AUTOMATIC AESTHETIC VALUE ASSESSMENT IN PHOTOGRAPHIC IMAGES

Wei Jiang *Alexander C. Loui* *Cathleen Daniels Cerosaletti*

Kodak Research Labs, Eastman Kodak Company, Rochester, NY
Email: {wei.jiang, alexander.loui, cathleen.cerosaletti}@kodak.com

ABSTRACT

The automatic assessment of aesthetic values in consumer photographic images is an important issue for content management, organizing and retrieving images, and building digital image albums. This paper explores automatic aesthetic estimation in two different tasks: (1) to estimate fine-granularity aesthetic scores ranging from 0 to 100, a novel regression method, namely *Diff-RankBoost*, is proposed based on RankBoost and support vector techniques; and (2) to predict coarse-granularity aesthetic categories (*e.g.*, visually "very pleasing" or "not pleasing"), multi-category classifiers are developed. A set of visual features describing various characteristics related to image quality and aesthetic values are used to generate multidimensional feature spaces for aesthetic estimation. Experiments over a consumer photographic image collection with user ground-truth indicate that the proposed algorithms provide promising results for automatic image aesthetic assessment.

Keywords— Aesthetic image value estimation, consumer photographic image

1. INTRODUCTION

The proliferation of digital cameras has led to an explosion in the number of digital images created, resulting in personal image databases large enough to require automated tools for efficient browsing, searching, and album creation. One possible method for automatic image management is to assess a collection of images according to characteristics such as image quality and aesthetic value, which is very useful for organizing and retrieving images, and for building digital albums and other creative outputs. In this paper we study how to automatically assess the aesthetic characteristics of images by machine learning methods.

There has been some recent work on characterizing photographs based on aesthetic quality as well as developing predictive algorithms [2, 3, 5, 7, 10]. In particular, the recent study by Cerosaletti and Loui [2] provides empirical understanding of the perceptual attributes of aesthetics by manipulating some important image variables within a set of consumer photographic images ranging in the sophistication of techniques used to capture the images.

In this paper we address the issue of automatic estimation of images' aesthetic values. Based on the previous empirical studies [2, 3, 5, 7, 10] a set of visual features describing various characteristics related to image quality and aesthetic values are used to generate multidimensional feature spaces, on top of which machine learning algorithms are developed to estimate images' aesthetic scales in two different tasks. In the first task we estimate fine-granularity aesthetic scores ranging from 0 to 100. To this end, a novel regression method, the *Diff-RankBoost* algorithm, is proposed based on RankBoost [6] and support vector techniques [11]. In the second task, our aim is to predict coarse-granularity aesthetic categories. That is, we care about five categorical aesthetic scales (*e.g.*, visually "very pleasing" or "not pleasing"), but not the exact aesthetic scores. Multi-category classifiers can be developed to solve this problem. In addition, the second classification task (that requires less training resources than the first regression task in learning a good model) enables us to use the limited training data for studying face and non-face images separately, which is often necessary since people tend to use different criteria to judge aesthetic values for images with or without faces [2].

We evaluate our algorithm over the consumer photographic image collection from [2], where ground-truth aesthetic values have been obtained through a comprehensive user study. In the rest of the paper, we provide details of our algorithms in the above-mentioned two tasks, followed by the image collection and visual features for aesthetic estimation. After that we will give experimental results and discussions.

2. AUTOMATIC AESTHETIC ASSESSMENT

We start with some terminologies. Assume that we have a data collection $\mathcal{D}=\{(x_n, y_n)\}$, $n=1,\ldots,N$, where each image x_n is represented by a d-dim feature vector $\mathbf{f}_n \in \mathbb{R}^d$, and y_n is the aesthetic score of image x_n ranging from 0 to 100. Our task is to learn a model based on data set $\mathcal{D}$ so that for a new input image x_0 we can predict its aesthetic score $\hat{y}_0$. In the following we will introduce two aesthetic estimation methods, one for predicting fine-granularity aesthetic scores by a Diff-RankBoost algorithm, and the other for estimating coarse-granularity aesthetic categories.

2.1. Diff-RankBoost with Relative Aesthetic Ranking

The RankBoost algorithm [6] tries to address the ranking problems by working on data points that are ranked as input. In this work we use the RankBoost framework to generate a list of relative ranking for a training or testing data, which are then used as features to feed into a *Support Vector Regression* (*SVR*) model

978-1-4244-7493-6/10/$26.00

[4] for predicting aesthetic scores.

Compared to direct SVR over raw features, the proposed Diff-RankBoost has the following advantages. First, human-annotated data based on ranking are more reliable than those based on exact scoring [1], due to the intrinsic subjectivity of human annotation. As a result, a model learned on top of relative data ranks, instead of raw scores, may rank test data better. Second, the outputs of Diff-RankBoost are a set of ranking differences, which can be used in several ways besides estimating exact aesthetic scores in this paper. For example, we can directly compare relative aesthetic rankings of new image pairs. Finally, Diff-RankBoost provides a natural way to combine different visual features for enhanced aesthetic estimation. In comparison, it may be difficult to appropriately normalize different features for combination in direct SVR.

2.1.1. Traditional RankBoost

Assume that $g(x_n)$ is the ground-truth rank of data x_n within the entire data set $\mathcal{D}$, where $g(x_n) < g(x_m)$ implies that data x_n has a better rank than data x_m. For each pair of data (x_n, x_m), define a feature $w(x_n, x_m)$ so that:

$$w(x_n, x_m) = \begin{cases} C_{nm}, \text{ if } g(x_n) > g(x_m) \\ 0 \quad , \text{ otherwise} \end{cases} \tag{1}$$

where w values are normalized to be a distribution, $\sum_{n,m} w(x_n, x_m)$ =1, and $C_{nm} \in [0, 1]$ determines the degree that x_m ranks better than x_n.

The output of RankBoost is a new ranking function, $H(x_n)$, which defines a linear ordering on data points, *i.e.*, $H(x_n) < H(x_m)$ if x_n has a better rank than x_m. We have the following cost function to minimize the missranking error:

$$\mathcal{L} = \sum\nolimits_{n,m} w(x_n, x_m) I\left[H(x_n) < H(x_m)\right]$$

The boosting process can be used to minimize cost $\mathcal{L}$ and the original RankBoost algorithm is summarized in Figure 1. The key issue in the algorithm is how to design the weak learner in each iteration t to generate the relative data ranking h^t.

Input: Training set $\mathcal{D} = \{(x_n, y_n)\}$, $n = 1, \ldots, N$. Initialize $w^1(x_n, x_m)$ according to Eqn. (1). Set $H^1(x_n)$=0, n, m=1, $\ldots, N$.

Iteration: for $t = 1, \ldots, T$

- Get a weak learner to obtain ranking $h^t(x_n)$, n=1, $\ldots, N$.
- Calculate the error measurement of the weak learner: $r^t = \sum_{n,m} w^t(x_n, x_m)\left[h^t(x_n) - h^t(x_m)\right]$; set $\alpha^t = \frac{1}{2}\log\frac{1+r^t}{1-r^t}$.
- Update: $w^{t+1}(x_n, x_m) = w^t(x_n, x_m) e^{\alpha^t\left[h^t(x_n) - h^t(x_m)\right]}$, and re-normalize weights so that $\sum_{n,m} w^{t+1}(x_n, x_m) = 1$.
- Update $H^{t+1}(x_n) = H^t(x_n) + h^t(x_n)$, $n = 1, \ldots, N$.

Fig. 1. The original RankBoost algorithm

.

2.1.2. The Proposed Diff-RankBoost Algorithm

We propose a weak learner model trained based on data differences, described as follows. Each data x_n is represented by a feature vector $\mathbf{f}_n$, and for each pair of data (x_n, x_m) we generate a new data point $\hat{x}_{nm}$, where the feature vector representation for this new data point is the difference between $\mathbf{f}_n$ and $\mathbf{f}_m$ *i.e.*, $\mathbf{f}_n - \mathbf{f}_m$, and the label for the new data point is $\hat{y}_{nm} = 1$ if x_m ranks better than x_n and -1 otherwise. Each $(\hat{x}_{nm}, \hat{y}_{nm})$ pair forms a new data sample, and each data sample is associated with weight $w(x_n, x_m)$. We can train a binary classifier based on these new data samples and their weights (we have in total of $N(N-1)/2$ samples generated from N original training data) to predict the relative ranking of data samples.

The idea of using difference-based classification has been successfully used in face recognition [9], where a multi-class face recognition problem is cast into a binary intrapersonal/extrapersonal classification task. The major advantage of such difference-based classification is the well-clustered data points in the vast and sparsely populated high-dimensional space, resulting in better density modeling/estimation for classification.

In this work we train a kernel-based SVM classifier [11] based on the generated difference data samples. For each pair of data, SVM gives the output prediction $d(x_n, x_m)$ where $d(x_n, x_m) > 0$ if x_m is predicted to rank better than x_n. Value d indicates how much higher or lower x_m ranks better than or worse than x_n, *i.e.*,

$$d(x_n, x_m) = h(x_m) - h(x_n) \tag{2}$$

Using Eqn. (2), the original RankBoost algorithm described in Figure 1 turns to our Diff-RankBoost algorithm in Figure 2. The output of Diff-RankBoost is the difference of data ranking between pairs of data points, $H(x_n)$–$H(x_m)$, $n, m = 1, \ldots, N$. So for each input data x_0, through Diff-RankBoost we can get a set of prediction about the ranking differences between this data and all the training data: $H(x_0) - H(x_n)$, $n = 1, \ldots, N$. In the next Section 2.1.3, this set of ranking differences generates an input feature to fit an SVR model for estimating aesthetic scores.

Input: Training set $\mathcal{D} = \{(x_n, y_n)\}$, $n = 1, \ldots, N$. Generate a new data set with $N(N-1)/2$ new samples where each new data $\hat{x}_{nm}$ is associated with feature vector $\mathbf{f}_n$–$\mathbf{f}_m$ and a binary label $\hat{y}_{nm}$. Initialize weight $w^1(x_n, x_m)$ according to Eqn. (1), and set $H^1(x_n)$–$H^1(x_m) = 0$, $n, m = 1, \ldots, N$.

Iteration: for $t = 1, \ldots, T$

- Train an SVM classifier based on the new data samples and generate $d^t(x_n, x_m) = h^t(x_m) - h^t(x_n)$.
- Calculate the error measurement of the weak learner: $r^t = -\sum_{n,m} w^t(x_n, x_m) d^t(x_n, x_m)$; set $\alpha^t = \frac{1}{2}\log\frac{1+r^t}{1-r^t}$.
- Update: $w^{t+1}(x_n, x_m) = w^t(x_n, x_m) \exp\{-\alpha^t d^t(x_n, x_m)\}$, and re-normalize so that $\sum_{n,m} w^{t+1}(x_n, x_m) = 1$.
- Update $H^{t+1}(x_n) - H^{t+1}(x_m) = H^t(x_n) - H^t(x_m) - d^t(x_n, x_m)$, $n, m = 1, \ldots, N$.

Fig. 2. The Diff-RankBoost algorithm

2.1.3. SVR to Estimate Aesthetic Scores

An SVM [11] constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification and regression. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Specifically, given a set of labeled data

$\mathcal{D}=\{(x_n,y_n)\}$, $n=1,\ldots,N$, the goal of SVR [4] is to find a function $F(x_n)=\mathbf{w}^T\phi(x_n)+b$ that has at most ϵ deviation from the actual targets y_n for all the training data, and at the same time maintains the large-margin property. $\phi(\cdot)$ maps the original data into a high-dimension feature space. By introducing slack variables ξ, ξ^*, the cost function of SVR is:

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{n=1}^{N}(\xi_n+\xi_n^*) \quad (3)$$
$$s.t.\ \ \mathbf{w}^T\phi(x_n)+b \le y_n+\epsilon+\xi_n,$$
$$\mathbf{w}^T\phi(x_n)+b \ge y_n-\epsilon-\xi_n^*, \xi_n, \xi_n^* \ge 0, n=1,\ldots,N$$

The constant $C>0$ determines the trade-off between the large-margin constraint and the amount up to which deviations larger than ϵ are tolerated. In our case, each input x_n (which can be a training or test sample) is represented by a feature vector $[\tilde{f}_1(x_n),\ldots,\tilde{f}_N(x_n)]^T$ where each $\tilde{f}_m(x_n)=H(x_n)-H(x_m)$, $m=1,\ldots,N$. That is, the feature vector consists of the set of ranking distances between x_n and all the training data, predicted by Diff-RankBoost in Figure 2. This feature is used to fit the SVR model in Eqn. (3) to estimate aesthetic scores.

2.2. SVM Classification with Quantized Categories

In this subsection, we consider the situation where we only care about the coarse aesthetic categorization, but not the exact aesthetic score. There are several reasons why we want to study such a problem. First, sometimes we do not have enough training data to learn a precise regression model for estimating the fine-granularity aesthetic values, and in comparison, the coarse-granularity aesthetic categorization is an easier issue to tackle where fewer training samples are required to obtain a relatively satisfactory classifier. Second, usually people do not care about exact aesthetic values, *e.g.*, whether an image is 96% visually pleasing or 91% visually pleasing is not that important. It matters more that whether an image "is very much visually pleasing" or just "looks fine". That is, it may be necessary to better use the limited training data to study the categorization problem, which can be more important than the difficult aesthetic value estimation issue. Third, from the user study, people tend to use different criteria to judge aesthetic scores for images with or without faces [2]. We need to study face and non-face images separately. In the previous fine-granularity aesthetic score estimation problem, if we separate face and non-face images and study each subsets individually, the issue of insufficient training data may be even worse. In comparison, the coarse-granularity aesthetic category classification problem enables us to study face and non-face subsets separately.

Specifically, we quantize the aesthetic scores into 5 categories: "very bad" for $y_n \le 20$; "medium bad" for $20 < y_n \le 40$; "neutral" for $40 < y_n \le 60$; "medium good" for $60 < y_n \le 80$; and "very good" for $y_n > 80$. At the same time, the entire data set is separated into face and non-face subsets by a face detection tool from Omron® (http://www.omron.com/). Then a five-category SVM classifier [11] is trained over the face and non-face subsets individually in the one-vs.-all manner. And finally for new input data, the five-category SVM classifier can predict the aesthetic categories of these new data.

3. DATA SET FOR EVALUATION

We evaluate the proposed algorithms over 450 real consumer photographic images [2], selected from a number of different sources: Flickr®, Kodak Picture of the Day, study observers, and an archive of recently captured consumer image sets. Half of the images contain people or animals as the main subject. For both people and no people images, there are two levels of main subject size ("small-medium" and "medium-large"), and six levels of type of perspective cue ("center", "vertical", "horizontal", "down", "up", and "none"). This design balances the size of the main subject. A "small-medium" main subject size is defined as the main subject size consuming less than or equal to 10% of the image area and a "medium-large" main subject size consumes greater than 10% of the image area. An attempt was made to manipulate the perception of perspective within the image set. Perspective cues are mainly lines or angles in the image that draw the eye in a specific direction or give the impression of directionality. For the "center", "vertical", and "horizontal" perspective cue cases, the image lines point to the center or mostly vertical or horizontal. For "up" and "down "perspective cues, the camera is looking down on or looking up at the main subject, respectively. Finally, the images that have no perceived perspective cues are categorized as "none".

These images were selected with careful attention to including indoor and outdoor and natural and man-made subject matter. The image set includes a range of apparent sophistication of the photographer which ensured differentiation in overall image quality. Consumers who are amateur photographers carefully captured some of the images using sophisticated techniques. But, other images are snapshots (see Figure 3). A number of these snapshots are of typical "consumer quality" and therefore, contain a variety of consumer picture-taking problems such as exposure problems, poor framing, and lack of interesting subject matter. Consumers range greatly in their picture-taking skill and this mixed image collection in total provides a balanced platform to understand the perception of aesthetics by consumers in consumer picture sets.

Fig. 3. Study image examples from a consumer image set and from the Kodak Picture of the Day (POD) archive.

The ground-truth aesthetic values over the 450 images were obtained through a user study from 30 observers. The study data collection consisted of two, two-hour sessions for each observer. All observers are digital camera users, who capture photographs casually and share photographs with other people. The following instructions were provided to the users: "Rate each image individually on a 0 to 100-point scale. The scale is bi-anchored at the extreme ends with 'lowest imaginable' and 'highest imaginable' for overall 'artistically pleasing'. The

scale is defined as follows: 0 = 'lowest imaginable' and 100 = 'highest imaginable'. For example, the highest imaginable image may be seen in a coffee table book. The lowest imaginable image may be the result of accidentally pressing the shutter release button before capturing the real image. Base your judgments on how artistically pleasing the images are to you. Feel free to choose numbers between those marked on the table such as a 32 or a 76." A 0 to 100-point scale was used to ensure that observers can adequately differentiate and directly scale the images for artistic quality.

4. EXPERIMENTS

The above data set is randomly split in half for training and testing, respectively. For the first task of predicting fine-granularity aesthetic scores ranging from 0 to 100, the Square of MSE between the predicted aesthetic scores and the ground-truth scores is used as performance measure. For the second task where we predict coarse-granularity aesthetic categories, we use two measures, *Cross-Category Error* (*CCE*) and *Multi-Category Error* (*MCE*). Let $c_n = 1, \ldots, 5$ be the aesthetic category of x_n, and $\hat{c}_n$ be the aesthetic category predicted by the classifier for x_n. CCE is defined as: $CCE(i) = \frac{1}{N}\sum_{n=1}^{N} I(c_n - \hat{c}_n = i)$. Figure 4 gives the CCE of random guess over the test set. The ideal CCE without prediction error is a δ function centering at 0. MCE measures the overall misclassification error based on CCE, and is defined as: $MCE = \sum_{i=-4}^{4} |i| CCE(i)$, where $|i|$ is the absolute value of i.

4.1. Features for Aesthetic Assessment

In this work, we use several different features to estimate the aesthetic values or classify aesthetic categories of consumer photographic images. These features are selected based on the previous empirical studies [2, 5, 7, 8, 10] that are generally related to human perception in evaluating image quality and aesthetic values. These features are as follows:

Colorfulness: the colorfulness of image, which is a function of brightness and saturation (1 dimension).

Contrast: edge blur contrast in gray level, including edge concentration in horizontal axis, edge concentration in vertical axis, overall sharpness, normalized contrast, and overall brightness (5 dimensions).

Symmetry: 8 types of symmetry in image [5] (8 dimensions).

VP Hist: histogram of vanishing points' 3D positions, where the whole space is quantized to 48 cells (48 dimensions).

VP Position: the position of the most confident vanishing point in image (2 dimensions).

Ke's: features developed by Ke *et al.* in [7] for photo quality assessment, including the spatial distribution of high-frequency edges, the color distribution, the hue entropy, the blur degree, the color contrast, and the brightness (6 dimensions).

Technical IVI: a multidimensional Image Value Index (IVI) consisting of numerical quality values to measure the image's technical quality, which is developed in [8] (12 dimensions).

Face: the number of faces in image, the size of the most confidently detected face, the horizontal and vertical position of the most confidently detected face (4 dimensions).

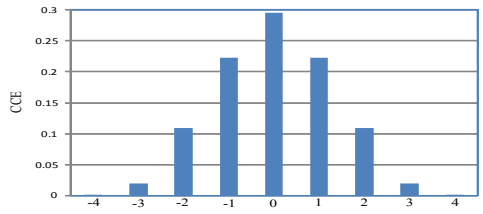


Fig. 4. CCE of random guess over the test set.

We extract these features in 3 different ways: over the entire image; over 3×3 image grids and then concatenate the features from the 9 image grids together; and over the face regions, which are usually especially interesting to consumers in evaluating aesthetic scales. The first type of features are named by adding a prefix "glob-", *e.g.*, glob-Contrast. The second type of features have the prefix "spatial-", *e.g.*, spatial-Contrast. The third type of features are applied to the most confidently detected face region (only to images where faces are detected), and have the prefix "face-", *e.g.*, face-Contrast.

4.2. Aesthetic Score Estimation

In this experiment, we evaluate our Diff-RankBoost algorithm in estimating the fine-granularity aesthetic scores of consumer photographic images. The straightforward SVR [4] is used as a baseline for comparison. We use the first type, features over global images, in this experiment. The other types are not used either because of the high dimensionality generated by concatenating features over spatial grids that adds difficulty to the regression task, or because of the incomplete feature over non-face images. To directly compare various features for aesthetic score estimation, we apply the baseline SVR over individual features, and the Square of MSE results are shown in Figure 5, where for "Early fusion" we concatenate all features into a long vector for regression, and for "Late fusion" we apply SVR over each individual feature and then averagely combine the regression results. From the figure, Technical IVI and Ke's features are much better than others in predicting the aesthetic scores. This is reasonable because both features are specially designed [7, 8] to describe the image quality and are important to consumers in evaluating the aesthetic scales. When we directly concatenate different features, the prediction performance can only improve a little.

Figure 6 gives the performance comparison between our Diff-RankBoost and the baseline SVR throughout different boosting iterations. For Diff-RankBoost, the algorithm described in Figure 2 is first applied to each individual feature to generate different sets of predicted ranking differences, and then these ranking differences are concatenated into a long vector to feed into an SVR model to estimate the final aesthetic score. From the result our Diff-RankBoost outperforms the baseline SVR consistently from the first iteration, and in practice 10 iterations can give a relatively good performance.

Figures 7 (a) and (b) show some example images with the best and worst aesthetic score estimation, respectively, using the Diff-RankBoost algorithm. From the figure we can see that the algorithm tends to make more mistakes in predicting some

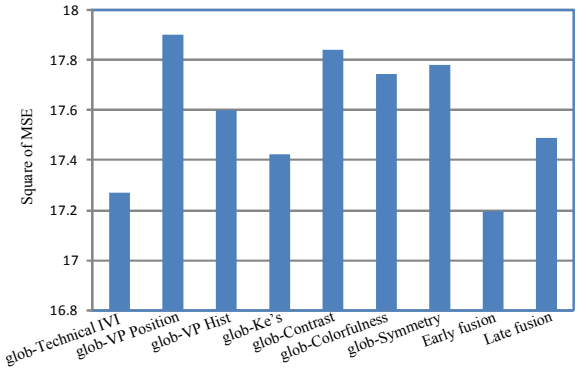


Fig. 5. Performance of SVR using various individual features for estimating fine-granularity aesthetic scores.

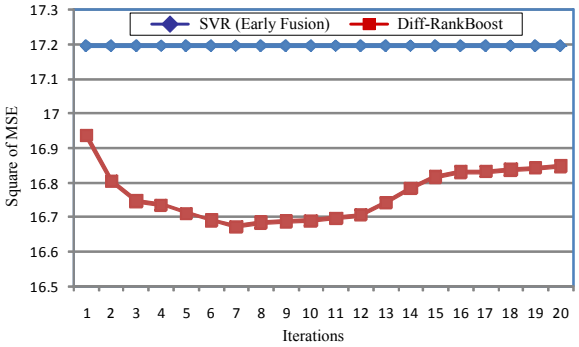
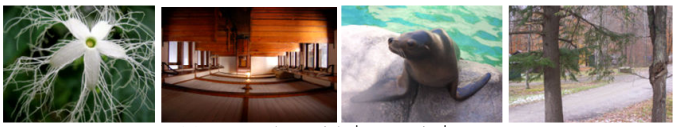


Fig. 6. Diff-RankBoost vs. SVR for aesthetic score estimation.

"very good" or "very bad" images. This is intuitively reasonable due to two reasons. First, such extreme-level aesthetic perception is more subjective compared to moderate-level aesthetic perception. For example, the first 2 images in Figure 7 (b) are professional images considered as visually very pleasing. However the aesthetic characteristics of such images can be very difficult to capture by low-level visual features. Second, usually there are only few images (so as training data) having extreme-level aesthetic scores, which are insufficient for learning a robust model.



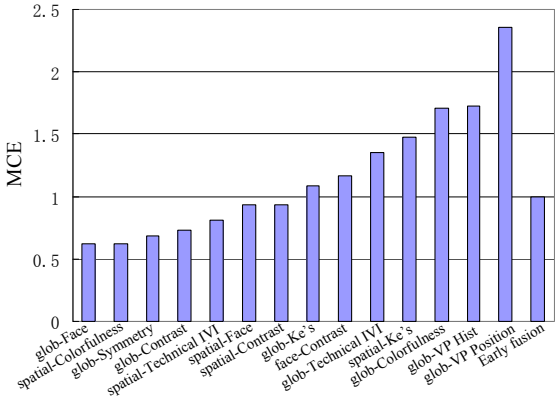
(a) examples with $|y_n - \hat{y}_n| < 5$.

(b) examples with $|y_n - \hat{y}_n| > 30$.

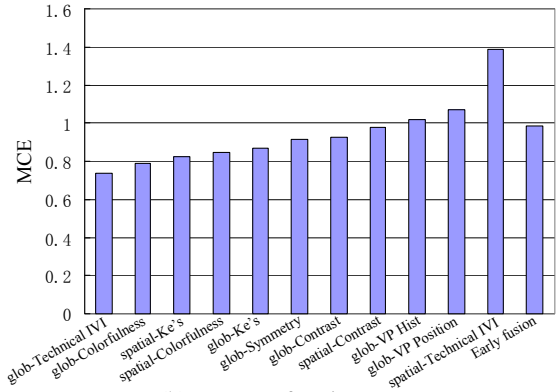
Fig. 7. Examples with best and worst score estimation. $\hat{y}_n$ and y_n are the predicted and ground-truth aesthetic scores of x_n, respectively. (a) Images with the best estimation usually have moderate-level scores, *i.e.*, $30 < y_n < 70$. (b) Images with the worst estimation usually have extreme-level scores, *i.e.*, $y_n < 25$ or $y_n > 75$.

4.3. Aesthetic Category Classification

In this experiment, we evaluate the SVM-based aesthetic category classification performance. As mentioned in Section 2.2 we separate the entire image set into face or non-face subsets and explore aesthetic category classification in each subset individually. The final test set contains 48 face and 177 non-face images, respectively. Figures 8 (a) and (b) show the MCE performances of multi-class SVM using different individual features for face images and non-face images, respectively. From the results, the classifier can achieve better performance over face images than non-face images. For face images, the global-Face feature works the best. For non-face images, global-Technical IVI works the best. Fusion of features can not generate any additional performance gain.



(a) over face images.

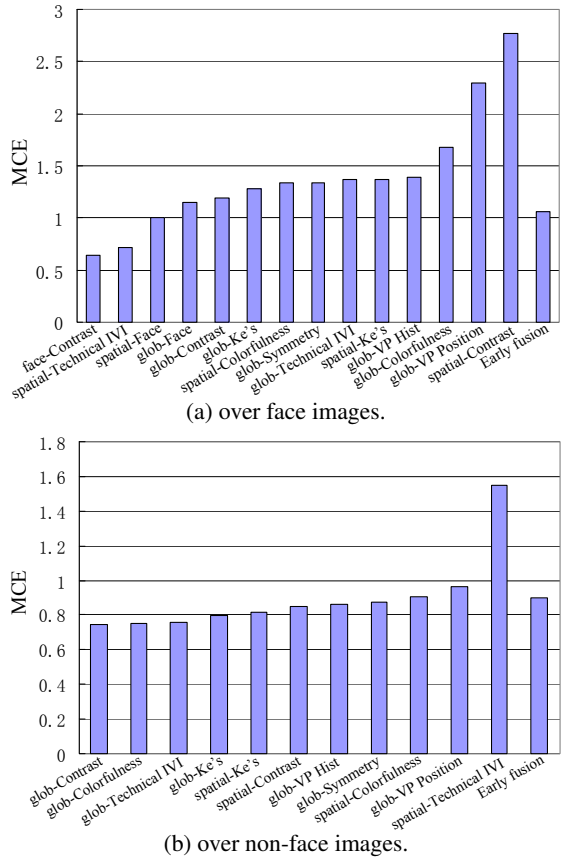


(b) over non-face images.

Fig. 8. Classifying aesthetic category with automatic face detection.

To obtain better understanding about the results, we also evaluate the MCE performance using ground-truth face annotation. The face regions in the 450 photographic images are manually cropped by users, according to which the images are separated into face and non-face subsets. The final test set contains 74 face and 151 non-face images, respectively. Then multi-class SVMs are learned over face and non-face subsets individually and we can get the final MCE result as shown in Figure 9. From this result, we can see that for face images, the face-related features are important for evaluating aesthetic categories and can give the best performances. For non-face images, in general global visual features work better than the spatial layout counterparts. Despite all the differences resulting from imperfect automatic face detection, both Figure 8 and Figure 9 suggest that global-Face and spatial-Technical IVI are effective for classifying the face subset, and global-Colorfulness and global-Technical IVI are effective for classifying the non-face subset.

Figures 10 (a) and (b) show the CCE performance over



(a) over face images.

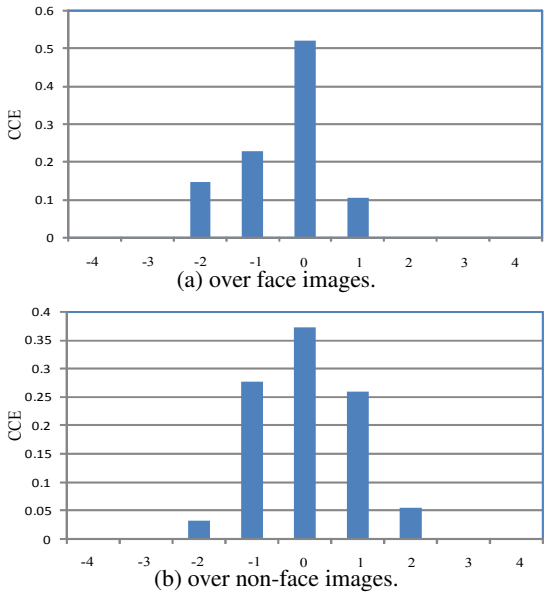
(b) over non-face images.

Fig. 9. Classifying aesthetic category with manual face annotation.

face (using global-Face) and non-face subsets (using global-Technical IVI), respectively, with automatic face detection. From the result, for the face subset, over 50% images are classified into the right categories, and for the non-face subset, over 35% images are correctly classified. For both face subset and non-face subset, over 85% images have no or small class misplacement (only misclassified by at most 1-category difference, *e.g.*, from "very good" to "medium good" at most). Also, for the face subset, the classifier sometimes predicts some images to better categories than they should be, *e.g.*, from "medium bad" to "medium good", resulting in about 15% $CCE(-2)$ misplacement. The classifier does not have such misclassification at all toward the other direction, *i.e.*, $CCE(i)=0$ for $i=2,3,4$. In general, for both face and non-face images, the classifiers do not make too severe mistakes (*e.g.*, predicting "very bad" to "medium good" or "very good"), and $CCE(i)=0$ for $i\leq -3$ or $i\geq 3$.

5. CONCLUSION

We study automatic assessment of the aesthetic value in consumer photographic images. A set of visual features related to characteristics of image quality and aesthetic values are used for aesthetic estimation in two tasks. For fine-granularity aesthetic score prediction, a novel Diff-RankBoost algorithm is proposed to predict images' aesthetic values ranging from 0 to 100. For coarse-granularity aesthetic category classification, a multi-category classifier is developed to study face and non-face image subsets separately. Experiments over a real consumer photographic image collection demonstrate promising results of such automatic aesthetic assessment.



(a) over face images.

(b) over non-face images.

Fig. 10. CCE performance with automatic face detection.

6. REFERENCES

[1] C. Callison-Burch *et al.*, "(Meta-) evaluation of machine translation," *ACL Workshop on Statistical Machine Translation*, pp.136-158, 2007.

[2] C.D. Cerosaletti and A.C. Loui, "Measuring the perceived aesthetic quality of photographic images," *IEEE QOMEX*, 2009.

[3] R. Datta *et al.*, "Studying aesthetics in photographic images using a computational approach," *Lecture Notes In Computer Science*, 3953:288-301, Springer, 2006.

[4] H. Drucker *et al.*, "Support vector regression machines," *NIPS*, pp.155-161, 1996.

[5] E. Fedorovskaya, C. Neustaedter, and W. Hao, "Image harmony for consumer images," *IEEE ICIP*, pp.121-124, 2008.

[6] Y. Freund *et al.*, "An efficient boosting algorithm for combining preferences," *Journal of Machine Learning Research*, 4:933-969, 2003.

[7] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," *IEEE CVPR*, 2006.

[8] A.C. Loui *et al.*, "Multidimensional image value assessment and rating for automated albuming and retrieval," *IEEE ICIP*, pp.97-100, 2008.

[9] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE TPAMI*, 19(7):696-710, 1997.

[10] A. Savakis *et al.*, "Evaluation of image appeal in consumer photography," *SPIE Human Vision & Electronic Imaging*, pp.111-120, 2000.

[11] V. Vapnik. *Statistical learning theory*. Wiley-Interscience, New York, 1998.