Aesthetic Image Classification Based on Multiple Kernel Learning

Ningning Liu[1], Xin Jin[2(✉)], Hui Lin[1], and De Zhang[3]

[1] School of Information Technology and Management,
University of International Business and Economics,
Beijing 100029, People's Republic of China
ningning.liu@uibe.edu.cn, linhuivicky@foxmail.com
[2] Department of Computer Science and Technology,
Beijing Electronic Science and Technology Institute,
Beijing 100070, People's Republic of China
jinxinbesti@foxmail.com
[3] Department of Automation,
Beijing University of Civil Engineering and Architecture, Beijing 100044, China
zhangde@bucea.edu.cn

Abstract. Aesthetic image classification aims at predicting the aesthetic quality of photos automatically, *i.e.* whether the photo elicits a high or low level of affection in a majority of people. To solve the problem, one challenge is to build features specific to image aesthetic perceptions, and another one is to build effective learning models to bridge the "semantic gap" between the emotion related concepts and the extracted visual features. In this paper, we present an approach for aesthetic image classification based on Multiple Kernel Learning (MKL) method, which seeks for maximizing the classification performance without explicit feature selection steps. The experiments are conducted on a large diverse database built from online photo sharing website, and the results demonstrated the advantages of MKL in terms of feature selection, classification performance, and interpretation, for the aesthetic image classification task.

Keywords: Aesthetic quality · Image classification · Multiple kernel learning

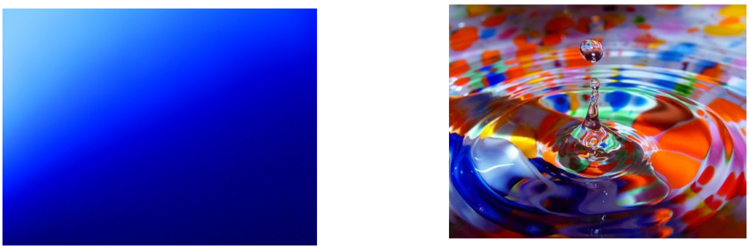
1 Introduction

Aesthetics is a sub discipline of philosophy and axiology dealing with the nature of beauty, art, and taste. The assessment or prediction of aesthetic value in images is considered to be of subjectivity and universality. The subjective feature suggests that the judgment relies on individual personal feelings, and there is no single agreement on what it exactly belongs to. In contrast, the universality indicates that certain features in photographic images are believed to please humans more than others. In conclusion, though the evaluation of beauty and other aesthetic qualities of photographs is highly subjective, still they have certain stability and generality across different people and cultures as a universal validity to classify images in terms of aesthetic quality [2]. Figure 1 shows two photos from an online website, and according to the ratings by web users, it is confirmed that the photograph (b) can inspire higher aesthetic feelings than the left one (a) for most people.

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 229–236, 2015.
DOI: 10.1007/978-3-662-48570-5_22

There could be many applications making use of an algorithm for photo quality assessment. For example, a search engine can merge a photo aesthetic factor into its ranking stage to get most relevant and better looking photos. An advertiser can make a choice referring to the most beautiful photos selected by the aesthetic quality assessment tools. Photo management solutions, like Picasa and iPhoto, can analyze the quality of one’s holiday snapshots and automatically present the best ones.

Research in the field of aesthetic image classification focuses on designing representation from various aspects, e.g., color, composition, lighting, and subjects. Recently, the impressive work made by R. Datta [2], Y. Ke [3] and M.Nishiyama et al. [4] have made a progress to this important issue. R. Datta [2] proposed 56 features based on the 'rules of thumb in photography'. Classification and linear regression on a community-based database showed that there is a significant correlation between various visual properties of photographs and their aesthetics ratings. Y. Ke [3] firstly proposed high level features based on a group of principles, including simplicity, realism and basic photographic technique, and the test provided a classification rate of 72% on a database. M. Nishiyama assess the aesthetic quality of a photo based on color harmony feature, namely ‘bags-of-color-patterns, and their results show that the performance of aesthetic image classification is improved by combining our color harmony feature with blur, edges, and saliency features. Meanwhile, there also other people [5, 6] simply employed the traditional low-level color, shape and texture features. Above works have designed various visual representations to characterize beauty in the form of photo art, but without considering the classifier or combination at all. For example, the authors in [2] use 5 cross-validation SVM accuracy score to rank and then select the top 15 descriptive features from the 56 proposed feature set, which requires explicit cross-validation steps for selecting features while optimizing the classifier parameters, and thus suffers from heavy computational complexities.



(a) (b)

Fig. 1. Example photos (a) and (b) received an average aesthetic rating of 2.4 from 222 votes and of 8.6 from 137 votes from a photo sharing website [6] respectively.

In this paper, we study the aesthetic image classification by applying multiple kernel framework, which can learns the feature representation weights and corresponding classifier in an intelligent way simultaneously. The main contributions of this paper included: (1) we investigate and implement visual features related to aesthetics, and also propose mid-level features to describe the dynamism and harmony in a photo; (2) we build a MKL scheme to perform aesthetic image classification, and received a good performance compared to the state-of-the-arts.

The rest of this paper is organized as follows: Section 2 introduces the image features used in this paper. Section 3 introduces our MKL framework for the aesthetic image classification. In Section 4, the experimental setup and results are reported. Finally, the conclusion and future work are presented in Section 5.

2 Image Features for Aesthetic Classification

Image feature extraction is a key issue for concept recognition in images. Features should be designed to carry sufficient information to be able to recognize the different concepts. In this paper, we complement low-level visual features based on color, texture and shape with higher level features such as color harmony and dynamism. Moreover, we make use of features based on aspects of a photograph appealing from a population and statistical standpoint [2], as well as representations based on perceptual factors that distinguish between professional photos and snapshots [3], and the aesthetic features based on color harmony [4]. The list of the features is given in Table 1.

2.1 Color, Texture and Shape

Studies have shown that the HSV (Hue, Saturation, and Value) color space is more related to human color perception than others such as traditional RGB. Moreover, different colors have different emotional meanings. Indeed, red is associated with happiness, dynamism and power whereas its opposite color, green, is associated with calmness and relaxation [7]. In this paper, different methods based on HSV color space are employed to describe color contents in images such as moments of color, color histograms.

The spatial gray-level difference statistics, known as co-occurrence matrix, can describe the brightness relationship of pixels within neighborhoods, and the local binary pattern (LBP) descriptor is a powerful feature for image texture classification. In this paper, these texture features are employed to contribute to aesthetic quality assessment.

Studies on artistic paintings have brought to the fore semantic meanings of shape and lines, and it is believed that shapes in a picture also influence the degree of aesthetic beauty perceived by humans [7]. Therefore, we make use of the Hough transform to build a histogram of line orientations in 12 different orientations.

2.2 Mid-Level

According to Itten's color theory [7], color combinations can produce effects such as harmony, non-harmony, calmness and excitation. Indeed, visual harmony can be obtained by combining hues and saturations so that an effect of stability on human eye can be produced. Itten has proposed to organize colors into a chromatic sphere where contrasting colors have opposite coordinates according to the center of the sphere. In case of harmony, color positions on Itten sphere are connected thanks to regular polygons. Therefore, by projecting the dominant image colors into the sphere and by comparing the distance between the polygon center and the sphere center, a value characterizing the image harmony can be obtained. At last, we extract the harmony features in 11 parts by dividing the image in (1, 2x2, 1x3, 3x1) sunblock's and concatenate them into one feature vector, by this way, it can include the spatial information.

Table 1. Summary of the features in this work.

Category	*Feature name*	#	*Short Description*
Color	Color moments	144	Three central moments (Mean, Standard deviation and Skewness) on HSV channels.
	Color histogram	64	$4^3 = 64$ bin histogram is created based on each HSV channel.
Texture	Grey level Co-occurrence matrix	16	GLCM, described by *Haralick* (1973), defined over an image to be the distribution of co-occurring values at a given offset.
	Local binary pattern(LBP)	256	A compact multi-scale texture descriptor analysis of textures with multiple scales by combining neighborhoods with different sizes.
Shape	Histogram of line orientations	12	12 different orientations by using Hough transform
Mid-level	Harmony	11	Try to describe color harmony of images based on Itten's color theory [7].
	Dynamism	11	The ratio of oblique lines against horizontal and vertical ones. Indeed, oblique lines communicate dynamism and action whereas horizontal or vertical lines rather communicate calmness and relaxation.
Others	Y. Ke	5	Features by Y. Ke [3] were chosen to measure criteria including: spatial distribution of edges, color distribution, hue count, blur, contrast and brightness.
	R.Datta	44	Features by R. Datta [2] including: exposure of light and colorfulness, saturation and hue, the rule of thirds, familiarity measure, wavelet-based texture, size and aspect ratio, region composition, low depth of field indicators, shape convexity. Note that we implement most of the features (44 of 56) except those (some from familiarity measure and Region composition) that are related to IRM (integrated region matching) technique [4].
	M. Nishiyama	200	Features by M. Nishiyama [4] namely "bags-of-color-patterns" based on the photo's color harmony the sum of color harmony scores computed from the local regions of a photograph is closely related to its aesthetic quality.

Lines also carry important semantic information in images: oblique lines communicate dynamism and action whereas horizontal or vertical lines rather communicate calmness and relaxation. To characterize dynamism in images, the ratio is computed between the numbers of oblique lines with respect to the total number of lines in an image. At last the dynamism features are obtained by extracting in 11 parts just as the harmony feature.

3 MKL for Image Aesthetic Classification

MKL refers to set methods that learn an optimal linear or non-linear combination of a predefined set of kernels. The reasons we build our image aesthetic classification based on MKL include: a) the ability to select an optimal kernel and parameters from a larger set of kernels, without an explicit feature selection step and b) combining data from different types of feature (e.g. color and texture) that have different notions of similarity and thus require different kernels. Moreover, instead of creating a new kernel, multiple kernel algorithms can be used to combine kernels which are already established for each individual features. All of these can improve the classification performance and makes the interpretation of the results straightforward. MKL has earlier been applied for visual object classification in [9], and we are the first to introduce it into image aesthetic classification. Our experimental results demonstrate the advantages of the MKL framework in image aesthetic classification.

According to the works [10, 12], we employ the Lasso MKL as our kernel learning method for it's simple and efficient. The algorithm formulates an alternating optimization method and updates the kernel weights η_m as follows:

$$\eta_m = \frac{\|\omega_m\|^2}{\sum_{h=1}^{P}\|\omega_h\|^2} \tag{1}$$

where $\|\omega_m\|^2 = \eta_m^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j \mathrm{K}_m(x_i^m x_j^m)$ is from the duality conditions. K_m denotes the kernel function calculated on the mth feature representation. P is the number of kernels or feature representations (P = 10 in our case), and $\sum_{m=1}^{P}\eta_m = 1$.

After updating the kernel weights in equation (1), the algorithm then solves a classical SVM problem by maximizing SVM dual formulation with the combined kernel $\mathrm{K} = \sum_{m=1}^{P}\eta_m \mathrm{K}_m$ as follows:

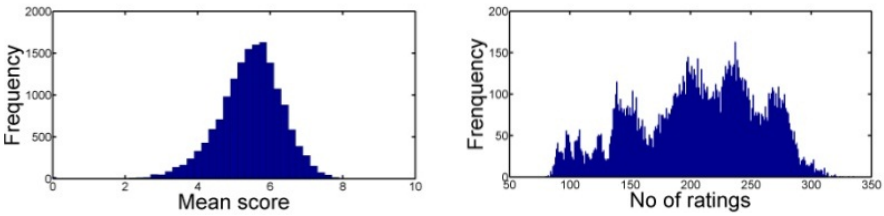
$$W(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \mathrm{K}(x_i, x_j) \tag{2}$$

subject to the constraints: $0 \le \alpha_i \le C$ for all $\mathrm{i} = 1, \dots, \mathrm{N}$ and $\sum_{i=1}^{N}\alpha_i y_i = 0$, where C is the regularization parameter and y_i is the label (±1) of training sample x_i. The two steps alternate until convergence.

4 Experiments and Results

4.1 Database

Previously, due to copyright restrictions, there is few public available database for photo aesthetic quality analysis. An exception is the preliminary work in [2] where photos have been collected from three Web-based sources [5, 6], in which photos have been rated by users of its community. Unfortunately, becasue photographs have been removed, it is hard to collect the same dataset as R. Datta [2] (about 15% has changed). Therefore, we have chosen to build a large and diverse training and testing database based on the Web source DPChallenge.com [6], which was created in January 2002 by Drew Ungvarsky and Langdon Oliver. To date, 180,255 users have submitted 318,599 photographs to 2086 challenges. Thus, we have collected a total of 60000 photographs by random crawling. Each photo is rated by at least 115 users with a mean average of 185 users, and the mean scores of all images are 5.6 with a std. dev. of 0.72. Figure 2 shows the distribution of average score and number of ratings. In order to reduce noise in the experiments, the top 10% and bottom 10% mean score of the photos were chosen and assigned as high and low aesthetic quality photo set respectively. From each set, half of the photos (3000) were used for training and the other half for testing. Some of the photos, especially the high quality ones, contain borders which we removed using a simple color counting algorithm in order to reduce bias in our results.



(a) (b)

Fig. 2. The distribution of mean score (a) and number of ratings (b).

4.2 Results

Experimental Setup. Our experiments are conducted as follows: Firstly, for each set, half of the photos (3000) were used for training and the other half for testing. To obtain the ground truth labels for the SVM classifier, we adopt the top 10% photos as the positive class, whereas those with bottom 10% are treated as the negative class. We conduct experiments in order to (1) compare visual features that show correlation with community-based aesthetics scores. For this, SVM is run 20 times per feature, and using a 5-fold cross-validation; (2) build a classification model based on MKL such that there is no need to select features and generalization performance is near optimal. For the MKL parameters, we set the regularization parameter C as $C = 1$, the kernel width $s = 2\sqrt{D}, D$ is the feature dimension size and the alternating iterations for inference as 20 times.

Results. Figure 3 shows us the accuracy performance of different features. We see that the features from R. Datta [2] receive 65% accuracy as the first place among the 10 features described in Section 2. But considering the size of feature vector, the feature from Y.Ke [3] with 5 dimensions belongs to the most effective one. The color and texture-based features achieve better results compared to the shape ones (dynamism and line histogram). SVM_all simply concatenates all the 10 features to a single feature, and receives a result around 70% better than other single features. This confirms our belief that by employing fusion method, we can improve the accuracy of aesthetic classification as it can provide complementary information to represent photo aesthetics.

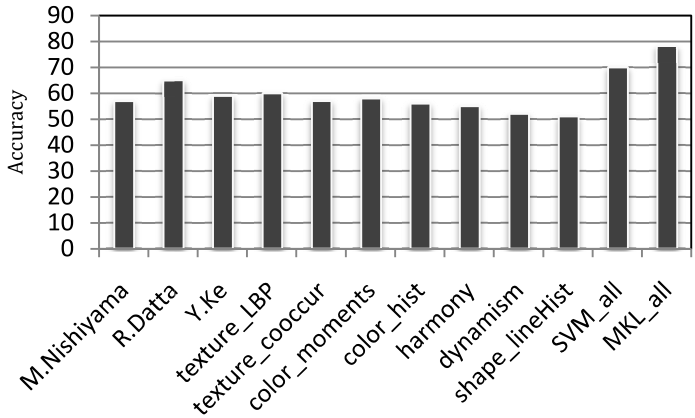


Fig. 3. The performance of different features.

Table 1 shows the comparison with other works. We can see that our method based on MKL scheme received the best performances. One should be noted that our database and Y. Ke's [3] are different, but are collected from the same web source and with the same training setting. Considering the nature of this problem, these classification results are indeed promising.

Table 2. The comparison with other works.

	Data size	*Selection / combine method*	*Performance*
R.Datta[2]	3581	filer and wrapper	70.12%
Y. Ke[3]	≈ 6000	naïve Bayers classifier	76%
Our method	**6000**	**MKL**	**78.3%**

Note: The data source [2] is from Photo.net [5].

5 Conclusion and Future Work

In this paper, we have presented an approach for aesthetic image classification based on MKL, which can make use of different feature representations simultaneously such that it jointly learns the feature weights and the corresponding classifier, by seeks for maximizing the classification performance without explicit feature selection steps. The experiments are conducted on a large diverse database built from online photo sharing website, and the results demonstrated the advantages of MKL in terms of feature selection, classification performance, and interpretation, for the aesthetic image classification task.

In future works, we believe that following effort can further enhance the performance: (1) proposing higher level visual features by combing visual saliency information, which indicated the region of interesting (ROI) in the image, (2) investigating the photograph metadata such as exposure time, aperture and ISO, and (3) introducing effective combination or regression techniques such as, the evidence theory and sparse logistic regression methods.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China (No. 0803603) (No. 61402021), Science and Technology Program of the State Archives Administration (2015-B-10), Beijing Nature Science Foundation (No. 4144070) (No. 4144072) and supported in part by the Fundamental Research Funds for the Central University in UIBE (14QD21).

References

1. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. IEEE Trans. on PAMI **23**(9), 947–963 (2001)
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
3. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: Proceedings of CVPR (2006)
4. Nishiyama, M., Okabe, T., Sato, I., Sato, Y.: Aesthetic quality classification of photographs based on color harmony. In: CVPR (2012)
5. Photo.net. http://photo.net
6. DPChallenge. http://www.dpchallenge.com
7. Columbo, C., Del Bimbo, A., Pala, P.: Semantics in visual information retrieval. IEEE Multimedia **6**(3), 38–53 (1999)
8. Datta, R., Li, J., Wang, J.: Algorithmic inferencing of aesthetics and emotion in natural images: an exposition. In: Proceedings of ICIP (2008)
9. Bucak, S.S., Jin, R., Jain, A.K.: Multiple Kernel Learning for Visual Object Recognition: A Review. T-PAMI (2013)
10. Zhang, H., Yang, Z., Gönen, M., Koskela, M., Laaksonen, J., Honkela, T., Oja, E.: Affective abstract image classification and retrieval using multiple kernel learning. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013, Part III. LNCS, vol. 8228, pp. 166–175. Springer, Heidelberg (2013)
11. Yu, et al.: L2-norm multiple kernel learning and its application to biomedical data fusion. BMC Bioinformatics **11**, 309 (2010)
12. Bach, F.: Consistency of the group lasso and multiple kernel learning. Journal of Machine Learning Research **9**, 1179–1225 (2008)