ICIC Express Letters
Part B: Applications
Volume **7**, Number **1**, January **2016**
 ISSN 2185-2766
pp. **57–62**

AESTHETIC QUALITY IMAGES CLASSIFICATION BY USING MULTIPLE KERNEL LEARNING APPROACH

Ningning Liu[1], Lei Qin[1], Li Zhang[1], Hui Lin[1] and De Zhang[2]

[1]School of Information Technology and Management
University of International Business and Economics
No. 10, Huixin East Road, Chaoyang District, Beijing 100029, P. R. China
ningning.liu@uibe.edu.cn

[2]Department of Computer Science and Technology
Beijing Electronic Science and Technology Institute
No. 1, ZhanLanGuan Road, Xicheng District, Beijing 100070, P. R. China

Received June 2015; accepted September 2015

Abstract. *Aesthetic quality image classification seeks to automatically classify the aesthetic quality of photos, i.e., whether the photo elicits a high or low level of affection in a majority of people. To address the problem, there are obviously two challenges: one is to build features specific to image aesthetic perceptions, and the other one is to build effective learning approaches to bridge the "semantic gap" between the emotion related concepts and the extracted visual features. In this paper, we propose an approach for aesthetic quality image classification based on Multiple Kernel Learning (MKL) method, which seeks for maximizing the classification performance without explicit feature selection steps. We carried out the experiments on a large diverse database built from online photo sharing website, and the results demonstrate the advantages of MKL in terms of feature selection, classification performance, and interpretation for the aesthetic quality 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. Therefore, the evaluation of beauty and other aesthetic qualities of photographs is highly subjective. However, 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. In practice, 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 photos.

Recent research in this field focuses on designing representation from various aspects, e.g., color, composition, lighting, and subjects. R. Datta et al. [2] proposed 56 features based on the 'rules of thumb in photography'. Y. Ke et al. [3] firstly proposed high level features based on a group of principles, including simplicity, realism and basic photographic technique. M. Nishiyama et al. [4] assessed the aesthetic quality of a photo based on color harmony feature, namely 'bags-of-color-patterns. Above works have designed various visual representations to characterize beauty attribute in the photo art, but they do not take account of the classifier and combination for the performance of classification. For example, the authors in [2] use 5 cross-validation support vector machines (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

(a) (b)

FIGURE 1. Example photos (a) and (b) received an average aesthetic rating of 3.5 from 159 votes and of 6.4 from 195 votes from a photo sharing website [6] respectively.

while optimizing the classifier parameters, and thus suffers from heavy computational complexities.

In this paper, we study the aesthetic quality image classification by applying MKL framework, which can learn the feature representation weights and corresponding classifier in an intelligent way simultaneously. The main contributions of this paper included: 1) we investigate visual features related to aesthetics, and also propose mid-level features to describe the dynamism and harmony in a photo; 2) we build an MKL scheme to perform aesthetic image classification, and achieve 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. Section 3 introduces the MKL framework for the image aesthetic classification. In Section 4, the experimental setup and results are reported. Finally, the conclusions and future work are presented in Section 5.

2. **Image Features for Aesthetic Classification.** In this paper, we implement low-level visual features such as color, texture, and shape, and mid-level features such as color

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), is 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 [7,13].
	Dynamism	11	Ratios of oblique lines against horizontal and vertical ones in 11 sub-blocks
Others	Y. Ke et al.	5	Features by Y. Ke et al. [3] were chosen to measure criteria including: spatial distribution of edges, color distribution, hue count, etc.
	R. Datta et al.	44	Most of the features (44 of 56) were extracted except those (some from familiarity measure and region composition) [2].
	M. Nishiyama et al.	200	Features by M. Nishiyama et al. [4] were computed from the local regions of a photograph related to its aesthetic quality.

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.** As different colors have different emotional meanings, and HSV (Hue, Saturation, and Value) color space has closer relationship with human color perception, compared with traditional RGB space [7], we employ different methods based on HSV color space to describe color contents in images such as moments of color, and 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 [7,13], colors can be organized into a chromatic sphere where contrasting colors have opposite coordinates according to the center of the sphere. To compute harmony, color positions on Itten sphere are first connected as regular polygons. Then, 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, 2×2, 1×3, 3×1) sub-block's and concatenate them into one feature vector, which include the spatial information. Meanwhile, lines also carry important semantic information in images: oblique lines communicate dynamism and action whereas horizontal or vertical lines rather convey calmness and relaxation [7]. To characterize dynamism in images, we compute a ratio 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 sub-blocks 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 make 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 work [10,12], we employ the Lasso MKL as our kernel learning method for it is simple and efficient. The algorithm formulates an alternating optimization method and updates the kernel weights η_m as follows:

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

where $\|\omega_m\|^2 = \eta_m^2 \Sigma_{i=1}^N \Sigma_{j=1}^N \alpha_i \alpha_j y_i y_j K_m \left(x_i^m x_j^m\right)$ 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 $\Sigma_{m=1}^P \eta_m = 1$.

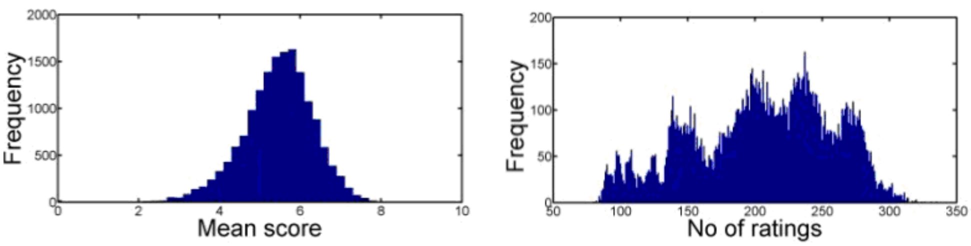
After updating the kernel weights in Equation (1), the algorithm then solves a classical SVM problem by maximizing SVM dual formulation with the combined kernel $K = \Sigma_{m=1}^P \eta_m K_m$ as follows:

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

subject to the constraints: $0 \le \alpha_i \le C$ for all $i = 1, \ldots, N$ and $\Sigma_{i=1}^N \alpha_i y_i = 0$, 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. To evaluate our approach, we build a large and diverse database based on the Web source DPChallenge.com [6], which was created in January 2002 by Drew Ungvarsky and Langdon Oliver. A total of 60000 photographs were collected 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.



(a) (b)

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

4.2. Experimental setup. In order to obtain the ground truth labels for the classifier, and to reduce noise in the experiments, we choose the top 10% and bottom 10% mean score of the photos, which are then assigned as high and low aesthetic quality photo set respectively. For each set, half of the photos (3000) were used for training and the other half for testing. In order to reduce bias, a border-remove preprocess is employed for some images just as Y. Ke et al.'s work [3]. Experiments were conducted to: 1) evaluate different visual features based on 5-fold cross-validation using an SVM classifier; 2) combine various features based on the MKL approach, setting the regularization parameter C as $C = 1$, the kernel width s as $s = 2\sqrt{D}$, where D is the feature dimension size; 3) compare with different combining methods including early fusion, majority voting, and mean score fusion.

4.3. Results. Figure 3 shows the average classification accuracy for different features described in Section 2. We can clearly see that, the features from R. Datta et al. [2] received 65% accuracy and ranked first, followed by texture LBP, color moments and Y. Ke et al.'s work etc. Considering the feature dimension, Y. Ke et al.'s work [3] is among the most effective one. The color and texture-based features achieved better classification performance compared to the shape ones (dynamism and line histogram). This also confirms

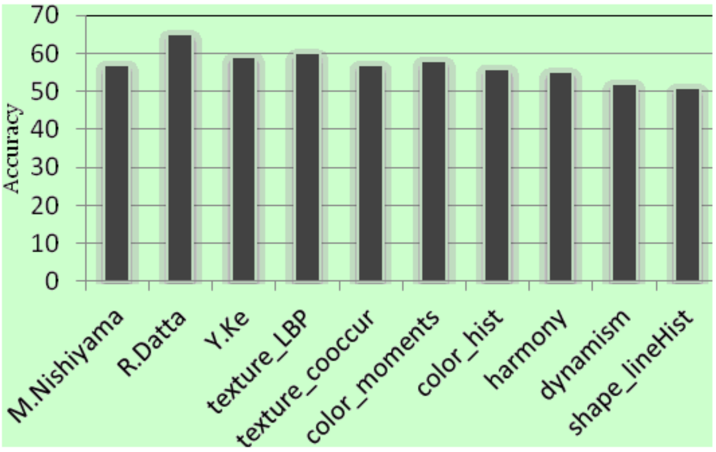


Figure 3. The performance of different features

Table 2. The comparison with other work

	classifier/combining method	*Performance*
SVM_MKL	MKL	**78.3%**
SVM_all	early fusion	**70%**
Majority voting	SVM	**71.5%**
Mean score	SVM	**73.2%**
R. Datta et al. [2]	filer and wrapper	70.12%
Y. Ke et al. [3]	naïve Bayers classifier	76%

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

most of the studies (e.g., [3,7,10,14]) that colors of an image are the most informative features for effect detection.

Table 2 shows the results of our approach based on MKL and the comparison with other fusion methods. It is clear that our method (SVM_MKL) based on MKL scheme received the best performances, followed by mean score method, and majority voting, and early fusion SVM_all, which just concatenates all the 10 features as a single input. This confirms our belief that by fusing with right features, we can improve the accuracy of aesthetic classification as they have provided complementary information to represent photo aesthetics. One should be noted that our database and Y. Ke et al.'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.

5. **Conclusions and Future Work.** In this paper, we have presented an approach for image aesthetic 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 work, we believe that following effort can further enhance the performance: 1) proposing higher level visual features by combining visual saliency information, which indicated the region of interesting (ROI) in the image; 2) introducing effective combination or regression techniques such as evidence theory and sparse logistic regression methods.

Acknowledgment. This work is partially supported by the Fundamental Research Funds for the Central University in UIBE (14QD21), Beijing Natural Science Foundation (No. 414 4070) (No. 4144072) and National Social Science Foundation (13BTQ027). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Z. Wang, J. Li and G. Wiederhold, SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, *IEEE Trans. PAMI*, vol.23, no.9, pp.947-963, 2001.

[2] R. Datta, D. Joshi, J. Li and J. Wang, Studying aesthetics in photographic images using a computational approach, *Proc. of ECCV*, 2006.

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

[4] M. Nishiyama, T. Okabe, I. Sato and Y. Sato, Aesthetic quality classification of photographs based on color harmony, *CVPR*, 2012.

[5] *Photo.net*, http://photo.net, 2015.

[6] *DPChallenge*, http://www.dpchallenge.com, 2015.

[7] C. Columbo, A. D. Bimbo and P. Pala, Semantics in visual information retrieval, *IEEE Multimedia*, vol.6, no.3, pp.38-53, 1999.

[8] R. Datta, J. Li and J. Wang, Algorithmic inferencing of aesthetics and emotion in natural images: An exposition, *Proc. of ICIP*, 2008.

[9] S. S. Bucak, R. Jin and A. K. Jain, Multiple kernel learning for visual object recognition: A Review, *T-PAMI*, 2013.

[10] H. Zhang, Z. Yang, M. Gönen, M. Koskela, J. Laaksonen, T. Honkela and E. Oja, Affective abstract image classification and retrieval using multiple kernel learning, *ICONIP*, no.3, pp.166-175, 2013.

[11] S. Yu et al., L2-norm multiple kernel learning and its application to biomedical data fusion, *BMC Bioinformatics*, 2010.

[12] F. Bach, Consistency of the group lasso and multiple kernel learning, *Journal of Machine Learning Research*, vol.9, pp.1179-1225, 2008.

[13] J. Itten, *The Art of Color: The Subjective Experience and Objective Rationale of Color*, John Wiley & Sons, New York, 1973.

[14] K.-C. Peng, A. Sadovnik, A. Gallagher and T. Chen, A mixed bag of emotions: Model, predict, and transfer emotion distributions, *CVPR*, 2015.