Personalized Photograph Ranking and Selection System Considering Positive and Negative User Feedbacks

Abstract

In this paper, we propose a novel personalized ranking system for amateur photographs. 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. Photographs are described using 20 image features which can be categorized into three types: photo composition, color and intensity distribution, and features for personal preferences. We adopt RBF-ListNet as the ranking algorithm. RBF-ListNet is based on an efficient algorithm, ListNet, using radial basis functions. The performance of our system is evaluated in terms of Kendall's tau rank correlation coefficient, precision-recall diagram, and binary classification accuracy. The Kendall's tau value (0.434) is higher than those obtained by ListNet and support vector regression (SVR). The precision-recall diagram and binary classification accuracy (93%) is close to the best results to date for both overall system and individual features. To realize personalization in ranking, we propose three approaches: feature-based, example-based, and list-based approach. User studies indicate that all three approaches are effective in both aesthetic and personalized ranking. In particular, the example-based approach obtained the highest user experience rating among all three.

Index Terms

I. 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. There have been several photographic topics using computation-based methods. For example, *computational aesthetics* is proposed to predict the emotional response to works of art [28], [29]. Several papers propose to use computation-based methods to enhance photographs or videos for better aesthetics [1], [6], [20], [25], [32], [41]. Photograph assessment, which targets to select "professional" or "high quality" photographs, is a popular research topic. Computation-based methods are also applied to photograph assessment to improve the effectiveness of selection. There have been several feasible solutions for automatic photograph ranking and selection [10], [17], [21], [37]. However, any solution based on computation will face challenges and difficulties since the judgement of aesthetics involves sentiments and personal taste [10], [24]. Everyone has his or her unique way to rank photographs. A fixed ranking list simply cannot meet everyone's requirements, in the same way that there is no universally preferred interior design for individual houses.

In this paper, we propose to apply personalization in the traditional photograph assessment problem. 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. Figure 1 shows photographs that are top-ranked and bottom-ranked by the proposed system. To take personal preferences into ranking, we propose an interactive user interface with three approaches: feature-based, example-based, and list-based approach.

II. Related Work

Various papers propose to select "high quality" photos. A number of papers [34], [38], [39] assess photos concerning image qualities such as degradation caused by noise, distortion, and artifacts. On the other hand, Tong *et al.* [37] and Datta *et al.* [10] try to classify professional photos and non-professional photos with low-level features utilized in image retrieval. Ke *et al.* [17] analyzed the human perception in aesthetic experience and designed high level feature descriptors such as simplicity, blurriness, and color distribution. The performance of features is often evaluated by the accuracy of binary classification, which labels photographs into two classes, such as professional and non-professional. In Ke's method, 72% classification rate is achieved on a set of 3,000 photos. Our previous work [43] ranks a photo by nine rules based on aesthetics. 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

(a)

(b)

Fig. 1. Ranking results by our system without any manually adjustments (a) Top-ranked photographs (b) Bottom-ranked photographs

blurriness. Accuracy of 81% is achieved on a set of 2,000 photos. Luo *et al.* [21] treats foreground and background differently instead of extracting features from the entire photograph, achieving over 93% classification rate using 12,000 photos.

The work described above can be categorized as rule-based approaches, which extract features based on photographic aesthetic guidelines. Recently, image descriptors have also been adopted in aesthetic modeling besides rule-based approaches. Image descriptors such as GIST [26], bag-of-features [9], and the Fisher Vector [27], have been widely applied to semantic tasks such as object/scene retrieval, images classification/annotation, and object localization. Its effectiveness is also proven in aesthetic recognition [23]. The advantage of image descriptor approaches is that they can cover more implicit aesthetic rules which cannot be covered completely by rule-based approaches. However, the weakness arises from the lack of flexibility in customization, which can provide unique results according to personal preferences. Therefore, our framework still adopts the rule-based approach.

In previous work, performance is often evaluated by the accuracy of binary classification. However, even within two-class photographs, there are still ranks in photographs. San Pedro *et al.* [31] uses Kendall's tau coefficient [19] to measure the similarity between their ranking results and the groundtruth. 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 improve the value to 0.48 by combining tag information of photographs. 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 selection. However, one of the most challenging aspects is that the results tend to be subjective. The judgement of aesthetics involves sentiments and personal taste [10], [24]. Everyone has his or her unique way to rank photographs, and a fixed ranking list simply cannot meet everyone's requirements. Sun *et al.* adopted the idea of personalization in which personalized photograph assessment is achieved by incorporating user preference [36]. 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 introduce a proposed system for ranking photographs according to individual preferences.

III. System Overview

Figure 2 illustrates the semantic diagram of the proposed system. The proposed system requires training to learn a model for aesthetic ranking. A set of training photographs, which have been scored, is collected to train the model. Features are extracted from each of the photographs according to the aesthetic rules which are discussed in Section IV. Then the features of the training photographs and their scores are taken as the input data to the RBF-ListNet algorithm to train a ranking model. The details of the RBF-ListNet will be described in Section V. After training, the ranking model takes the features of a new photograph as input and it outputs its estimated ranking score. The ranking scores of a photograph set can be re-ranked through the user interface, which provides three approaches: feature-based, example-based, and list-based approach. The feature-based approach provides users who understand the twenty features with the functionalities that can emphasize some features over others by manually adjusting the weighting for corresponding features. Using the example-based approach, users can select some of the photographs they like and have the system update the weighting based on these few example photographs. For the list-based approach, users provide personal photographs with corresponding rankings to re-rank

the results. Finally, a personalized ranking result can be obtained.

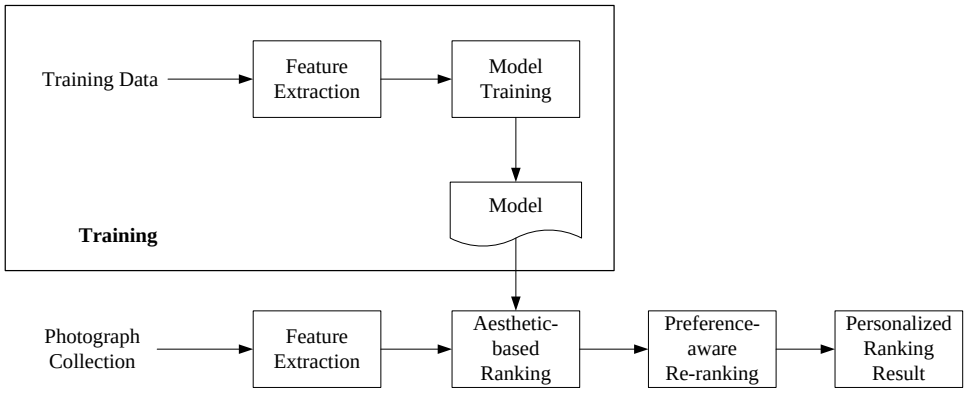


Fig. 2. System overview

IV. Rules of Aesthetics

Rules of aesthetics in photography describe how to arrange different visual elements inside an image frame. We categorize these rules into three major categories: photographic composition, color and intensity distribution, and personalized features.

A. Photographic Composition

Composition involves 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 principles which suggest a composition that will be pleasing for most people. We use two principles: rule of thirds and simplicity.

1) Rule of Thirds: The rule of thirds is the most well-known photograph composition guideline [13], [18]. The idea is to place main subjects at roughly one-third of the horizontal or vertical dimension of the photograph. An example is shown in Figure 3.

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 [11]. Figure 4(b) illustrates the segmented results of the example photograph shown in Figure 4(a). Simultaneously, we extract the salient map from the photograph by using the approach proposed by Harel *et al.* [14], Graph-Based Visual Saliency. The saliency map is

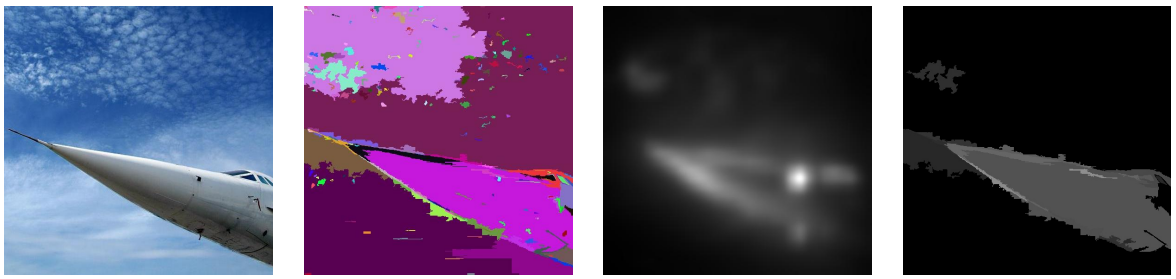
Fig. 3. Example of rule of thirds: the flower is located at one of the "power points"

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

We measure the rule of thirds in a photograph by weighted averaging the probability of euclidean distances between patches and corresponding closest power points. The model is formulated as:

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

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

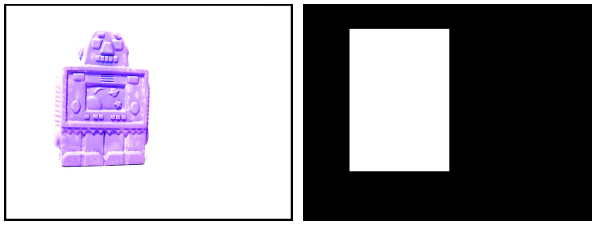


(a) (b) (c) (d)

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

2) Simplicity: Simplicity in a photograph is a distinguishing factor in determining whether a photograph is professional or not [17]. We use two kinds of features to measure the simplicity of the photograph: the size of saliency segments and the simplicity feature proposed by Luo *et al.* in [21].

(a)



(b)

Fig. 5. Saliency Area size feature (a) Large saliency region, depicted as the white area in the right frame (b) Small saliency region

A binary saliency map is created by applying a threshold to the saliency map of the photograph.

$$B_{Saliency} = \begin{cases} 1, & \text{if } x < \alpha \cdot \max_{Saliency}, \quad \alpha = 0.67 \\ 0, & \text{otherwise.} \end{cases}$$

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

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

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 saliency segments, we also include one of the features from [21] 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 histogram (H) of 4096 bins

is generated for the photographs. The simplicity feature is defined as:

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

where $s = \{i|H(i) \geq \gamma h_{max}\}$, and $\gamma = 0.01$. Table III(b) shows that our implementation performs with 89.48% accuracy which is an improvement over the 73% accuracy of Luo's method.

(a) (b)

Fig. 6. Simplicity feature (a) High simplicity (b) Low simplicity

B. Color and Intensity Distribution

Lighting condition is an important factor for photography since it affects the realism of photograph due to the compositions in colors and illumination. We use five principles to measure the color and intensity distribution of photographs: texture, focus, color harmony, intensity balance, and contrast.

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 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 [30]. 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. It uses the Gabor filter to evaluate the convolution responses of the image under different scales and orientations [4], [22].

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

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

where $v = 0, ..., v_{max} - 1$, $u = 0, ..., u_{max} - 1$. 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 $\{30°, 60°, 90°, 120°, 150°, 180°\}$. This texture feature performs well in classification accuracy (84.15%) as shown in Table III(b).

2) Focus: 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 [17], [37]. 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 $(t = 2)$.

$$f_{focus} = \frac{\textit{number of clear pixels}}{\textit{total pixels}} \tag{4}$$

Depth of field (DOF) is the distance between the nearest and farthest objects in a scene which are focused in the image. Professional photographers usually put main objects in the depth of field for emphasis, while background and minor objects are out of the depth of field. We manage to describe the usage of DOF by partitioning a photograph into grids and applying blur detection on them.

$$f_{DOF} = \frac{\textit{number of clear grids}}{\textit{total grids}}$$

We exclude grids with low color variations because they sometimes produce an erroneous evaluation of low quality on what is really a high quality image.

3) Color Harmony: 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 [7] is:

(a) (b)

Fig. 7. Focus feature (a) High focus (b) Low focus

$$F(X,(m,\alpha)) = \sum_{p \in X} \left\| H(p) - E_{T_{m(\alpha)}}(p) \right\| \cdot S(p) \tag{5}$$

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)

Fig. 8. Color Harmonization feature (a) Harmonic color (b) Less harmonic color

4) Intensity Balance: Balance provides a sense of equilibrium and is also a fundamental principle of visual perception in that our 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.

$$f_{balance} = \left| \sqrt{\sum_{i=1}^{k} (E_{left} - E_{right})} \right| \tag{6}$$

(a) (b)

Fig. 9. Intensity balance feature (a) balanced (b) left-right unbalanced

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:

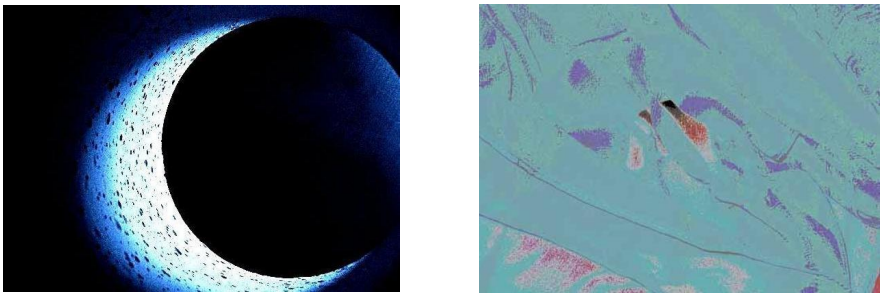
$$f_{WeberContrast} = \frac{1}{width} \frac{1}{height} \sum_{x=0}^{width} \sum_{y=0}^{height} \frac{I(x,y) - I_{avg}}{I_{avg}} \tag{7}$$

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 CIE2000 [33] to determine color contrast.

The image segmentation method is applied to photographs and the mean color is computed for each segment [11]. 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.

$$f_{ColorContrast} = \sum_{i=0}^{n} \sum_{j=i+1}^{n} (1 - D(i,j)) \frac{C(i,j)}{M_i M_j} \tag{8}$$

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



(a) (b)

Fig. 10. Contrast feature (a) High contrast (b) Low contrast

accuracy (84.12%), as shown in Table III(b).

C. Personalized features

Although photographs can be assessed based on aesthetic rules, these rules do not fully capture personal taste. For example, some people may prefer photographs with a specific color style, or high color saturation, or high intensity, etc. Others 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.

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 more friendly user interface than that using 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.

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.

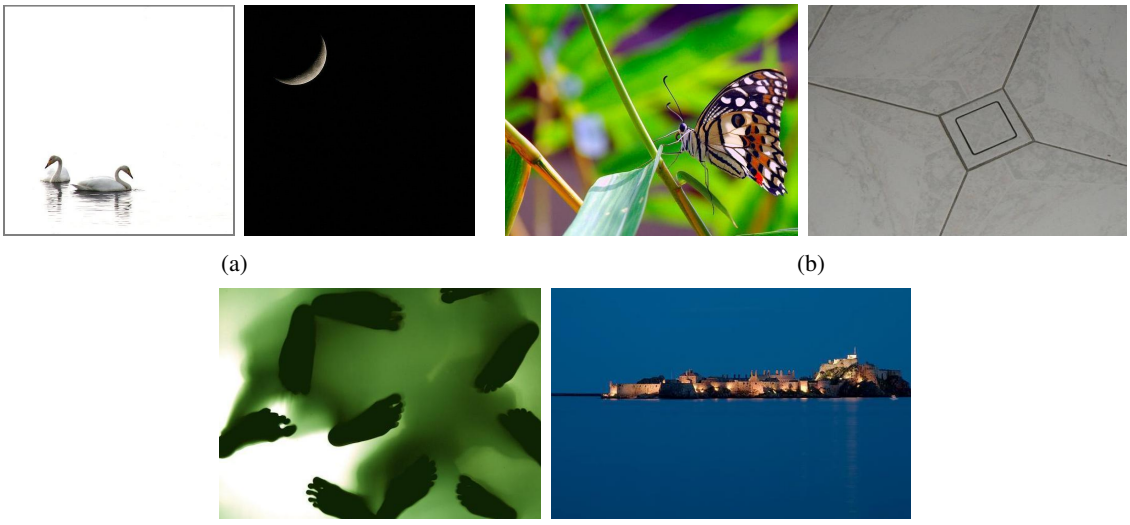


Fig. 11. Color preference (a) High brightness and low brightness (b) High saturation and low saturation (c) Color style (when green and blue are selected)

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) 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.

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

$$f_{AspectRatio} = \frac{width}{height} \tag{9}$$

V. Aesthetic Learning and Personalization

A. Learning to Rank

Related to the classification problem, ranking generates an ordered list according to certain criteria, such as utility function. A ranking algorithm assigns a relevant score to each object, and the score order represents the relevance to the goal function. A model can be trained to predict scores and ranks by ranking algorithms with a set of training data. The training procedure 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 $\mathbf{D} = \{d_1, d_2, ..., d_N\}$, 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 $\mathbf{Y} = \{y_1, y_2, ..., y_N\}$, where y_i is the relevance score of photograph d_i. A feature vector, denoted $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^M)$ where M is the number of dimensions, is extracted from each photograph based on the rules described in Section IV. A ranking algorithm f is trained to predict the scores of test data by leveraging the co-occurrence patterns among feature $\mathbf{X}$ and score $\mathbf{Y}$. During training the ranking algorithm, a list of predicted scores, denoted $\mathbf{S} = \{s_1, s_2, ..., s_N\} = \{f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_N)\}$, is obtained for the set D of training photographs. The ranking algorithm f is optimized by minimizing the loss function $L(\mathbf{Y}, \mathbf{S})$.

The state-of-the-art algorithms can be categorized into pairwise approaches and listwise approaches [2], [40]. The pairwise approaches formulate the ranking problem as a classification problem by judging if object pairs are correctly or incorrectly ranked. RankSVM is one of the pairwise approaches based on the support vector machines [15]. However, the pairwise approach will encounter the problem of effectiveness due to the pair instance learning (N^2 pairs), which is time consuming while N is large.

The listwise approach, such as ListNet, overcomes the shortcomings of the pairwise approach and is efficient [40]. ListNet employs cross-entropy between two probability distributions of input scores and predicted scores as a listwise loss function (Eq. 10).

$$L(\mathbf{Y}, \mathbf{S}) = -\sum_{i=1}^{N} P(y_i) log(P(s_i)) \tag{10}$$

and

$$P(s_i) = \frac{\Phi(s_i)}{\sum_{j=1}^{N} \Phi(s_j)} \tag{11}$$

where function Φ is an increasing and strictly positive function; we adopt the exponential function in our work. 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.

$$z_i = f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x_i} \tag{12}$$

where $\mathbf{w} = (w_1, w_2, ..., w_M)$ is the weighting vector of features.

B. RBF-ListNet

A problem of ListNet is the oversimplied assumption in the score function, which only considers the linear relationship of features (Eq. 12). Previous literature shows that the characteristics of multidimensional data such as music are better modeled with a non-linear function such as radial basis function (RBF) [42]. A RBF neural network can approximate well any function under certain mild conditions [5]. Therefore, RBF-ListNet is proposed as a ranking algorithm, which adapts the RBF neural network model as the ranking model of ListNet [42]. In our work, we adopt the RBF-ListNet as the ranking algorithm, and a comparison with results of ListNet will be shown in Section VI.

The score function for $\mathbf{x}_i$ is based on a cosine radial basis function and is defined as

$$\begin{aligned} f(\mathbf{x}_i) &= \sum_{k=1}^{K} \lambda_k h_k(\mathbf{x}_i) \\ h_k(\mathbf{x}_i) &= \frac{\alpha_k}{(||\mathbf{x}_i - \mathbf{v}_k||^2 + \alpha_k^2)^{1/2}} \end{aligned} \tag{13}$$

where λ_k is the weight of the kth h_k function, $h_k(\mathbf{x}_i)$ is the response of the function located at the kth prototype $\mathbf{v}_k$ to an input vector $\mathbf{x}_i$, α_k is the reference distance for function h_k, and K is the number of radial basis functions.

The loss function to be minimized is the same as Eq. 10. A gradient descent algorithm is used to update the value of λ_k, $\mathbf{v}_k$, and α_k.

$$\begin{aligned} \lambda_k &\leftarrow \lambda_k - \eta \cdot \Delta\lambda_k \\ \mathbf{v}_k &\leftarrow \mathbf{v}_k - \eta \cdot \Delta\mathbf{v}_k \\ \alpha_k &\leftarrow \alpha_k - \eta_2 \cdot \Delta\alpha_k \end{aligned} \tag{14}$$

$\Delta\lambda_k$, $\Delta\mathbf{v}_k$, and $\Delta\alpha_k$ are derived as:

$$\begin{aligned} \Delta\lambda_k &= \sum_{i=1}^{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i) \\ \Delta\mathbf{v}_k &= \frac{\lambda_k}{\alpha_k^3} \sum_{i=1}^{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i)^3 (\mathbf{x}_i - \mathbf{v}_k) \\ \Delta\alpha_k &= \frac{\lambda_k}{\alpha_k} \sum_{i=1}^{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i)(1 - h_k(\mathbf{x}_i)^2) \end{aligned} \tag{15}$$

where $\delta(f(\mathbf{x}_i), y_i) = P(f(\mathbf{x}_i)) - P(y_i)$ is the estimation error for $\mathbf{x}_i$. η is the learning rate for λ_k and $\mathbf{v}_k$, and η_2 is another learning rate for α_k. The value of η_2 is set one order of magnitude lower than η according to [16]. The weights λ are all initialized to zero, the $\mathbf{v}$ are determined by K-mean clustering, and the reference distance α are computed with $\alpha_k = \min_{t \neq k} ||\mathbf{v}_k - \mathbf{v}_t||$.

The flow of RBF-ListNet algorithm is:

Algorithm 1: RBF-ListNet Algorithm

Input: N training photos and scores $((\mathbf{x}_i, s_i), i = 1 \sim N)$, K (the number of radial basis function), η (learning rate for $\lambda, \mathbf{v}$), η_2 (learning rate for α), δ (convergent threshold)

Output: RBF neural network model parameters $\lambda, \mathbf{v}, \alpha$

while true **do**

Compute $f(\mathbf{x}_i)$ according to Eq. 13

Transform the ranking scores to probabilities using Eq. 11

Compute the loss L according to Eq. 10

if changes in L is smaller than δ **then**

break

end if

Update parameters at rates η and η_2 according to Eq. 14 and 15

end while

C. Personalization

After training, a ranking list of photograph collections can be estimated by the trained model. To meet each individual's tastes, re-ranking according to personal preferences is necessary. Therefore, a user interface is proposed for personalized ranking. Users are able to show their own preference through three approaches: feature-based, example-based, and list-based approach.

1) Feature-based approach: Since features can be extracted from each photograph according to the rules described in Section IV, we allow users to re-rank the photographs depending on personal preference

rules. The users can increase the weightings of preferred rules and decrease those of disliked rules. First, we calculate the z-scores of all features dimension for each photograph. The z-score function is defined as

$$z(x_i^j) = \frac{x_i^j - \mu_j}{\sigma_j} \tag{16}$$

where μ_j is the mean value, and σ_j is the standard deviation of the feature $\mathbf{x}^j$.

Personalized ranking is further realized based on the z-score. We define the function $s_i^{'}$ to scale the score estimated by the ranking algorithm.

$$s_i^{'} = f^{'}(\mathbf{x}_i) = \prod_{j=1}^{M} w_j z(x_i^j) f(\mathbf{x}_i) \tag{17}$$

where w_j is the weighting of feature $\mathbf{x}^j$. The weighting is initially set to 1 for all features and can be manually adjusted by users. After the adjustments, scores $s^{'}$ are updated and used for re-ranking the photographs.

2) Example-based approach: In addition to the feature-based approach, we provide an example-based approach which allows users to re-rank photographs by providing preferred examples. In these examples, the features whose values are over or under the mean value of all photographs will be taken into the weighting update. The update is according to the feature z-scores in the example photographs, and then re-ranking on photographs is performed according to the new scores scaled by the updated weightings as in Eq. 17. The weightings $\mathbf{w}$ are updated as follows:

$$w_j = w_j + c \cdot z(x_i^j) \tag{18}$$

where c is a constant value parameter indicating the step size of updating $\mathbf{w}$. After updating the weights, scores $s^{'}$ are calculated according to Eq. 17. Finally, photographs are re-ranked according to $s^{'}$. In our system, example-based re-ranking can perform quite well by selecting a minimum number of three to five example photographs.

3) List-based approach: Using a set of photographs ranked by the public for training, we can obtained a general model which assesses photographs according to public preference. Similarly, a personal preferred model can also be obtained using a set of photographs ranked personally. However, if the number of

provided photographs is insufficient, the trained model may overfit with the data and generate biased results. A solution to avoid overfitting is to take both the personal and public preferences into consideration. Therefore, in the proposed list-based approach, we allow users to provide ranked photographs, and a preference-aware model is then trained by the customized RBF-ListNet algorithm. In our system, the list-based re-ranking can show its effectiveness by providing at least 10 personal ranked photographs.

In the customized RBF-ListNet algorithm, the listwise loss $L^{'}$ of personal ranked photographs is computed according to Eq. 10 in addition to the loss L of training data. At each iteration, the change in $L + \omega L^{'}$ is checked. The values of λ_k, $\mathbf{v}_k$, and α_k are updated according to Eq. 14, while $\Delta\lambda_k$, $\Delta\mathbf{v}_k$, and $\Delta\alpha_k$ are replaced by $\Delta\lambda_k^{'}$, $\Delta\mathbf{v}_k^{'}$, and $\Delta\alpha_k^{'}$ defined as:

$$
\begin{aligned}
\Delta\lambda_k^{'} &= \gamma \left(\sum_{i \in Training} \frac{1}{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i) + \omega \sum_{j \in User} \frac{1}{M} \delta(f(\mathbf{x}_j), y_j) h_k(\mathbf{x}_j) \right) \\
\Delta\mathbf{v}_k^{'} &= \gamma \cdot \frac{\lambda_k}{\alpha_k^3} \left(\sum_{i \in Training} \frac{1}{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i)^3 (\mathbf{x}_i - \mathbf{v}_k) + \omega \sum_{j \in User} \frac{1}{M} \delta(f(\mathbf{x}_j), y_j) h_k(\mathbf{x}_j)^3 (\mathbf{x}_j - \mathbf{v}_k) \right) \\
\Delta\alpha_k^{'} &= \gamma \cdot \frac{\lambda_k}{\alpha_k} \left(\sum_{i \in Training} \frac{1}{N} \delta(f(\mathbf{x}_i), y_i) h_k(\mathbf{x}_i)(1 - h_k(\mathbf{x}_i)^2) + \omega \sum_{j \in User} \frac{1}{M} \delta(f(\mathbf{x}_j), y_j) h_k(\mathbf{x}_j)(1 - h_k(\mathbf{x}_j)^2) \right)
\end{aligned}
\tag{19}
$$

where γ is a scalar and ω is the weight of user-ranked photographs, and where ω is determined with a weighting function based on the probability density function of the exponential distribution. The value of ω increases while the number of personal photographs grows. The iteration will be terminated when the change is smaller than the threshold.

The customized RBF-ListNet algorithm is:

Algorithm 2: Customized RBF-ListNet Algorithm for List-based approach

---

Input: N training photos and scores $((\mathbf{x}_i, s_i), i = 1 \sim N)$, M user-ranked photographs and scores $((\mathbf{x}_j, s_j), j = 1 \sim M)$, K (the number of radial basis function), η (learning rate for $\lambda, \mathbf{v}$), η_2 (learning rate for α), δ (convergent threshold)

Output: RBF neural network model parameters $\lambda, \mathbf{v}, \alpha$

while true **do**

Compute $f(\mathbf{x}_i)$ according to Eq. 13

Transform the ranking scores to probabilities using Eq. 11

Compute the loss L for N training photos and $L^{'}$ for M according to Eq. 10

if changes in $L + \omega L^{'}$ is smaller than δ **then**

break

end if

Update parameters at rates η and η_2 according to Eq. 14 and 19

end while

VI. Experiments and User Study

A. Dataset

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 as that was used in [21].

B. Ranking Performance

We evaluate our ranking results using Kendall's Tau-b coefficient:

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

where 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. The τ_b coefficient equals 1 for perfect agreement, -1 for total disagreement, and 0 if the rankings are independent.

Three thousand top ranked photographs and three thousand bottom ranked photographs are selected as the training data. The corresponding score for each photograph is its rank. The remaining six thousand photographs are used for testing. The parameters of RBF-ListNet are set as described below. The convergent threshold δ is set to $\frac{1}{2}e - 3$, the learning rate η is set to 10, the learning rate η_2 is set

to 1, and the number of radial basis functions K is set to 8. As a result, a Kendall's Tau-b coefficient of 0.434 is derived from the predicted score list of test data. We also compare the performance of ListNet and support vector regression (SVR) [8] with different kernel functions. LIBSVM [3] is utilized to perform SVR experiments[1]. Table I shows the results and the highest value can be obtained by RBF-ListNet.

TABLE I
KENDALL'S TAU-B COEFFICIENTS OF SVR AND LISTNET WITH LINEAR AND RBF KERNEL

Method	**Kernel Function**	**Kendall's Tau-b**
SVR	Linear	0.384
SVR	RBF	0.402
ListNet	Linear	0.423
RBF-ListNet	RBF	0.434

A sensitivity test is also conducted on the parameters of RBF-ListNet. Kendall's Tau-b coefficients are evaluated under different settings on K and δ. Table II shows the results. It can be observed that a small δ overfits the data, while large δ underfits the data. In Table II, the entry marked with $*$ is the setting used in our system.

TABLE II
KENDALL'S TAU-B COEFFICIENTS OF RBF-LISTNET UNDER DIFFERENT PARAMETER SETTINGS

$\delta \backslash \mathbf{K}$	$\mathbf{K=5}$	$\mathbf{K=8}$	$\mathbf{K=10}$	$\mathbf{K=20}$
$\mathbf{1e-3}$	0.289	0.277	0.335	0.430
$\mathbf{\frac{1}{2}e-3}$	0.424	0.434*	0.421	0.426
$\mathbf{1e-4}$	0.405	0.393	0.397	0.387

C. 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 [35], 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% in accuracy. This compares favorably

[1]The context of a footnote.

with what was reported by Luo *et al.* [21] who used three different approaches (Bayes, SVM, Gentle Adaboost), and achieved the best result of above 93% with Gentle Adaboost [12].

In Figure 12, we compare the results of our approach to those by Ke *et al.*'s [17], Luo *et al.*'s [21]. The diagram for Ke *et al.*'s and Luo *et al.*'s are from [21] directly since the results in [21] are not easy to replicate as reported in [23]. Direct comparison is of limited utility since Luo *et al.* is using Bayesian based and ours is using RBF-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 III, 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 features), perform better even compared to the blur factor.

TABLE III
SVM CLASSIFICATION ACCURACY OF SINGLE FEATURE (A) LUO'S FEATURES (B) OUR FEATURES

(a) Luo's features [21]

Features	**Accuracy**
Composition	79%
Clarity	77%
Simplicity	73%
Color Combination	71%
Lighting	62%

(b) Our features

Features	**Accuracy**
Simplicity(modified)	89.5%
Texture	84.2%
Contrast	84.1%
Intensity Average	75.2%
Region Blur	71.0%

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.

D. User Study

We conducted a user study to evaluate the effectiveness of the three proposed approaches. Fifteen participants including eleven males and four females are recruited. Five of them adore photography, nine are common users of commercial cameras, and one seldom uses cameras. Each participant was trained around 20 minutes to be familiar with the user interface of our system, and the formal experiment was

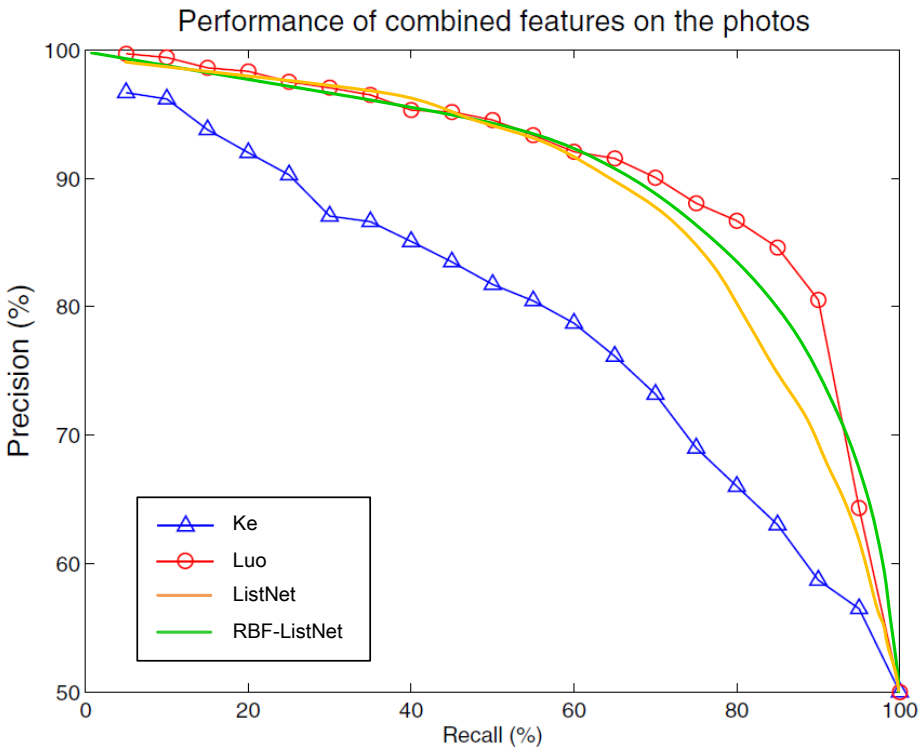


Fig. 12. Precision Recall curve of four methods, where Ke's and Luo's use Bayes classifier, and ours uses ListNet and RBF-ListNet. Notice that the diagram for Ke *et al.*'s and Luo *et al.*'s are from the paper [21] directly.

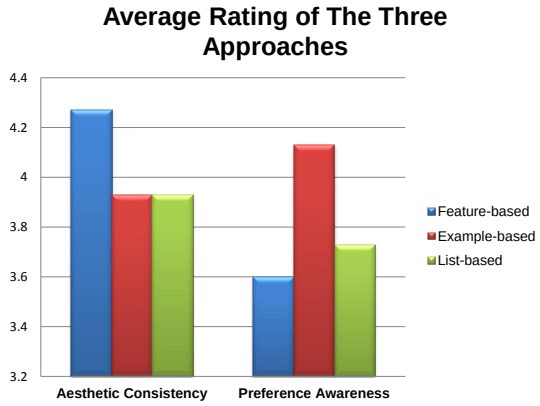
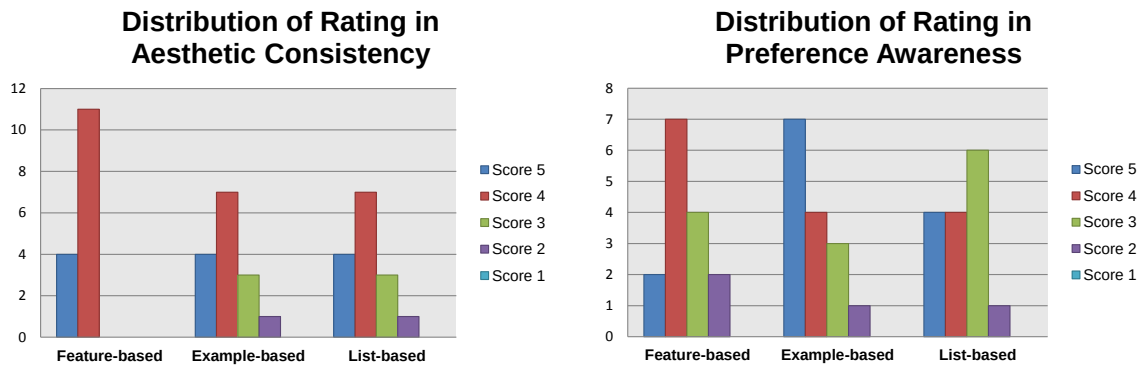


Fig. 13. The average rating of the two performance evaluation of the three proposed approach

done within 10 minutes. Our prototype was implemented using PHP and participants manipulated the system through the web browser, Google Chrome. The photograph collection used was the same 6000 test data as in previous experiments.



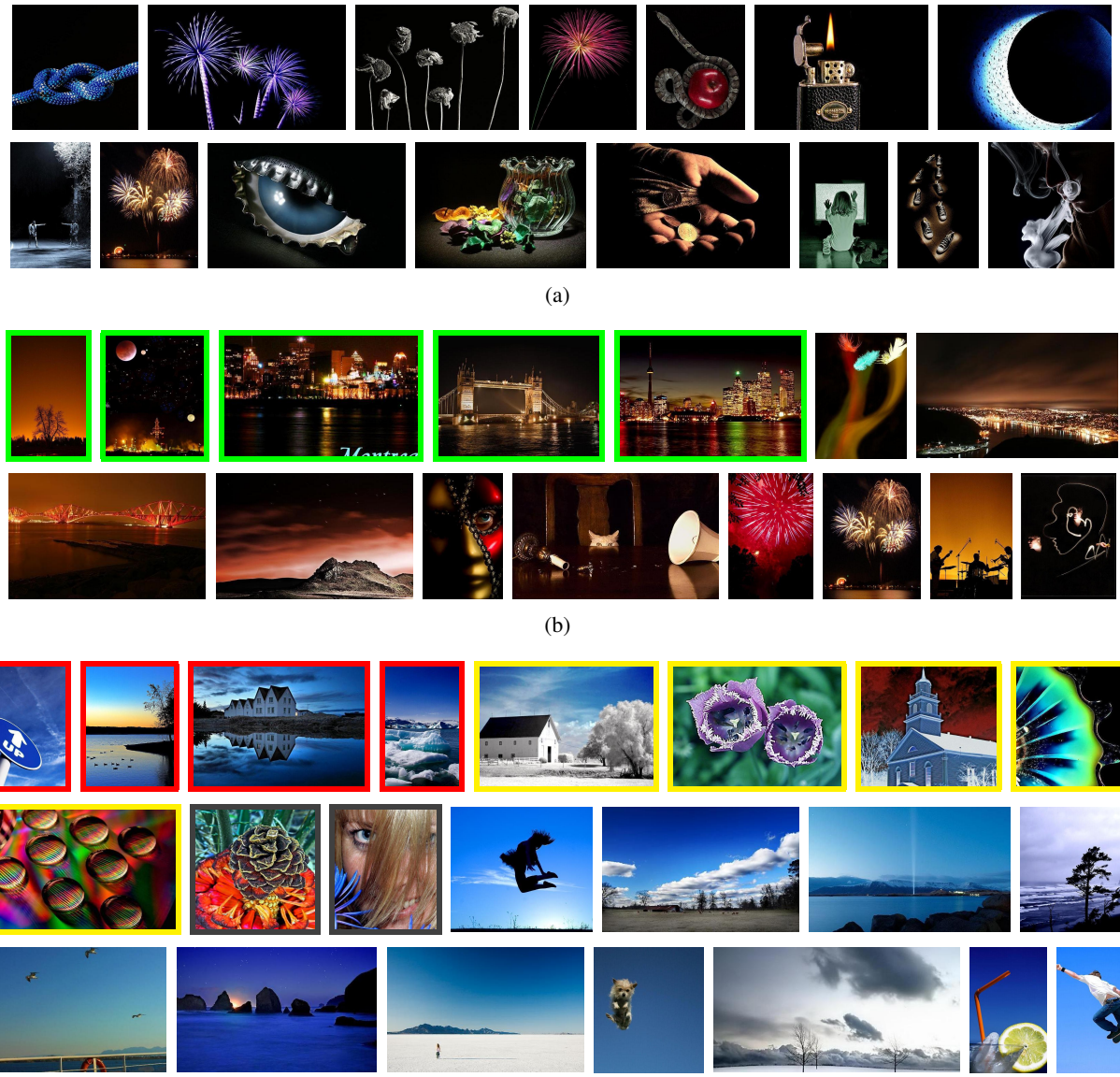
(a) (b)

Fig. 14. The distributions of the rating (a) Aesthetic Consistency (b) Preference Awareness

Participants were asked to re-rank the photographs separately through the proposed three approaches. Each operation generated a new ranking list for the photograph collection. After experiencing our system, each participant was asked to score the three approaches according to two performance measured: the aesthetic consistency and preference-awareness. The aesthetic consistency evaluates if the photographs are still ranked according to visual aesthetics after the adjustments by participants. The judgement in aesthetic consistency only considered visual aesthetics and did not involve personal tastes. Therefore, participants were further asked to judge the effectiveness of our system in ranking photographs according to personal preferences (preference awareness). Participants scored 5 (best) to 1 (worst) for the two judgements of each of the three approach. Figure 13 shows the average rating for the 15 participants, and Figure 14 illustrates the distributions of rating. The results show that the feature-based approach can preserve aesthetic ranking the most. However, the feature-based approach is given a lower average rating than the example-based and list-based approaches in preference-awareness. The reason is that example-based and list-based approaches are more intuitive for personalized ranking than the feature-based approach, since users might not understand and analyze all rules quite in their favorite photographs. The example-based approach received the highest rating due to its effectiveness and intuitiveness, which are important factors in taking personal preferences into ranking photographs.

VII. Conclusion

In this paper, we propose an interactive personalized ranking system for amateur photographs. We address the photograph assessment as a ranking problem rather than the traditional binary classification



(a)

(b)

(c)

Fig. 15. Some results re-ranked by the proposed approaches (a) The feature-based approach: Re-ranked photographs by the Weber contrast feature. (b) The example-based approach: The photographs with green frames are selected examples, and the others are results with top ranking. (c) The list-based approach: The photographs with red frames are user-specified top-ranked ones, those with yellow frames are middle-ranked, those with gray frames are bottom-ranked, and the others are results with top-ranking.

problem. The extracted features from photographs include both aesthetic rules and those for personal preferences. The performance of our system is shown in terms of a Kendall's tau coefficient which is 0.434, higher than those obtained by ListNet and SVR. The three approaches in the user interface, feature-based, example-based, and list-based approaches, provide an interactive and intuitive way to re-

rank photographs according to personal tastes. The user study shows the effectiveness of the proposed system.

However, one limitation of the proposed system is that its effectiveness highly depends on the performance and accuracy of the feature extractor algorithms, since the three approaches for personalization are based on the extracted feature. For example, the feature "rule of thirds" are extracted based on saliency information. The accuracy of the rule of thirds will be proportional to the accuracy of the saliency information. As a future work, more features or better feature extractors can be integrated into the system. The proposed three approaches are also only one kind of implementation for personalization, and we would like to discover more in the future.

REFERENCES

[1] S. Bhattacahrya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the international conference on Multimedia*, MM '10, pages 271–280, New York, NY, USA, 2010. ACM.

[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] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] 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.

[5] T. Chen and H. Chen. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *Neural Networks, IEEE Transactions on*, 6(4):904 –910, jul 1995.

[6] B. Cheng, B. Ni, S. Yan, and Q. Tian. Learning to photograph. In *Proceedings of the international conference on Multimedia*, MM '10, pages 291–300, New York, NY, USA, 2010. ACM.

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

[8] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995. 10.1007/BF00994018.

[9] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[10] 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.

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

[12] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK, 1995. Springer-Verlag.

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

[14] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19*, pages 545–552. MIT Press, 2007.

[15] 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.

[16] N. Karayiannis and M. Randolph-Gips. On the construction and training of reformulated radial basis function neural networks. *Neural Networks, IEEE Transactions on*, 14(4):835 – 846, july 2003.

[17] 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.

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

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

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

[21] 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.

[22] 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.

[23] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1784 –1791, nov. 2011.

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

[25] 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.

[26] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42:145–175, May 2001.

[27] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1 –8, june 2007.

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

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li. Listwise approach to learning to rank: theory and algorithm. In *Proceedings of the 25th international conference on Machine learning*, ICML '08, pages 1192–1199, New York, NY, USA, 2008. ACM.

[41] Y. Y. Xiang and M. S. Kankanhalli. Automated aesthetic enhancement of videos. In *Proceedings of the international conference on Multimedia*, MM '10, pages 281–290, New York, NY, USA, 2010. ACM.

[42] Y.-H. Yang and H. Chen. Ranking-based emotion recognition for music organization and retrieval. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(4):762 –774, may 2011.

[43] 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.