

Towards Aesthetics: a Photo Quality Assessment and Photo Selection System

Congcong Li
School of ECE, Cornell University
Ithaca, NY, USA
cl758@cornell.edu

Alexander C. Loui
Eastman Kodak Company
Rochester, NY, USA
alexander.loui@kodak.com

Tsuhuan Chen
School of ECE, Cornell University
Ithaca, NY, USA
tsuhan@ece.cornell.edu

ABSTRACT

Automatic photo quality assessment and selection systems are helpful for managing the large amount of consumer photos. In this paper, we present such a system based on evaluating the aesthetic quality of consumer photos. The proposed system focuses on photos with faces, which constitute an important part of consumer photo albums. The system has three contributions: 1) We propose an aesthetics-based photo assessment algorithm, by considering different aesthetics-related factors, including the technical characteristics of the photo and the specific features related to faces; 2) Based on the aesthetic measurement, we propose a cropping-based photo editing algorithm, which differs from prior works by eliminating unimportant faces before optimizing photo composition; 3) We also incorporate the aesthetic evaluation with other metrics to select quintessential photos for a large collection of photos. The entire system is delivered by a web interface, which allows users to submit images or albums, and returns promising results for photo evaluation, editing recommendation, and photo selection.

Categories and Subject Descriptors

I.4.9 [Image Processing and Computer Vision]: Applications

General Terms

Algorithms, Experimentation, Measurement.

Keywords

Aesthetic quality assessment, photo selection, photo editing.

1. INTRODUCTION

The proliferation of consumer photos raises an issue that manually managing a large number of photos would be time-consuming. It becomes necessary and important to develop automated tools for efficient browsing, selecting, and managing photos. It has raised great attention in recent research, and automated tools based on different photo characteristics have been proposed [1-4] for better organizing and retrieving of photos. Among these solutions, one promising direction is to assess a collection of photos based on their *aesthetic* visual quality, which plays an important role in the photo selection process of common people. In this paper we present an automated system, which can evaluate the aesthetic

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'10, October 25–29, 2010, Firenze, Italy.
Copyright 2010 ACM 978-1-60558-933-6/10/10...\$10.00.

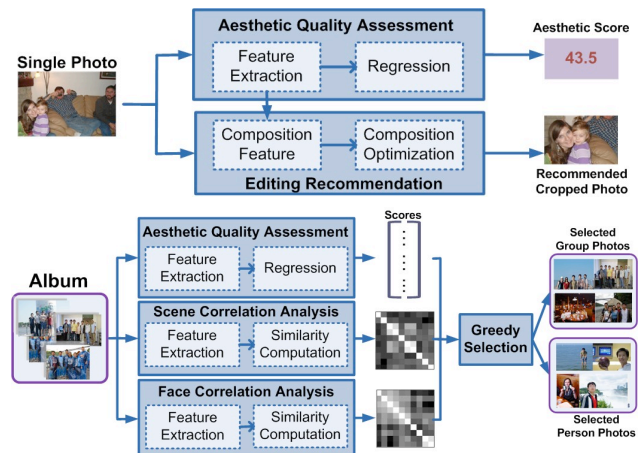


Figure 1. Overview of the proposed system. The upper part shows the flow chart for a single photo submission. The bottom part shows the flow chart for an album submission.

visual quality of photos, provide recommendations for photo editing, and select quintessential photos for large collections.

Previous works in aesthetics-based visual quality assessment have shown great potential in using machine learning algorithms to predict the subjective photo quality [3-6]. Based on the related research in [3], An efficient system *Acquino* has been published, which allows users to submit photos and automatically returns the prediction on the photo's aesthetic quality. However, since the algorithm in [3] is mainly trained to discriminate professional photos and unprofessional photos, it could be not sensitive enough to consumer photos with different qualities. Moreover, only predicting the quality scores of photos may not fully satisfy the users' needs in an easier album management. Therefore, in this work, we build up an aesthetic quality evaluation system that aims at evaluating consumer photos, more specifically, consumer photos with faces. Based on the quality assessment, our system also provides functions to recommend cropping-based editing for photos and select quintessential photos for large albums.

Different from most prior works that measure statistic features across the entire image, or focusing on a general main subject region discovered by object-free methods, our aesthetic evaluation method emphasizes the features related to the facial regions. This is from our common sense that faces are often the main intention when they are included in nonprofessional photos. Therefore, when providing recommendations on editing a photo, our solution also focuses on the facial regions and takes into account the importance of different faces. We currently provide an automatic cropping recommendation for refining the face composition. Recommendations on editing the lighting, color, or other

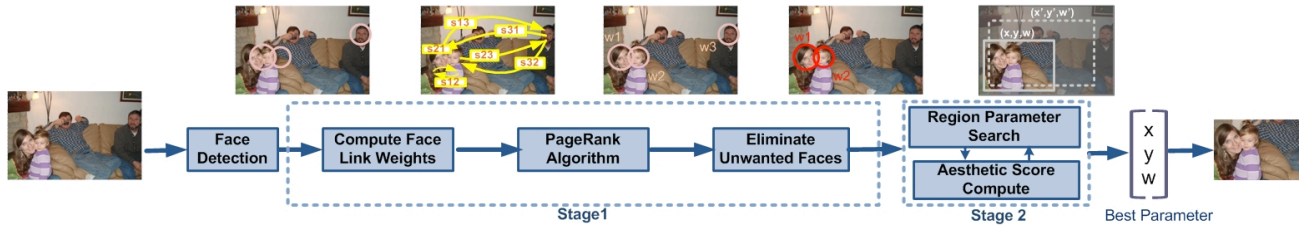


Figure 2. The composition optimization method. Stage 1: discover the importance of different faces and eliminate the unwanted ones. Stage 2: Optimize the composition aesthetic score. Details are given in Section 3.2.

properties will be part of our future work. Another feature in our system is to help select quintessential photos for large albums. In this photo selection process, the aesthetic quality of photos serves as an important metric. Besides the photo appeals, our system also takes into account the representation of the scenes and persons, and performs a greedy algorithm to optimize the selection output.

2. OVERVIEW

Figure 1 gives an overview of the proposed system. The system can handle two types of inputs: a single photo or a photo album. If the input is a single photo, the system automatically returns a quality score (on 0–100 scale) by performing aesthetic quality assessment on the image. Furthermore, if the score of the photo is lower than 70, the system will provide an auto-crop editing recommendation for the image to achieve better composition and higher score. If the input is a photo album, the system automatically performs photo selection on the collection. The top ten essential group photos and the top ten essential individual photos will be displayed on the screen. The system can also return the ranks for all images in the collection, separately in two sets: the group photo set and the individual photo set. Algorithms for these three functions are introduced in Section 3.

3. ALGORITHMS

3.1 Aesthetic quality assessment

A commercial face detector and the Active Shape Model [7] are used to detect faces and locate the positions of 82 characteristic facial points¹. To represent the aesthetic quality of a photo, we consider the following factors and design computational features.

Color and lighting. Color and lighting play an important role in photography and highly affect the viewer’s impression towards an image. Instead of using the color and lighting statistics across the entire image, we focus on measuring the color and lighting statistics in the face region and the lighting and color contrasts between the face region and background region.

Composition. Composition is the special arrangement of visual elements in the photo. It is essential in the creation of a variety of artistic works. The most important elements in the photos we consider here are the faces. Therefore, we measure the spatial organization of faces and their relationship towards the classic photography rules, such as the rule of thirds, visual balance, etc. We will further discuss this in Section 3.2.

Face characteristics. As discussed earlier that a face draws much of the viewer’s attention, it is also intuitive to consider features related to the faces, such as individual face expressions, individual face poses, between-face distances, the consistency of expressions on multiple faces, etc. These features not only depict the individual faces, but also discover the relationship between people in the photo, which could affect the viewer’s emotion towards the

photo appeals. More details about the feature computations are described in [8].

Based on the extracted features, we use the *sparsity-embedded linear regression* model to map a feature vector onto an aesthetic score. The sparsity in the model is introduced by adding a Laplace prior on the model parameter [9]. The ground-truth data used for training are collected in an earlier rating survey on 500 photos [8] through Amazon Mechanical Turk. Instead of using all features proposed in [8], the sparse model here automatically determines a small subset of important features, and improves the efficiency of the evaluation algorithm. It makes possible turning the algorithm into an online application, which we will discuss in Section 4. After selection, the total number of features is 35 for group photos (face number > 1) and 29 for individual photos (face number = 1).

3.2 Aesthetics-based photo editing

When a user submits a photo for evaluation and obtains a low score, the user could be interested in how to edit the image to improve the photo’s appeal. Commercial software such as *Photoshop* provides tools for refining an image, but it often requires the user to know well about design concepts and photography theory. Our system aims to automatically provide some recommendations on editing the photo to achieve better aesthetic quality. Currently, we focus on improving the composition of a photo with an auto-cropping algorithm.

There have been some attempts in automatic image cropping. For example, Zhang et al. [10] utilize composition rules to crop photos by using face detection. Nishiyama et al. [11] apply a statistically built quality classifier to choose a good cropping candidate. Liu et al. [12] develop a crop-and-retarget operator to modify the composition aesthetics of an image. Some commercial software such as Picasa has also provided an automatic cropping function in its interface. Most of these existing algorithms that give special attention to face regions often tend to retain all of the faces of the original image. However, it is common that unwanted faces appear in consumer photos. Therefore, we propose a two-stage algorithm to optimize the face composition in an image. As shown in Figure 2, the first stage is to *eliminate unwanted faces* and the second stage is to search a region with optimal aesthetic composition. Detailed steps of the algorithm are given below:

1. With the face detection result, compute the weights for the directed links between faces: $s(i, j) = size(j) / dist(i, j)$, where $s(i, j)$ is the score for link $i \rightarrow j$, $size(j)$ is the size of the face j and $dist(i, j)$ is the distance between the two faces.
2. Perform the PageRank algorithm [13]. The output PageRank value is used as the importance weight for each face: $w(i)$.
3. Eliminate the faces whose weight is small than 25% of the maximal $w(i)$.
4. Search the best region that optimizes the composition aesthetic score. The computation aesthetics is measured as: $E = \alpha_1 E_{RT} + \alpha_2 E_{VB} + \alpha_3 E_{SZ}$, where E_{RT} measures how well

¹ These modules are provided by Kodak Research Laboratories.

the face composition fits the Rule of Thirds, E_{VB} measures the visual balance, and E_{SZ} measures the influence of the relative face sizes compared to the image. The computation of E_{RT} , E_{VB} and E_{SZ} can be referred to [12], by replacing the saliency regions and their weights with the face regions and the face importance weights found in the first stage.

To find the best region, the search is performed in the space of the parameters that define the region. To limit the dimensionality of the search space, we keep the aspect ratio of the cropped image the same as that of the original image. We also maintain the relative positions of objects in the original image. Therefore, we have only three parameters: x and y position of the cropped image and its width. Similarly to prior work [10,12], we use the Particle Swarm Optimization method (PSO) [14] to seek the optimal solution. The cropped image will be shown to the user as a photo editing recommendation.

3.3 Aesthetics-based photo selection

Our system can take in a large collection of photos and help the user to select the most quintessential ones. We assume the submission is a personal photo collection of the user. It can be a collection of photos taken during a trip with friends, a series of recent family events, and so on. The photos might contain different individuals, and can be taken in different places and by different cameras. The goal of the photo selection function is to select the most quintessential photos for the collection. In our system, we measure the quintessence with two properties: appeal and representation. Considering the photo appeal, we expect the selected photos to have high scores predicted by our aesthetic quality assessment algorithm. Considering the representation, we expect the selected photos to cover as many scenes and as many people that have appeared in the collection. Therefore, in addition to the aesthetic quality evaluation, we also measure the scene similarity and face identity similarity between the photos.

We first divide the photos into two subsets. Photos with more than one face are considered to be group photos, while the remainder is considered to be photos of individuals. The subsequent processes are performed separately on the two subsets.

For group images, we mainly consider the aesthetic quality and scene coverage while selecting quintessential images. For individual images, we also consider the coverage of faces in the selected images. Before the selection stage, we compute the scene similarity and person similarity between images. Features we use to represent a scene of a photo include a 512-dimension gist feature (extracted in 4 scales and 8 orientations) and a 12-dimension background color feature (the mean and variance values of the background pixels in RGB space and YCrCb space). Features we use to represent a face identity are the 100-dimension projection of the face region onto a pre-trained eigen-space.

Take the group photo subset as an example. Assume that Γ is the entire set of group photos and E is the set of selected photos. The first photo added into E is the one that has the highest aesthetic score in the set Γ . Then we repeatedly adding photos one by one according to Equation (1), where a_i is the aesthetic score of the i^{th} photo in the candidate set $(\Gamma - E)$, and \bar{s}_i represents the scene features of the corresponding photo. $SIM(\cdot, \cdot)$ is a similarity function of two vectors. We perform a greedy search by optimizing Equation (1) in every iteration. Each time we select a photo that has the best combination of a high aesthetic score and a low scene correlation with the previous selected photos.

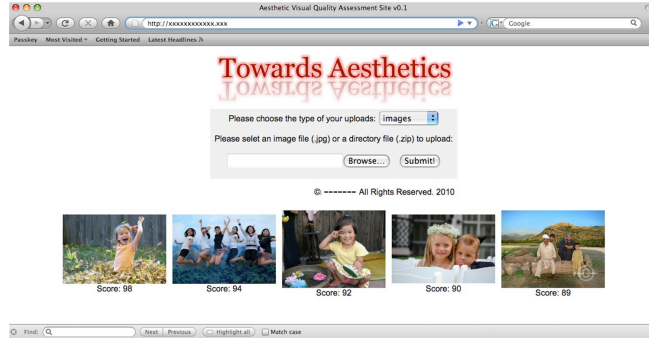


Figure 3. A snapshot of the website interface of the system

$$\max_{i \in (\Gamma - E)} \left[a_i + \min_{j \in E} \left[1 - SIM(\bar{s}_i, \bar{s}_j) \right] \right] \quad (1)$$

Similarly, we perform photo selection in the individual photo subset according to Equation (2). The only difference here is that we also take into account the coverage of different faces. The selected photo is preferred to have low face correlation with those in the previous selected photos. \bar{f}_i represents the face identity features of the corresponding photo.

$$\max_{i \in (\Gamma - E)} \left[a_i + \min_{j \in E} \left[1 - SIM(\bar{s}_i, \bar{s}_j) \right] + \min_{k \in E} \left[1 - SIM(\bar{f}_i, \bar{f}_k) \right] \right] \quad (2)$$

We repeat the selection process until each image gets its rank. The system displays the top 10 photos of each subset in its interface and returns to the user's two sets of photos, which are named by their ranks in the corresponding set.

4. EXPERIMENTS

4.1 Web interface

We set up a web interface for the system, as shown in Figure 3. (Website link: <http://chenlab.ece.cornell.edu/projects/aesthetics/>.) It allows the user to choose uploading "image" or "directory." After the "submission" button is clicked, the system will perform the corresponding functions as described in Section 2. Figure 4 and Figure 5 are the result interfaces for a single image submission example and a photo collection submission example.²

4.2 Results

Aesthetic quality assessment. We evaluate our aesthetic quality assessment algorithm by a 50-fold cross validation test. We use the 500 images collected in [8] with human rating scores as ground truth. We leave out 10 images for testing each time and use the remaining for training the *sparse* linear regression model. We use the residual sum-of-square error to measure the score predication performance. We get Res = 228 for our algorithm, compared to Res = 317 for random guess. The 28% error reduction confirms that our method is able to predict the human-rated aesthetic score with some success, considering the subjective challenge involved.

Aesthetics-based photo editing. We evaluate the photo-editing algorithm by comparing with the auto-crop function in the Picasa software. We randomly select 100 images with automated aesthetic scores lower than 70. We upload these images onto

² The faces shown in the photos are smoothed for privacy protection. The blurring does not represent their original quality.



Figure 4. Result Interface for a single image submission

Picasa and perform the auto-crop function. *Picasa* often returns three candidates for the user to select. Our algorithm only returns one recommendation. We present 100 groups of images to 8 nonprofessional photographers, in random order. Each group contains five images: the original image, the recommended cropped image from our method, and three images from the *Picasa* auto-crop function. We request the subjects to select the photo with the best composition in each group. Among the 800 results, 51% voted for our proposed method, 37% voted for one of the three *Picasa* recommendations, and the remaining 12% voted for the original image. This shows our method can provide competitive results with popular tools.

Figure 4 shows the result interface for a single image submission. The aesthetic score of the uploaded image is shown under the uploaded image. The photo shown next to the original image is created by the composition optimization algorithm, which is introduced in Section 3.2. A “Download” button is provided to the user for downloading the auto-edited photo. The interface also shows another four photos from the existing database, which attained scores nearest to that of the submitted photo. The interface allows the user to provide opinions about whether these score-similar photos look better than the submitted one. Although we have not used the user feedback in the current algorithm, we believe the pair-wise comparison feedback would provide us useful information to retrain the algorithm, or to develop online learning algorithms in the future.

Aesthetics-based photo selection. Figure 5 shows the resulting interface for a photo collection submission. In this example, we submit a collection including 84 photos. Faces are detected in 80 photos. A total 38 photos contain only one face for each, and the remaining photos contain more than one face. The interface shows the top 10 quintessential group images and the top 10 quintessential individual images. From the results we can see that the selected group images include different scenes and different persons. The interface also allows the user to download results in a text file, which lists the file names of the submitted photos and their corresponding ranks in the group/individual subset.

5. CONCLUSIONS

In this paper, we propose an automated system for photo evaluation, photo editing recommendation, and photo selection, based on the assessment of the aesthetic visual quality of



Figure 5. Result Interface for a photo collection submission

consumer photos. The system provides promising results in statistical performance evaluation, human study evaluation, and also shows nice visualized effects.

The system currently focuses on a specific genre of photos: consumer photos with faces, which constitute an important part of personal photo collections. We would like to extend the system to cover more genres in our future work. The system has also set up a mechanism to collect users’ ideas on the automatic quality assessment results. This allows us to use dynamic information to further refine our learning algorithm in the future.

6. REFERENCES

- [1] A. Loui, M. D. Wood, A. Scalise, and J. Birkelund. Multidimensional image value assessment and rating for automated albuming and retrieval. In *Proc. ICIP 2008*.
- [2] W. Chu, C. Lin, and J. Yu. Feature classification for representative photo selection. In *Proc. ACM Multimedia 2009*.
- [3] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Proc. ECCV 2006*.
- [4] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Proc. CVPR 2006*.
- [5] C. Li, and T. Chen. Aesthetic visual quality assessment of paintings. *IEEE Journal of Selected Topic in Signal Processing*, vol.3, no.2, pp. 236–252, April, 2009.
- [6] Y. Luo, and X. Tang. Photo and video quality evaluation: focusing on the subject. In *Proc. ECCV 2008*.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995.
- [8] C. Li, A. Gallagher, A. C. Loui, T. Chen. Aesthetic quality assessment of consumer photos with faces. In *Proc. ICIP 2010*.
- [9] H. Lee, A. Battle, R. Raina, A.Y. Ng. Efficient sparse coding algorithms. In *Proc. NIPS 2007*.
- [10] M. Zhang, L. Zhang, Y. Sun, L. Feng, and W. Y. Ma. Auto cropping for digital photographs. In *Proc. ICME 2005*.
- [11] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *Proc. ACM Multimedia 2009*.
- [12] L. G. Liu, R.J. Chen, L. Wolf, and D. Cohen-Or. Optimizing photo composition. In *Proc. Eurographics 2010*.
- [13] S. Brin and L. Page. The anatomy of a large-scale hyper-textual web search engine. In *Proc. WWW 1998*.
- [14] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proc. IEEE Conf. on Neural Networks 1995*.