Personalized Photograph 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

August 23, 2010

Abstract

In this paper, we propose a novel personalized ranking system for amateur photographs. Although some of the features used in our system are similar to previous work, new features, such as texture, RGB color, portrait (through face detection), and black-and-white, are included for individual preferences. Our goal of automatically ranking photographs is not intended for award-wining professional photographs but for photographs taken by amateurs, especially when individual preference is taken into account. 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. Two personalized ranking user interfaces are provided: one is feature-based and the other is example-based. Although both interfaces are effective in providing personalized preferences, our user study showed that example-based was preferred by twice as many people as feature-based.

I.4.9Image Processing and Computer VisionApplications H.5.2Information Interfaces and PresentationUser Interfaces [Graphical user interfaces (GUI)]

Algorithms, Performance

Photograph Ranking, Personalized ranking, Example-based re-ranking, Aesthetic Rules, Photograph Composition, Color Distribution, Ordinal Ranking

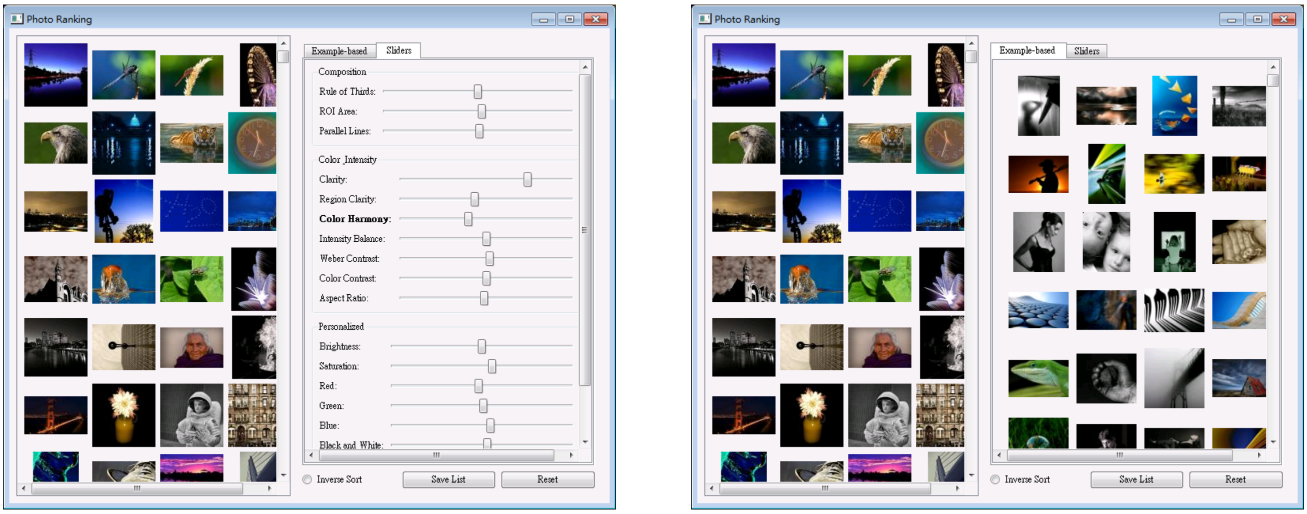
1 Introduction

With the current widespread use of digital cameras, the process of selecting and maintaining personal photographs is becoming an onerous task. To address the growing number of photographs and browsing time, it is desirable to discard unattractive photographs while retaining visually pleasing ones. Due to the time-consuming nature of this process, it would be useful to have computation-based solutions to assist in photograph maintenance. However, since the evaluation of photographs is subjective and involves personal taste, any solution based on computation will face challenges and difficulties. Notwithstanding these shortcomings, *computational aesthetics* is proposed to predict the emotional response to works of art [17, 18]. Also, there are other topics using a similar approach, such as photograph optimization and photograph assessment. Photograph optimization based on aesthetics has been proposed by several authors [21, 16, 12]. In this paper, we will focus on photograph assessment and ranking.

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

In this work, performance is often evaluated by the accuracy of binary classification. However, even within two-class photographs, there are still ranks in photographs. 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, where 1 indicates perfect agreement between the two rankings, and -1 means perfect disagreement. Their work results in a Kendall's tau value of 0.25 for the ranking based on visual features of 70,000 photographs collected from the Flickr website. This value indicates that there is only weak agreement between the ranking list and groundtruth, and thus the authors combine tag information of photographs to improve the value to 0.48. The results of our work are also evaluated using Kendall's tau; we achieve a value of 0.43 without using tag information of photographs.

These various efforts indicate that there are feasible solutions for automatic photograph ranking and selecting. How-



(a) (b)

Figure 1: Our system for personalized photograph ranking, where 1,000 ranked photographs are shown on the left side of the window in both figures (a) Re-ranking photographs by adjusting the feature weightings (b) Re-ranking photographs by selecting a few example photographs from the right part.

ever, one of the most challenging aspects is that the results tend to be subjective. The judgement of aesthetics involves sentiments and personal taste [5, 15]. Everyone has his or her unique way to rank photographs. A fixed ranking list simply cannot meet everyone's requirements, just like there is no universally preferred interior design of individual houses.

Sun *et al.* adopted the idea of personalization [25] in which personalized photograph assessment is achieved by incorporating user preference. However, the assessment is based only on the proportion of the saliency region that is covered by a predefined region, and uses only 600 photographs and three subjects in their experiments.

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

2 System Overview

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

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

The score of each photograph 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, photographs can be ranked according to their scores. However, these weightings are generated from the training set, and they might not agree with individual user's personal preferences. Therefore, the system enables users to combine their personal tastes with a trained model to produce results tailored to each individual.

Two methods are provided for weighting adjustments: feature-based and example-based. We provide 18 features that users can use to customize their ranking lists. Users who understand the features can emphasize some over others by manually adjusting the weighting for corresponding features Using the example-based approach, users can select some of the photographs they like from our database and have the system update the weighting based on these few example photographs.

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 categories: photograph composition and color distribution.

3.1 Photograph Composition

Composition is the placement or arrangement of visual elements in a photograph. Although there are no absolute rules that guarantee perfect composition for all photographs, there are nonetheless some heuristic principles which when applied properly suggest a composition that will be pleasing for most people.

3.1.1 Rule of Thirds

The rule of thirds is the most well-known photograph composition guideline [7, 10]. The idea is to place main subjects

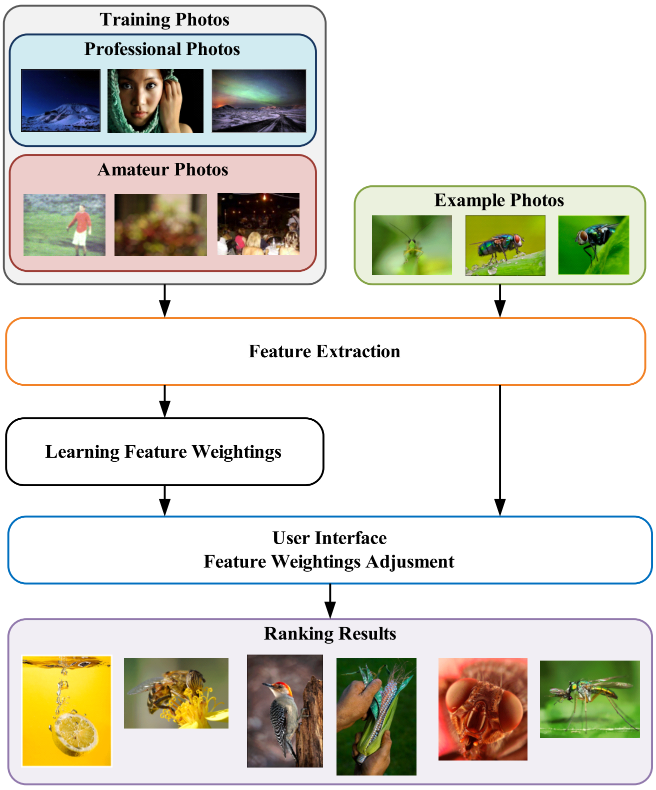


Figure 2: System overview

at roughly one-third of the horizontal or vertical dimension of the photograph. An example is shown in Figure 3.

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

To measure how close the main subjects are placed near these "power points", the position of main subjects should be located in each picture. First, each photograph is segmented into homogeneous patches using a graph-based segmentation technique [6]. Figure 4(b) illustrates the segmented results of the example photograph shown in Figure 4(a). Then a saliency value is assigned to each pixel based on Achanta's method [1], where the saliency value for a pixel is the difference between the color vector of the pixel and the average color vector for the entire image, in "Lab" color space:

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

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 saliency value is then assigned to each patch by averaging the saliency for the pixels that covered by the patch. The saliency map is shown in Figure 4(c). The combined segmented photograph and saliency map is shown in Figure 4(d).

The rule of thirds is then measured by the model:

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

where A_i is the patch size, S_i is the saliency value of the patch, and D_i is closest distance from the patch center to one of the four power points ($\sigma = 0.17$). If main subjects are closer to the four points, the value of f_{ROT} is larger.

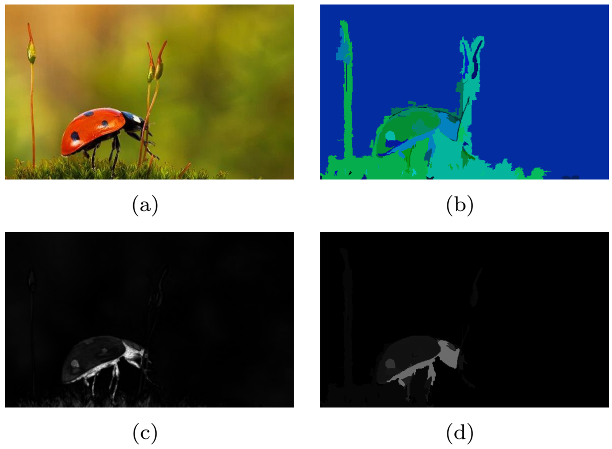


Figure 4: Locating subject (a) Original photograph (b) Segmented photograph (c) Saliency map (d) Combination of segmented photograph and saliency map

3.1.2 Simplicity

Simplicity in a photograph is a distinguishing factor in determining whether a photograph is professional or not [9]. We use two kinds of features to measure the simplicity of the photograph: size of ROI segments and the simplicity feature proposed by Luo *et al.* in [13].

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

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

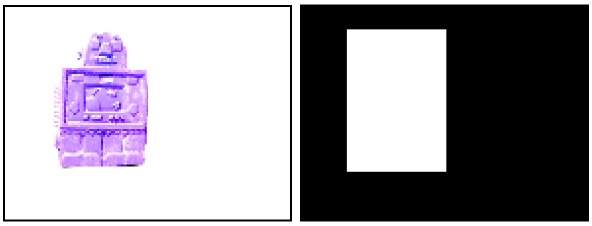
After obtaining the binary ROI map, bounding boxes are generated for each of the non-overlapping saliency regions and the area for all bounding boxes is summed:

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

where w and h are the width and height of the photograph, respectively. An example is shown in Figure 5.

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". An example is shown in Figure 6. We extract the subject region of a photograph and what remains is the background region and we use the color distribution of the background to evaluate the simplicity of the photograph. The RGB channels are quantized respectively into 16 different levels and the

(a)



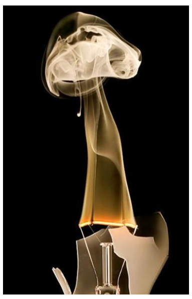
(b)

Figure 5: Region of Interest (ROI) Area size feature (a) Large ROI region, depicted as the white area in the right frame (b) Small ROI region

histogram (H) of 4096 bins is generated for the photographs. 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 $\gamma = 0.01$. Table 1(b) shows that our modified simplicity feature performs with 89.48% accuracy which is an improvement over the 73% accuracy of Luo's method.



(a)

(b)

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

3.2 Color and Intensity Distribution

3.2.1 Texture

We include texture as a feature, even though it is not included in any of the other photograph-ranking related papers [5, 9, 12, 13, 20, 26].

Texture is one of the important features for image retrieval, and it also conveys the idea of repetitive patterns or similar orientations among photograph components. Photographers also consider texture richness as a positive feature since repetitions and similar orientations not only extend viewers' perspective depth but also reflect a sense of harmony.

We use the homogeneous texture descriptor defined in the MPEG-7 standard to extract and describe the texture richness of the photographs [19]. The MPEG-7 homogeneous texture descriptor is based on the property of the human brain to decompose the spectra into perceptual channels that are bands in spatial frequency and it 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

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

$v = 0, ..., v_{max} - 1$, $u = 0, ..., u_{max} - 1$. the 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

Photographs that are out of focus are usually regarded as poor photographs, and previous work has included blurriness as one of the most important features for determining the quality of the photographs [26, 9]. Figure 7 shows an example. The photographs are transformed from spatial domain to frequency domain by a Fast Fourier Transform, and the pixels whose values surpass a threshold are considered as sharp pixels ($t = 2$).

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

However, bokeh describes the rendition of out-of-focus points of light and is an important techniques used by professional photographers to emphasize the main objects. We manage to detect bokeh by partitioning a photograph into grids and applying blur detection on them.

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

Since bokeh is a combination of sharp and blurred grids, we do not consider bokeh for photographs that are either entirely sharp or entirely blurred. We also exclude grids with low color variations because they sometimes produce an erroneous evaluation of low quality on what is really a high quality image.

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

(a) (b)

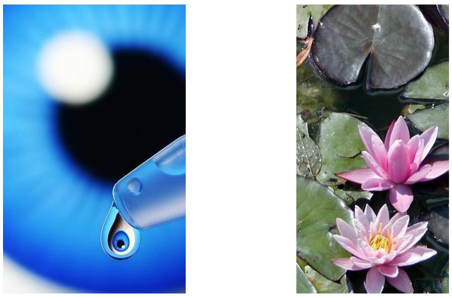
Figure 7: Clarity feature (a) High clarity (b) Low clarity

3.2.3 Color Harmonization

Harmonic colors are known to be aesthetically pleasing in terms of human visual perception, and we use this to measure the quality of color distribution for the photographs. Figure 8 shows an example. The optimization function defined by [4] is:

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

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



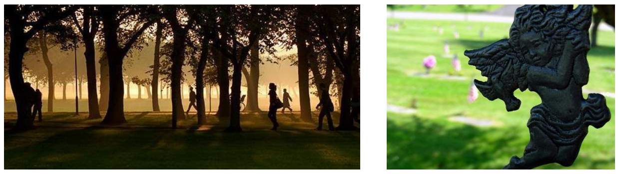
(a) (b)

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 in that the eye seeks to balance the elements within a photograph. Photographic composition involves organizing the positions of objects within the image and balancing them with respect to lines or points that establish the harmony. Figure 9 shows an example. The weight for each pixel is given according to its intensity. Two sets of histograms are produced for the left and right portions of the image. The histograms are later converted into chi-square distributions to evaluate the similarities between them.

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



(a) (b)

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

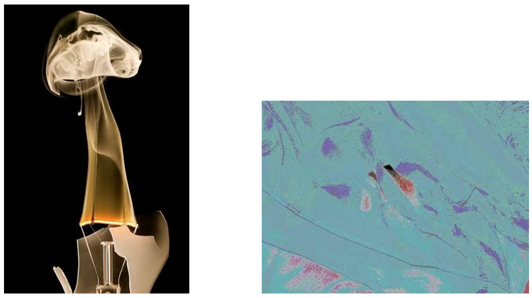
3.2.5 Contrast

Contrast can be defined as the dissimilarity between components within a picture. Figure 10 shows an example. In our system, we measure two types of contrasts: Weber contrast and color contrast. Weber contrast for any given image is defined as:

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

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

The image segmentation method is applied to photographs 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 then normalized by the number and the size of color segments.



(a) (b)

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

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

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 yields features with good accuracy (84.12%), as shown in Table 1(b).

3.3 Personalized features

Although photographs can be assessed based on aesthetic rules, these rules do not fully capture personal taste. For example, some may prefer photographs with a specific color style, or high color saturation, or high intensity, etc. Some even prefer portraits over scenic photographs. Although these properties are not suitable for assessing photographs, it is still necessary to include them as features. These personalized features are described in this section.

3.3.1 Color preference

Color can be represented by brightness, saturation, and hue. Some photograph selection is based on a specific color style. For example, the color green contributes more than other colors in plant photographs, whereas the color blue plays a dominant role in sea and sky photographs. An example is shown in Figure 11. To meet each user's preference in color style of photographs, we add three color preference features to our system: brightness, saturation, and RGB channels.

Brightness, also referred to as intensity, records the average intensity of whole pixels in each photograph. The saturation of whole pixels is averaged as a feature. RGB channels are used as features since this provides a friendlier user interface than the hue feature. Average values of whole pixels are calculated separately for each of red, green, and blue channels. Grayscale pixels are omitted. Consequently, the ratio of each of red, green, and blue divided by the sum of the three channels, is calculated and assigned as a feature.

3.3.2 Black-and-white ratio

Appropriate color arrangements can make photographs more attractive and outstanding. However, for black and white photography, composition is the primary determining factor. To distinguish black and white photographs from color photographs, one feature descriptor is added to indicate if a photograph 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 photographs and faces are also selected as one of personalized features since users may prefer photographs of human figures.

3.3.4 Aspect Ratio

The aspect ratio of photographs can affect photograph composition. The aspect ratio of 4:3 and 16:9 are often used.

$$f_{AspectRatio} = \frac{width}{height} \tag{10}$$

4 Personalized Ranking

4.1 Ranking and ListNet

Related to the 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 predict ranking results. The training procedure of ranking algorithms is commonly referred to as *learning to rank*.

In our work, a set of photographs is selected as training photographs; we denote the set by $D = (d_1, d_2, ..., d_N)$, where d_i is the i-th photograph, and N is the number of training photographs. For each training photograph in the set, there is a corresponding score, forming a set of scores denoted by $Y = (y_1, y_2, ..., y_N)$, where y_i is the relevance score of photograph d_i. A feature vector, denoted $X_i = (x_i^1, x_i^2, ..., x_i^M)$ where M is the number of dimensions, is extracted from each photograph 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 scores, denoted $Z = (z_1, z_2, ..., z_N) = (f(X_1), f(X_2), ..., f(X_N))$, is obtained for the set D of training photographs. The ranking algorithm f is optimized by minimizing the loss function $L(Y, Z)$.

We adopted ListNet in our work since it has been shown in [8, 29] that ListNet is efficient and even outperforms conventional approaches, such as RankSVM. ListNet employs cross-entropy between two probability distributions of input scores and predicted scores as a 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 sum of linear weighted features is the predicted score.

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

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

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

Each w_j, for $j = 1$ M is initially assigned to zero. In each iteration, w_j is updated by

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

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, the scores of new photographs are generated and a ranked list is produced based on the scores. Personalized ranking is further realized by manually modifying the weightings, so called feature-based.

Example-Based: Our system also provides weighting adjustment by example photographs. A weighting vector is

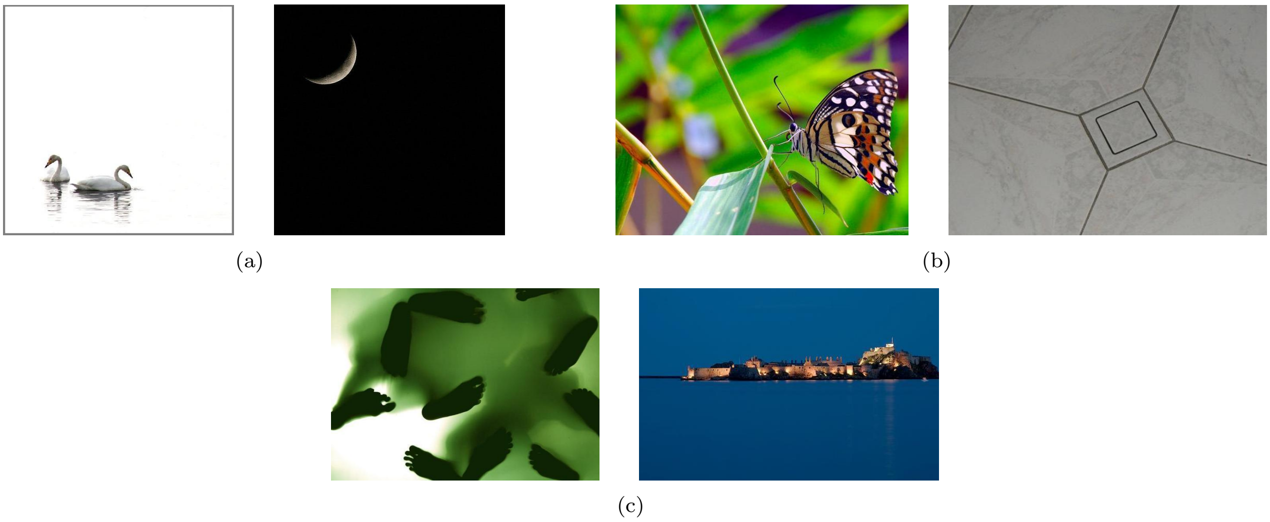


Figure 11: Color preference (a) High brightness and low brightness (b) High saturation and low saturation (c) Color style (when green and blue are selected)

associated with each example photograph where each entry of the weighting update vector is defined as:

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

where

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

x_i^j is the j-th feature value for the photograph i, m^j is the mean value of feature j from all training photographs, 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 photographs. Function F is a voting mechanism, which determines whether selected photographs are consistent in features. If two or more photographs contradict each other in a specific feature, the feature will not be updated.

5 Experiments and User Study

All data are selected from a photograph contest website, DPChallenge.com, which contains diverse types of photographs taken by different photographers. Each photograph is rated from 1 to 10 by a minimum of 200 users so as to reduce the influence of the outliers. We used the 6,000 highest-rated and 6,000 lowest-rated photographs for our experiments, the same data that was used in [13].

5.1 Ranking

3,000 top ranked photographs and 3,000 bottom ranked photographs are selected to train our system by the ranking algorithm, ListNet. The corresponding score for each photograph is its rank. After the weightings of features are learned, the remaining 6,000 photographs are used for testing. We evaluate our ranking results using Kendall's Tau-b coefficient.

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

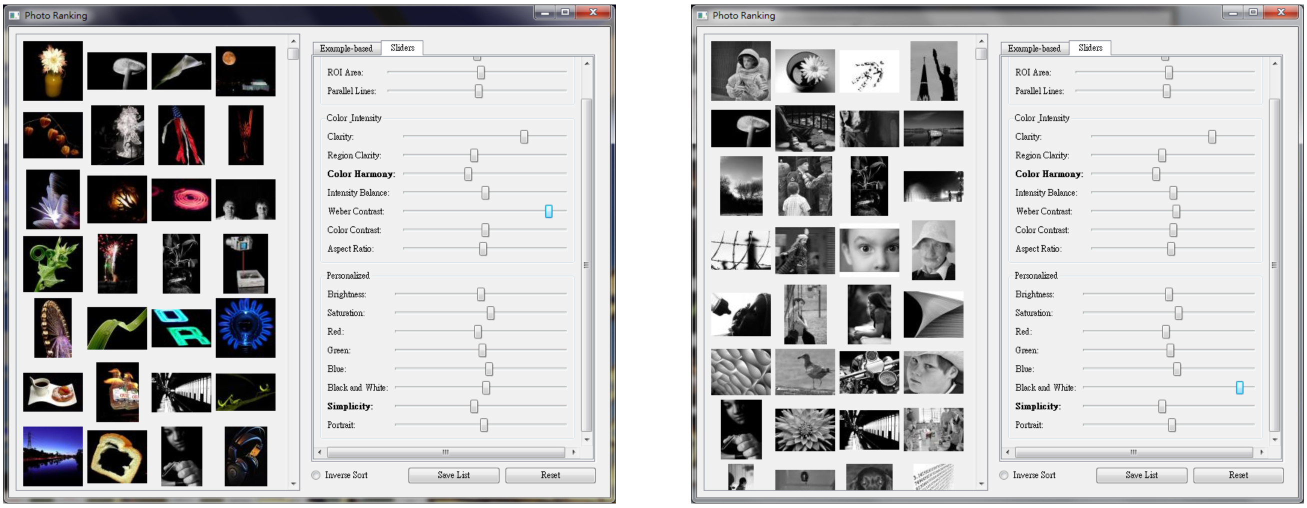
n_0 is the number of all pairs, n_c is the number of concordant pairs, n_d is the number of discordant pairs in the lists, t_1 is the number of pairs tied in the first list, and t_2 is the number of pairs tied in the second list. A Kendall's Tau-b *where* value of 0.4228 is derived from the predicted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

5.2 Binary Classification

With so many features, we need to address the issue of how to combine them in the binary classification problem. We use the “late fusion” technique [24], where a “voting strategy” is used, with the voting weighting of each feature determined by the training phase accuracy. We used the best three features (simplicity, texture, and contrast) in voting, and our result is 93% accuracy. This compares favorably with what was reported by Luo *et al.* [13] who used three different approaches (Bayes, SVM, Gentle Adaboost), and achieved the best result of above 93% with Gentle Adaboost.

In Figure 14, we compare the results of our approach to those by Ke *et al.*'s [9], Luo *et al.*'s [13]. Direct comparison is of limited utility since Luo *et al.* is using Bayesian based and ours is using ListNet, while Ke *et al.*'s has a much smaller database (2,000 for training). We use the same dataset of 12,000 photographs (6,000 for training) as Luo *et al.* does. Nonetheless, the features proposed in our approach have been effective and the overall difference is small: both systems are 93% in binary class classification.

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



(a) (b)

Figure 12: Ranking results with feature-based UI, where the left side of the window is the ranked result, and the right side is for user manipulation. (a) Re-ranking photographs by the contrast feature (b) Re-ranking photographs by the black-and-white feature

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%

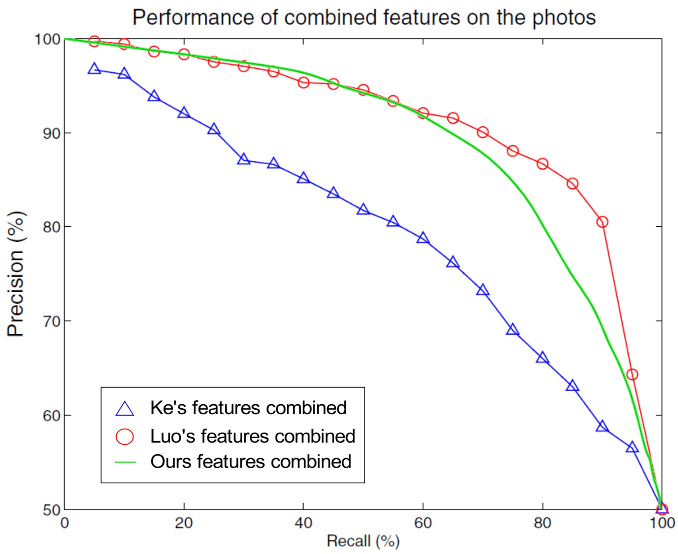


Figure 14: Precision Recall curve of three methods, where Ke's and Luo's use Bayes classifier, and ours uses ListNet.

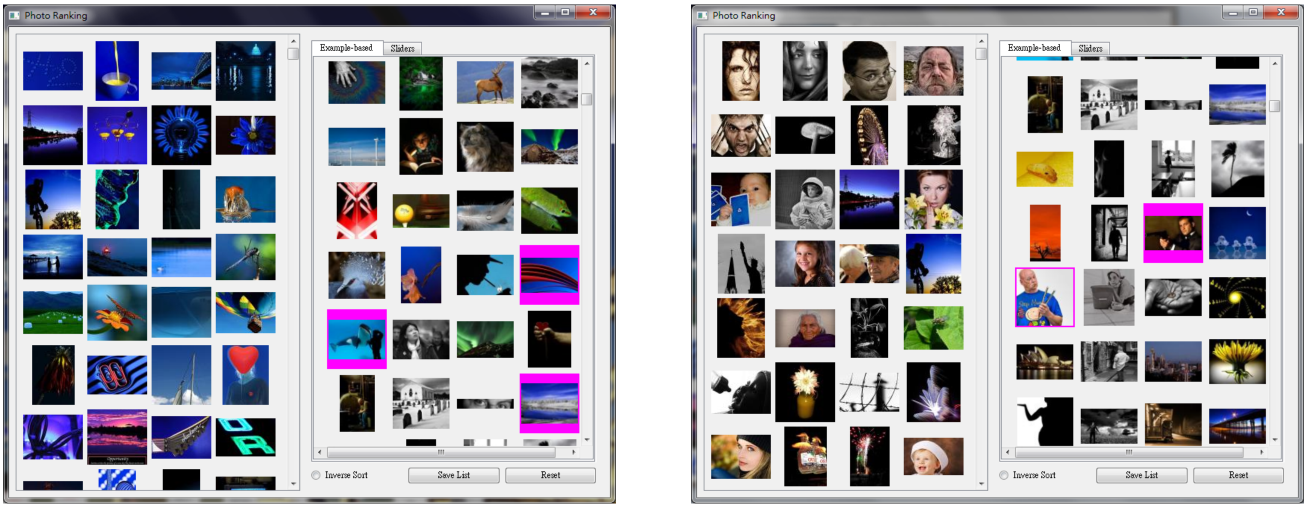
Some features, such as RGB colors, portrait (via face detection), and black-and-white, may not perform well as individual feature in a two-class classification problem, but they are important for individual preference. Thus, some of the features used in previous work have proven effective, but are insufficient for personal preference.

5.3 User Study

We conducted two user studies to evaluate the effectiveness of our system. In the first user study, each subject was asked to adjust weightings using slider bars to generate a new ranked list of photographs. The newly-generated list was compared with the previous list to verify the effectiveness of our personalization process. Subjects were asked if the new list was closer to their preference and four options were provided for their choice: "very good", "good", "bad", and "very bad". In the second user study, each subject was asked to select a few (typically two to five) preferred photographs and our system then re-ranked the list accordingly. The same four options were provided to examine their results.

Two thousand photographs, comprising a thousand highest-rated and a thousand lowest-rated from DPChallenge.com, were used in the two experiments, with half of them used as the training set and the other half used as the testing set. A total of twelve subjects participated in both experiments, with each subject taking an average of 25 minutes.

The results for the four levels ("very good", "good", "bad", and "very bad") were: (8.3%, 91.7%, 0%, 0%) for the first user study and (0%, 83.3%, 16.7%, 0%) for the second user study. The results from the two experiments shows that our system can re-rank the list closer to user preference.



(a) (b)

Figure 13: Ranking results with example-based UI, where the left side of the window is the ranked result, and the right side is for example selection. (a) Re-ranking photographs by blue color (b) Re-ranking photographs by portrait

In addition to the two user studies, participants were also asked which of the two approaches, updating each feature manually or selecting example photographs, was the more effective and intuitive way for re-ranking the list: the example-based UI was preferred by 66.7% of the users and 33.3% of the users preferred the feature-based UI.

6 Conclusion and Future Work

We propose a novel personalized ranking system for amateur photographs. Although automatically ranking award-wining professional photographs may not be a sensible pursuit, such an approach is reasonable for photographs taken by amateurs, especially when taking individual preference into account. 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. Two personalized ranking user interfaces are provided: the feature-based and example-based. Both are effective in providing personalized preferences, and in our user study, twice as many people preferred example-based than feature-based.

In our study, more than 18 features were proposed and tested for ranking prediction, as described in 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. Similarly, our implementation of example-based UI is just one kind of implementation, and we would like to see more.

7 Project Page

A demo and supplementary materials can be downloaded from the project page:
http://www.cmlab.csie.ntu.edu.tw/project/photorank/

8 Acknowledgement

We appreciate the DPChallenge.com users for sharing their images that were used in this paper. We are grateful to the anonymous reviewers for valuable feedback, and helpful discussion with Professor Winston H. Hsu and Dr. Ming-Fang Weng. Special thanks go to Hong-Cheng Kao and Wai-Seng Ng for their pioneering work. This project is funded in part by NSC of Taiwan (NSC 98-2622-E-002-001-CC2) and Cyberlink Inc.

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, J. Proenaa, J. Jorge, and M. Sousa. 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.