, and vertical standing objects, as shown in Figure 8. We take vertical standing objects as subject areas. The technical details are skipped and can be found in [40].

C. Human-Based Subject Area Extraction

We employ face detection [43] to extract faces from human photos. For images where face detection fails, we use human detection [44] to roughly estimate the locations of faces. See examples in Figure 8.

V. Regional Features

We have developed regional features in accordance with human aesthetic judgements to work together with the proposed subject area extraction methods. We propose a new dark channel feature to measure both the clarity and the colorfulness of subject areas. Clarity and Lighting contrast features are computed to evaluate the clarity and the lighting condition of subject areas and background. We use the composition geometry feature to assess the location of a subject, and specially design a set of features for "human" photos to measure clarity, brightness, and lighting effects of faces. Complexity features are proposed to measure the complexities of subject areas and background.

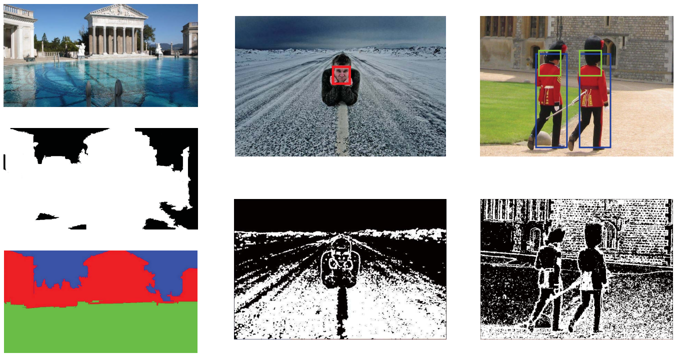


Fig. 8. (a): From top downwards: The input photo; result of clarity-based subject area extraction (white region); result of layout-based subject area extraction (red region). (b),(c): First row: input photos with face and human detection results. Second row: clarity-based subject area extraction results.

A. Dark Channel Feature

Dark channel was introduced by He *et al.* [45], [46] for haze removal. The dark channel of an image I is defined as:

$$I_{dark}(i) = \min_{c \in R,G,B} (\min_{i' \in \Omega(i)} I_c(i'))$$

where I_c is a color channel of I and $\Omega(i)$ is the neighborhood of pixel i. We choose $\Omega(i)$ as a 10×10 local patch. We normalize the dark channel value by the sum of RGB channels to reduce the effect of brightness. The dark channel feature of a photo I is computed as the average of the normalized dark channel values in the subject areas:

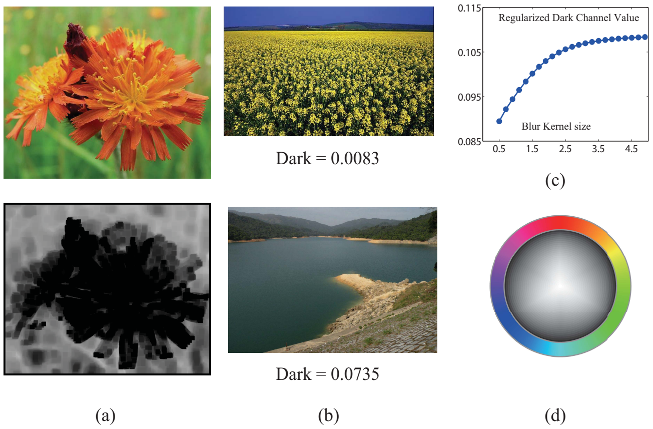
$$\frac{1}{\|S\|} \sum_{(i) \in S} \frac{I_{dark}(i)}{\sum_{c \in R,G,B} I_c(i)},$$

where S is the subject area of I.

The dark channel feature is a combined measurement of clarity, saturation, and hue composition. Since dark channel is essentially a minimum filter on RGB channels, blurring the image would average the channel values locally and thus increase the response of the minimum filter. Figure 9 (c) shows that the dark channel value of an image increases with the degree to which it is blurred. The subject area of a shallow depth of field image show lower dark channel values than the background as shown in Figure 9 (a). For pixels of the same hue value, those with higher saturation gives lower dark channel values (Figure 9 (d)). As shown in Figure 9 (b), a low-quality photograph with dull color gives a higher averaged dark channel value. In addition, different hue values gives different dark channel values (Figure 9(d)). So the dark channel feature also incorporates hue composition information.

B. Clarity Contrast Feature

To attract the audience's attention to a subject and to isolate the subject from the background, professional photographers sometimes keep the subject in focus and make the background out of focus. High-quality photographs of certain categories, such as "animal", "human", and "static", are neither entirely clear nor entirely blurred. To characterize this property, we



Dark = 0.0083

(c)

Dark = 0.0735

(a)

(b)

(d)

Fig. 9. (a) A close-up on plant and its dark channel. (b) Landscape photographs with different color composition. (c) Average dark channel value of the input photo from (a) blurred by Gaussian kernel. (d) For any point in the circle, the hue wheel indicates the hue, the radius equals to the saturation, and the brightness is the normalized dark channel value.

first extract a rectangular bounding box for the subject area of a photo belonging to the categories mentioned above, since the subject area of a photo of these categories tends to concentrate to one rectangular region. Given the clarity mask U, we project it onto the x and y axes to get the horizontal and vertical density,

$$U_x(i) = \sum_j U(i,j), \; U_y(i) = \sum_i U(i,j).$$

On the x axis, we find x_1 and x_2 such that the energy in $[0, x_1]$ and the energy in $[x_2, W-1]$ are each equal to $(1-\alpha)/2$ of the total energy in U_x, where W is width of the image. Similarly, we compute y_1 and y_2 in the y direction. The bounding box R is the region $[x_1, x_2] \times [y_1, y_2]$. We choose $\alpha = 0.9$ in our experiments. The clarity contrast feature is computed as

$$f = \frac{\|M_R\|/\|R\|}{\|M_I\|/\|I\|},$$

where $|R|$ and $|I|$ are the areas of region R and image I, and

$$\begin{gathered} M_I = \{(u,v) \,|\, |F_I(u,v)| > \beta \max\{F_I(u,v)\}\}, \\ M_R = \{(u,v) \,|\, |F_R(u,v)| > \beta \max\{F_R(u,v)\}\}, \\ F_I = FFT(I), \; F_R = FFT(R). \end{gathered}$$

$FFT()$ is the fast Fourier transform. u and v are spatial frequencies in x and y directions. $|M_R|$ and $|M_I|$ measure the degrees of concentration of high frequency components in the subject area and in the whole image. We choose $\beta = 0.2$ in our experiments. f is high if the subject area is in focus and the background is out of focus.

C. Lighting Contrast Feature

Since professional photographers often use different lightings on the subject and the background, the brightness of the subject is significantly different from that of the background. However, most amateurs use natural lighting and let the camera automatically adjust the brightness of a photo, which usually reduces the brightness difference between the subject and the background. We thus formulate the lighting contrast feature as

$$f = \ln(B_f/B_b),$$

where B_f and B_b are average lightings of the subject area and the background, respectively.

D. Composition Geometry Feature

Good geometrical composition is a basic requirement for high-quality photos. One of the most well-known principles of photographic composition is the *Rule of Thirds*. If we divide a photo into nine equal-size parts by two equally-spaced horizontal lines and two equally-spaced vertical lines, the rule suggests that the four intersections of the lines should be the centers for subjects. Studies have shown that when viewing images, people usually look at one of the intersection points rather than the center of the image. To formulate this criterion, we define a composition feature as

$$f = \min_i \sqrt{(C_x - P_{ix})^2/W^2 + (C_y - P_{iy})^2/H^2}$$

where (C_x, C_y) is the centroid of the subject region, (P_{ix}, P_{iy}), $i \in 1,2,3,4$ are the four intersection points, and W and H are the width and height of image I, respectively.

E. Complexity Features

Professional photographers tend to keep background composition simple in order to reduce its distraction. We use the segmentation result and the color distribution of the background to measure both spatial and hue complexities. A photo is over-segmented into super-pixels. Let N_s and N_b be the numbers of super-pixels in the subject area and the background, $\|S\|$ and $\|B\|$ be the areas of the subject area and the background. Then the following spatial complexity features are defined,

$$g_1 = N_s/\|S\|, \; g_2 = N_b/\|B\|, \; g_3 = N_s/N_b.$$

We also quantize each of the RGB channels of the background into 16 values, creating a histogram H of 4096 bins. Letting h_{max} be the maximum count among all the bins of the histogram, the hue complexity feature is defined as:

$$g_4 = \|S\|/4096, \;\; S = \{i | H(i) \; = \gamma h_{max}\}.$$

We choose $\gamma = 0.01$ in our experiments.

F. Human Based Features

Faces in high-quality human portraits usually occupy reasonable portions of photos, have high clarity, and show professional employment of lightings. Therefore, we extract the following features to assess the quality of human photos: the ratio of face areas, the ratio of shadow areas, the clarity of faces, and the average lighting of faces.

Let I be a grayscale photo and X_k be a detected face region. The ratio of face areas is computed as

$$f_1 = \frac{1}{\|I\|} \sum_k \|X_k\|,$$

where $\|I\|$ and $\|X_k\|$ are the areas of the photo and the faces.

Lighting plays an essential role in portrait photography. Portrait photographers use special light settings in their studios to highlight the face and create shadows. To evaluate the lighting effect in artistic portraits, we compute the area S_k of shadow on a face region X_k as following,

$$S_k = \|\{i \mid i \in X_k \ \ \& \ \ I(i) < 0.1 \max_i I(i)\}\|.$$

The ratio of shadow areas on faces is extracted as a feature,

$$f_2 = \sum_k S_k / \sum_k \|X_k\|.$$

The clarity of face regions is computed through the Fourier transform by measuring the ratio of the area of high frequency components to that of all frequency components. Let $\widetilde{X}_k$ be the Fourier transform of X_k and $M_k = \{(u, v) \mid |\widetilde{X}_k(u, v)| > \beta \max \widetilde{X}_k(u, v)\}$. We choose $\beta = 0.2$ in our experiments. The face clarity feature is

$$f_3 = \sum_k \|M_k\| / \sum_k \|X_k\|.$$

The average lighting of faces is computed as

$$f_4 = \frac{\sum_k \sum_{i \in X_k} I(i)}{\sum_k \|X_k\|}.$$

VI. Quality Assessment without the Information of Photo Categories

Each of the proposed features has different effectiveness on photos with different visual content. Therefore, a natural way of improving the performance is to train a classifier (such as SVM) for each photo category separately. For a test photo, a proper classifier is chosen to combine features according to its category label. However, if only visual information is available, the problem of classifying a test photo into one of the seven categories defined in Section I-C is challenging. Instead, we propose to online train an adaptive classifier from neighboring samples whose visual content is similar to the test photo. It is likely for the neighboring samples to be in the same category as the test photo.

The features proposed in this work cannot well characterize visual content and scene categories. Therefore, a different set of features including Edge Orientation Histograms [47], Histogram of Oriented Gradients [44] and GIST [48], which were used for image search in [49], [50], are used to retrieve neighboring low- and high-quality training samples for the test photo.

For each test photo, we employ the mentioned visual features and find its K nearest neighbors using the kd-tree [51] in both low- and high-quality photo training sets. $K = 100$ in our experiment. We then use the majority labels in the returned $2K$ samples to determine which of the following three groups the test photo belongs to: 1) "animal", "static", "plant", "night"; 2) "architecture", "landscape"; 3) "human". We extract regional features using clarity-based subject area extraction, layout-based subject area extraction or face/human detection according to this classification result. Examples of test photos and returned neighboring high- and low-quality samples are shown in Figure 10.

For each test photo, we online train a linear SVM classifier to combine features of assessing photo quality using the returned training samples. The online training is very efficient because the size of our proposed features is very small. Online training and K-NN search take around 30ms in total for a test photo on a regular PC. The detailed experimental results are discussed in Session VII.

VII. Experimental Results

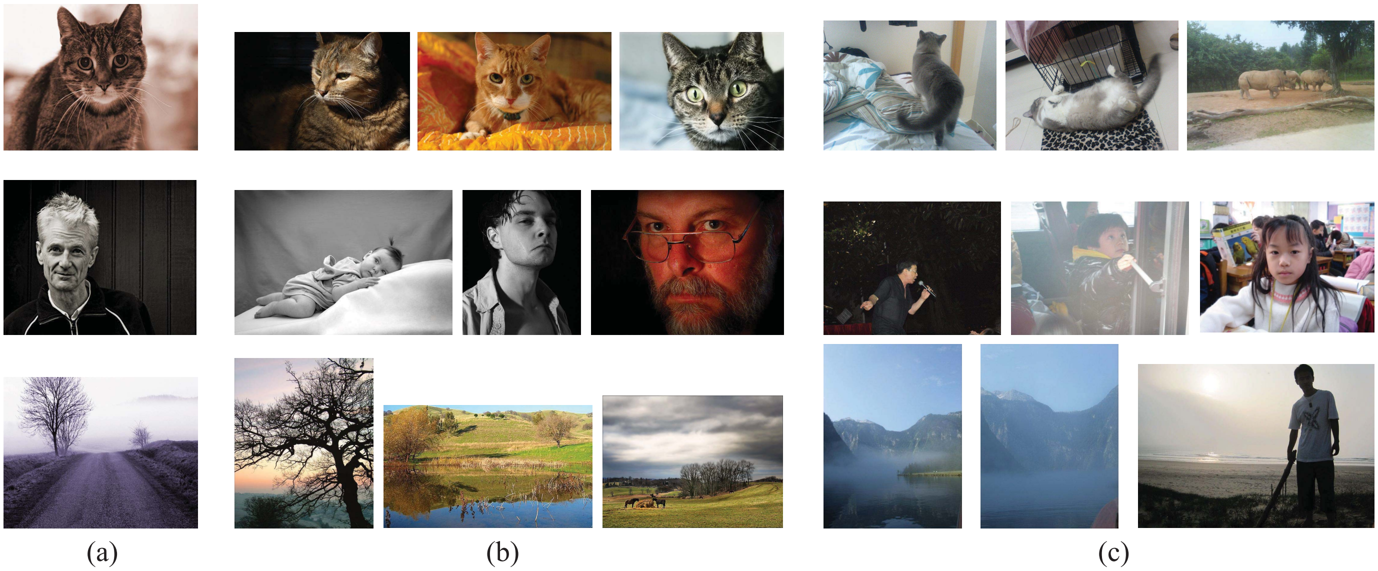
In this section, we evaluate our features and other state-of-the-art features on a benchmark dataset with photos of different categories. The results show that our features significantly outperform other existing features and the effectiveness of different features highly depends on the visual content of photos. We also compare different ways of combining different types of features: learning a classifier for each photo category separately; and learning an adaptive classifier for a test photo without knowing its category. It shows that the later approach achieves comparable performance with the former one and is even more effective on some categories.

A. Database description

The initial database consists of $32,097$ photos acquired from professional photography websites and contributed by amateur photographers. They are divided into seven categories according to photo content (see Table I) and are labeled by ten independent reviewers into three classes: *high quality*, *low quality*, and *uncertain about quality*. A photo is classified as high or low quality only if eight out of the ten reviewers agree on its assessment. Other photos, on which the reviewers have more diversified opinions, are not included in the benchmark dataset. Finally the benchmark dataset has $17,673$ photos for evaluation.

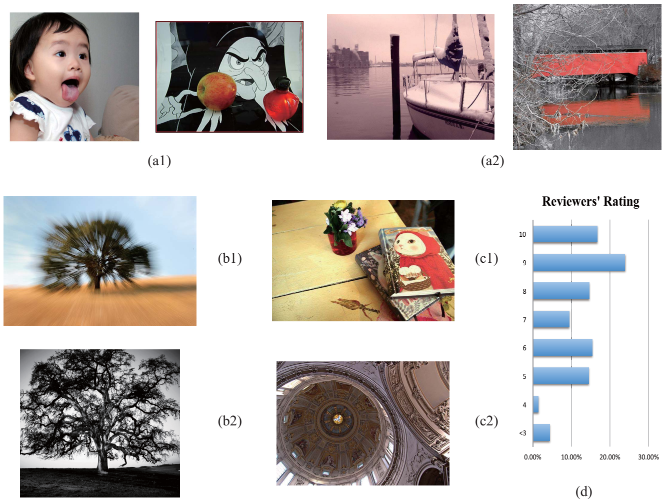
Of all the photos, 55% are with consensus quality rating and are included in the benchmark dataset (see Figure 11 (d)). This confirms that there exist general criteria for quality assessment. Although the photos without consensus quality assessment are not included in the dataset, they give us insights on the process of photo quality evaluation. It is observed that reviewers assess photo quality mainly from the following three aspects. If a photo only satisfies the criteria in one or two aspects, reviewers may have different opinions on it, since they may put different emphasis on the three aspects.

Photo topic. Some photos have interesting topics, such as attractive faces, interesting arrangement of objects, or intriguing concepts (see Figure 11 (a1)). However, they are not voted as high quality unanimously for the lack of photographic skills. In fact, interestingness of images is regarded as a different topic from aesthetic quality assessment [35]. Some photos tend to invoke viewers specific emotions but may not be qualified on certain criteria for high-quality photos. For instance, a photo that inspires nostalgic feelings (see Figure 11 (a2)) might be relatively poor in colorfulness.



(a) (b) (c)

Fig. 10. K-NN search results using features characterizing visual content. (a) Test photos (b) Returned high-quality neighboring samples. (c) Returned low-quality neighboring samples.



(a1) (a2) (b1) (c1) (b2) (c2) (d)

Fig. 11. Examples of photos to be labeled by reviewers. (a1) and (a2) are amateur snapshots with interesting topics. (b1) and (b2) are photos with special effects (motion blur and infrared imaging). (c1) and (c2) are photos taken by advanced photographic equipments however lack of photographic skills and clear topics. They may not be rated as high-quality photos with high consensus. (d) is the statistics of reviewers' rating result. The histogram of photos with the maximum k out of 10 reviewers giving consensus quality rating. k is from 3 to 10.

Photographic skills. Photos taken with special photographic skills might be perceived as high quality. It is true that such photos require higher shooting skills. However, skills cannot make up for bad topics. In the meantime, the special effects may backfire: motion blurred photos (see Figure 11 (b1)) may be considered as lacking details instead of stressing dynamics; infrared imaging photos (see Figure 11 (b2)) can be regarded as dull in color.

Quality of equipment. Advanced photographic instruments are becoming more accessible to the public. Expert-level cameras surely help to produce high-quality photos. However, photography demands more than equipments. We find a large amount of such photos that cannot be rated as high-quality unanimously because of badly chosen topics or lack of photographic skills (see Figure 11 (c1) and (c2)).

B. Experimental Settings

We are able to select out $17,673$ photos with labels according to the criterion mentioned in Section VII-A. The size of the dataset is shown in Table I. Features are tested separately or combined with a linear SVM. When combining features with SVM, for each category, we randomly sample half of the high- and low-quality photos as the training set and keep the other half as the test set. If the information of photo categories is assumed to be known, the classifiers for different categories are trained separately. Otherwise, for a test photo, its adaptive classifier is online trained from its neighboring samples in the training set as described in Section VI. The random partition repeats for ten times and the averaged test results are reported. The performance of features is measured with the area under the ROC curve. Four groups of features are compared in Table I: the proposed regional features; the proposed global features; the selected regional features and selected previous global features previously proposed in [7], [8], [14]. For each photo category, the best performance achieved by a single feature is underlined and marked bold. Reasonably good suboptimal results achieved by other features are also marked bold.

C. Result Analysis

All the tested features show different performance for photos with different content. Generally speaking, in the categories

Category		**Animal**	**Plant**	**Static**	**Architecture**	**Landscape**	**Human**	**Night**	**Overall**
Number of high-quality photos		948	594	531	595	820	678	352	4517
Number of low-quality photos		2224	1803	2004	1290	1947	2536	1352	13156
Regional features	Our proposed regional features								
	Dark Channel	**0.8393**	0.7858	**0.8335**	**0.8869**	**0.8575**	0.7987	0.7062	0.8189
	Clarity Contrast	**0.8074**	0.7439	0.7309	0.5348	0.5379	0.6667	0.6297	0.6738
	Lighting Contrast	0.7551	0.7752	0.7430	0.6460	0.6226	0.7612	0.5311	0.7032
	Geometry Composition	0.7425	0.7308	0.5920	0.5806	0.4939	0.6828	0.6075	0.6393
	Complexity Combined	**0.8212**	**0.8972**	0.7491	0.7219	0.7516	0.7815	**0.7284**	0.7817
	Face Combined	N.A	N.A	N.A	N.A	N.A	**0.9521**	N.A	N.A
	Combined	0.8632	0.9102	0.8437	0.8966	0.8931	0.9612	0.8382	0.8820
	Previously proposed regional features								
	Low Depth-of-Field [8]	0.7231	0.7646	0.6930	0.5204	0.5841	0.7277	0.5642	0.6711
	Central Saturation [8]	0.6844	0.6615	0.6771	0.7208	0.7641	0.6707	0.5974	0.6857
	Combined	0.7861	0.7638	0.7174	0.7386	0.7753	0.7694	0.6421	0.7792
Global features	Our proposed global features								
	Hue Composition	0.7861	**0.8316**	**0.8367**	**0.8376**	**0.8936**	0.7909	**0.7214**	0.8165
	Scene Composition	0.7003	0.5966	0.7057	0.6781	0.6979	0.7923	**0.7477**	0.7056
	Combined	0.7891	0.8350	0.8375	0.8531	0.8979	0.8081	0.7744	0.8282
	Previously proposed global features								
	Blur [7]	0.7566	0.7963	0.7662	0.7981	0.7785	0.7381	0.6665	0.7592
	Brightness [7]	0.6993	0.7337	0.6976	**0.8138**	0.7848	0.7801	**0.7244**	0.7464
	Hue Count [7]	0.6260	0.6920	0.5511	0.7082	0.5964	0.7027	0.5537	0.6353
	Visual balance [14]	N.A	N.A	N.A	0.6204	0.6373	N.A	0.6537	N.A
	Combined	0.7751	0.8093	0.7829	0.8526	0.8170	0.7908	0.7321	0.7944
Proposed features combined		0.8867	0.9004	0.9041	0.9100	0.9266	0.9662	0.8403	0.9121
Previous features combined		0.8129	0.8127	0.8010	0.8547	0.8411	0.8392	0.7406	0.8241
All features combined		0.8937	0.9182	0.9069	0.9275	0.9468	0.9740	0.8463	0.9209

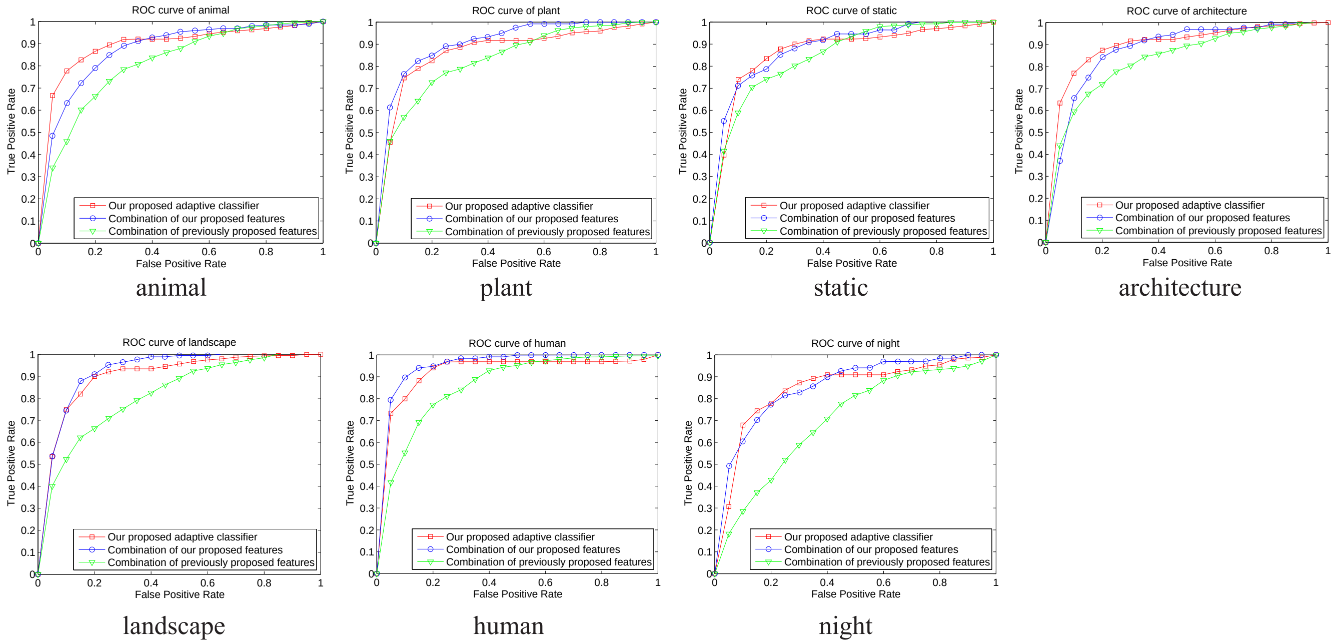
TABLE I

Overview of feature performance on our database. The best performance achieved by a single feature is underlined and marked bold. Reasonably good suboptimal results achieved by other features are also marked bold. For features previously proposed in [7], [8], [14], we select those with the best performance.

of "animal", "plant", and "static", the subject areas of high-quality photos often exhibit high contrast with background and can be well detected. Therefore regional features are more effective for them. For outdoor photos in the categories of "architecture", "landscape", and "night", subject areas may not be well detected and global features are more robust. For the photos in "human", specially designed features for faces are the best performers. Assessing the quality of photos in the category of "night" is very challenging. The features previously proposed in [7], [8], [14] perform slightly better than random guess. Although our proposed features perform much better, the result is still not satisfactory. There is a large room to improve in the future work. Some new features need to be developed considering the special photographic skills used at night. Combining different types of features with SVM can improve the performance.

Our proposed features significantly outperform the existing features in general. We also observe some detailed differences of their performance on different types of photos. The dark channel feature measures the clarity and the colorfulness of photos and is very effective in most categories. It achieves the best performance in the categories of "animal" and "architecture" and its performance is close to the best in the categories of "static" and "landscape". It outperforms previously proposed clarity features such as "blur" [7]. The Clarity Contrast feature has high performance when the subject area can be well fitted with a rectangular region. It performs well in the category of "animal" especially. Lighting Contrast performs acceptably well in the categories of "animal", "static", and "plant". Our complexity feature achieves the best performance in the category of "static" and its performance is close to the best in the categories of "animal" and "plant". The high-quality photos in both categories usually have high complexity in subject areas and low complexity in the background. Our proposed face features are very effective for "human" photos and enhance the best performance (0.78) obtained by previously proposed features to 0.95.

The hue composition feature is a very effective measurement of color composition quality. It achieves the best performance on "static" and "landscape" and its performance is close to



animal plant static architecture

landscape human night

Fig. 12. Performance comparison of photo quality assessment on seven categories of photos.

the best on "plant", "architecture", and "night". Our scene composition feature has the best performance on "night". It outperforms previous relevant features such as "visual balance" [14] in most categories.

Previously proposed features show mixed performance across categories. For example, the shallow depth-of-field feature proposed in [8] works reasonably well on "animal", and "plant", where the assumption that the subject is at the central area is reasonable to some degree. However, their performance greatly decreases on "static", "architecture", "landscape", "human", and "night".

In Figure 12, we show the ROC curves of combining previously proposed features, combining our features, and the adaptive classifiers proposed in Section VI. In the first two approaches, the information of photo categories is known. It shows that our features outperform previously proposed features. We also show that combining all the features together leads to the best performance in Table I. Using the adaptive classifiers, we achieve performance comparable to training a classifier for each photo category separately and assuming the information of photo categories is known. In some categories, the adaptive classifiers even achieve higher performance because the retrieved training samples in local regions are more relevant to the test photos than roughly partitioning all the photos into seven categories.

VIII. Conclusions and Discussions

In this paper, we propose the content-based photo quality assessment framework, together with a set of new subject area detection methods and new global/regional features. Extensive experiments on a large benchmark database show that the subject area detection methods and features have very different effectiveness on different types of photos. Therefore we should extract subject areas in different ways and train different classifiers of combining features for different photo categories separately. Our proposed new features significantly outperforms existing features. The performance of this framework can be further improved by incorporating more new features in the future. We also propose adaptive classifiers without knowing the information of photo categories. A different set of features characterizing visual content are first employed to retrieve both high- and low-quality training photos whose visual content is similar to the test photo. Then the retrieved training samples are used to train an adaptive classifier to combine the visual features used to assess photo quality. It achieves comparable performance to the case of knowing the category of each test photo.

However, photo quality assessment still has many challenging and interesting problems to be solved in the future work. For some types of photos such as as "night", the performance of current features is still not satisfactory and new features need to be developed. Although photo assessment has some general criteria, it is still a highly subjective task. Viewers may have very different opinions on the same photo. In our experiments, only the photos with consensus among reviewers are selected into the benchmark databset. However, it is also interesting to study how and why reviewers have different opinions on the remaining photos. For example, viewers have different personal preference on photos. Someone may weight photo topics more than photographic skills in the process of their assessment. Someone like "animals" more than "landscape". Professional photographers and amateurs view photos in different ways. Learning personalized classifiers for photo assessment has interesting applications to photo recommendation and image retrieval. It is also possible to develop computing algorithms which automatically classify viewers into professional and amateurs or cluster viewers into different groups according to how viewers select their preferred photos. How to integrate photo quality assessment into other applications, such as image search [50], [52], is another important problem to be explored yet.

Besides photo quality, there is some recent research work on other types of high-level feelings on images, such as interestingness [15], memorability [53] and attractiveness [54]. They are related to photo quality but not the same. For example, Isola *et al.* [53] found that some high-quality landscape photos were actually the least memorable. It is interesting to explore what types of features are shared or different when predicting these properties.

Acknowledgments

This work is partially supported by Hong Kong SAR through GRF projects CUHK416510, CUHK417110, CUHK417011, National Natural Science Foundation of China (61005057), and by Guangdong Province through Introduced Innovative R&D Team of Guangdong Province 201001D0104648280.

References

[1] B. Manav. Color-emotion associations and color preferences: A case study for residences. *Color Research Application*, 2007.

[2] X.P. Gao, J.H. Xin, T. Sato, A. Hansuebsai, M. Scalzo, K. Kajiwara, S.S. Guan, J. Valldeperas, M.J. Lis, and M. Billger. Analysis of cross-cultural color emotion. *Color Research Application*, 2007.

[3] H. Mante and E.F. Linssen. *Color design in photography*. Focal Press, 1972.

[4] M. Freeman. *The complete guide to light & lighting in digital photography*. Lark Books (NC), 2006.

[5] M. Freeman. *The photographer's eye: composition and design for better digital photos*. Focal Pr, 2007.

[6] B. London and J. Stone. *Short Course in Photography*. Addison-Wesley Educational Publishers, 1998.

[7] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.

[8] R. Datta, D. Joshi, J. Li, and J. Wang. Studying aesthetics in photographic images using a computational approach. In *Proc. European Conf. Computer Vision*, 2006.

[9] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *Proc. European Conf. Computer Vision*, 2008.

[10] R. Datta, J. Li, and J.Z. Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *Proc. IEEE Int'l Conf. Image Processing*, 2008.

[11] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *Proc. ACM Multimedia*, 2009.

[12] L.K. Wong and K.L. Low. Saliency-enhanced image aesthetics class prediction. In *Proc. IEEE Int'l Conf. Image Processing*, 2009.

[13] X. Jin, M. Zhao, X. Chen, Q. Zhao, and S.C. Zhu. Learning Artistic Lighting Template from Portrait Photographs. In *Proc. European Conf. Computer Vision*, 2010.

[14] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proc. ACM Multimedia*, 2010.

[15] S. Dhar, V. Ordonez, and T.L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[16] W. Luo, X. Wang, and X. Tang. Content-based photo quality assessment. In *Proc. Int'l Conf. Computer Vision*, 2011.

[17] J. Carucci. *Capturing the Night with Your Camera: How to Take Great Photographs After Dark.* Amphoto Books, 1995.

[18] L. White. *Infrared Photography Handbook*. Amherst Media, Inc., 1995.

[19] C. Grey. *Master Lighting Guide for Portrait Photographers.* Amherst Media, Inc., 2004.

[20] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via plsa. In *Proc. European Conf. Computer Vision*, 2006.

[21] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results. http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html.

[22] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proc. ACM Multimedia*. ACM, 2010.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 2010.

[24] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity, 1993.

[25] A.M. DijK, J.B. Martens, and A.B. Watson. Quality assessment of coded images using numerical category scaling. *Advanced Image and Video Communications and Storage Technologies*, 1995.

[26] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik. Image quality assessment based on a degradation model. *IEEE Trans. on Image Processing*, 2000.

[27] Y. Wang, T. Jiang, W. Ma, and W. Gao. Novel spatio-temporal structural information based video quality metric. *IEEE Trans. on Circuits and Systems for Video Technology*, 2012.

[28] X. Li. Blind image quality assessment. In *Proc. IEEE Int'l Conf. Image Processing*, 2002.

[29] H.R. Sheikh, A.C. Bovik, and L. Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE Trans. on Image Processing*, 2005.

[30] H. Tong, M. Li, H.J. Zhang, C. Zhang, J. He, and W.Y. Ma. Learning no-reference quality metric by examples. *Proc. Int'l Conf. Multimedia Modeling*, 2005.

[31] H. Tong, M. Li, H.J. Zhang, J. He, and C. Zhang. Classification of Digital Photos Taken by Photographers or Home Users. In *Proc. Pacific Rim Conf. Multimedia*, 2004.

[32] M. Nishiyama, Sato I. Okabe, T., and Y. Sato. Aesthetic quality classification of photographs based on color harmony. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[33] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based Photo Cropping. In *Proc. ACM Multimedia*, 2009.

[34] L. Lo and J. Chen. A statistic approach for photo quality assessment. In *Proc. Information Security and Intelligence Control*, 2012.

[35] S. Dhar, V. Ordonez, and T. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

[36] M. Yeh and Y. Cheng. Relative features for photo quality assessment. In *Proc. IEEE Int'l Conf. Image Processing*, 2012.

[37] M. Tokumaru, N. Muranaka, and S. Imanishi. Color design support system considering color harmony. In *Proc. IEEE International Conference on Fuzzy Systems*, 2002.

[38] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.Q. Xu. Color harmonization. In *Proc. ACM SIGGRAPH*, 2006.

[39] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002.

[40] D. Hoiem, A.A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision*, 2007.

[41] A. Levin. Blind motion deblurring using image statistics. *Proc. Advances in Neural Information Processing Systems*, 2007.

[42] Xiaofeng Ren and Jitendra Malik. Learning a classification model for segmentation. In *Proc. Int'l Conf. Computer Vision*, 2003.

[43] R. Xiao, H. Zhu, H. Sun, and X. Tang. Dynamic cascades for face detection. In *Proc. Int'l Conf. Computer Vision*, 2007.

[44] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2005.

[45] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009.

[46] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33:2341–2353, 2011.

[47] W. Freeman and M. Roth. Orientation histogram for hand gesture recognition. In *Proc. Int'l Workshop on Automatic Face and Gesture Recognition*, 1995.

[48] A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context-based vision system for place and object recognition. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2003.

[49] X. Wang, K. Liu, and X. Tang. Query-specific visual semantic spaces for web image re-ranking. In *Proc. Int'l Conf. Computer Vision*, 2011.

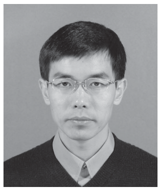
[50] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang. Intentsearch:capturing user intention for one-click internet image search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34:1342–1353, 2012.

[51] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *Proc. VISAPP Int'l Conf. Computer Vision Theory and Applications*, 2009.

[52] Jingyu Cui, Fang Wen, and Xiaoou Tang. Real time google and live image search re-ranking. In *Proc. ACM Multimedia*, 2008.

[53] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2011.

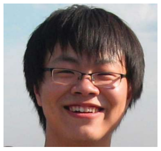
[54] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. In *Proc. ACM SIGGRAPH*, 2008.



Xiaoou Tang (S'93-M'96-SM'02-F'09) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, and the M.S. degree from the University of Rochester, Rochester, NY, in 1991. He received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is a Professor in the Department of Information Engineering and Associate Dean (Research) of the Faculty of Engineering of the Chinese University of Hong Kong. He worked as the group manager of the Visual Computing Group at the Microsoft Research Asia from 2005 to 2008. His research interests include computer vision, pattern recognition, and video processing.

Dr. Tang received the Best Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009. He is a program chair of the IEEE International Conference on Computer Vision (ICCV) 2009 and an Associate Editor of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and International Journal of Computer Vision (IJCV). He is a Fellow of IEEE.



Wei Luo received the B.S. degree from Tsinghua University in Electronic Engineering in 2010. He is currently an M.Phil. student in the Department of Information Engineering at the Chinese University of Hong Kong. His research interests include image processing, computer vision and machine learning.

Xiaogang Wang (S'03-M'10) received the B.S. degree from University of Science and Technology of China in Electrical Engineering and Information Science in 2001, and the M.S. degree from Chinese University of Hong Kong in Information Engineering in 2004. He received the PhD degree in Computer Science from the Massachusetts Institute of Technology. He is currently an assistant professor in the Department of Electronic Engineering at the Chinese University of Hong Kong. His research interests include computer vision and machine learning.