Photo and Video Quality Evaluation: Focusing on the Subject

Conference Paper · October 2008

DOI: 10.1007/978-3-540-88690-7_29 · Source: DBLP

Xiaoou Tang
The Chinese University of Hong Kong
427 PUBLICATIONS **54,965** CITATIONS

Photo and Video Quality Evaluation: Focusing on the Subject

Yiwen Luo and Xiaoou Tang

Department of Information Engineering
The Chinese University of Hong Kong, Hong Kong
{ywluo6,xtang}@ie.cuhk.edu.hk

Abstract. Traditionally, distinguishing between high quality professional photos and low quality amateurish photos is a human task. To automatically assess the quality of a photo that is consistent with humans perception is a challenging topic in computer vision. Various differences exist between photos taken by professionals and amateurs because of the use of photography techniques. Previous methods mainly use features extracted from the entire image. In this paper, based on professional photography techniques, we first extract the subject region from a photo, and then formulate a number of high-level semantic features based on this subject and background division. We test our features on a large and diverse photo database, and compare our method with the state of the art. Our method performs significantly better with a classification rate of 93% versus 72% by the best existing method. In addition, we conduct the first study on high-level video quality assessment. Our system achieves a precision of over 95% in a reasonable recall rate for both photo and video assessments. We also show excellent application results in web image search re-ranking.

1 Introduction

With the popularization of digital cameras and the rapid development of the Internet, the number of photos that can be accessed is growing explosively. Automatically assessing the quality of photos that is consistent with human's perception has become more and more important with the increasing need of professionals and home users. For example, newspaper editors can use it to find high quality photos to express news effectively; home users can use such a tool to select good-looking photos to show from their e-photo albums; and web search engines may incorporate this function to display relevant and high quality images for the user. Fig. 1 shows two example photos. Most people agree that the left photo is of high quality and the right one is not. To tell the differences between high quality professional photos and low quality photos is natural to a human, but difficult to a computer.

There have been a number of works on image quality assessment concerning image degradation caused by noise, distortion, and compression artifacts [1], [2], [3]. Different from these works, we consider photo quality from an aesthetic point of view and try to determine the factors that make a photo look good in human's perception. The most

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part III, LNCS 5304, pp. 386–399, 2008.

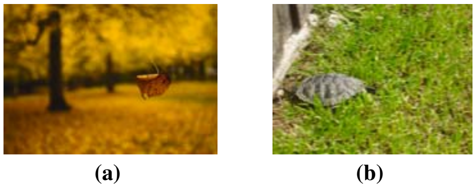

(a)
(b)

Fig. 1. Most people may agree that (a) is of higher quality than (b)

related work is published in [4], [5], and [6]. Tong *et al.* [4] and Datta *et al*. [5] combined features that are mostly used for image retrieval previously with a standard set of learning algorithms for the classification of professional photos and amateurish photos. For the same purpose, Ke *et al.* designed their features based the spatial distribution of edges, blur, and the histograms of low-level color properties such as brightness and hue [6]. Our experiments show that the method in [6] produce better results than that in [4] and [5] with much less number of features, but it is still not good enough with a classification rate of 72% on a large dataset.

The main problem with existing methods is that they compute features from the whole image. This significantly limits the performance of the features since a good photo usually treats the foreground subject and the background very differently. Professional photographers usually differentiate the subject of the photo from the background to highlight the topic of the photo. High quality photos generally satisfy three principles: *a clear topic, gathering most attention on the subject, and removing objects that distract attention from the subjec*t [7], [8], [9]. Photographers try to achieve this by skillfully manipulating the photo composition, lighting, and focus of the subject. Motivated by these principles, in this paper, we first use a simple and effective blur detection method to roughly identify the focus subject area. Then following human perception of photo qualities we develop several highly effective quantitative metrics on subject clarity, lighting, composition, and color. In addition, we conduct the first study on video quality evaluation. We achieve significant improvement over state of the art methods reducing the error rates by several folds. We also apply our work to on-line image reranking for MSN Live image search results with good performance.

In summary, the main contributions of this paper include: 1) Proposed a novel approach to evaluate photo and video quality by focusing on the foreground subject and developed an efficient subject detection algorithm; 2)Developed a set of highly effective high-level visual features for photo quality assessment; 3) Conducted the first study of high-level video quality assessment and build the first database for such study; 4) First studied visual quality re-ranking for real world online image search.

2 Criteria for Assessing Photo Quality

In this section, we briefly discuss several important criteria used by professional photographers to improve photo quality. Notice that most of them rely on different treatment of the subject and the background.

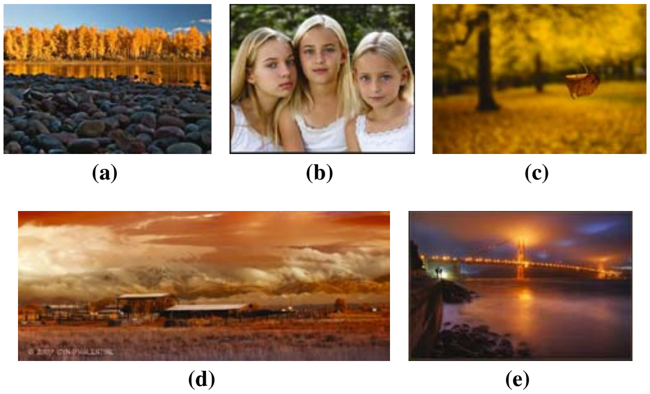


Fig. 2. (a) “Fall on the Rocks” by M. Marjory, 2007. (b) “Mona Lisa Smiles” by David Scarbrough, 2007. (c) “Fall: One Leaf at a Time” by Jeff Day, 2007. (d) “Winter Gets Closer” by Cyn D. Valentine, 2007. (e) “The Place Where Romance Starts” by William Lee, 2007.

2.1 Composition

Composition means the organization of all the graphic elements inside a photo. Good composition can clearly show the audience the photo’s topic and effectively express photographer’s feeling. The theory of composition is usually rooted in one simple concept: contrast. Professional photographers use contrast to awaken a vital feeling for the subject through a personal observation [10]. Contrast between light and dark, between shapes, colors, and even sensations, is the basis for composing a photo. The audience can often find the obvious contrast between the cool and hard stones in the foreground and the warm and soft river and forest in the background in Fig. 2a.

2.2 Lighting

A badly lit scene ruins the photo as much as poor composition. The way a scene is lit changes its mood and the audience’s perception of what the photo tries to express. Lighting in high quality photos makes the subjects not appear flat and enhances their 3D feeling, which is helpful to attract the audience’s attention to the subjects. Good lighting results in strong contrast between the subject and the background, and visually distinguishes the subject from the background. The lighting in Fig. 2b isolates the girls from the background and visually enhances the 3D feeling of them.

2.3 Focus Controlling

Professional photographers control the focus of the lens to isolate the subject from the background. They blur the background but keep the subject in focus, such as Fig. 2c. They may also blur closer objects but sharpen farther objects to express the depth of the scene, such as Fig. 2d. More than capturing the scene only, controlling the lens can create surrealistic effects, such as Figs. 2c and 2e.

2.4 Color

Much of what viewers perceive and feel about a photo is through colors. Although their color perception depends on the context and is culture-related, recent color science study shows that the influence on human emotions or feeling from a certain color or a certain color combination is usually stable in varying culture background [11], [12]. Professional photographers use various exposure and interpreting methods to control the color palette in a photo, and use specific color combination to raise viewers' specific emotion, producing a pleasing affective response. The photographer of Fig. 2a uses the combination of bright yellow and dark gray to produce an aesthetic feeling from the beauty of nature. The photographer of Fig. 2b uses the combination of white and natural skin color to enhance the beauty of chasteness from the girls.

3 Features for Photo Quality Assessment

Based on the previous analysis, we formulate these semantic criteria mathematically in this section. We first separate the subject from the background, and then discuss how to extract the features for photo quality assessment.

3.1 Subject Region Extraction

Professional photographers usually make the subject of a photo clear and the background blurred. We propose an algorithm to detect the clear area of the photo and consider it as the subject region and the rest as the background.

Levin *et al.* [13] presented a scheme to identify blur in an image when the blur is caused by 1D motion. We modify it to detect 2D blurred regions in an image. Let us use Fig. 3 as an example to explain the method. Fig. 3a is a landscape photo. We use a kernel of size $k \times k$ with all coefficients equal to $1/k^2$ to blur the photo. Figs. 3b, 3c and 3d are the results blurred by 5×5, 10×10, and 20×20 kernels, respectively. The log histograms of the horizontal derivatives of the four images in Fig. 3 are shown in Fig. 3e, and the log histograms of the vertical derivatives of the four images are shown in Fig. 3f. It is obvious that the blurring significantly changes the shapes of the curves in the histograms. This suggests that the statistics of the derivative filter responses can be used to tell the difference between clear and blurred regions.

Let f_k denotes the blurring kernel of size $k \times k$. Convolving the image I with f_k, and computing the horizontal and vertical derivatives from $I * f_k$, we have the distributions of the horizontal and vertical derivatives:

$$p_{xk} \propto \mathit{hist}(I * f_k * d_x), \quad p_{yk} \propto \mathit{hist}(I * f_k * d_y) \tag{1}$$

where $d_x = [1, -1]$, and $d_y = [1, -1]^T$. The operations in Eq. (1) are done 50 times with $k = 1, 2, ..., 50$.

For a pixel (i, j) in I, we define a log-likelihood of derivatives in its neighboring window $W_{(i,j)}$ of size $n \times n$ with respect to each of the blurring models as:

$$l_k(i,j) = \sum_{(i',j') \in W_{(i,j)}} (\log p_{xk}(I_x(i',j')) + \log p_{yk}(I_y(i',j'))), \tag{2}$$

where $I_x(i', j')$ and $I_y(i', j')$ are the horizontal and vertical derivatives at pixel (i', j'), respectively, and $l_k(i, j)$ measures how well the pixel (i, j)'s neighboring window is explained by a $k \times k$ blurring kernel. Then we can find the blurring kernel that best explains the window's statistics by $k^*(i, j) = \arg\max_k l_k(i, j)$. When $k^*(i, j) = 1$, pixel (i, j) is in the clear area; otherwise it is in the blurred area. With $k^*(i, j)$ for all the pixels of I, we can obtain a binary image U to denote the clear and blurred regions of I, defined as:

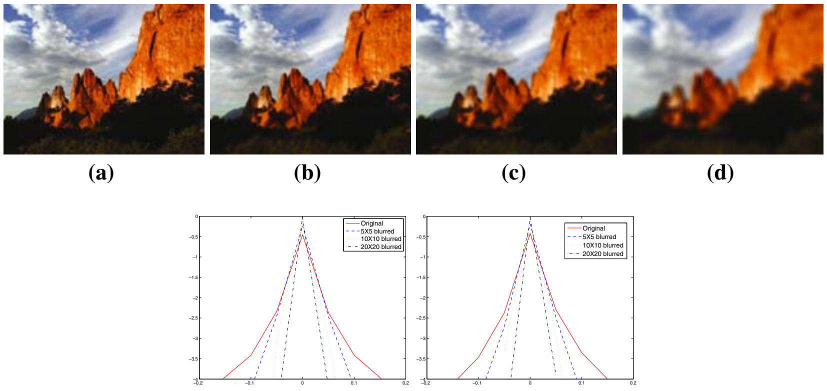
$$U(i, j) = \begin{cases} 1, & k^*(i, j) = 1 \\ 0, & k^*(i, j) > 1. \end{cases} \tag{3}$$

Two examples of such images are show in in Figs. 4a and 4b with the neighboring window size of 3 × 3. Next, we find a compact bounding box that encloses the main part of the subject in an image.

Projecting U onto the x and y axes independently, we have

$$U_x(i) = \sum_j U(i, j), \quad U_y(j) = \sum_i U(i, j). \tag{4}$$

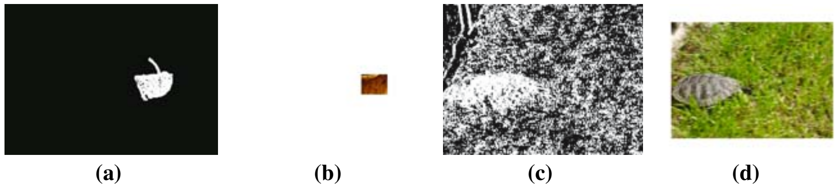
On the x axis, we find x_1 and x_2 such that the energy in $[0, x_1]$ and the energy in $[x_2, N - 1]$ are each equal to $\frac{1-\alpha}{2}$ of the total energy in U_x, where N is the size of the image in the x direction. Similarly, we can find y_1 and y_2 in the y direction. Thus, the subject region R is $[x_1 + 1, x_2 - 1] \times [y_1 + 1, y_2 - 1]$. In all our experiments, we choose $\alpha = 0.9$. Two examples of subject regions corresponding to Figs. 1a and 1b are given in Figs. 4c and 4d.



(a) (b) (c) (d)

(e) (f)

Fig. 3. Images blurred by different blurring kernels. (a) Original Image. (b) Result blurred by the 5 × 5 kernel. (c) Result blurred by the 10 × 10 kernel. (d) Result blurred by the 20 × 20 kernel. (e) Log histograms of the horizontal derivatives of the original image and the images blurred by the 5 × 5, 10 × 10, and 20 × 20 kernels, respectively. (f) Log histograms of the vertical derivatives of the original image and the blurred images by 5 × 5, 10 × 10, and 20 × 20 kernels, respectively.



(a) (b) (c) (d)

Fig. 4. (a) The clear regions (white) of Fig. 1a. (b) The subject region of Fig. 1a. (c) The clear (white) regions of Fig. 1b. (d) The subject region of Fig. 1b.

3.2 Clarity Contrast Feature

To attract the audience's attention to the subject and to isolate the subject from the background, professional photographers usually keep the subject in focus and make the background out of focus. A high quality photo is neither entirely clear nor entirely blurred. We here propose a clarity contrast feature f_c to describe the subject region with respect to the image:

$$f_c = (\|M_R\|/\|R\|)/(\|M_I\|/\|I\|), \tag{5}$$

where $\|R\|$ and $\|I\|$ are the areas of the subject region and the original image, respectively, and

$$M_I = \{(u,v) \mid |F_I(u,v)| > \beta \max\{F_I(u,v)\}\}, \tag{6}$$

$$M_R = \{(u,v) \mid |F_R(u,v)| > \beta \max\{F_R(u,v)\}\}, \tag{7}$$

$$F_I = FFT(I), \quad F_R = FFT(R). \tag{8}$$

A clear image has relatively more high frequency components than a blurred image. In Eq. (5), $\|M_R\|/\|R\|$ denotes the ratio of the area of the high frequency components to the area of all the frequency components in R. The similar explanation applies to $\|M_I\|/\|I\|$. In all our experiments, we choose $\beta = 0.2$. For the two images in Fig. 1, their clarity contrast features are 5.78 and 1.62, respectively. From our experiments, we have found that the clarity feature of high quality and low quality photos mainly fall in $[1.65, 20.0]$ and $[1.11, 1.82]$, respectively.

3.3 Lighting Feature

Since professional photographers often use different lighting on the subject and the background, the brightness of the subject is significantly different from that of the background. However, most amateurs use natural lighting and let the camera automatically adjust a photo's brightness, which usually reduces the brightness difference between the subject and the background. To distinguish the difference between these two kinds of photos, we formulate it as:

$$f_l = |\log(B_s/B_b)|, \tag{9}$$

where B_s and B_b are the average brightness of the subject region and the background, respectively. The values of f_l of Fig. 1a and Fig. 1b are 0.066 and 0.042, respectively. Usually, the values of f_l of high quality and low quality photos fall in $[0.03, 0.20]$ and $[0.00, 0.06]$, respectively.

3.4 Simplicity Feature

To reduce the attention distraction by the objects in the background, professional photographers make the background simple. We use the color distribution of the background to measure this simplicity. For a photo, we quantize each of the RGB channels into 16 values, creating a histogram His of 4096 bins, which gives the counts of quantized colors present in the background. Let h_{max} be the maximum count in the histogram. The simplicity feature is defined as:

$$f_s = (\|S\|/4096) \times 100\%, \tag{10}$$

where $S = \{i|His(i) \geq \gamma h_{max}\}$. We choose $\gamma = 0.01$ in all our experiments. The values of f_s of Fig. 1a and Fig. 1b are 1.29% and 4.44%, respectively. Usually, the simplicity features of high quality and low quality photos fall in $(0, 1.5\%]$ and $[0.5\%, 5\%]$, respectively.

3.5 Composition Geometry Feature

Good geometrical composition is a basic requirement for high quality photos. One of the most well-known principle of photographic composition is the *Rule of Thirds*. If we divide a photo into nine equal-size parts by two equally-spaced horizontal lines and two equally-spaced vertical lines, the rule suggests that the intersections of the two lines should be the centers for the subject (see Fig. 5). Study has shown that when viewing images, people usually look at one of the intersection points rather than the center of the image. To formulate this criterion, we define a composition feature as

$$f_m = \min_{i=1,2,3,4} \{\sqrt{(C_{Rx} - P_{ix})^2/X^2 + (C_{Ry} - P_{iy})^2/Y^2}\}, \tag{11}$$

where (C_{Rx}, C_{Ry}) is the centroid of the binary subject region in U(see Section 3.1), $(P_{ix}, P_{iy}), i = 1, 2, 3, 4$, are the four intersection points in the image, and X and Y are the width and height of the image. For Figs. 1a and 1b, the values of f_m are 0.11 and 0.35, respectively.

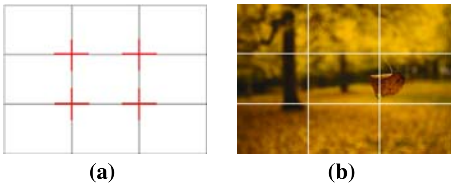


Fig. 5. (a) An illustration of the *Rule of Thirds*. (b) A high quality image obeying this rule.

3.6 Color Harmony Feature

Harmonic colors are the sets of colors that are aesthetically pleasing in terms of human visual perception. There are various mathematical models for defining and measuring the harmony of the color of the image [14], [15]. We have tried to measure the color harmony of a photo based on previous models [15], and found that the single feature classification rate is low. Here we develop a more accurate feature to measure the color harmony of a photo in terms of learning the color combinations (coexistence of two color in the photo) from the training dataset. For each photo, we compute a 50-bin histogram for each of the hue, saturation, and brightness. The value of the color combination between hue i and hue j is defined as $H_{hue}(i) + H_{hue}(j)$. The definitions for saturation combination and brightness combination are similar. For the high quality and low quality photos in the training database, we can obtain the histograms of hue combinations with

$$H_{high,hue}(i,j) = Average(H_{high,hue}(i) + H_{high,hue}(j)), \tag{12}$$

$$H_{low,hue}(i,j) = Average(H_{low,hue}(i) + H_{low,hue}(j)). \tag{13}$$

where $H_{high,hue}$ ($H_{low,hue}$) is the histogram of hue from high (low) quality training photos. Similarly, we can have the histograms of saturation combinations and brightness combinations, $H_{high,sat}(i,j)$, $H_{low,sat}(i,j)$, $H_{high,bri}(i,j)$, and $H_{low,bri}(i,j)$.

We design a feature f_h to measure whether a photo is more similar to the high quality photos or the low quality photos in the color combinations, which is formulated as:

$$f_h = Hue_s \times Sat_s \times Bri_s, \tag{14}$$

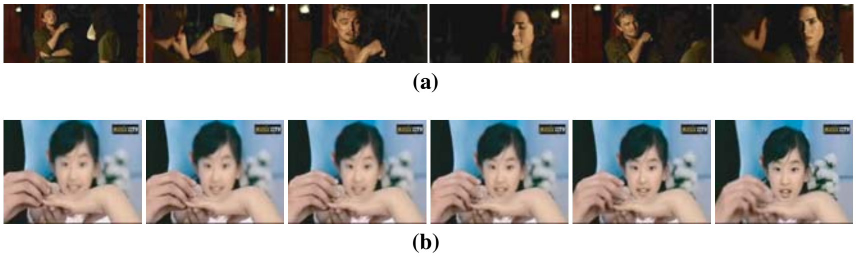
where $Hue_s = Hue_{high}/Hue_{low}$, $Sat_s = Sat_{high}/Sat_{low}$, $Bri_s = Bri_{high}/Bri_{low}$, and Hue_{high} is the cross product distance between $H_{high,hue}$ and the histogram of hue of the input photo, Hue_{low}, Sat_{high}, Sat_{low}, Bri_{high}, and Bri_{low} are computed similarly. For Figs. 1a and 1b, the values of f_h are 1.42 and 0.86, respectively. Usually, the color combination features of high quality photos fall in $[1.1, 1.6]$, and those of low quality photos are in $[0.8, 1.2]$.

4 Features for Video Quality Assessment

A video is a sequence of still images, and so the features proposed to assess photo's quality are also applicable for video quality assessment. Since a video contains motion information that can be used to distinguish professional videos from amateurish videos, we design two more motion-related features in this section.

4.1 Length of Subject Region Motion

Experienced photographers usually adjust the focus and change the shooting angle to tell the story more effectively. For example, Fig. 6a shows a sequence of conversation in the movie "Blood Diamond". The photographers change the shooting angle and focus continually to show the audience not only the speaking man's expression but also the



(a)

(b)

Fig. 6. (a) A sequence of screenshots in "Blood Diamond". (b) A sequence of screenshots in "Love Story". Both of them show how the professional photographers change the shooting angle and focus when taking the videos.

listening woman's. In Fig. 6b, the photographer moves the focus from the ring to the girl's face, showing the girl's expression and feeling when she sees the ring. However, amateurish photographers seldom change the shooting angle and focus when taking videos.

Since the change of shooting angle and focus usually changes the subject region in the frames, we evaluate these changes by average moving distance of the subject region between neighbor frames. We sample frame groups from a video, each of which contains P frames with a rate of 5 frames per second. Then this feature is defined as:

$$f_d = (\sum_{i=2}^{P} \sqrt{(C_{i,x} - C_{i-1,x})^2/X^2 + (C_{i,y} - C_{i-1,y})^2/Y^2})/(P-1), \tag{15}$$

where $(C_{i,x}, C_{i,y})$ is the centroid of the binary subject region of frame i, and X and Y are the width and height of the frame. Usually, the values of f_d of high quality and the low quality photos fall in [0.05, 0.6] and [0.003, 0.2], respectively.

4.2 Motion Stability

Camera shake is much less in high quality videos than in low quality videos. This feature can be used to distinguish between these two kinds of videos. Various methods have been proposed to detect shaking artifacts [16]. We use Yan and Kankanhalli's method [16] for this work due to its simplicity. We sample Q groups of frames from a video, each of which contains several successive frames. Then this feature is defined as:

$$f_t = Q_t/Q, \tag{16}$$

where Q is the total number of successive three frames in all the groups, and Q_t is the number of successive three frames that are detected as shaky frames.

Here we briefly explain how to detect shaky frames. We select the subject region from the first frame of a group as the target region, and then iteratively compute the best motion trajectory of the region, which results in a set of motion vectors. From three successive frames in the group, we have two motion vectors that form an angle. If the angle is larger than 90 degree, these three successive frames are considered shaky.

5 Experiments

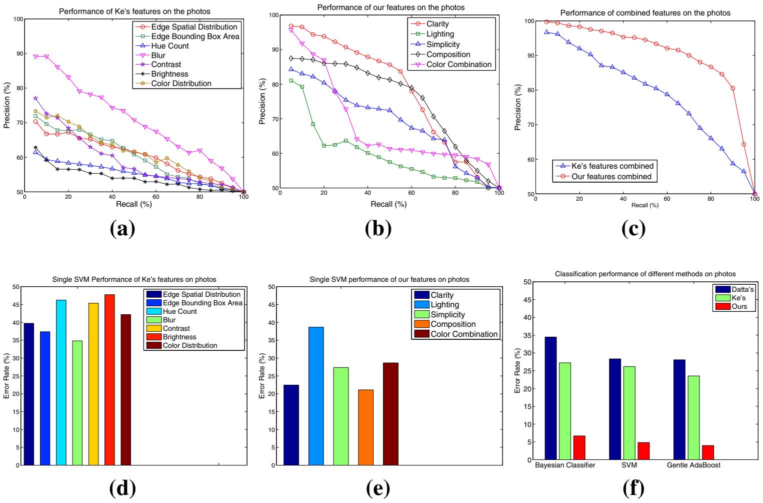
In this section, we demonstrate the effectiveness of our features using the photo database collected by Ke *et al.* [6], and a large and diverse video database collected from professional movies and amateurish videos. To further test these features' usability, we apply them on the images searched by MSN Live Search to give better rankings. To compare different features, we use three popular classifiers including the Bayes classifier which is also used in [6], the SVM [17] and the Gentle AdaBoost [18].

5.1 Photo Assessment

We compare the performance of our features with Ke *et al*'s features [6], and Datta *et al.*'s features [5] on the database collected by Ke *et al.* [6]. The database was acquired by crawling a photo contest website, DPChallenge.com, which contains a diverse set of high and low quality photos from many different photographers. The obtained 60000 photos were rated by hundreds of users at DPChallenge.com. The top 10%, total 6000 photos, were rated as high quality photos, and the bottom 10%, total 6000 photos, were rated as low quality photos. We randomly choose 3000 high quality and 3000 low quality photos as the training set, and choose the remaining 3000 high quality and 3000 low quality photos as the testing set. This design of the experiment is the same as that in [6].

We first give the classification results of individual features, and then the combined result using the Bayes classifier. For each feature, we plot a precision-recall curve to show its discriminatory ability.

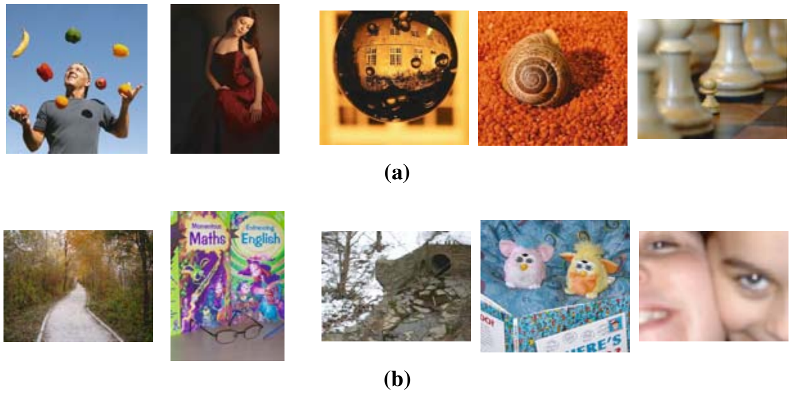
Fig. 7b shows the performance of our features. For comparison, Fig. 7a shows the performance of the features proposed by Ke *et al.* [6]. In low recall rates, the precisions



(a) (b) (c)

(d) (e) (f)

Fig. 7. Photo classification performance comparisons. Bayes classifier performance of (a) Ke's features, (b) our features, (c) combined features. One-dimensional SVM performance of (d) Ke's features, (e) our features. (f) Classification performance of different methods.



(a)

(b)

Fig. 8. (a) Five samples from the 1000 top ranked test photos by our features using the Bayesian classifier. (b) Five samples from the 1000 bottom ranked test photos by our features using the Bayesian classifier.

of all our four features are over 80%, but only the precision of the blur feature in Ke *et al.*'s method is over 80%. The clarity is the most discriminative of all the features. Fig. 7c shows two curves denoting the performances of Ke *et al.*'s features combined and our features combined, respectively. It is easy to see that our method outperforms Ke *et al.*'s. Part of the training and testing samples can be found in the supplementary materials.

To further test the performance of our features, we perform one-dimensional SVM on individual features. Figs. 7d and 7e show that four of our five features' classification error rates are below 30%, and those of the Ke's features' are all above 30%. We use the Bayesian classifier, SVM and gentle AdaBoost to test the performance of Datta *et al.*'s, Ke *et al.*'s and our methods. The results are given in Fig. 7f, from which we can clearly see that our algorithm greatly outperforms the other two algorithms. To show some examples, we randomly pick 5 samples from the ranked test photos and display them in Fig. 8. It is easy to tell the quality difference between the two groups.

One reason why our features perform much better than Ke *et al.*'s and Datta *et al.*'s is that we extract the subject region from a photo first and then define the features based on this region and the entire photo, while they developed their features from the whole photo only. Another reason is that we design our features mostly based on professional photography techniques.

5.2 Video Assessment

To demonstrate the effectiveness of our video quality assessment method, we collect a large and diverse video database from a video sharing website, YouTube.com. There are 4000 high quality professional movie clips and 4000 low quality amateurish clips. We randomly select 2000 high quality clips and 2000 low quality clips as the training set, and take the rest as the test set.

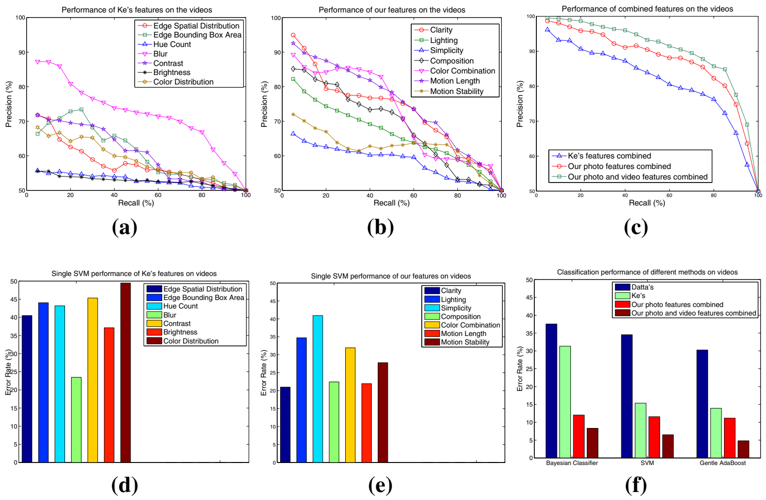


Fig. 9. Video classification performance comparisons. Bayes classifier performance of (a) Ke's features, (b) our features, (c) combined features. One-dimensional SVM performance of (d) Ke's features, (e) our features. (f) Classification performance of different methods.

To apply the features for photo assessment to a video, we select a number of frames from the video in a rate of one frame per second, and take the average assessment of these frames as the assessment of the video. Similar to the photo experiment, we first use Bayesian classifier to plot the precision-recall curve for each of our features and Ke *et al.*'s feature. Fig. 9a shows the performances of Ke *et al.*'s features, and Fig. 9b shows the performances of our features. Most of our features perform better than Ke *et al.*'s. Fig. 9c shows the performances of Ke *et al.*'s features combined, our photo features combined, and our photo and video features combined. Then we use the one-dimensional SVM to test individual features. Figs. 9d and 9e show the experiment results. We apply the three classifiers to compute the classification error rates with Datta *et al.*'s, Ke *et al.*'s and our features. The results are shown in Fig. 9f. The improvement of our method over the other is obvious.

5.3 Web Image Ranking

To further evaluate the usability of our image assessment method, we use it to rank the images retrieved by MSN Live Search. 50 volunteers aged between 18 and 30 took part in the experiment. They used 10 keywords randomly selected from a word list to search for the images. The top 1000 images in each search were downloaded. Then the volunteers gave them scores, ranging from 1 to 5(5 is the best). We use Ke *et al.*'s classification method and ours to re-rank these images. Fig. 10c shows the average scores of the top 1 to top 50, top 51 to top 100, ..., top 951 to top 1000 images, respectively. From Figs. 13a and 13b, we can see that the MSN Live Search engine does not consider the quality of the images. After re-ranking of these images by our method, the top ranked

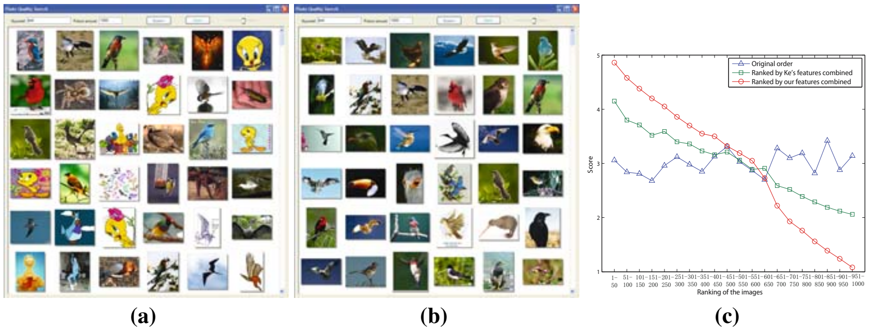


Fig. 10. (a) The first page of the images searched by MSN Live Search with the key word “bird”. (b) The first page of the images after re-ranking by our classification system. (c) The average scores of top rank images by different ranking system.

images are of higher quality. Ke *et al.*’s method can also improve the original ranking but does not perform as well as ours. Figs. 10a and 10b show an example of the images ranked by our system. More examples can be found in the supplementary materials.

It should be noticed that the photo quality ranking is not the only feature for image search re-ranking. Better photo quality does not mean more relevant. We plan to combine this work with other image search re-ranking work [19] in our future research.

6 Conclusion

In this paper, we have proposed a novel method to assess photo and video quality. We first extract the subject region from a photo, and then formulate a number of high level semantic features based on professional photography techniques to classify high quality and low quality photos. We have also conducted the first video quality evaluation study based on professional video making techniques. The performance of our classification system using these features is much better than the previous work. Our algorithm can be integrated into existing image search engines to find not only relevant but also high quality photos. Notice, one strength of our algorithm is that only using very simple features we achieve very good results. It is certainly possible to improve with more sophisticated design of features. The data used in this paper can be downloaded at http://mmlab.ie.cuhk.edu.hk.

References

1. Wang, Z., Sheikh, H.R., Bovik, A.C.: No-reference perceptual quality assessment of JPEG compressed images. ICIP (2002)
2. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Processing 13 (2004)
3. Sheikh, H., Bovik, A., de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans. Image Processing 14 (2005)

4. Tong, H., Li, M., Zhang, H., He, J., Zhang, C.: Classification of Digital Photos Taken by Photographers or Home Users. In: Proc. Pacific-Rim Conference on Multimedia (2004)
5. Datta, R., Joshi, D., Li, J., Wang, J.: Studying Aesthetics in Photographic Images Using a Computational Approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951. Springer, Heidelberg (2006)
6. Ke, Y., Tang, X., Jing, F.: The Design of High-Level Features for Photo Quality Assessment. In: CVPR (2006)
7. Freeman, M.: The Complete Guide to Light and Lighting. Ilex Press (2007)
8. Freeman, M.: The Photographer's Eye: Composition and Design for Better Digital Photos. Ilex Press (2007)
9. London, B., Upton, J., Stone, J., Kobre, K., Brill, B.: Photography, 8th edn. Pearson Prentice Hall, London (2005)
10. Itten, J.: Design and Form: The Basic Course at the Bauhaus and Later. Wiley, Chichester (1975)
11. Manav, B.: Color-Emotion Associations and Color Preferences: A Case Study for Residences. Color Research and Application 32 (2007)
12. Gao, X., Xin, J., Sato, T., Hansuebsai, A., Scalzo, M., Kajiwara, K., Guan, S., Valldeperas, J., Lis, M., Billger, M.: Analysis of Cross-Cultural Color Emotion. Color Research and Application 32 (2007)
13. Levin, A.: Blind motion deblurring using image statistics. In: NIPS (2006)
14. Tokumaru, M., Muranaka, N., Imanishi, S.: Color design support system considering color harmony. In: Proc. of the 2002 IEEE International Conference on Fuzzy Systems, vol. 1 (2002)
15. Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T., Xu, Y.: Color harmonization. ACM Transactions on Graphics (TOG) 25 (2006)
16. Yan, W., Kankanhalli, M.: Detection and removal of lighting & shaking artifacts in home videos. In: Proc. of the tenth ACM international conference on Multimedia (2002)
17. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, Heidelberg (2000)
18. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). Ann. Statist 28 (2000)
19. Cui, J., Wen, F., Tang, X.: Real Time Google and Live Image Search Re-ranking. In: Proc. of ACM international conference on Multimedia (2008)