Content-Based Photo Quality Assessment

Wei Luo[1], Xiaogang Wang[2,3], and Xiaoou Tang[1,3]
[1]Department of Information Engineering, The Chinese University of Hong Kong
[2]Department of Electronic Engineering, The Chinese University of Hong Kong
[3]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

`lw010@ie.cuhk.edu.hk xgwang@ee.cuhk.edu.hk xtang@ie.cuhk.edu.hk`

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 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 subject areas and the background regions are combined with global features to assess the photo quality. Since professional photographers may adopt different photographic techniques and may have different aesthetic criteria in mind when taking different types of photos (e.g. landscape versus portrait), we propose to segment regions and extract visual features in different ways according to the categorization of photo content. Therefore we divide the photos into seven categories based on their content and develop a set of new subject area extraction methods and new visual features, which are specially designed for different categories. This argument is supported by extensive experimental comparisons of existing photo quality assessment approaches as well as our new regional and global features over different categories of photos. Our new features significantly outperform the state-of-the-art methods. Another contribution of this work is to construct a large and diversified benchmark database for the research of photo quality assessment. It includes $17,613$ photos with manually labeled ground truth.

1. Introduction

Automatic assessment of photo quality based on aesthetic perception gains increasing interest in computer vision community. It has important applications. For example, when users search images on the web, they expect
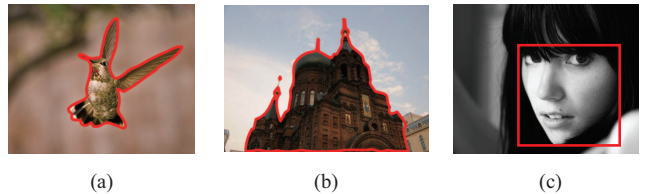

Figure 1. Subject areas of photos. (a) Close-up for a bird. (b) Architecture. (c) Human portrait.

the search engine to rank the retrieved images according to their relevance to the queries as well as their quality. Various methods of automatic photo quality assessment were proposed in recent years [16, 18, 11, 5, 12, 20, 10]. In early works, only global visual features, such as global edge distribution and exposure, were used [11]. However, later studies [5, 12, 20] showed that regional features lead to better performance, since human beings perceive subject areas differently from the background (see examples in Figure 1). 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. Both Regional and global features will be used in our work.

One major problem with the existing methods is that they treat all photo equally without considering the diversity in photo content. It is known that professional photographers adopt different photographic techniques and have different aesthetical criteria in mind when taking different types of photos [2, 19]. For example, for close-up photographs (e.g. Figure 1 (a)), viewers appreciate the high contrast between the foreground and background regions. In human portraits photography (e.g. Figure 1 (c)), professional photographers use special lighting settings [6] 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, the subject areas of different types of photos should

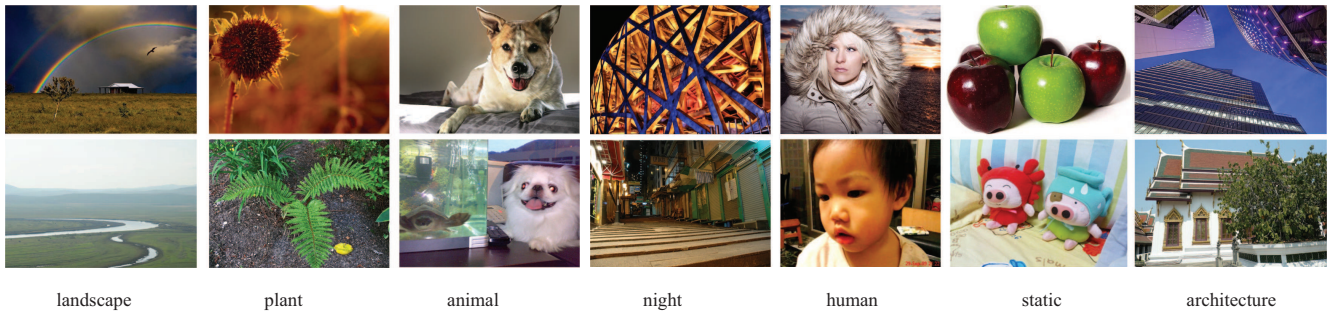| landscape | plant | animal | night | human | static | architecture |

Figure 2. Photos divided into seven categories according to content. First row: high quality photos; Second row: low quality photos.

be extracted in different ways. In a close-up photo, the subject area is emphasized using the low 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 in this type of photos.

1.1. Our Approach

Motivated by these considerations, we propose content-based photo quality assessment. Photos are manually divided into seven categories based on photo content: "animal", "plant", "static", "architecture", "landscape", "human", and "night". See examples in Figure 2. 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 region detection* combines blur kernel estimation with image segmentation to accurately extract the clear region as the subject area.

- *Layout based region detection* analyzes the layout structure of a photo and extracts vertical standing objects.

- *Human based detection* locates faces in the photo with a face detector or a human detector.

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

- *Dark channel feature* measures the clearness and the colorfulness of the subject areas.

- *Complexity features* use the numbers of segmentations to measure the spatial complexity of the subject area and the background.

- *Human based features* capture the clarity, brightness, and lighting effects of human faces.

In addition, two types of new global features are proposed.

- *Hue composition feature* fits photos with color composition schemes.

- *Scene composition features* capture the spatial structures of photos from semantic lines.

These new methods and features are introduced in Section 3-5, which emphasize on *dark channel feature*, *hue composition feature*, and *human based features*, since they lead to the best performance in most categories. Through extensive experiments on a large and diverse benchmark database, the effectiveness of different subject area extraction methods and different features on different photo categories are summarized in Table 1. These features are combined by a SVM trained on each of the categories separately. Experimental comparisons show that our proposed new features significantly outperform existing features. To the best of our knowledge, it is the first systematic study of photo quality features on different photo categories.

2. Related Work

Existing methods of assessing photo quality from the aesthetic point of view can be generally classified into using global features and using regional features. Tong *et al.* [18] used boosting to combine global low-level features for the classification of professional and amateurish photos. However, these features were not specially designed for photo quality assessment. To better mimic human aesthetical perception, Ke *et al.* [11] designed a set of high-level semantic features based on rules of thumb of photography. They measured the global distributions of edges, blurriness, hue, and brightness.

Some approaches employed regional features by detecting subject areas, since human beings percept subject areas differently from the background. Datta *et al.* [5] divided a photo into 3×3 blocks and assumed that the central block
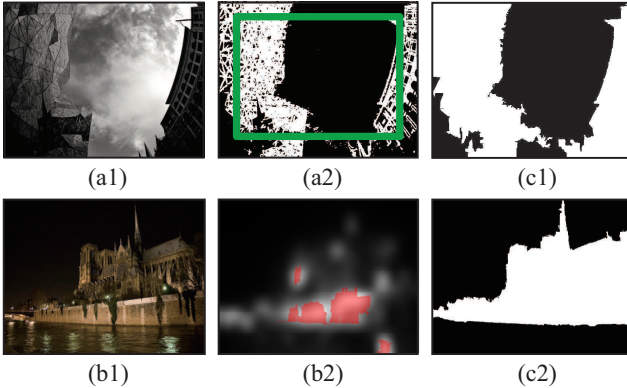
| (a1) | (a2) | (c1) |
| (b1) | (b2) | (c2) |

Figure 3. (a1) and (b1) are input photos. (a2) is the subject area (green rectangle) extracted by the method in [12]. The green rectangle cannot accurately represent the subject area. (b2) saliency map with the subject area (red regions) extracted by the method in [20]. Because of the very high brightness in the red regions, other subject area is ignored. (c1) and (c2) are the subject areas (white regions) extracted by our clarity based region detection method described in Section 4.1.

is the subject area. Luo *et al.* [12] assumed that in a high quality photo the subject area has a higher clarity than the background. Therefore, clarity based criterions were used to detect the subject area, which was fitted by a rectangle. Visual features of clarity contrast, lighting contrast, and geometry composition extracted from the subject areas and the background were used as regional features. Although it worked well on some types of photos, such as "animal", "plant", and "static", it might fail on the photos of "architecture" and "landscape" whose subject areas and background both have high clarity. Also a rectangle is not an accurate representation of the subject area and may decrease the performance. Wong *et al.* [20] and Nishiyama *et al.* [14] used 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 examples in Figure 3.

3. 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 solution and to invoke certain feeling among the viewers of their artworks. They also try to arrange objects 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.

3.1. Hue Composition Feature

Proper arrangement of colors engages viewers and creates inner sense of order and balance. Major color templates [13, 17] can be classified as *subordination* and *coordination*. Subordination requires the photographer 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° color scheme and the Complementary color scheme, which leads to aesthetically pleasing images. With *coordination*, the color composition is created with 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 [3]. 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.

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 ($\alpha_1, s_1, \alpha_2, s_2$), which includes average hue and saturation centers, are extracted from the two sectors. Based on the
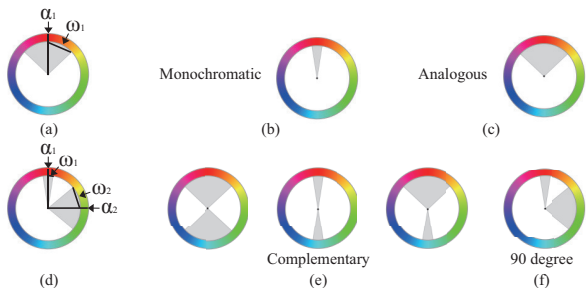
Figure 4. Harmonic templates on the hue wheel used in [3]. An image is considered as harmonic if most of its hue fall within the gray sectors(s) on the template. The shapes of templates are fixed. Templates may be rotated by an arbitrary angle. The templates correspond to different color schemes.

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. 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, 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.

3.2. Scene Composition Feature

High quality photos show well-arranged spatial composition to hold attention of the viewer. Long continuous lines often bear semantic meanings, such as the horizon and the surface of water, in those photos. They can be used to compute scene composition features. For example, the location of the horizon in outdoor photos was used by Bhattacharya *et al.* [1] to assess the visual balance. We characterize scene composition by analyzing the locations and orientations of semantic lines. The prominent lines in photos are extracted by the Hough transform and are classified into horizontal lines and vertical 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.
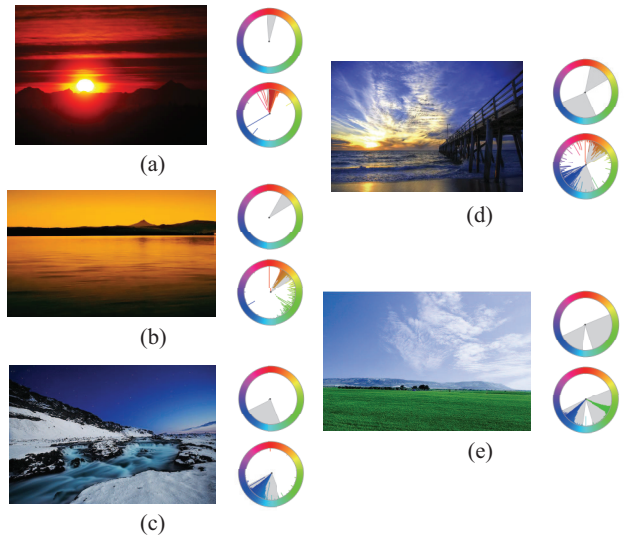

Figure 5. (a),(b),(c): Mixture components for images best fitted with single sector templates. Color wheels on top right side show the mixture components. The center and width of each gray sector are set to mean and standard deviation of each mixture component. Color wheels on down right side show hue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

4. Subject Area Extraction Methods

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 attracted by human faces. In outdoor photography, architectures, mountains, and trees are often the main subjects.

We propose a clarity based method to find clear regions in low depth of field images, which take the majority of high-quality photographs in the categories of "animal", "plant", and "static". We adopt a layout based method [9] 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 human detector and face detector to locate faces.

4.1. Clarity based region detection

A clarity based subject area detection method was proposed in [12]. Since it used a rectangle to represent the subject area and fitted it to pixels with high clarity, the detection results were not accurate. We improve the accuracy by oversegmentation. We first obtain a mask U_0 of the clear area using a method proposed in [12], which labels each pixel as clear or blur. The mask is improved by an iterative
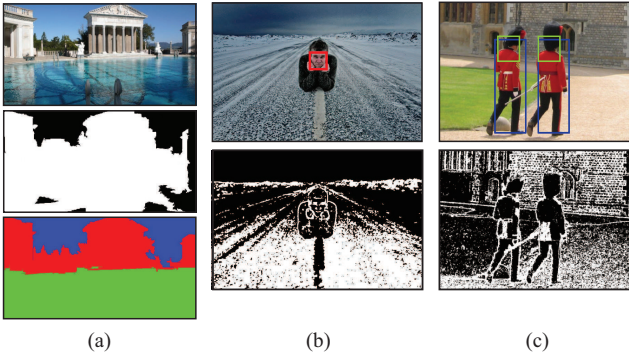
Figure 6. (a): From top downwards: The input photo; result of clarity based detector (white region); result of layout based detector (red region). (b),(c): First row: face and human detection result. Second row: clarity based detection results.

procedure. 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 segmented into super-pixels [15]. A super-pixel is labeled as clear if more than half of its pixels are labeled as clear. The comparison of the method in [12] and ours can be found in Figure 3.

4.2. Layout based region detection

Hoiem *et al.* [9] proposed a method to recover the surface layout from an outdoor image. The scene is segmented into sky regions, ground regions, and vertical standing objects as shown in Figure 6. We take vertical standing objects as subject areas.

4.3. Human based region detection

We employ face detection [21] to extract faces from human photos. For images where face detection fails, we use human detection [4] to roughly estimate the locations of faces. See examples in Figure 6.

5. Regional Features

We have developed new regional features to work together with our proposed subject area detectors. We propose a new dark channel feature to measure both the clarity and the colorfulness of the subject areas. We also specially design a set of features for "human" photos to measure clarity, brightness, and lighting effects of faces. New features are proposed to measure the complexities of the subject areas and the background.

5.1. Dark Channel Feature

Dark channel was introduced by He *et al.* [7, 8] for haze removal. The dark channel of an image I is defined as:

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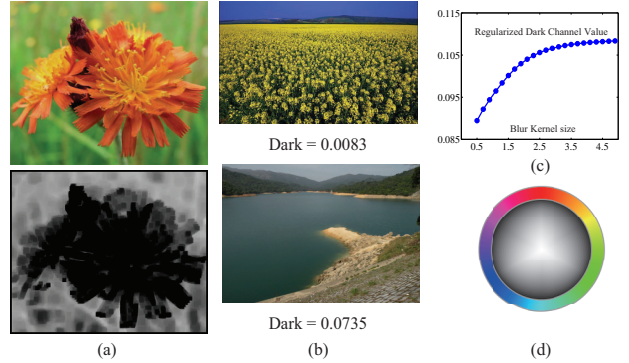

Figure 7. (a) A close-up on plant and its dark channel. (b) Landscape photographs with different color composition. (c) Average dark channel value of input photo from (a) blurred by Gaussian kernel. (d) For each point on the circle: its hue is indicated by the hue wheel, saturation is equal to the radius, and normalized dark channel value is presented by its pixel intensity.

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:

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

with S 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 7 (c) shows that the dark channel value of an image increases with the degree it is blurred. Subject area of low depth of field images show lower dark channel value than the background as shown in Figure 7 (a). For pixels of the same hue value, those with higher saturation gives lower dark channel values (Figure 7 (d)). As shown in Figure 7 (b), low-quality photograph with dull color gives higher average dark channel value. In addition, different hue values gives different dark channel values (Figure 7(d)). So the dark channel feature also incorporates hue composition information.

5.2. Human based Feature

Faces in high-quality human portraits usually possess a reasonable portion of the photo, have high clarity, and show professional employment of lighting. Therefore, we extract the features of the ratio of face areas, the average lighting of faces, the ratio of shadow areas, and the face clarity to assess the quality of human photos.

The ratio of face areas to the image area is computed as feature f_1. The average lighting of faces is computed as f_2.

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,

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

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

$$f_4 = \sum_k \|M_k\| / \sum_k \|X_k\|.$$

5.3. Complexity Feature

Professional photographers tend to keep background composition simple to reduce its distraction. Previous works [11, 12] on complexity features focused on overall distribution of hue and ignored the spatial complexity. We use the segmentation result to measure the spatial complexity. A photo is oversegmented 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 complexity features are defined,

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

6. Experiments

We compare our features with the state-of-the-art features [5, 11, 12, 1] for photo quality assessment on our database . The database consists of photos acquired from the professional photography websites and contributed by amateur photographers. It is divided into seven categories according to photo content (Table 1). They are labeled by ten independent viewers. A photo is classified as high or low quality only if eight out of the ten viewers agree on its assessment. Other photos (40% of labeled photos), on which the viewers have different opinions, are not included in the benchmark database. Features are tested separately or combined with a linear SVM. For each category, we randomly sample half of the high- and low- quality images

as the training set and keep the other half as the test set. The classifiers for different categories are trained separately. The random partition repeats 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 1: proposed regional features; proposed global features; selected previous regional features and selected previous global features. For each 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.

All tested features show different performance for photos with different contents. Generally speaking, in the categories of "animal", "plant", and "static", the subject areas of high-quality photos often exhibit strong contrast with background and can be well detected. Therefore regional features are very 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 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. Previous features 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. Combining different types of features can improve the performance.

Our proposed features significantly outperform the existing features in general. 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 previous clarity features including "clarity contrast"[12] and "blur"[11]. It also outperforms the "color combination" feature[12], which is a color composition measure. Our complexity feature achieves the best performance in the category of "static" and its performance is close to the best in the category of "animal". The high-quality photos in both categories usually have high complexity in subject areas and low complexity in the background. Our complexity features outperform previous complexity features such as "simplicity"[12] and "hue count"[11]. Our proposed face features are very effective for "human" photos and enhanced the best performance (0.78) got by previous features to 0.95.

The hue composition feature is very effective to measure color composition quality. It achieves the best performance on "static" and "landscape" and its performance is close to the best on "plant", "architecture", and "night". It outperforms previous "color combination" feature [12] in all categories except for "animal". Our scene composition feature has the best performance on "night". It outperforms previ-

Category	Animal	Plant	Static	Architecture	Landscape	Human	Night	Overall
Number of high quality photos	947	594	531	595	820	678	352	4517
Number of low quality photos	2224	1803	2004	1290	1947	2536	1352	13156

		Animal	Plant	Static	Architecture	Landscape	Human	Night	Overall
Regional features	Proposed regional features								
	Dark Channel	**_0.8393_**	0.7858	**_0.8335_**	**_0.8869_**	**0.8575**	0.7987	0.7062	0.8189
	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	0.8581	0.9105	0.8667	0.8926	0.8821	0.9599	0.8214	0.8889
	Previous best performing regional features								
	Clarity Contrast [12]	**0.8074**	0.7439	0.7309	0.5348	0.5379	0.6667	0.6297	0.6738
	Lighting [12]	0.7551	0.7752	0.7430	0.6460	0.6226	0.7612	0.5311	0.7032
	Geometry Composition [12]	0.7425	0.7308	0.5920	0.5806	0.4939	0.6828	0.6075	0.6393
	Simplicity [12]	0.6478	0.7450	0.7849	0.5582	0.6918	0.7752	0.4954	0.6865
	Color Combination [12]	**0.8052**	0.7846	0.7513	0.7194	0.7280	0.6513	0.5873	0.7244
	Central Saturation [5]	0.6844	0.6615	0.6771	0.7208	0.7641	0.6707	0.5974	0.6857
	Combined	0.8161	0.8238	0.8174	0.7386	0.7753	0.7794	0.6421	0.7792
Global features	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
	Previous best performing global features								
	Blur [11]	0.7566	0.7963	0.7662	0.7981	0.7785	0.7381	0.6665	0.7592
	Brightness [11]	0.6993	0.7337	0.6976	**0.8138**	0.7848	0.7801	**0.7244**	0.7464
	Hue Count [11]	0.6260	0.6920	0.5511	0.7082	0.5964	0.7027	0.5537	0.6353
	Visual balance [1]	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		0.8712	0.9147	0.8890	0.9004	0.9273	0.9631	0.8309	0.9044
Previous features combined		0.8202	0.8762	0.8230	0.8647	0.8412	0.8915	0.7343	0.8409
All features combined		0.8937	0.9182	0.9069	0.9275	0.9468	0.9740	0.8463	0.9209

Table 1. 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.

ous relevant features such as "geometry composition"[12] and "visual balance"[1] in most categories.

Previous features show mixed performance across categories. For example, the regional features proposed in [12] work reasonably well on "animal", "plant", and "static", where their clarity-based subject area detection generally works. However, their performance greatly decrease on "architecture", "landscape", "human", and "night".

In Figure 8, we show ROC curves of combining regional features proposed in [12], combining global features proposed in [11], combined all the previous features mentioned in Table 1 and combining our proposed features. It shows that our features outperform previous features. We also show that combining all the features together leads to the best performance in Table 1.

7. Conclusions and Discussions

In this paper, we propose content based photo quality assessment together with a set of new subject area detection methods, new global and 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 features in different ways and train different classifiers for different photo categories separately. Our proposed new features significantly outperform existing features. In this work we focus on feature extraction and assume that the category of a photo is known. In some cases, such information is available, e.g. some websites already categorize their photos, but not in all the cases. There is a
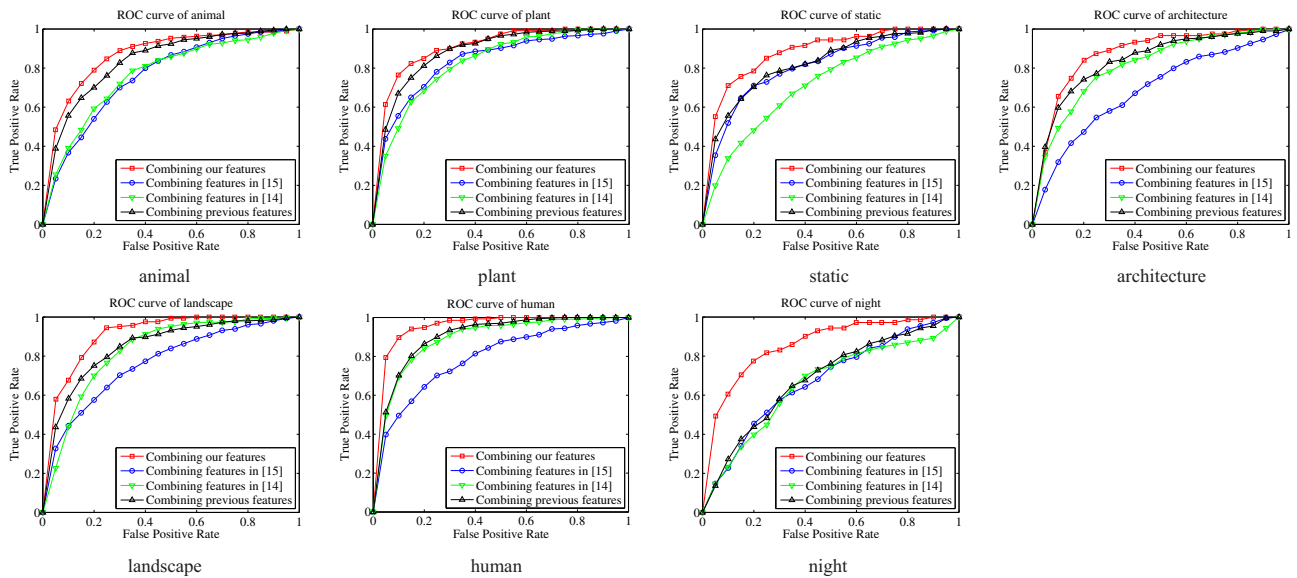
Figure 8. Photo quality assessment performance comparisons on seven categories of photos.

huge literature on automatic image categorization based on visual and textual features. Image categorization has been greatly advanced in the past years and the problem can be solved reasonable well especially when more textual information is available. We will leave the integration of automatic photo categorization and quality assessment as the future work.

References

[1] S. Bhattacharya, R. Sukthankar, and M. Shah. A Framework for Photo-Quality Assessment and Enhancement based on Visual Aesthetics. In *Proc. ACM MM*, 2010. 4, 6, 7

[2] J. Carucci. *Capturing the Night with Your Camera: How to Take Great Photographs After Dark*. Amphoto, 1995. 1

[3] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. In *Proc. ACM SIGGRAPH*, 2006. 3, 4

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, 2005. 5

[5] R. Datta, D. Joshi, J. Li, and J. Wang. Studying aesthetics in photographic images using a computational approach. In *Proc. ECCV*, 2006. 1, 3, 6, 7

[6] C. Grey. *Master Lighting Guide for Portrait Photographers*. Amherst Media, Inc., 2004. 1

[7] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *Proc. CVPR*, 2009. 5

[8] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Trans. on PAMI*, 2010. 5

[9] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *Int'l Journal of Computer Vision*, 2007. 4, 5

[10] X. Jin, M. Zhao, X. Chen, Q. Zhao, and S. Zhu. Learning Artistic Lighting Template from Portrait Photographs. In *Proc. ECCV*, 2010. 1

[11] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Proc. CVPR*, 2006. 1, 2, 6, 7

[12] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *Proc. ECCV*, 2008. 1, 3, 4, 5, 6, 7

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

[14] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *Proc. ACM MM*, 2009. 3

[15] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proc. ICCV*, 2003. 5

[16] A. Savakis and S. Etz. Method for automatic assessment of emphasis and appeal in consumer images, Dec. 30 2003. US Patent 6,671,405. 1

[17] M. Tokumaru, N. Muranaka, and S. Imanishi. Color design support system considering color harmony. In *Proc. IEEE International Conference on Fuzzy Systems*, 2002. 3

[18] H. Tong, M. Li, H. Zhang, J. He, and C. Zhang. Classification of digital photos taken by photographers or home users. In *Proc. PCM*, 2004. 1, 2

[19] L. White. *Infrared Photography Handbook*. Amherst Media, Inc., 1995. 1

[20] L. Wong and K. Low. Saliency-enhanced image aesthetics class prediction. In *Proc. ICIP*, 2009. 1, 3

[21] R. Xiao, H. Zhu, H. Sun, and X. Tang. Dynamic cascades for face detection. In *Proc. ICCV*, 2007. 5