Vol.2011-CVIM-177 No.33
2011/5/20

Aesthetic Quality Classification of Photographs

Masashi Nishiyama[†1,†2] and Yoichi Sato[†1]

The aesthetic quality of a photograph is a significant cue for inferring the level of appreciation that arises when people view digital content on various types of computer systems. It would be useful to classify the aesthetic quality in photographs to more easily edit large photograph collections that are available because of the widespread use of digital cameras and the Internet. The research community is currently tackling this challenging problem of aesthetic quality classification, which automatically assesses whether a photograph generates aesthetic appreciation. This paper focuses on developing techniques for aesthetic quality classification of photographs. In particular, we discuss detecting of multiple-subject regions and assessing color harmony in order to enhance the performance of aesthetic quality classification. These techniques play an important role in determining aesthetic quality. We statistically build these quality classifiers using large photograph databases collected on websites where users manually provide quality labels to photographs. Furthermore, we demonstrate that our aesthetic quality classifier emulates how people edit large photograph collections.

1. Introduction

The demand for human-centered computer systems is increasing due to the ever-increasing diversity of user needs. To achieve the goal of providing fast services suited to different people and environment and enabling these systems to interact more naturally with people, human-centered computer systems have been in a state of continuous development for the past few decades.

To build human-centered computer systems, two types of user information must be automatically recognized: external and internal. External information is recognized from visible human surfaces, *e.g.*, the face, body, and hands. Internal information is inferred from physiological and/or behavioral signs of internal conditions, such as impressions, appreciation, and preferences, and is more difficult to recognize clearly.

Pattern recognition[2] and computer vision[48] are the core technologies that support the recognition of both types of user information in human-centered computer systems. Recently, useful real-world applications that recognize external information have gradually emerged. For instance, significant advances have been made in face recognition[3,21,52], a function that is used in video surveillance, watch-list monitoring, and digital cameras, on the basis of pattern recognition and computer vision techniques. Similarly, major technologies for recognizing external information have been established, and many researchers still continue to develop them. On the other hand, the development of technologies for recognizing internal information, which has only recently begun, has become an active research area in the fields of pattern recognition and computer vision.

One of the most important aspects of recognizing internal information is the aesthetic appreciation that arises when people view digital content on various types of computer systems, *e.g.* mobile phones, TV monitors, PCs, and digital signages. Many people now have various opportunities to view large collections of digital content such as the large number of photographs they have owing to the widespread use of digital cameras and the Internet.

To manipulate such large photo collections on human-centered computer systems, it would be useful to evaluate the aesthetic appreciation these photographs generate. In particular, the aesthetic quality of photographs serves as a significant cue for inferring human appreciation. For instance, we can readily identify photographs with high (or low) aesthetic quality from those shown in Figure 1. Researchers are currently trying to solve this challenging problem of aesthetic quality classification, which automatically assesses whether a photograph will gain people's appreciation.

In this paper, we focus on developing techniques for aesthetic quality classification that distinguish between high- and low-quality photographs. In particular, we studied the detection of multiple-subject regions and the assessment of color harmony to enhance aesthetic quality classification of photographs. We aim to construct an aesthetic quality classifier that represents the opinions of a majority of people. Furthermore, we demonstrate that our aesthetic quality classifiers

†1 The University of Tokyo
†2 Toshiba Corporation

Vol.2011-CVIM-177 No.33
2011/5/20

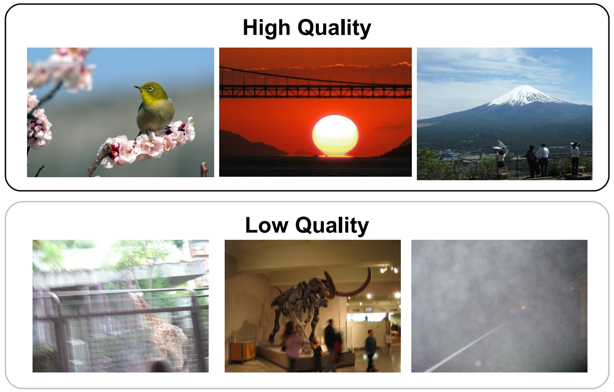


Fig. 1 Examples of high- and low-quality photographs. We present an aesthetic quality classification method that automatically distinguishes high- or low- quality photographs.

emulate how people organize large photo collections. Our results indicate that the quality classifiers above play an important role in determining the aesthetic quality of photographs.

1.1 Definition of aesthetic quality

We define high- or low-aesthetic-quality photograph by seeing whether the photograph generates a lot or little aesthetic appreciation in a majority of people. Photographs with both aesthetic qualities can be collected from databases available on websites, *e.g.* DPChallenge[10], and Photo.net[38]. Various users often manually provide aesthetic quality scores to photographs in these large photo databases. We have used the average value of a quality score for each photograph to determine its aesthetic quality label: high and low average values determined high- and low-quality photographs. Note that we do not consider highly artistic photographs to be any different from other photographs in terms of aesthetic quality.

1.2 Outline

First of all, Section 2 describes an aesthetic quality classifier with multiple subjects. First, we detect multiple-subject regions that contain attention grabbing salient pixels in a photograph. We also detect the background region. From these detected regions, we extract features based on the rule of thumb for photography. We combine these features to show the relationship among multiple-subject regions and the background region. The combined features are then used to classify the aesthetic quality of a photograph. The prior works on this issue considered only a single-subject region. The advantage of our method is that it can deal with photographs containing multiple-subjects, for example, a flower among leaves or individual buildings in a landscape. Our method can extract more detailed features from multiple-subject regions than from a single-subject region.

Next, Section 3 describes the aesthetic quality classifier based on color harmony assessment of photographs. Color harmony plays an important role in various aspects determining the perceived quality of a photograph; furthermore, color harmony should be taken into account to enhance the ability of automatic aesthetic quality classification. In this section, we tackle the challenging problem of evaluating the color harmony of photographs, particularly for aesthetic quality classification. A key observation is that a photograph can be seen as a collection of local regions whose color variation is relatively simple. This has led us to develop a method for assessing aesthetic quality of a photograph on the basis of a photograph's color harmony.

In Section 4, we describe an application using aesthetic quality classification. We design a technique for automatically cropping a photograph using a quality classifier that assesses whether the cropped region is agreeable to users. We first trim the original image and then decide on the candidates for cropping. We find the cropped region with the highest quality score by applying the quality classifier to the candidates. Section 5 concludes the paper and describes possible future directions.

2. Aesthetic quality classifier with multiple subjects

To enhance the performance of an aesthetic quality classifier, features extracted from a photo play an important role. The existing methods of aesthetic quality classification extract features from a whole photo, or a single-subject region detected in a photo. However, such features often perform poorly when classifying aesthetic quality because often there is more than one subject in a photo.

In this section, we present an aesthetic quality classification method that uses the features extracted from multiple-subject regions. Detecting multiple regions can be useful to enhance the recognition performance of aesthetic quality classification. Our method detects multiple-subject regions and a background region from the whole photo and extracts the features based on the rule of thumb for photography. After combining these features, we classify the aesthetic quality labels. We train an aesthetic quality classifier using photos with manually labeled aesthetic quality. Our experiments demonstrate that our method with multiple subjects outperforms the existing methods in large photo datasets.

2.1 Multiple-subject regions

We study the detection of multiple-subject regions for aesthetic quality classification. Luo *et al.*[27] have proposed a method for determining a subject region using the amount of blur at each pixel. This section also assumes that a background region is blurred to emphasize the subject region. This assumption is valid for a clear subject such as the dragonfly in Figure 2 (a), but is invalid for a vague subject such as the landscape in (d). The amount of blur sometimes makes it very difficult to separate a subject region from a background region, *e.g.*, Figure 2 (e). These regions are unsuitable for feature extraction. Furthermore, there is not always only one subject. Loui et al.[24] uses a saliency map instead of the amount of blur. However, they utilize only a single subject region to extract features.

We present a new quality classifier that is based-on the rules of thumb for photography exploiting multiple subjects to improve the performance of the quality classification. Figures 2 (c) and (f) show examples of multiple-subject regions. The subjects in these figures can capture the contents of the photos in more detail than the one in (b) and (e), *e.g.*, a dragonfly, leaves, and buildings. The compositions of multiple subjects could drastically influence the quality. Thus, we believe that the features extracted from multiple subjects are of a stronger quality classification than the ones from a single subject.

We aim to construct an aesthetic quality classifier that represents a consensus from the majority of people. This quality classifier is trained from large photo databases where various people insert quality scores to various photos. These databases consist of photo collections that are available on the Internet (DPChallenge[10] & Photo.net[38]). In the DPChallenge and Photo.net, various

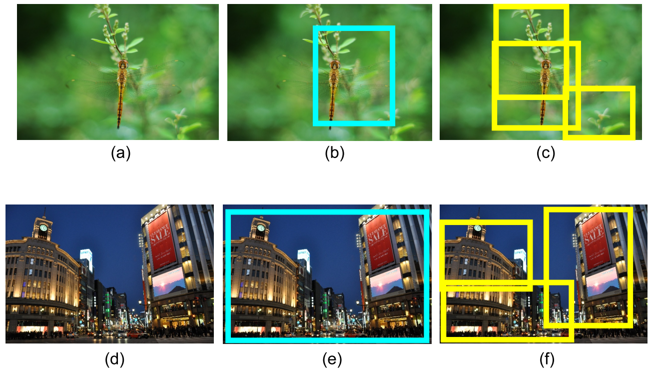


Fig. 2 Examples of subject and background regions. Figures (a) and (d) are original images. The blue boxed regions in Figures (b) and (e) are the subjects detected by Luo's work. The yellow ones in Figures. (c) and (f) are the subjects detected by our method. A background region is set to a region other than that of the subjects. Our method aims to extract the features for quality classification from the compositions of multiple subjects.

people insert their scores to various photos. Note that we do not consider highly artistic photos any differently than other photos in terms of quality.

2.2 Procedure of classifier with multiple subjects

Our quality classifier is composed of the four steps shown in Figure 3: (1) given a photo we detect multiple-subject regions using a saliency map, (2) we next extract the features representing the basic techniques for photography from each region, (3) we then compute a posterior probability that the feature is matched as high quality, and (4) finally, we determine the actual quality using the combined probabilities. Each step is described in detail below.

2.2.1 Detecting multiple-subject regions

Our method detects multiple-subject regions using the low-level saliency map proposed by Itti *et al.*[16]. We use the k-Means clustering method[11] against a saliency map to acquire the subject regions. In the clustering, we use a vector as

$$\mathbf{v} = (n(x), n(y), n(a_{x,y})) \ , \qquad (1)$$

where $a_{x,y}$ is a saliency value at coordinate (x, y), and $n(\ )$ is a function that normalizes the range of each value. From k clusters divided by the k-Means method, the subject regions are assigned to $m(< k)$ clusters whose average saliencies are

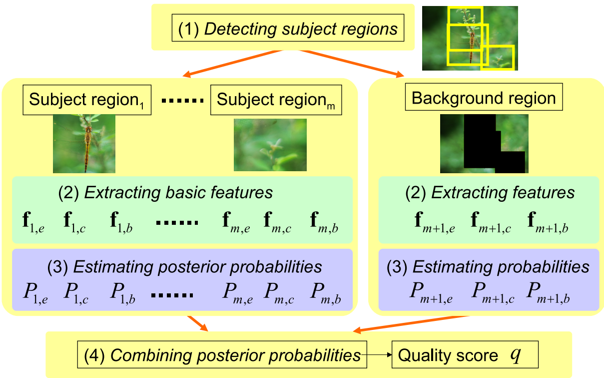


Fig. 3 Our quality classifier is composed of four steps: (1) detecting multiple-subject regions and a background region, (2) extracting features from the regions, (3) estimating a posterior probability against each feature and (4) determining the quality score using the combined probabilities.

higher. A subject region R_i is set by fitting a bounding rectangle against the (x, y) coordinates of each cluster. A background region R_{m+1} is set to a region other than the subject regions. Then, the k, m parameters are empirically determined.

2.2.2 Extracting features

We extract the features representing the basic techniques for photography for each region $R_i (i = 1, \ldots, m+1)$, *e.g.*, no camera shakes and adequate exposure. We design the edge, color, and blur features by referring to the rules of thumb described in photography handbooks.

- Edge $\mathbf{f}_{i,e}$ is a histogram of 256 bins generated from the vertical and horizontal Sobel filter outputs.
- Color $\mathbf{f}_{i,c}$ is a histogram of 512 bins generated from $8\times8\times8$ segmented values in the RGB color space.
- Blur $\mathbf{f}_{i,b}$ is a 1024 dimensional vector of amplitude values calculated by using the discrete Fourier transform and a 32×32 resampling in the Frequency domain.

2.2.3 Estimating posterior probabilities

We estimate the posterior probability representing a rate in which a feature is matched as high quality. The probability is computed from an output that is estimated using the Support Vector Machines (SVM)[50] using training samples with high- or low aesthetic quality labels. We use photos with higher aesthetic quality scores given by people for the photo collections as high aesthetic quality label training samples, and vice-versa.

Given the feature $\mathbf{f}_{i,j} (j = e, c, b)$ of a subject region $R_i (i = 1, \ldots, m)$, the output $s_{i,j}$ is defined as

$$s_{i,j} = SVM_{subject,j}(\mathbf{f}_{i,j}) \ . \tag{2}$$

The sign of output represents the label of the high/low quality for each feature. An output for a background region is also defined as

$$s_{m+1,j} = SVM_{background,j}(\mathbf{f}_{m+1,j}) \ . \tag{3}$$

Since the output is a normalized distance from a separating hyperplane, the posterior probability $P_{i,j}$ is calculated by fitting the output $s_{i,j}$ to the Sigmoid function as

$$P_{i,j}(high|s_{i,j}) = \frac{1}{1 + \exp(As_{i,j} + B)} \ , \tag{4}$$

where A, B are constants determined from training samples by using a previously reported technique[22),40)]. The probability $P_{i,j}$ of a high-quality photo is close to 1 , and the one of a low-quality photo is close to 0. See the references for more details.

2.2.4 Combining posterior probabilities

We determine the quality score by combining the posterior probabilities. For instance, Meynet *et al.*[30] have combined the probabilities by using sum, product, or median of them. The use of sum, product, or median often performs poorly in classifying aesthetic quality.

To represent the relationship between multiple-subject and background regions, we use a combined feature consisting of the posterior probabilities and the product of a pair of probabilities as

$$\begin{aligned} \mathbf{f}_{all} = (&P_{1,e},\ P_{1,c},\ P_{1,b},\ \ldots, P_{m+1,e},\ P_{m+1,c},\ P_{m+1,b}, \\ &P_{1,e} \cdot P_{2,e},\ P_{2,e} \cdot P_{3,e},\ \ldots\ \ P_{m-1,b} \cdot P_{m,b},\ P_{m,b} \cdot P_{m+1,b}) \ . \end{aligned} \tag{5}$$

A quality score is defined as

$$q = SVM_{all}(\mathbf{f}_{all}) \ . \tag{6}$$

2.3 Effectiveness of multiple subjects

We evaluated the performance of our quality classifier on several photo collections (DPChallenge, and Photo.net). Each database has a different tendency in the quality scores given by people. We mixed the databases to create quality diversification. The mixed database consisted of 15,220 ($= 13,420 + 1,800$) photos. From DPChallenge, the top and bottom 10% of quality scores were assigned as the high- and low-quality photos in the reference[18]. From Photo.net, the top and bottom 20% were assigned as the high- and low-quality photos in the reference[7]. After dividing each photo collection in half, one was used for the training samples and the other was used as the test samples.

We compare the performance of the quality classifiers using the following methods for feature extraction.

- **Pixel value**: a feature is extracted by raster-scanning the RGB value after down-sampling to 32×32 pixels.
- **Edge**: a feature is an edge histogram of the Sobel filter outputs $\mathbf{f}_{whole,e}$ extracted from a whole photo.
- **Color**: a feature is a color histogram of RGB values $\mathbf{f}_{whole,c}$ extracted from a whole photo.
- **Blur**: a feature is amplitude values calculated by using the discrete Fourier transform $\mathbf{f}_{whole,b}$ computed from a whole photo.
- **Linking**: a feature is extracted by simply concatenating as $\mathbf{f}_{linking} = (\ \mathbf{f}_{1,e}{}^{\mathrm{T}},\ \mathbf{f}_{1,c}{}^{\mathrm{T}},\ \mathbf{f}_{1,b}{}^{\mathrm{T}},\ \ldots,\ \mathbf{f}_{m+1,e}{}^{\mathrm{T}},\ \mathbf{f}_{m+1,c}{}^{\mathrm{T}},\ \mathbf{f}_{m+1,b}{}^{\mathrm{T}}\ )$.
- **BoK**: a feature is extracted by using *Bags-of-keypoints*[6] for generic object recognition and image retrieval.
- **Whole image**: a feature is extracted from an entire photo such as that in the references[7),18)]. Our features $\mathbf{f}_{whole,e}, \mathbf{f}_{whole,c}, \mathbf{f}_{whole,b}$ were extracted from an entire photo instead of the features described in the references[7),18)]. Posterior probabilities were computed from the features and were combined using our approach.
- **Single subject**: a feature is extracted from a single-subject region and background region such as in Luo et al.[27]. We applied the same method for detecting a subject region as that described in the reference[27], but used only our features $\mathbf{f}_{i,j}$. The posterior probabilities were computed from the features and were combined using our approach.
- **Multiple subjects (simple)**: a feature is extracted from multiple-subject regions and background region such as our method. But, we simply combine the posterior probabilities of blur, color, and edge as
$$\mathbf{f}_{all'} = (P_{1,e},\ P_{1,c},\ P_{1,b},\ \ldots, P_{m+1,e},\ P_{m+1,c},\ P_{m+1,b})\ . \qquad (7)$$
- **Multiple subjects**: our feature $\mathbf{f}_{all}$ (see Sec. 2.2.4). We detect multiple-subject regions and a background region from a whole photo and extract blur, color, and edge features representing the rule of thumb of photography. After combining those features, we perform aesthetic quality classification.

A SVM was applied to each extracted feature. A SVM without using kernels was used since it has approximately the same recognition performance as a SVM with non-linear kernels (polynomial, sigmoid, rbf). We obtained a better performance when using $k = 12, m = 5$ in this experiment.

Figure 4 shows the classification performance as a recognition rate: the probability that the quality estimated using each classifier is matched to the correct quality. The 'Pixel value' performance is nearly equal to a random guess. The 'Color', 'Edge', and 'Blur' obtain improved performance compared with 'Pixel value'. The 'Whole image', 'Single subject', and 'Multiple subjects' performances are superior to the 'Linking' one. Our 'Multiple subjects' quality classifier is superior to the 'Single subject' one. Our classifier achieved about a 71% accuracy for this difficult task. More experimental results were demonstrated in the reference[35].

3. Aesthetic quality classifier based on color harmony assessment of photographs

The perceived quality of a photograph depends on various aspects, e.g. color, composition, lighting, and subjects appearing in the photograph. In this section, we focus on color harmony assessment because the colors of photos significantly affect their perceived quality.

The existing models[15),17),29),32)] proposed in the field of color science consider the color harmony of a simple color pattern such as the ones shown in Figure 5 (a). These models can be used to assess colors of many man-made objects such as cars, clothes, and websites. Unfortunately, however, they perform poorly in assessing

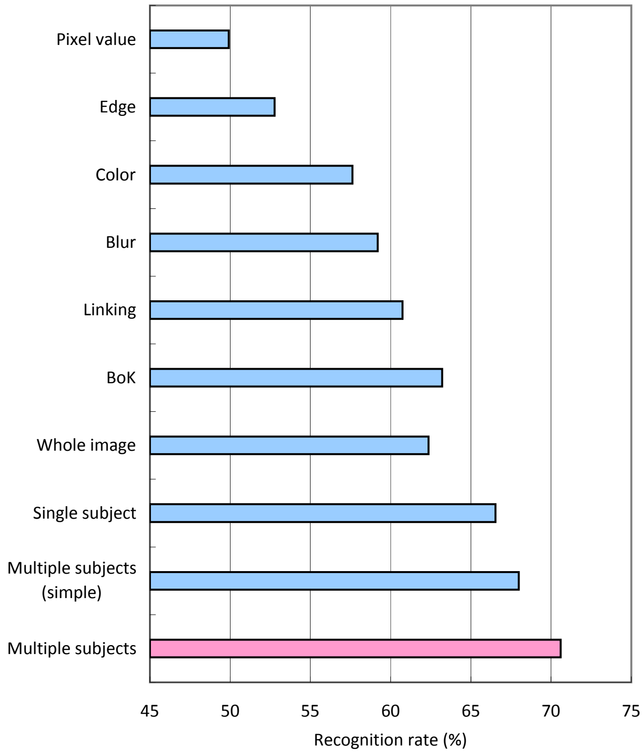


Fig. 4 Comparison of aesthetic quality classification performance for large photo collections. Our 'Multiple subjects' quality classifier is superior to 'Single subject' classifiers.

Fig. 5 Complexity of color harmony of photos. We wanted to bridge the gap between simple color patterns and actual colors appearing in photos and develop a method for automatically assessing the color harmony of photos.

the color harmony of a photo such as shown in Figure 5 (b). This is because the color distribution of a photo is often significantly more complex in both color and spatial domains than those considered in the existing color harmony models. Thus, the existing models cannot be used for assessing the color harmony of photos with complex color distributions.

The color harmony of photos has been largely ignored in the existing methods of aesthetic quality classification[7),18),24),27),35)]. In those methods, a global color histogram computed from an entire image is employed as one of the image features for evaluating the aesthetic quality of the image. Only recently has the research community started to look at the color harmony of photos, e.g. inferring affective words from photos of a limited category[45)].

This section addresses the challenging problem of assessing the color harmony of photos. We propose a method for automatically evaluating it to enhance the performance of aesthetic quality classification. Our key observation is that a photo can be seen as a collection of local regions whose color variations are relatively simple. Our preliminary experiments show that the sum of color harmony scores computed from the local regions of a photograph is closely related to its aesthetic quality. Based on this observation, we developed a method for color harmony assessment of photos. In our method, a color harmony model is applied to each local region of an image to evaluate distributions of relative values of hue, lightness, and chroma to the dominant color in the region, and then the result is integrated to describe the entire image in the *bag-of-features* framework. Our experimental results demonstrate that our aesthetic quality classification method that explicitly takes into account the color harmony of a photo outperforms the existing methods.

Our present work is closely related to two topics: aesthetic quality classification and color harmony models. This section summarizes prior work in each of them.

Aesthetic quality classification:

Some researchers have proposed aesthetic quality classification methods[7)–9),18),24),27),43)] and the use of the aesthetic quality for applications such as image cropping and recomposition[1),23),35)]. In contrast to the conventional image quality measures[51),53)], aesthetic image quality measures are advantageous because they are more closely correlated with our impressions of images.

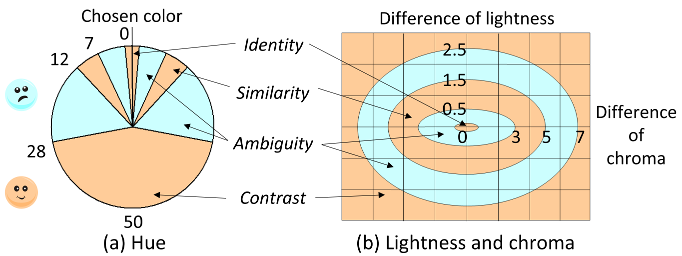


Fig. 6 The Moon-Spencer model for color harmony assessment. The model assesses the relationship between two colors by comparing a chosen color value with a certain color value. When the relative value between them does not appear in the region of 'Ambiguity', the model tells us that the two colors are in harmony.

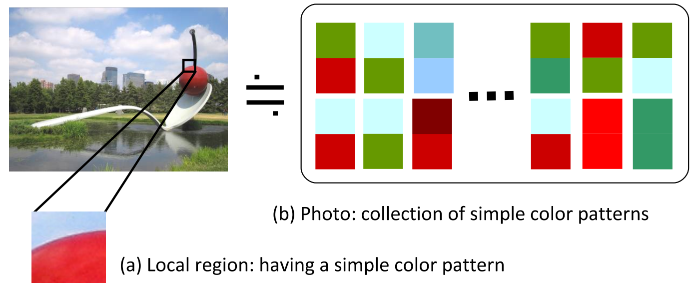


Fig. 7 Assumptions on assessing the color harmony of photos. We observed that (a) a local region is regarded as a simple color pattern and assume that (b) a photo consists of a collection of simple color patterns. We developed a method for assessing the color harmony of a photo by using a collection of local regions.

The existing aesthetic quality classification methods distinguish high- and low-quality photos using a classifier trained from a large collection of image samples, typically collected from the Internet, with manually provided aesthetic quality scores. For better classification accuracy, it is essential to choose good image features. In the previous methods, each photo is described with empirically chosen features such as the rule of thirds, color histogram, and the size and position of a salient region in the image. Here, we argue that simple color features like color histograms are insufficient, and it is essential to consider color harmony when evaluating aesthetic quality properly.

Color harmony models:

There have been two major models proposed in the field of color science for evaluating color harmony: the Moon-Spencer model[32] and the Matsuda model[29].

The Moon-Spencer model handles a simple color pattern consisting of two colors such as in Figure 5 (a). This model is based on psychological experiments by evaluating the relationship between a color pattern and its affection when the pattern is shown as a stimulus. The model computes the relative value between two colors in the Munsell color system. As Figure 6 illustrates, when the relative values of hue, chroma, and lightness do not appear in 'Ambiguity', the two colors are considered to be in harmony. The model presents three types of color harmony. 'Contrast' is a certain color being largely different from a chosen color, 'Similarity' is a resembling color, and 'Identity' is the same color.

The Matsuda model was developed for designing clothes based on simple patterns with a few colors. Matsuda presented nine harmonic templates that define ranges where colors are in harmony on the hue circle. This model also uses relative hue values such as the Moon-Spencer model, and it has been used in certain computer graphics and pattern recognition applications. For instance, Cohen-Or *et al.*[4] proposed a method for color transfer while maintaining color harmony by fitting the model to a hue histogram counted in an object region. Existing color transfer methods[14),19),39),41] does not take into consideration the agreeability that a transferred color distribution give people. Tokumaru *et al.* have proposed a color design support system by using the model[49]. Moorthy *et al.*[33] exploited the model to assess the color harmony of videos by comparing the model with hue histograms counted from the whole image.

Unfortunately, we cannot apply those methods of color harmony for the problem of aesthetic quality classification of images. This is because the color distribution of images is significantly more complex than a simple combination of few color patches. As shown later in Figure 12, the use of the Moon-Spencer and the Matsuda models for a whole photo is not suitable for color harmony assessment of photos.

3.1 Assessing color harmony of photos

As shown in Figure 5, the distribution of colors in a photo is more complicated than simple color patterns in both the color and spatial domains. Therefore, there would be a large gap between the previously proposed color harmony models for simple color patterns and that for a photo.

To cope with the difficulty, we exploit the fact that, when we observe a local region of a photo, the distribution of colors within the region is relatively simple, as illustrated in Figure 7 (a). Therefore, we could consider each region to have a simple color pattern and evaluate the color harmony of the region by using the color harmony models for simple color patterns. Because a photo is a collection of all the local regions, the color harmony of the entire photo could somehow be computed from the color harmony scores of the local regions.

Our preliminary experiments demonstrate that we can represent a photo as a collection of local regions with color patterns and use the collection to assess the whole photo. The sum of the color harmony scores computed from local regions of a photo by using the Moon-Spencer model is positively correlated with aesthetic quality of the photo. In addition, a photo with high (low) aesthetic quality often contains a large number of local color patterns with high (low) color harmony scores.

Based on these observations, we developed a method for assessing the color harmony of photos. Our method assumes that a photo is described by a collection of simple color patterns as shown in Figure 7 (b), and classifies the aesthetic quality of the photo on the basis of the frequency of appearance, *i.e.* the histogram of those color patterns (Figure 8). Specifically, local regions are sampled from a photo, and each region is described by a feature based on the color harmony models for simple color patterns. Then, these features are quantized, and the photo is represented by the histogram of those quantized features. Finally, the aesthetic quality of the photo is determined by a classifier trained with labeled photos.

Our method has analogical aspects to the technique called *bags-of-features*[6),12),20),36),46),54)] for generic object recognition and image retrieval. One of the difficulties in these research areas is that the appearance of an object drastically varies depending on imaging conditions such as camera viewpoint, object pose, and illumination. To alleviate this difficulty, the technique represents an image as a set of local features insensitive to imaging conditions. Thus, in this sense, our method for assessing the color harmony of photos can be termed

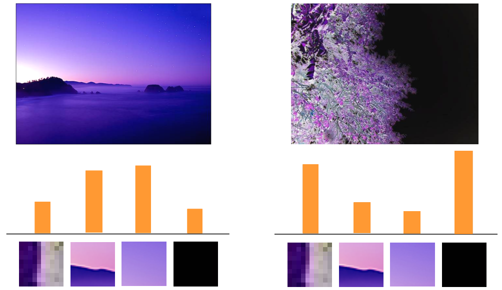


Fig. 8 Histogram features representing color harmony. Our method prepares representative local regions as prior knowledge for assessing the color harmony of photos and counts the frequency of appearances of local regions that are similar to the representative regions in a photo. The histogram features tell us the difference in the color harmony between photos.

bags-of-color-patterns. The descriptor of a local region computed from the color harmony models is a counterpart of the descriptor of local features such as SIFT for generic object recognition.

3.2 Bags-of-color-patterns

The proposed method (1) samples local regions of a photo, (2) describes each local region by features based on color harmony models for simple color patterns, (3) quantizes these features, and (4) represents the photo as a histogram of quantized features. Finally, it uses an SVM classifier[50)] trained by using sample photos with aesthetic qualities labeled by various people who assessed the color harmony of the whole photo. The steps are described below.

3.2.1 Sampling local regions

Our method uses a grid-sampling technique to extract a set of local regions. The operators used in existing methods for object recognition, e.g. Difference of Gaussian[25)], Maximally Stable Extremal Regions[28)], Harris-Hessian[31)], and Features from Accelerated Segment Test[42)] are not suitable for color harmony assessment because not only the colors around the edges and corners but also those in uniform regions affect the perceived aesthetic quality of a photograph.

The grid-sampling technique extracts local regions with a fixed size from equally spaced positions. We empirically determine the size of the regions and the sampling density to determine the number of the regions cropped from a single photo.

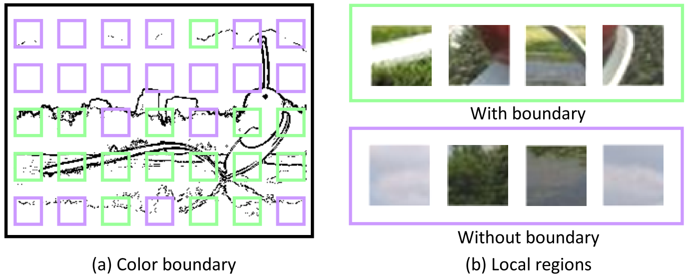


Fig. 9 Sampling local regions. We extract local regions from a whole photo using a grid-sampling technique, and distinguish the regions that contain color boundaries. The pixel value in (a) represents a color boundary computed from a photo shown in Figure 5 (b). We make two sets of local regions with/without color boundaries in (b).

We experimentally confirmed that aesthetic quality classification is not necessarily sensitive to the parameters for grid-sampling.

The Moon-Spencer model shown in Figure 6 tells us that a simple color pattern is in harmony when colors in the pattern are 'Identity', 'Similarity', or 'Contrast'. This means that uniform regions are usually in harmony but regions around edges and corners are in harmony only when the colors within a region are similar or contrasting. In order to incorporate the reason why the color of a local region is in harmony into our method, we distinguish uniform regions from regions around edges and corners, and treat them separately. We divide a whole photo into segments by using mean shift segmentation[5], and detect color boundaries by using discriminant analysis[37], as shown in Figure 9 (a). We split a set of local regions into those with/without color boundaries as illustrated in Figure 9 (b).

3.2.2 Describing local regions

The Moon-Spencer model shows that the color harmony of simple color patterns such as one with two colors can be described by the difference between two colors: the hue, chroma, and lightness values relative to the chosen color. Motivated by the Moon-Spencer model, we find the dominant color of a local region and describe the local region by using the hue, chroma, and lightness values relative to the dominant color. It is worth nothing that the existing methods for aesthetic quality classification use absolute values, *i.e.* the usual pixel values in RGB channels, and not the relative values.

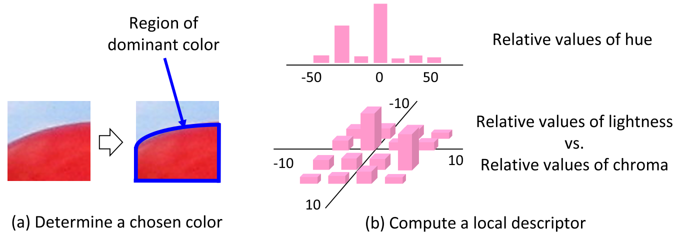


Fig. 10 Local descriptor for assessing color harmony in a local region. To compute relative values, we determine the dominant color in a local region, such as red in (a). Next, we plot the two histograms of (b) by using relative colors with respect to the dominant color.

Specifically, we describe local regions based on the Moon-Spencer model as follows. First, we convert pixel values within a local region from RGB color space to the Munsell color system⋆1. Then, we find the dominant color of the region based on the hue values. We compute the hue values relative to the dominant color by subtracting the hue value of the dominant color from that of each pixel, and obtain the histogram of the relative hue values as illustrated on the top side of Figure 10 (b). Motivated by the analogy with the difference of hue values in the Moon-Spencer model shown in Figure 6 (a), we use this histogram for describing a local region. In addition, we compare the average values of chroma and lightness on the pixels with the dominant color, and subtract these average values from chroma and lightness of each pixel. Thus, we obtain 2D histogram of the relative chroma and lightness values illustrated on the bottom side of Figure 10 (b). This histogram is an analogy to the differences of chroma and lightness in Figure 6 (b). Finally, we concatenate the histogram of relative hue values and the 2D histogram of relative chroma and lightness into a single vector, and use it as the descriptor of a local region.

3.2.3 Quantizing local descriptors

As is often the case with *bags-of-features* for generic object recognition and

---

⋆1 We used a table published by Berns *et al.* of the Rochester Institute of Technology *http://www.cis.rit.edu/mcsl/online/munsell.php*. The table provides values going from the Munsell color system to CIE xyY color space. Note that we calculated values not in the table by using a linear interpolation technique.

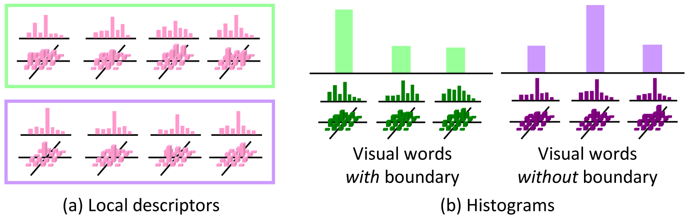


Fig. 11 Quantizing local descriptors. We compute local descriptors of (a) from local regions with/without color boundaries to generate codebooks. We plot the two histograms of (b) by using visual words in the codebooks. We concatenate the histograms to make a feature representing a whole photo.

image retrieval, we also quantize local descriptors by using visual words in codebooks. First, we obtain a large number of local descriptors from training samples as shown in Figure 11 (a). Then, we generate two codebooks for local descriptors with/without color boundaries by using the k-means clustering method[11)].

3.2.4 Representing a whole photo

By using those codebooks, we compute the frequency of appearance, *i.e.* the histograms of quantized local descriptors with/without color boundaries as illustrated in Figure 11 (b). We concatenate the histograms for local regions with/without color boundaries into a single vector. Unfortunately, however, the vector cannot represent the distribution of colors in spatial domain because the descriptors of local regions have no information about their spatial locations.

Accordingly, in order to incorporate spatial information into the descriptor for a whole photo, we divide a photo into rectangular segments and obtain a vector (histogram) from each segment. Finally, we concatenate those vectors into a single vector and use it for representing a whole photo.

3.3 Color harmony assessment for photos

We tested our method using a photo collection available on the Internet (DPChallenge[10)]). On the web site, various people have given aesthetic scores to various photos. We collected 124,664 photos in 14 categories and removed all the sepia-tone, and black and white photos from them. Let us first provide the detail of those photo collection. In our experiments, the top and bottom 10% of scores were deemed the high- and low-quality photos. We trained our system using half the photo collection and tested our system using the other half.

We compared the performance of our method in two aspects: (1) considering color harmony based on a 'Whole' photo or dividing it into a set of 'Local' regions, and (2) evaluating harmony in color using its 'Absolute' or 'Relative' color values. A feature representing color harmony of a photo is computed using either existing models of color harmony[29),32)], absolute values in color of the photo, or its relative values. All of these methods are tested for 'Whole' and 'Local' cases. More detail on our test cases are summarized as follows: an SVM with a linear kernel was used for all cases except 'Matsuda' and 'Moon-Spencer' cases, $2,304$ local regions were extracted from a photo, and the size of each region was 32×32. We compare the following methods.

- **Whole, Matsuda**: a feature is designed from the harmonic templates of the Matsuda model[29)]. We computed a hue histogram in CIE LCH color space from a whole photo and evaluated the similarity between the templates and the histogram using the technique described in the references[4),33)].
- **Whole, Moon-Spencer**: a feature is extracted from a whole photo by using the Moon-Spencer model[32)]. We computed the histograms of relative values of hue, chroma, and lightness, and summed bins corresponding to the 'Contrast', 'Similarity', and 'Identity' in Figure 6.
- **Local, Moon-Spencer**: a feature is the sum of color harmony scores using the Moon-Spencer model from local regions of a photo.
- **Whole, Absolute (Chroma)**: a histogram of chroma values extracted from a whole photo in CIE LCH color space.
- **Whole, Absolute (Hue)**: a histogram of hue values extracted from a whole photo in CIE LCH color space.
- **Whole, Absolute (RGB)**: a histogram of pixel values extracted from a whole photo in RGB color space.
- **Local, Absolute (Chroma)**: a feature extracted from a collection of local regions. A local region was described by using chroma histogram. We used our method described in Section 3.2.3 to combine the descriptors.
- **Local, Absolute (Hue)**: a feature computed from a collection of local regions. A local region was described by using hue histogram.
- **Local, Absolute (RGB)**: a feature computed from a collection of local regions. A local region was described by using RGB histogram.

- **Local, Relative (Hue)**: our color harmony feature described in Section 3.2. We used only hue values relative to the chosen color for describing a local region. The feature was extracted from a collection of local regions. The number of dimensions of our local descriptor was 100.
- **Local, Relative (Hue, Chroma, Lightness)**: our color harmony feature described in Section 3.2. We used hue, chroma, and lightness values relative to the chosen color for describing a local region. The feature was extracted from a collection of local regions. The number of dimensions of our local descriptor was 200.

Figure 12 shows the classification performance as a recognition rate: the probability that the quality inferred using each feature matched the correct quality. The plot shows the average rate among the 14 categories. In this Figure, we clearly see that the 'Local' feature using a set of local regions is superior to the 'Whole' feature extracted from the whole photo and it is effective for existing models of color harmony. Also, we see that the features of 'Relative' values are superior to the one of 'Absolute' value. Overall, our 'Local, Relative' method outperforms the alternative features and achieved about 66% accuracy in this difficult task.

Figure 13 shows the classification results by using the proposed method: (a) photos with high color harmony and (b) photos with low color harmony. A comparison showed that the photos in (a) have more pleasant color distributions and spatially better balanced. More experimental results were demonstrated in the reference[34].

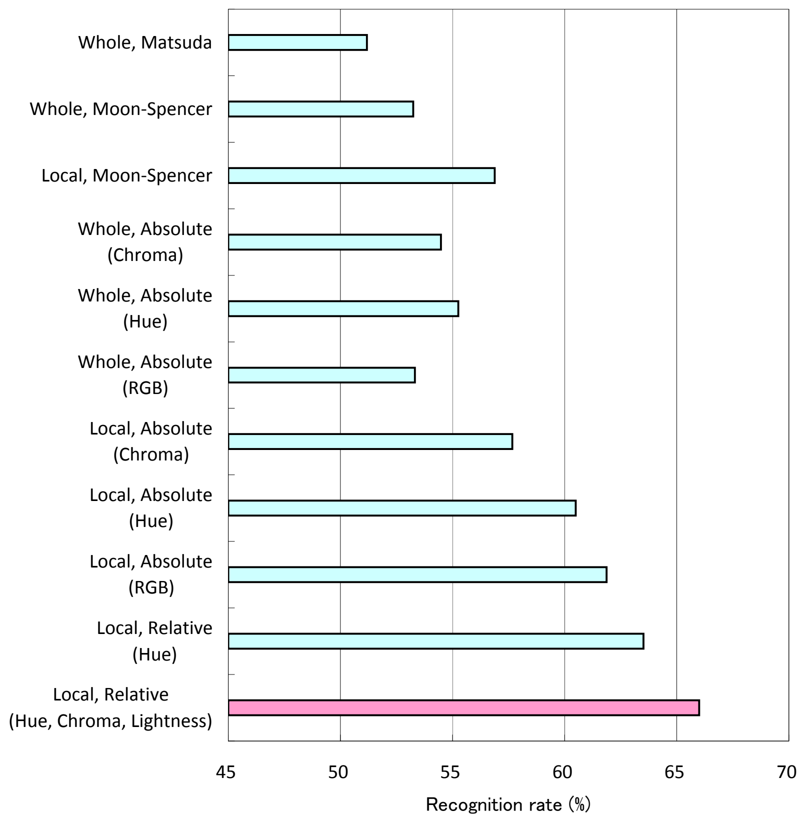


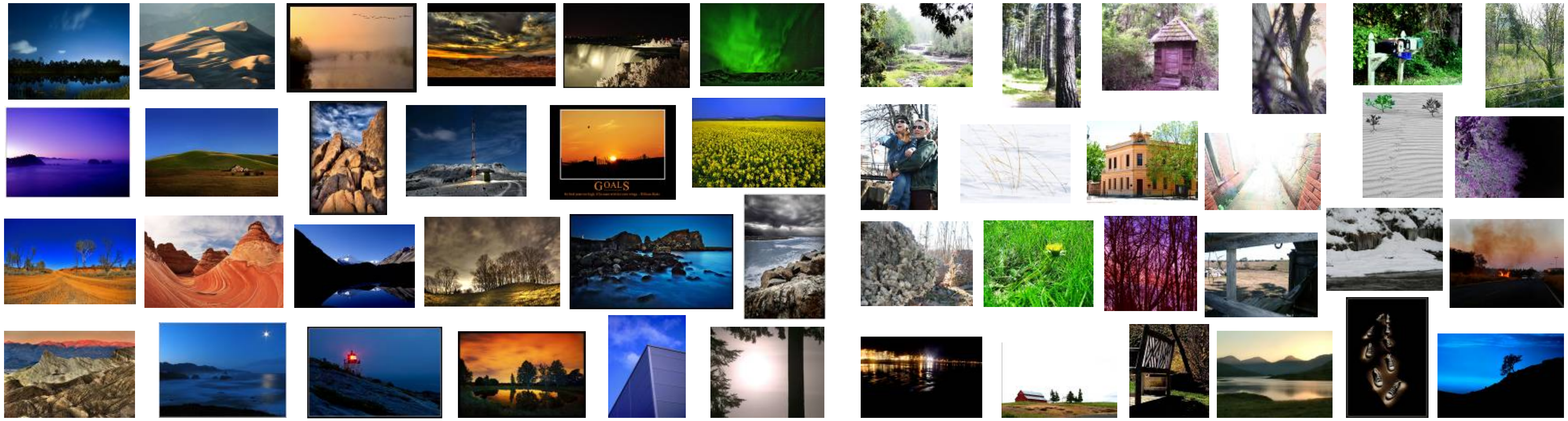
Fig. 12 Performance of assessing the color harmony of photos in the DPChallenge dataset.

4. Application using aesthetic quality classification

In this section, we present a novel method for automatically editing a photo using a quality classifier that assesses whether the cropped region is agreeable to users.

Cropping is a technique used for removing the unwanted subjects and irrelevant details from a photo, to change its aspect ratio, and to improve its overall composition. The technique plays an important role in various photo editing tasks, *e.g.*, making a thumbnail for easily visualizing a large number of photos or printing a digital photo of an arbitrary size on paper of a specific size. Large photo collections are now available with the widespread use of digital cameras and the Internet. Automating photo cropping is essential for editing such a large amount of photos without requiring iterative user operation.

Prior work on automatic photo cropping has taken only the attention grabbing regions that consist of salient pixels in an original photo into consideration. The relevant papers thus only address how to estimate where the region of attention lies in a photo. Suh *et al.*[47] made their estimates using a low-level saliency map. The saliency is small in a uniformly textured region, and large in a region



(a) (b)

Fig. 13 Examples of photos classified from their color as being of high quality (a) and low quality (b).

with a complex texture, edge and/or corner. Santella et al.[44] semi-automatically determined the region of attention by estimating the gaze of a user looking at each photo. Luo[26] determined the region by estimating the contents in a photo using the structural features, *e.g.*, the centrality and shape, and the semantic features, *e.g.*, the face, or the sky. This section calls these approaches *attention-based cropping*.

Although attention-based cropping is effective for emphasizing the regions of attention, it does not take into consideration the agreeability that the cropped regions give users. Thus, the regions are not necessarily agreeable to users. This sometimes causes a problem in that users feel the regions are low quality and do not satisfy the outputs of the automatic photo cropping.

We propose a novel automatic photo cropping method with a quality classifier that automatically distinguishes between the high- and low-quality regions of a photo to ensure there is a higher level of agreeability for the cropped region. We call our method *sensation-based cropping*. To the best of our knowledge, the work presented in this section is the first effort to introduce aesthetic quality classification into automatic photo cropping. Experiments demonstrated that our method, which uses the quality classifier for the automatic photo cropping, outperforms an existing cropping method[47].

4.1 Sensation-based photo cropping

We start with an overview of our method, which is illustrated in Figure 14. To find a region for cropping, the candidates for a region $I_{x,y,w,h}$ with the top-left corner coordinates (x, y) and a width w and a height h of a rectangle are generated from the original photo by trimming it. An aesthetic quality score $q_{x,y,w,h}$ is estimated by applying the quality classifier to each candidate. As the quality score increases, a given region is considered high quality. Finally, a cropped region is determined by finding the candidate with the highest aesthetic quality score. Currently, this is done by a brute-force search, but other more sophisticated methods could also be used.

An aesthetic quality classifier is required to achieve sensation-based cropping. We built a classifier with multiple subjects described in Section 2.2. This classifier detects multiple-subject regions by using a saliency map, extracts the features representing the basic techniques for photography from each region, computes a posterior probability that the feature is matched as high quality, and determines

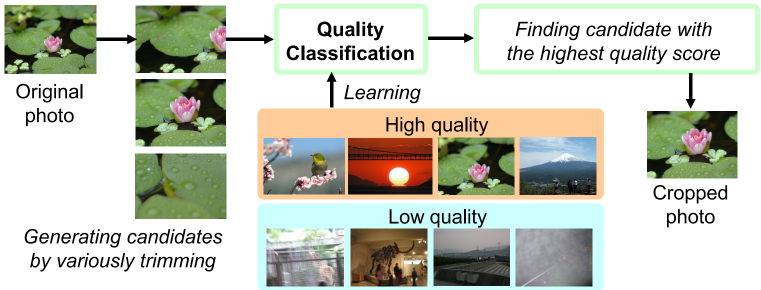


Fig. 14 Overview of sensation-based cropping. This section introduces a quality classifier for automatic photo cropping. To find a cropped region, we generate the candidates for a region by variously trimming the original photo, and then by estimating the aesthetic quality score by applying a quality classifier to each candidate. A cropped region is determined by finding the candidate with the highest quality score.

the actual aesthetic quality using the combined probabilities.

4.2 Experiments of automatic photo cropping

We demonstrated our automatic photo cropping by conducting a subjective assessment between the regions cropped by our method and ones by Suh *et al.*'s method[47]. We used 200 photos downloaded from Flickr[13] by searching for photos with a 'wide angle' tag. Note that the photos of human faces were removed since we believe that cropping using a face detector performs better than our approach. We showed a pair of regions cropped by our method and Suh *et al.*'s method to 30 people in random order. We asked each person to select within a 3 to 5 second period which result they preferred per pair. For each photo we defined the selection rate that represents the percentage of the people preferring our result.

Figure 15 show some examples of the original photos and the cropped regions. The photos in those Figures obtained a high selection rate expected a photo in Figure 15 (f) that obtained a low selection rate. In (a)-(e), the cropped region using our method was more agreeable to users than that compared with the one using Suh *et al.*'s method. In our unoptimized implementation on a single core 2.8-GHz processor, cropping took several minutes per photo.

Figure 16 shows a stacked bar graph in terms of the selection rate given by 30 people. As we can see, our method obtains a more significantly improved performance than Suh *et al.*'s method[47]. More experimental results were demonstrated in the reference[35].

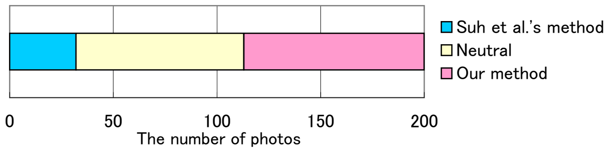


Fig. 16 Subjective assessment of automatic photo cropping. We grouped 200 photos into 'Our method' obtaining over a 65% selection rate, and 'Suh *et al.*'s method', which obtained a selection rate under 35%, and the rest was 'Neutral'.

5. Conclusions

In this paper, we presented the development of techniques for classifying aesthetic quality. Specifically, we presented the following two ideas to improve the aesthetic quality classification performance: detecting multiple-subject regions and assessing color harmony. We showed that these ideas play an important role in determining the aesthetic quality of photographs through evaluations of classification performance.

5.1 Contributions

The salient contributions of this paper are summarized as follows:

- **Aesthetic quality classification with multiple subjects**
  We presented an aesthetic quality classifier that is based on rules of thumb for photography to improve the performance of aesthetic quality classification. We detected multiple-subject regions and a background region and extracted features from these regions. The merit of this technique is that the quality classifier deals with photographs containing multiple subjects, for instance, a flower among leaves or individual buildings in a landscape. Our technique extracts more detailed features from multiple-subject regions than from a single-subject region. We showed that our technique with multiple subjects for aesthetic quality classification outperforms existing methods in the case of large photograph datasets available on the Internet.
- **Aesthetic quality classifier based on color harmony assessment of photographs**

Vol.2011-CVIM-177 No.33
2011/5/20

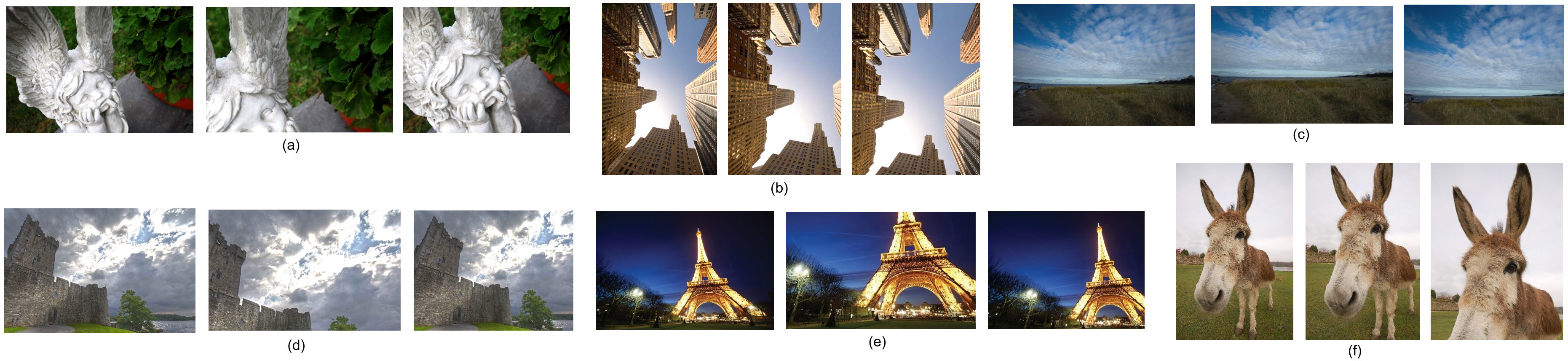


Fig. 15 Examples of automatic photo cropping. In (a)-(f), we show that the original photo (right) is automatically cropped to the regions using Suh *et al.*'s method (center) and our method (left). Our results had a higher selection rate for the subjective assessment than Suh *et al.*'s except for the results in (f).

Color harmony plays an important role in enhancing the effectiveness of aesthetic quality classification. To overcome the difficulty in evaluating the color harmony of a photograph owing to the complexity of color distribution, we introduced the *bags-of-color-patterns* representation. A key observation in this paper is that a photograph can be seen as a collection of local regions whose color variation is relatively simple. This led us to develop a technique for assessing the aesthetic quality of a photograph by using the *bags-of-features* framework. We showed that our technique substantially more accurately classifies aesthetic quality than the existing methods in the case of a large dataset in the DPChallenge.

- **Sensation-based photo cropping**
  We presented a technique for automatically cropping a photograph using an aesthetic quality classifier. Our technique ensures a higher level of agreeability for the cropped region. To the best of our knowledge, this is the first effort that considers human appreciation for photograph manipulation. We showed that the regions cropped using our technique generate greater appreciation than those cropped using existing methods.

5.2 Future Directions

We propose several avenues for future research, which are presented as follows:

- **Aesthetic quality classification based on composition assessment**
  The perception of the aesthetic quality of a photograph is based on many photographic properties such as prominent subject regions, color distribution, and composition balance. In this paper, we discussed detecting multiple subject regions and assessing of color harmony. To further improve the aesthetic quality classification of photographs, a method needs to be devised that enables photographic composition to be assessed. We are planning to analyze a method that evaluates composition of a photograph by detecting the spatial structure of its constituent objects.
- **Scale factor feature for aesthetic quality classification**
  The resolution of a photograph plays an important role in determining its aesthetic quality. If a photograph is resized, an individual's perception of its aesthetic quality will change. In Section 3, we discussed extracting features from local regions that have a fixed scale size for aesthetic quality classifica-

tion. To further improve the classification performance, we intend to extend our method by considering the scale variations that occur in the sampling of the local regions. Furthermore, we will investigate extracting features invariant to changes in photographic resolution.

- **Personalization of aesthetic quality classification**
  We define high-aesthetic-quality photographs as those a majority of people will like. Building human-centered computer systems that can provide personalized services to individuals will require aesthetic quality to be personalized. For instance, the computer system could provide photographs sorted by personalized quality by collecting information on a person's preferences when the person uses the systems. We intend to develop a method for personalizing aesthetic quality classification by using online machine learning techniques.
- **Regression technique for inferring aesthetic quality score**
  In this paper, we presented techniques that assign aesthetic quality labels to photographs; these techniques classify photographs as being either high or low quality. In addition to aesthetic quality classification, we must consider developing an aesthetic quality regression technique that can directly estimate an aesthetic score for a photograph. This technique can be used to develop an automatic subjective assessment system that infers aesthetic quality scores to images.
- **Further development of human-centered computer systems with aesthetic quality analysis**
  Aesthetic quality analysis has an immense potential for use in real-world applications such as automatic aesthetic quality assessment of videos, retargeting, color transfer, and automatic white balance. We are planning to extend its potential uses further by exploring the development of human-centered computer systems that can assess the aesthetic quality of content. We plan one of the systems to be a digital camera system that makes it convenient for users to capture high-aesthetic-quality photographs.

References

1) Bhattacahrya, S., Sukthankar, R. and Shah, M.: A Framework for Photo-Quality Assessment and Enhancement based on Visual Aesthetics, *Proc. ACM Int'l Conf. Multimedia*, pp.271–280 (2010).
2) Bishop, C.M.: *Pattern Recognition And Machine Learning (Information Science and Statistics)*, Springer (2006).
3) Blanz, V. and Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.9, pp. 1063 – 1074 (2003).
4) Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T. and Xu, Y.Q.: Color Harmonization, *ACM Trans. Graphics*, Vol.25, No.3, pp.624–630 (2006).
5) Comaniciu, D. and Meer, P.: Mean shift: A robust approach toward feature spaceanalysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.5, pp. 603–619 (2002).
6) Dance, C., Willamowski, J., Fan, L., Bray, C. and Csurka, G.: Visual Categorization with Bags of Keypoints, *Proc. ECCV Int'l Workshop Statistical Learning in Computer Vision*, pp.1–22 (2004).
7) Datta, R., Joshi, D., Li, J. and Wang, J.Z.: Studying Aesthetics in Photographic Images Using a Computational Approach, *Proc. European Conf. Computer Vision*, Vol.3953, pp.288–301 (2006).
8) Datta, R., Li, J. and Wang, J.Z.: Learning the Consensus on Visual Quality for Next-Generation Image Management, *Proc. ACM Int'l Conf. Multimedia* (2007).
9) Datta, R., Li, J. and Wang, J.Z.: Algorithmic Inferencing of Aesthetics and Emotion in Natural Images: an Exposition, *Proc. Int'l Conf. Image Processing*, pp.105 – 108 (2008).
10) DPChallenge: *http://www.dpchallenge.com.*
11) Duda, R.O., Hart, P.E. and Stork, D.G.: *Pattern classification second edition*, John Wiley & Sons (2000).
12) Fei-Fei, L., Fergus, R. and Perona, P.: Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories, *Proc. Workshop Generative-Model Based Vision* (2004).
13) Flickr: *http://www.flickr.com.*
14) Huang, T.W. and Chen, H.T.: Landmark-Based Sparse Color Representations for Color Transfer, *Proc. Int'l Conf. Computer Vision*, pp.199–204 (2009).
15) Itten, J.: *The Art of Color*, Van Nostrand Reinhold Company (1960).
16) Itti, L., Koch, C. and Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Trans. PAMI*, Vol.20, No.11, pp.1254 – 1259 (1998).
17) Judd, D.B. and Wyszecki, G.: *Color in Business, Science, and Industry*, John Wiley & Sons (1975).
18) Ke, Y., Tang, X. and Jing, F.: The Design of High-Level Features for Photo Quality Assessment, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.419–426 (2006).
19) Lalonde, J.F. and Efros, A.A.: Using Color Compatibility for Assessing Image Realism, *Proc. Int'l Conf. Computer Vision*, pp.1–8 (2007).

20) Lazebnik, S., Schmid, C. and Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Vol.2, pp.2169–2178 (2006).
21) Li, S.Z. and Jain, A.K.: *Handbook of Face Recognition*, Springer (2005).
22) Lin, H.T., Lin, C.J. and Weng, R.C.: A note on Platt's probabilistic outputs for support vector machines, *Technical report, National Taiwan University* (2003).
23) Liu, L., Chen, R., Wolf, L. and Cohen-Or, D.: Optimizing Photo Composition, *Computer Graphic Forum*, Vol.29, No.2, pp.469–478 (2010).
24) Loui, A., Wood, M. D., Scalise, A. and Birkelund, J.: Multidimensional Image Value Assessment and Rating for Automated Albuming and Retrieval, *Proc. IEEE Int'l Conf. Image Processing*, pp.97–100 (2008).
25) Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *Int'l J. Computer Vision*, Vol.60, No.2, pp.91–110 (2004).
26) Luo, J.: SUBJECT CONTENT-BASED INTELLIGENT CROPPING OF DIGITAL PHOTOS, *Proc. Int'l Conf. Multimedia & Expo*, pp.2218 – 2221 (2007).
27) Luo, Y. and Tang, X.: Photo and Video Quality Evaluation: Focusing on the Subject, *Proc. European Conf. Computer Vision*, Vol.5304, pp.386–399 (2008).
28) Matas, J., Chum, O., Urba, M. and Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions, *Proc. British Machine Vision Conf.*, pp.384–396 (2002).
29) Matsuda, Y.: *Color design*, Asakura Shoten (1995).
30) Meynet, J., Popovici, V., Sorci, M. and Thiran, J.P.: Combining SVMs for Face Class Modeling, *Proc. 13th European Signal Processing Conf.* (2005).
31) Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L.V.: A Comparison of Affine Region Detectors, *Int'l J. Computer Vision*, Vol.65, No.1/2, pp.43–72 (2005).
32) Moon, P. and Spencer, D.E.: Geometric Formulation of Classical Color Harmony, *J. Opt. Soc. Amer.*, Vol.34, pp.46–50 (1944).
33) Moorthy, A.K., Obrador, P. and Oliver, N.: Towards Computational Models of the Visual Aesthetic Appeal of Consumer Videos, *Proc. European Conf. Computer Vision*, Vol.6315, pp.1–14 (2010).
34) Nishiyama, M., Okabe, T., Sato, I. and Sato, Y.: Aesthetic Quality Classification of Photographs Based on Color Harmony, *to appear in Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2011).
35) Nishiyama, M., Okabe, T., Sato, Y. and Sato, I.: Sensation-based Photo Cropping, *Proc. ACM Int'l Conf. Multimedia*, pp.669–672 (2009).
36) Nowak, E., Jurie, F. and Triggs, B.: Sampling Strategies for Bag-of-Features Image Classification, *Proc. European Conf. Computer Vision*, pp.490–503 (2006).
37) Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, *IEEE Trans. Systems, Man and Cybernetics*, Vol.9, No.1, pp.62–66 (1979).
38) Photo.net: *http://photo.net*.
39) Pitie, F. and annd R.Dahyot, A. C.K.: Automated Colour Grading using Colour Distribution Transfer, *Computer Vision and Image Understanding*, Vol.107, No.1-2, pp.123–137 (2007).
40) Platt, J.: Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, *Advances in Large Margin Classiers. MIT Press*, pp.61 – 74 (2000).
41) Reinhard, E., Ashikhmin, M., Gooch, B. and Shirley, P.: Color Transfer between Images, *Proc. IEEE Computer Graphics and Applications*, pp.34–41 (2001).
42) Rosten, E. and Drummond, T.: Machine learning for high-speed corner detection, *Proc. European Conf. Computer Vision*, Vol.1, pp.430–443 (2006).
43) Sachs, T., Kakarala, R., Castleman, S. and Rajan, D.: A data-driven approach to understanding skill in photographic composition, *Proc. Workshop on Computational Photography and Aesthetics* (2010).
44) Santella, A., Agrawala, M., DeCarlo, D., Salesin, D. and Cohen, M.: Gaze-Based Interaction for Semi-Automatic Photo Cropping, *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pp.771 – 780 (2006).
45) Shin, Y. and Kim, E.Y.: Affective Prediction in Photographic Images Using Probabilistic Affective Model, *Proc. ACM Int'l Conf. Image and Video Retrieval*, pp. 390–397 (2010).
46) Sivic, J. and Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos, *Proc. Int'l Conf. Computer Vision*, Vol.2, pp.1470–1477 (2003).
47) Suh, B., Ling, H., Bederson, B. and Jacobs, D.: Automatic Thumbnail Cropping and Its Effectiveness, *Proc. ACM Symp. User Interface Software and Technology*, pp.95 – 104 (2003).
48) Szeliski, R.: *Computer Vision: Algorithms and Applications (Texts in Computer Science)*, Springer (2010).
49) Tokumaru, M., Muranaka, N. and Imanishi, S.: Color Design Support System Considering Color Harmony, *Proc. IEEE Int'l Conf. Fuzzy Systems*, pp. 378–383 (2002).
50) Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer-Verlag (1995).
51) Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Processing*, Vol.13, No.4, pp.600–612 (2004).
52) Wright, J., Yang, A., Ganesh, A., Sastry, S. and Ma, Y.: Robust Face Recognition via Sparse Representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.31, No.2, pp.210 – 227 (2009).
53) Xiang, Y., Zou, B. and Li, H.: Selective color transfer with multi-source images, *Pattern Recognition Letters*, Vol.30, No.7, pp.682–689 (2009).
54) Yang, L., Jin, R., Sukthankar, R. and Jurie, F.: Unifying Discriminative Visual Codebook Generation with Classifier Training for Object Category Recognition, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp.1–8 (2008).