Personalized Photo Ranking and Selection System

Che-Hua Yeh[1], Yuan-Chen Ho[1], Brian A. Barsky[2], Ming Ouhyoung[1]

National Taiwan University[1]
Dept. of CSIE & GINM
Taipei 10617, Taiwan, R.O.C.
{chyei, goheel, ming}@csie.ntu.edu.tw

University of California, Berkeley[2]
Computer Science Division and School of Optometry
Berkeley, CA 94720-1776, USA
barsky@cs.berkeley.edu

ABSTRACT

We have proposed a novel personalized ranking system for amateur photos. While the features used in our system are similar to previous works, new features such as texture, RGB color, portrait (through face detection), and black-and-white are included for individual preferences. Although automatically ranking award-wining professional photos may not be a sensible pursuit, such an approach may be reasonable for photos taken by amateurs, especially when individual preference is taken into account. We show that (1) The performance of our system in terms of precision-recall diagram and binary classification accuracy (93%) is close to the best results to date for both overall system and individual features. (2) Two personalized ranking user interfaces are provided: the feature-based and example-based. Both are effective in providing personalized preferences, and twice more people prefer example-based in our user study.

Keywords

Photo Ranking, Personalized ranking, Example-driven re-ranking, Aesthetic Rules, Photo Composition, Color Distribution, Ordinal Ranking

1. INTRODUCTION

With the current widespread use of digital cameras, the process of selecting and maintaining personal photos is becoming an onerous task. To reduce the growing amount of photos and browsing time, visually pleasing photos are expected to be kept while unattractive ones tend to be discarded. Since this process is time consuming, computation-based solutions are needed to assist in photo maintenance. However, since the judgement involves subjectivity and personal taste, solutions based on computation will always face challenges and difficulties. Regardless of these shortcomings, *computational aesthetics* is proposed to predict the emotional response to works of art [17, 18]. There are other topics using similar approach, such as photo optimization and photo assessment. Photo optimization based on aesthetics has been proposed in previous works [21, 16, 12], in this paper, we will focus on photo assessment and ranking.

Various works have been proposed to select "high quality" photos. A number of works assess photos concerning image qualities such as degradation caused by noise, distortion, and artifacts [28, 27, 23]. On the other hand, Tong *et al.* [26] and Datta *et al.* [5] try to classify professional photos and non-professional photos with low-level features utilized in image retrieval. Ke *et al.* [9] used more visual features such as edges, blurriness, brightness and hue for classification. Then binary classification accuracy is used to evaluate the results. Photos are labeled into two classes, such as professional and non-professional, and then predicted by the system. The performance of the system is often determined by prediction accuracy. In Ke's method, 72% classification rate is achieved on a set of 3,000 photos. Our previous work ranks a photo by 9 rules based on aesthetics [30]. The rules include horizontal balance, line patterns, size of ROI (region of interest), merger avoidance, the rule of thirds, color harmonization, contrast, intensity balance and blurriness. 81% accuracy is achieved on a set of 2,000 photos. In the work of Luo *et al.* [13], they try to treat foreground and background differently instead of extracting features from the whole photo, and over 93% classification rate is achieved using 12,000 photos. Using the same data as in [13], our system also has a 93% classification rate, and additional personal preference in re-ranking is provided.

In these works, performance is often evaluated by the accuracy of binary classification. However, even within two-class photos, there are still ranks in photos. In the work of San Pedro *et al.*, Kendall's tau coefficient is utilized to measure the similarity between their ranking results and the groundtruth [20, 11]. Kendall's tau coefficient ranges from 1 to -1, which 1 stands for perfect agreement between the two rankings, and -1 stands for perfect disagreement. In their work, a Kendall's tau value 0.25 is derived for the ranking based on visual features of 70,000 photos collected from the Flickr website. The value gives a low agreement between the ranking list and groundtruth, so the authors improve the value to 0.48 by combining tag information of photos. In our work, Kendall's tau will be also utilized to evaluate the results, and 0.43 is achieved without using photos' tag information.
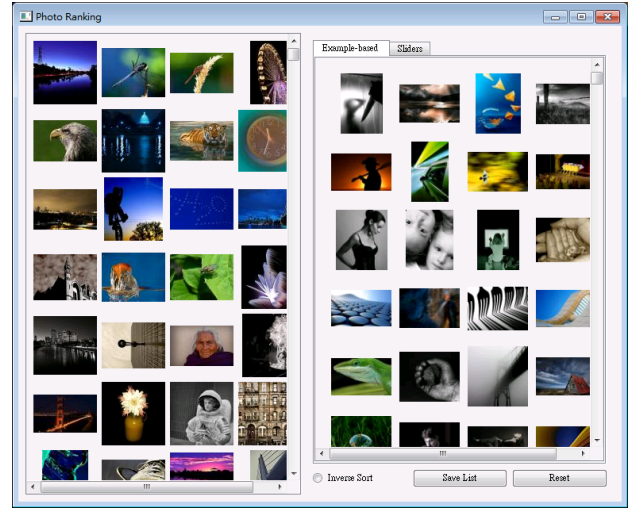
With efforts of these works, there seem to be feasible solutions for automatic photo ranking and selecting. However, the most challenged question is that the results tend to be subjective. The judgement of aesthetics involves sen-
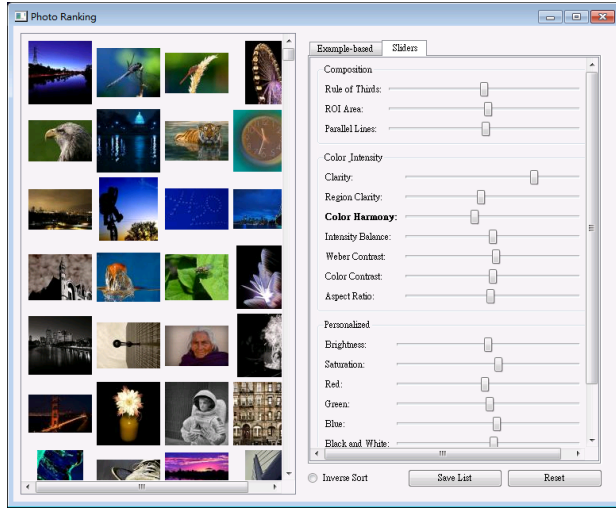
Figure 1: Our system for personalized photo ranking, where 1,000 ranked photos are shown on the left part of both figures (a) Re-ranking photos by adjusting the feature weightings (b) Re-ranking photos by selecting a few example photos from the right part.

timents and personal taste [5, 15]. Everyone has his or her unique ways to rank photos. A fixed ranking list just can not meet everyone's requirements, which is a situation similar to the interior design of individual houses. Sun *et al.*'s work adopted the idea of personalization [25]. Personalized photo assessment is achieved by taking user preference into account, but the assessment is only based on the percentage of the saliency region is that is covered by a predefined region, and a small size of 600 photos and 3 subjects are involved in their experiments.

In this paper, we propose a system to re-rank photos according to individual preferences. We use Listnet to derive the weightings of rules employed to rank photos [2]. By adjusting the weightings, photos can be re-ranked immediately. Example-based user interface can also be used as one's favorite style to modify the final results.

2. SYSTEM OVERVIEW

Figure 1 is the user interface panel of our personalized photo ranking system. Figure 1(a) is the scenario of re-ranking photos by adjusting the feature weightings and figure 1(b) shows re-ranking by selecting example photos using the photos on the right half. For this demo program, 1,000 photos are listed from high ranking scores to low scores.

Figure 2 is the overview of our system. Training photos are separated into two classes: preferred and non-preferred. Rules used for feature extraction will be covered in section 3.

The score of each photo can be considered as a linear combination of each feature and its corresponding weighting factor. After feature extraction, the ListNet is adopted to train the prediction model by finding the optimal weightings for each feature. Once the optimal weightings are found, photos can be ranked according to their scores. However, these weightings are generated from the training set, and they may not agree with individual user's personal preference. Therefore, we create a system for users to combine their personal

tastes with a trained model to produce the tailor-made results for each individual.

Two methods are provided for weighting adjustments: feature-based and example-driven. If users are well aware of the meaning of features to be emphasized, they can manually update the weighting for corresponding features. We provide 18 features for users to customize their ranking lists. On the other hand, user can also select some of the photos they like from our database and the system will update the weighting based on the few example photos.

3. RULES OF AESTHETICS

Rules of aesthetics in photography describe how to arrange different visual elements inside an image frame. We categorize these rules into two major parts: photo composition and color distribution.

3.1 Photo Composition

Composition is the placement or arrangement of visual elements in a photo. No absolute rules exist that guarantee perfect composition for all photos. However, there are still some heuristic principles that suggest an pleasing composition for most people when applied properly.

3.1.1 Rule of Thirds

The rule of thirds is the most well-known photo composition guideline [7, 10]. Photographers are encouraged to place main subjects around one third of the horizontal or vertical dimension of the photograph. Figure 3 shows one example.

To measure how close the main subjects are placed near power points, the position of main subjects should be located in each picture. First, each photo is segmented into homogenous patches using a graph-based segmentation technique [6]. Figure 4(a) shows the original photo, and figure 4(b) shows the segmented results. Then a salient value is assigned to each pixel based on Achanta's method [1]. The
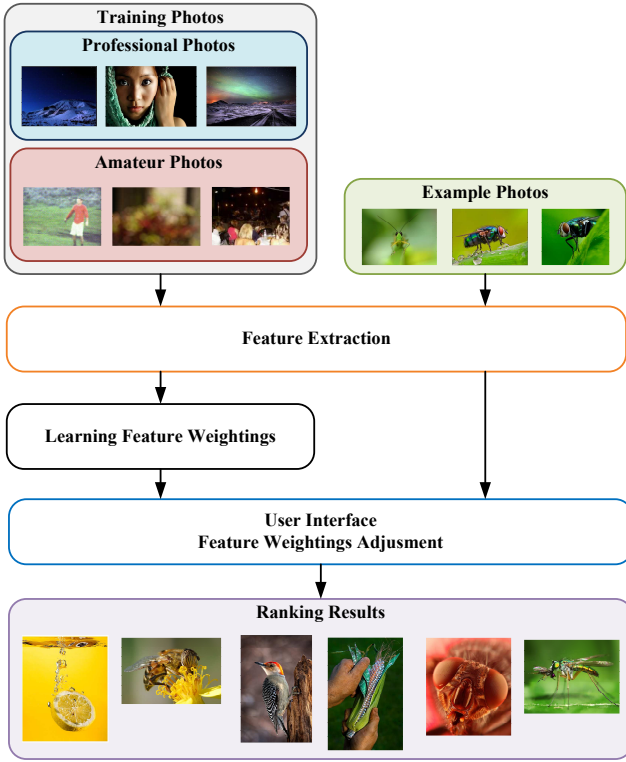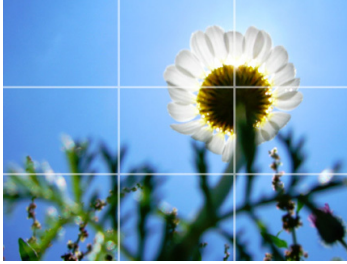
Figure 2: System overview

Figure 3: Example of rule of thirds: the flower is located at one of the "power points"

saliency value is the difference between the Lab color vector and the average Lab vector for the entire image:

$$S(x,y) = |I_u - I_{whc}(x,y)|$$

where I_u is the arithmetic mean pixel value of the image and I_{whc} is the Gaussian blurred version of the original image.

A salient value is then assigned to each patch by averaging the saliency for the pixels that it covers. Figure 4(c) shows the salient map and figure 4(d) shows the combined map with segmentation map (figure 4(b)).

The rule of thirds is then measured by the model:

$$f_{ROT} = \frac{1}{\sum_i A_i S_i} \sum_i A_i S_i e^{-\frac{D_i^2}{2\sigma}} \qquad (1)$$

where A_i is the patch size and S_i is it's average salient value. D_i is minimal one of four distance values from the patch center to the four power points ($\sigma = 0.17$). If main subjects

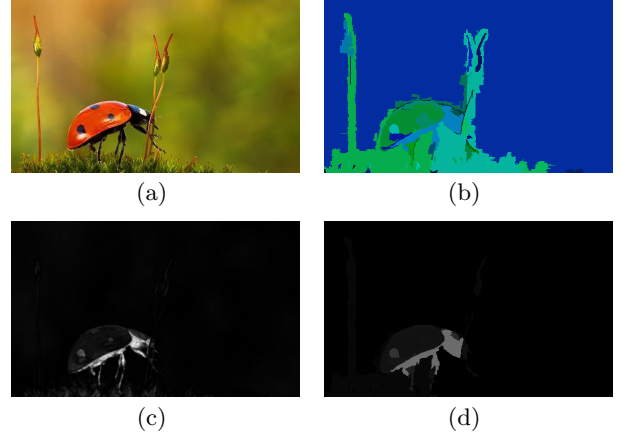are more close to the four points, the value of f_{ROT} is larger.

Figure 4: Locating subject (a) Original photo (b) Segmented photo (c) Saliency map (d) Combination of saliency map and segmented map

3.1.2 Simplicity

Simplicity in a photo is mentioned in [9] as a distinguishing factor to determine whether a photo is professional or not. In our modified version, two types of features are used to measure the simplicity with the photo: size of ROI segments and the simplicity feature proposed in Luo *et al.*'s work [13].

The ROI map of the photo is converted to a binary ROI map by applying the threshold :

$$B_{ROI} = \begin{cases} 1, & \text{if } x < \alpha \text{Max}_{ROI}, \quad \alpha = 0.67 \\ 0, & \text{otherwise.} \end{cases}$$

After obtaining the binary ROI map, bounding boxes are generated for each of the non-overlapping salient regions and the area for all bounding boxes is summed up:

$$f_{ROIArea} = \sum_{i=1}^{n} \frac{Area_i}{wh} \qquad (2)$$

where w and h are width and height of the photo, respectively.

In addition to the size of ROI segments, we also include one of the features from [13] which defines simplicity as the "attention distraction of the objects from the background". The subject region of a photo is extracted and the rest of the photo is considered as the background region. The color distribution of the background is used to evaluate the simplicity of a photo. The RGB channels are quantized respectively into 16 different levels and the histogram H of 4096 bins is generated for the photos. The simplicity feature is defined as:

$$f_{Simp} = \left(\frac{\|S\|}{4096} \right) \times 100\% \qquad (3)$$

where $s = \{i | H(i) \geq \gamma h_{max}\}$, and $\gamma = 0.01$. In table 1(b), it shows that our modified simplicity feature performs better with 89.45% over the original method (73%).

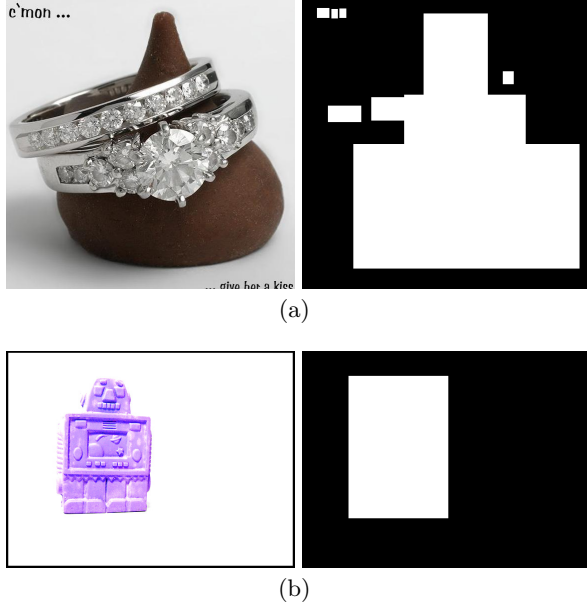## 3.2 Color and Intensity Distribution

Figure 5: Region of Interest (ROI) Area size feature (a) Large ROI region, where the white area in the right part means ROI area (b) Small ROI region
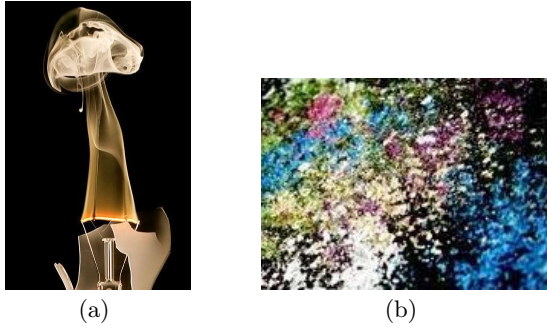

Figure 6: Simplicity feature (a) High simplicity (b) Low simplicity

3.2.1 Texture

Texture is one of the important features for image retrieval, and it also conveys the idea of repetitive patterns or similar orientations among photo components. Photographers also consider texture richness as a positive feature since repetitions and similar orientations not only extend viewers' perspective depth but also reflect the sense of harmony. Therefore, we include this feature, while it is missing from all the other photo-ranking related papers [5, 9, 12, 13, 20, 26].

We decide to use the homogeneous texture descriptor defined in MPEG-7 standard to extract and describe the texture richness of the photos [19]. MPEG-7 homogeneous texture descriptor utilizes the fact that the human brain decomposes the spectra into perceptual channels that are bands in spatial frequency and uses Gabor filter to evaluate the convolution responses of the image under different scales and orientations [3, 14].

The Gabor wavelets (kernels, filters) can be defined as

follow:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{\left(-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{izk_{u,v}} - e^{-\frac{\sigma^2}{2}} \right]$$

where

$$k_{u,v} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \phi_u \\ k_v \sin \phi_u \end{pmatrix}, \ k_v = \frac{f_{max}}{2^{\frac{v}{2}}}, \ \phi_u = u(\frac{\pi}{8}),$$

$v = 0, ..., v_{max} - 1, \ u = 0, ..., u_{max} - 1$. MPEG-7 homogeneous texture descriptor consists of mean and variance of the image intensity and the combination of five different scales $\{0, 1, 2, 3, 4\}$ and six different orientations $\{30°, 60°, 90°, 120°, 150°, 180°\}$. Actually this texture feature performs well (84.15%) as shown in table 1(b).

3.2.2 Clarity

Out-of-focus photos are usually regarded as poor photos, and previous works have included blurriness as one of the most important features for determining the quality of the photos [26, 9]. The photos are transformed from spatial domain to frequency domain by Fast Fourier Transform, and the pixels greater than the threshold are considered as clear pixels ($t = 2$).

$$f_{blur} = \frac{number \ of \ clear \ pixels}{total \ pixels} \qquad (4)$$

Besides, bokeh, as figure 4(c) shows, describes the rendition of out-of-focus points of light and is one of the common techniques employed by professional photographers to emphasize on the main objects. We manage to detect bokeh by partitioning a photo into grids and applying blur detection on them.

$$Q_{bokeh} = \frac{number \ of \ clear \ grids}{total \ grids}$$

Since bokeh is a combination of clear and blurred grids, photos are not bokehs if they are either entirely clear or entirely blurred. We also exclude grids with low color variations because they sometimes produce false alarm.

$$f_{bokeh} = \begin{cases} 1, & \text{if } 0.3 \leq Q_{bokeh} \leq 0.7 \\ 0, & \text{otherwise.} \end{cases} \qquad (5)$$

Figure 7: Clarity feature (a) High clarity (b) Low clarity

3.2.3 Color Harmonization

Harmonic colors are known to be aesthetically pleasing in terms of human visual perception, and it is used to measure the quality of color distribution for the photos. The optimization function defined by [4] is:

$$F(X, (m, \alpha)) = \sum_{p \in X} \left\| H(p) - E_{T_{m(\alpha)}}(p) \right\| \cdot S(p) \qquad (6)$$

where H and S are hue and saturation channels for a photo, respectively. X is the input image and each pixel in the image is denoted by p. The best color template m and the best offset α are chosen to minimize the optimization function yet to create the most pleasant visual result, and we define our color feature accordingly.
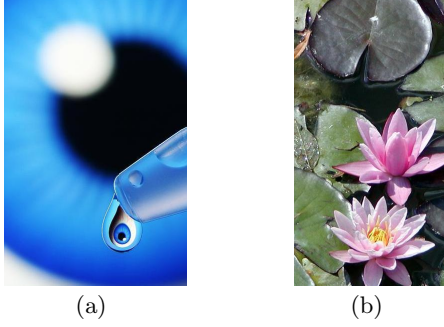

Figure 8: Color Harmonization feature (a) Harmonic color (b) Less harmonic color

3.2.4 Intensity Balance

Balance provides a sense of equilibrium and is also a fundamental principle of visual perception that eye seeks to balance the elements within a photo. Photo composition can be regarded as setting the positions of objects within a photo and balancing the objects with respect to lines or points that brings in the harmony. Weight for each pixel is given according to its intensity and two sets of histograms are produced for left and right parts of the image. The histograms are later converted into chi-square distributions to evaluate the similarities between them.

$$f_{balance} = \left| \sqrt{\sum_{i=1}^{k} (E_{left} - E_{right})} \right| \qquad (7)$$
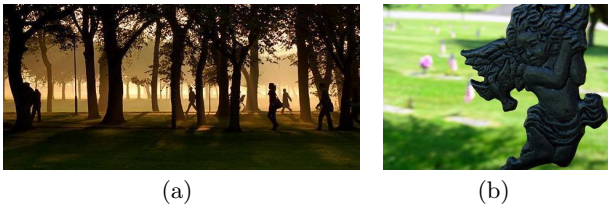

Figure 9: Intensity balance feature (a) balanced (b) left-right unbalanced

3.2.5 Contrast

Contrast is defined as the dissimilarity between components within a picture. In our system, we measure two types of contrasts: Weber contrast and color contrast. Weber contrast for any given image is defined as:

$$f_{WeberContrast} = \frac{1}{width} \frac{1}{height} \sum_{x=0}^{width} \sum_{y=0}^{height} \frac{I(x,y) - I_{avg}}{I_{avg}} \qquad (8)$$

where $I(x,y)$ represents the intensity at (x,y) position of the image and I_{avg} is the average intensity of the image. While Weber contrast measures the disparity between components in terms of intensity values, we would also like to consider the color dissimilarity within the photo. Therefore, we use the color difference equation by CIE 2000 to determine color contrast [22].

Image segmentation method is applied to photos and the mean color is computed for each segment [6]. Color disparity is calculated and summed for each pair of segments according to their mean colors and the sum is later normalized by the number and the size of color segments.
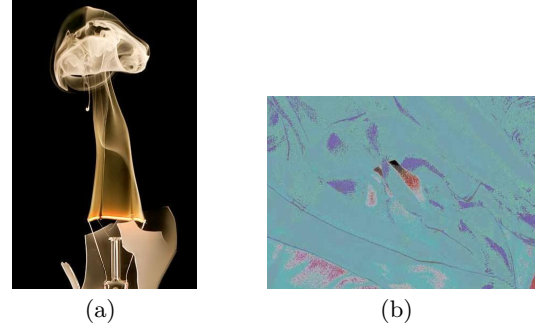

Figure 10: Contrast feature (a) High contrast (b) Low contrast

$$f_{ColorContrast} = \sum_{i=0}^{n} \sum_{j=i+1}^{n} (1 - D(i,j)) \frac{C(i,j)}{M_i M_j} \qquad (9)$$

where $D(i,j)$ is the relative distance between two segments and $C(i,j)$ is the color dissimilarity between the two segments. The combined result of Weber and CIE2000 contrasts give a rather good feature (84.12%) shown in table 1(b).

3.3 Personalized features

Photos can be assessed based on aesthetic rules, but only those rules can not meet personal taste completely. For example, someone may prefer photos based on a specific color style, or prefer photos with a high color saturation or high intensity. Some even prefer portrait over scenic photos. Although these properties are not suitable for assessing photos, it is still necessary to take them as features. These personalized features are described in subsections.

3.3.1 Color preference

Any color can be described by brightness, saturation, and hue. In photo selection, sometimes specific color style is required by users. For example, green color contributes more than other colors in plant photos, while blue color does in sea and sky photos. To meet each user's preference in color style of photos, three color preference features are added in our system. They are brightness, saturation, and RGB channels.

Brightness, also referred as intensity, records the average intensity of whole pixels in each photo. The saturation of whole pixels are averaged as a feature. Instead of hue, RGB channels are utilized as features since the interface is more friendly to users with RGB features than the hue feature. Average values of whole pixels are calculated separately in red, green, and blue channel, and grayscale pixels are omitted. Consequently, the ratios of red, green, and blue over the sum of 3 channels are calculated and assigned as features.

3.3.2 Black-and-white ratio

Appropriate color arrangements can make photos more attractive and outstanding. Relative to colorful photos, composition is the only determining factor in black and white photos. To distinguish these photos from color photos, one feature descriptor is added to indicate if a photo is colorful. The black and white feature is also treated as a personalized factor.

3.3.3 Portrait with face detection

Faces are treated as a part of region of interest in photos. Besides, faces are also selected as one of personalized features since users would prefer photos of human figures.

3.3.4 Aspect Ratio

The aspect ratio of photos can affect photo composition. The well-known 4:3 and 16:9 aspect ratios are sometimes considered as "golden ratio", however, other aspect ratios may also be chosen for viewing pleasure.

$$f_{AspectRatio} = \frac{width}{height} \qquad (10)$$

4. PERSONALIZED RANKING

4.1 Ranking and ListNet

Relative to classification problem, ranking generates an ordered list according to certain criteria, e.g. utility function. A ranking algorithm assigns a relevant score to each object, and the score order represents the relevance to the goal function. A ranking algorithm is trained with a set of data, to be utilized to predicts ranking results. The training procedure of ranking algorithms is commonly referred as *learning to rank*.

In our work, a set of photos is selected as training photos: $D = (d_1, d_2, ..., d_N)$, where N denotes the number of training photos and d_i denotes the i-th photo. There is a corresponding score for each photo in training photos: $Y = (y_1, y_2, ..., y_n)$, where y_i denotes the relevance score of d_i. A feature vector $X_i = (x_i^1, x_i^2, ..., x_i^M)$ of M dimensions is extracted from each photo based on the rules described in section 3. A ranking algorithm f is trained to predict the scores of test data by leveraging the co-occurrence patterns among feature X and score Y. While training the ranking algorithm, a list of predicted score of training photos D, $Z = (z_1, z_2, ..., z_N) = (f(X_1), f(X_2), ..., f(X_N))$, is obtained. The ranking algorithm f is optimized by minimizing the loss function $L(Y, Z)$.

Listnet is adopted in our work since existing works shows it is efficient and even outperforms the conventional approaches such as RankSVM [8, 29]. Listnet employs cross-entropy between two probability distributions of input scores and predicted scores as listwise loss function. The function is defined as:

$$L(Y, Z) = -\sum_{i=1}^{N} P(y_i) log(P(z_i))$$

The loss function is minimized with a linear neural network model. A weight is assigned to each feature and the predicted score is the sum of linear weighted features.

$$z_i = f(X_i) = W \cdot X_i$$

$W = (w_1, w_2, ..., w_M)$ is the weighting vector of features. The gradient with respect to each w is derived via gradient descend:

$$\Delta w_j = \frac{\partial L(Y, Z)}{\partial w_j} = \sum_{i=1}^{N} (P(z_i) - P(y_i)X_{ij}$$

Each w is initially assigned to zero. In each iteration, w is updated with

$$w_j = w_j - \eta \times \Delta w_j$$

where η is the learning rate. The iteration terminates if the change in W is less than a convergent threshold.

4.2 Personalization

After deriving the weightings for each feature, scores of new photos can be generated and a ranked list is produced based on the scores. Personalized ranking is further realized by modifying the weightings manually.

Example-Base: Weighting adjustment by example photos is also provided in our system. An weighting vector is associated with each example photo and each entry of the weighting update vector is defined as:

$$w_j = w_j + \sum_{i \in S} F(x_i^j)$$

where

$$\sum_{i \in S} F(x_i^j) = \begin{cases} \sum_{i \in S} \left\lfloor \frac{x_i^j - m^j}{\sigma_j} \right\rfloor * u, & \text{if voting members of } S \text{ "all" agree} \\ 0, & \text{if two or more voting members contradict to each other} \end{cases}$$

where x_i^j is the j-th feature value for the photo i, m^j is the mean value of feature j from all training photos, σ_j is standard deviation of feature j, $\lfloor \ \rfloor$ is a floor function, and u is a fixed step size. S is the set of selected example photos. Function F is a voting mechanism, which determines whether selected photos are consistent in features. If two or more photos contradict to each other in a specific feature, the feature will not be updated.

5. EXPERIMENTS AND USER STUDY

All data are selected from a photo contest website, DPChallenge.com, which contains diverse varieties of photos from different photographers. Each photo is rated, ranging from 1 to 10, by at least two hundred users to reduce the influence of the outliers. The 6,000 highest-rated and 6,000
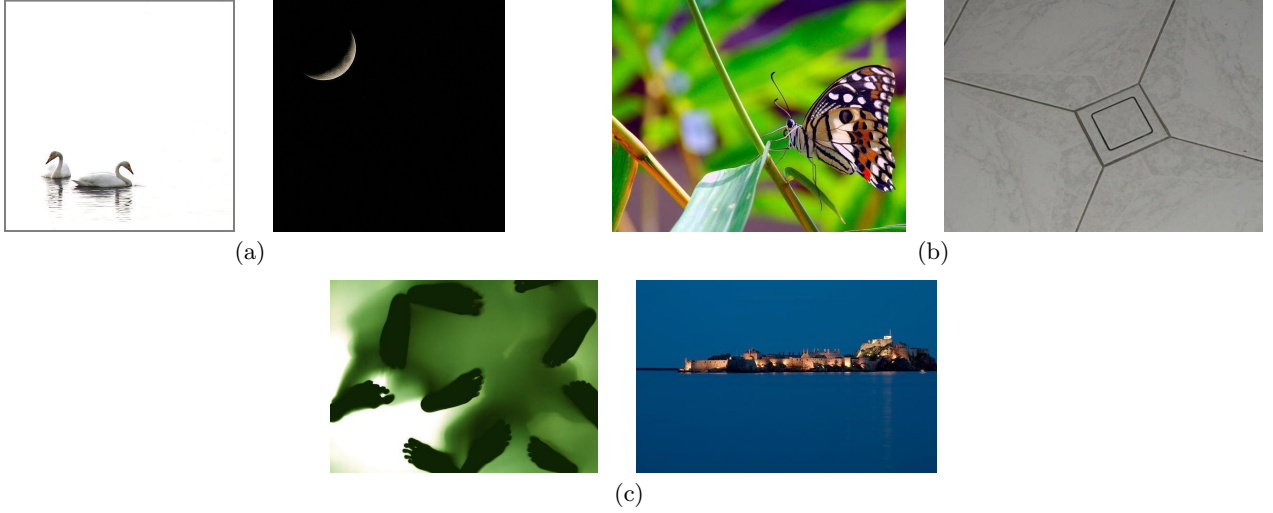
Figure 11: Color preference (a) High brightness and low brightness (b) High saturation and low saturation (c) Color style (when green and blue are selected)

lowest-rated photos are used for our experiments, the same as the data used in [13].

5.1 Ranking

3,000 top ranked photos and 3,000 bottom ranked photos are selected to train our system by the ranking algorithm, Listnet. The corresponding score for each photo is it's rank. After the weightings of features are learned, the rest 6,000 photos are used for testing. Kendall's Tau-b coefficient is utilized to evaluate our ranking results.

$$\tau_b = \frac{n_c - n_d}{\sqrt{(n_0 - t_1)(n_0 - t_2)}}$$

n_0 is the number of all pairs, n_c is the number of concordant pairs, n_d is the number of discordant pairs in the lists, t_1 is the number of pairs tied in the first list, and t_2 is the number of pairs tied in the second list. A Kendall's Tau-b value of 0.4228 is derived from the predicted score list of test data, and the value indicates the agreement between two lists is not weak.

5.2 Binary Classification

Given so many features, how to combine them together in the binary classification problem? We use the "late fusion" technique [24], where the "voting strategy" is used, with the voting weighting of each feature being determined by the training phase accuracy. We used the best three features (simplicity, texture, and contrast) in voting, and the result is 93%, similar to what is reported in Luo *et al.*'s work [13]. However, Luo *et al.* used three different approaches (Bayes, SVM, Gentle Adaboost), where Gentle Adaboost got the best result (above 93%).

In figure 14, we compare the results from Ke *et al.*'s [9], Luo *et al.*'s [13], and our approach. However, we have to point out that Luo *et al.*'s is using Bayesian based and ours is using ListNet, while Ke *et al.*'s has much smaller database (2,000 for training). Luo *et al.*'s and ours use the same dataset of 12,000 photos (6,000 for training). Therefore, direct comparison can be misleading, and we just want to

show that the features proposed in this paper have been effective and the overall difference is small, actually both system is 93% in binary class classification.

In table 1, for binary classification problem, we can see that individual features used in Luo *et al.*'s and ours have very similar performance. We noticed that two features: simplicity and texture(our new feature) perform better even compared to the blur factor.

**Table 1:** SVM classification accuracy of single feature (a) Luo's features (b) Our features

(a) Luo's features[13]	
Features	**Accuracy**
Composition	79%
Clarity	77%
Simplicity	73%
Color Combination	71%
Lighting	62%

(b) Our features	
Features	**Accuracy**
Simplicity(modified)	89.48%
Texture	84.15%
Contrast	84.12%
Intensity Average	75.23%
Region Blur	71.03%

Some features, such as RGB colors, portrait(via face detection), and black-and-white may not perform well as individual feature in two-class classification problem, however, it is important for individual preference. In short, some of the features used by previous works have been proven effective, yet not enough for personal preference.

5.3 User Study

Two user studies are conducted to evaluate the effectiveness of our system. In the first user study, each subject was asked to adjust weightings by slider bars to generate a
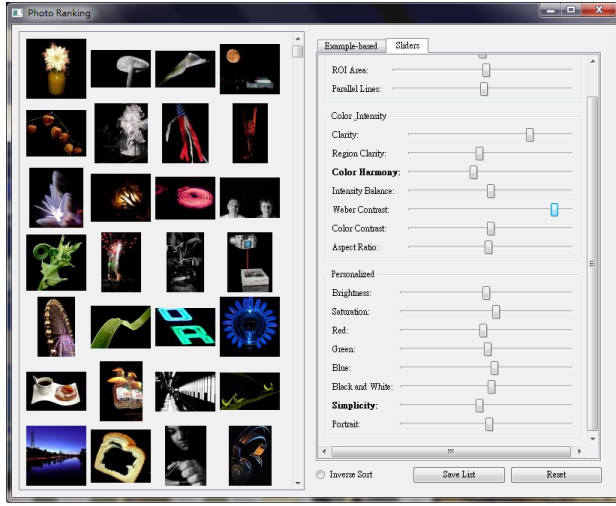
Figure 12: Ranking results with feature-based UI, where the left part is the ranked result, and the right part is for user manipulation. (a) Re-ranking photos by the contrast feature (b) Re-ranking photos by the black-and-white feature
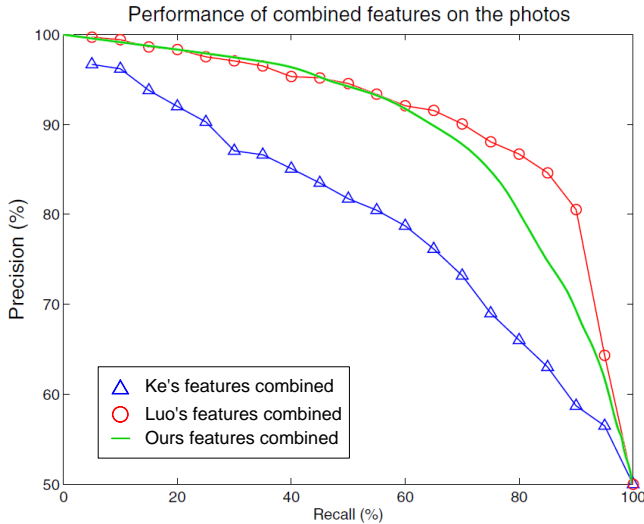

Figure 14: Precision Recall curve of three methods, where Ke's and Luo's use Bayes classifier, and ours uses Listnet.

new ranked list of photos. The newly-generated list is compared with the previous list to verify the effectiveness of our personalization process. Subjects are asked if the new list is closer to their preference and four options are given for their choice: very good, good, bad, and very bad. In the second user study, each subject was asked to select a few preferred photos (typically two to five) and our system will re-rank the list accordingly. The same four options are provide to examine their results.

Two thousand photos, one thousand highest-rated and one thousand lowest-rated from DPChallenge.com, are used in the two experiments with half of them as the training set and the other half as the testing set. A total of twelve subjects participated in both experiments, and each subject took 25 minutes on average.

For the first user study, the results for the four levels ("very good", "good", "bad", and "very bad") are: (8.3%, 91.7%, 0%, 0%), and for the second user study, the results are (0%, 83.3%, 16.7%, 0%). The results from the two experiments show that our system is able to re-rank the list more closer to users' preference.

In addition to two user studies, participants are also asked whether updating each feature manually or selecting example photos is the more effective and intuitive way for re-ranking the list; 66.7% of the users prefer example-based UI and 33.3% of the users prefer feature-based UI.

6. CONCLUSION AND FUTURE WORK

We have proposed a novel personalized ranking system for amateur photos. Although automatically ranking award-wining professional photos may not be a sensible pursuit, such an approach may be reasonable for photos taken by amateurs, especially when individual preference is taken into account. We show that (1) The performance of our system in terms of precision-recall diagram and binary classification accuracy (93%) is close to the best results to date for both overall system and individual features. (2) Two personalized ranking user interfaces are provided, the feature-based and example-based. Both are effective in providing personalized preferences, and twice more people prefer example-based in our user study.

In our study, more than 18 features are proposed and tested for ranking prediction, as given in and section 3. Three features are already very powerful, namely, simplicity(89.5%), texture(84%), and contrast(84%) as shown in table 1, and yet our current "late fusion" method can only provide 93% accuracy in binary classification. We will anticipate more sophisticated fusion in the future. Similary, our implementation of example-based UI is just one kind of realization, and we would like to see more.

7. PROJECT PAGE

The demo and supplementary materials can be down-
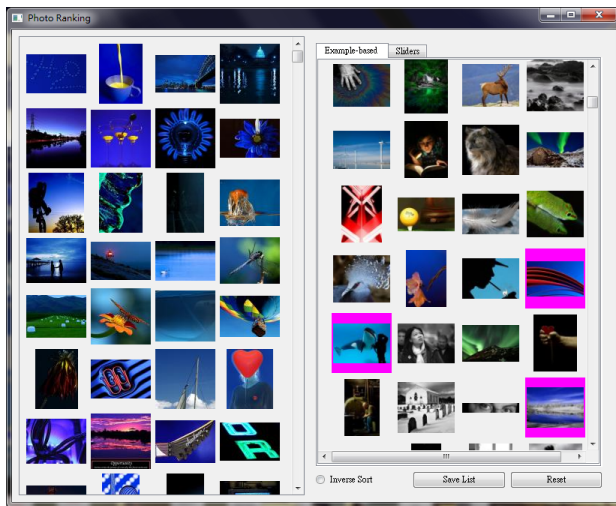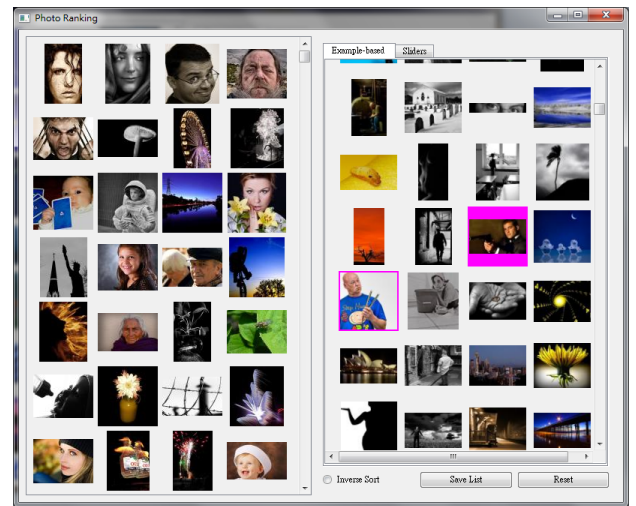
Figure 13: Ranking results with example-based UI, where the left part is the ranked result, and the right part is for example selection. (a) Re-ranking photos by blue color (b) Re-ranking photos by portrait

9. REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597 –1604, june 2009.

[2] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 129–136, New York, NY, USA, 2007. ACM.

[3] R. Chellappa. Two-dimensional discrete Gaussian Markov random field models for image processing. *Journal of the Institution of Electronics and Telecommunication Engineers*, 35(2):114–120, 1989.

[4] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu. Color harmonization. *ACM Trans. Graph.*, 25(3):624–630, 2006.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *In Proc. ECCV*, pages 7–13, 2006.

[6] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[7] T. Grill and M. Scanlon. *Photographic composition.* Amphoto Books, 1990.

[8] R. Herbrich, T. Graepel, and K. Obermayer. Support vector learning for ordinal regression. In *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)*, volume 1, pages 97 –102 vol.1, 1999.

[9] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 419 – 426, june 2006.

[10] B. Krages. *Photography: the art of composition.* Allworth Press, 2005.

[11] W. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, pages 814–861, 1958.

[12] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or. Optimizing photo composition. *Computer Graphic Forum (Proceedings of Eurographics)*, 29(2), 2010.

[13] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 386–399, Berlin, Heidelberg, 2008. Springer-Verlag.

[14] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(8):837 –842, aug 1996.

[15] B. Martinez and J. Block. *Visual forces: an introduction to design.* Prentice Hall, 1988.

[16] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 669–672, New York, NY, USA, 2009. ACM.

[17] G. Peters. Aesthetic primitives of images for visualization. In *Information Visualization, 2007. IV '07. 11th International Conference*, pages 316 –325, july 2007.

[18] V. RIVOTTI and J. PROENAA. J. JORGE J., SOUSA M.: Composition principles for quality depiction and aesthetics. In *The International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging*, pages 37–44, 2007.

[19] Y. Ro, M. Kim, H. Kang, B. Manjunath, and J. Kim. MPEG-7 homogeneous texture descriptor. *ETRI journal*, 23(2):41–51, 2001.

[20] J. San Pedro and S. Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 771–780, New York, NY, USA, 2009. ACM.

[21] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 771–780, New York, NY, USA, 2006. ACM.

[22] G. Sharma, W. Wu, and E. Dalal. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color research and application*, 30(1):21–30, 2005.

[23] H. Sheikh, A. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *Image Processing, IEEE Transactions on*, 14(12):2117 –2128, dec. 2005.

[24] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders. Early versus late fusion in semantic video analysis. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 399–402, New York, NY, USA, 2005. ACM.

[25] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 541–544, New York, NY, USA, 2009. ACM.

[26] H. Tong, M. Li, H. Zhang, J. He, and C. Zhang. Classification of digital photos taken by photographers or home users. *Lecture Notes in Computer Science*, pages 198–205, 2004.

[27] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600 –612, april 2004.

[28] Z. Wang, H. Sheikh, and A. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–477 – I–480 vol.1, 2002.

[29] Y. H. Yang, P. T. Wu, C. W. Lee, K. H. Lin, W. H. Hsu, and H. H. Chen. Contextseer: context search and recommendation at query time for shared consumer photos. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 199–208, New York, NY, USA, 2008. ACM.

[30] C.-H. Yeh, W.-S. Ng, B. A. Barsky, and M. Ouhyoung. An esthetics rule-based ranking system for amateur photos. In *SIGGRAPH ASIA '09: ACM SIGGRAPH ASIA 2009 Sketches*, pages 1–1, New York, NY, USA, 2009. ACM.