A User Preference-Based Aesthetic Image Recommendation Method (Postprint)

Authors: Xu Yongbo, Su Shimei, Fan Longqing

Date: 2018-10-11T00:00:00+00:00

Abstract

With the rapid development and widespread adoption of Internet and multimedia photography technologies, the volume of image resources has expanded dramatically. Locating users' most preferred images quickly and effectively among vast image information resources has become a critical challenge in the field of image recommendation. To address this problem, we propose an aesthetic image recommendation method based on user preferences. This approach extracts deep features from images using deep convolutional neural networks and obtains an image ranking score through SVMRank. Simultaneously, it computes aesthetic features using manually annotated aesthetic attributes (such as color harmony, composition rules, sharpness, and simplicity) to derive an aesthetic score. Finally, weighted cross-validation is employed to generate recommendation results that satisfy users. Experimental results demonstrate that this algorithm is an effective aesthetic preference recommendation method.

Full Text

Preamble

User-Specific Method for Aesthetic Image Recommendation

Xu Yongbo, Su Shimei, Fan Longqing
(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: With the rapid development and popularization of Internet and multimedia camera technology, the volume of various image resources has expanded dramatically. How to quickly and effectively find users' favorite images among numerous image information resources has become an important issue in the field of image recommendation. To address this problem, this paper proposes a user-preference aesthetic image recommendation method. The method extracts deep features of images using deep convolutional neural networks, obtaining an image ranking score through SVMRank. Simultaneously, it calculates aesthetic features from manually labeled image aesthetic factors (such as color harmony, composition rules, clarity, and simplicity) to obtain an aesthetic score. Finally, weighted cross-validation yields a recommendation result that satisfies users. Experiments demonstrate that this algorithm is an effective aesthetic preference recommendation method.

Keywords: deep convolutional neural network; aesthetic rules; user preferences

0 Introduction

Traditional image retrieval systems primarily fall into two categories: text-based image retrieval [1,2] and content-based image retrieval [2,4]. Text-based image retrieval, the earliest popular approach, matches user-input keywords with textual descriptions of images in the database and returns images with high similarity. However, this method suffers from semantic gap and cold start problems. Content-based image retrieval, currently the most commonly used approach, retrieves images based on features such as color, shape, and texture [5,6,7]. The retrieval process involves three main steps: extracting low-level image features, designing feature fusion methods, and performing similarity matching to return results.

With the vigorous development of mobile Internet and social networks, people can access massive information resources (including digital images) online. Moreover, digital cameras and camera phones enable everyone to capture scenes anytime and anywhere, causing individual image collections to grow exponentially. As an important information carrier, images offer many advantages over text. For instance, images present information more intuitively and richly, accurately reflecting object shapes while also conveying color and texture information. How to quickly and effectively find users' favorite images among numerous image information resources has become a critical problem in image recommendation.

Recent advances in neural networks [12,16,18,27] have enabled researchers to extract deep image features to replace handcrafted features. Lu et al. [16] proposed a two-column deep convolutional neural network to simultaneously capture global and local features, overcoming the inability of traditional handcrafted features to accurately express semantics. Wang et al. [27] utilized deep learning networks to automatically complete feature learning from different perspectives of the same image, obtaining more comprehensive image aesthetic feature descriptions. While neural networks accelerate feature extraction and improve classification accuracy, the absence of aesthetic factors remains a limitation.

As image processing has advanced, researchers have derived a series of modern aesthetic rules from high-quality photographs. Applying simple aesthetic and photographic rules to image classification and recommendation can effectively enhance computational performance in distinguishing high-quality from low-quality images. Researchers have conducted extensive experiments ranging from low-level visual features to high-level compositional features. Tong et al. [11] proposed a set of low-level visual features to differentiate professional from amateur photography. Datta et al. [12] computed a 56-dimensional feature vector to understand images. For more professional photography settings, Luo et al. [13] proposed an image subject region extraction scheme. Mavridaki et al. [14] first proposed and used pattern aesthetic attributes, achieving good results by combining simplicity and composition attributes. Dhar et al. [15] proposed describable features based on content, composition, and lighting to predict the interestingness of input images.

To improve user query efficiency, researchers have incorporated personalized user information [10,17~20,26] into recommendation systems to suggest relevant images that may interest users, making results more targeted and better aligned with individual needs. Li et al. [20] proposed a novel recommendation system that provides image sets related to personal preferences and interests. Zhang et al. [10] added personalized tags to represent user preferences, offering a new effective approach to overcome the semantic gap and achieve personalized recommendation. However, image descriptive information or tag settings carry user subjectivity, which is detrimental to personalized recommendation.

This paper proposes a user-preference image recommendation model that efficiently addresses these issues. The method uses user-selected images to represent preferences and employs deep convolutional neural networks (DCNN) to extract image features, ensuring full capture of intrinsic image diversity. Simultaneously, it uses widely recognized computational methods for image aesthetic rules to calculate aesthetic features, ensuring user aesthetic characteristics are satisfied. The main contributions are twofold: (a) proposing a user-preference image recommendation method that combines user preferences with DCNN; and (b) integrating aesthetic computational rules with DCNN into the image recommendation algorithm.

1 Recommendation Algorithm

The proposed method is illustrated in Figure 1 [Figure 1: see original paper]. The approach consists of two main components: (a) using DCNN to extract deep, indescribable features from input images, then applying SVMrank for classification and ranking experiments to obtain a DCNN score for each image; and (b) calculating aesthetic features of input images through aesthetic experiments to obtain a set of aesthetic vectors, which are then fed into an SVM classifier to produce an aesthetic feature scoring model. By multiplying with corresponding weight values, these two scoring models are cross-combined to obtain the final recommendation result. The detected images that match user preferences are input again into the system, yielding a set of images sorted by descending DAS scores, which are then recommended to users.

1.1 DCNN Scoring Algorithm

The DCNN network employed in this paper was designed and proposed by Krizhevsky et al. [21]. First, the training image dataset undergoes preliminary cropping to unify each input image to 227×227 pixels, and the cropped images are fed into the DCNN network, which outputs a deep feature set for the image dataset. On the other hand, user-selected preference images are similarly processed and input into the DCNN network to obtain a deep feature set for user input images. Figure 2 [Figure 2: see original paper] shows the effect of neural network convolution on images. Finally, similarity retrieval is performed between the feature set of user input test images and the dataset feature set, with weighted fusion of scores obtained from different models for the same image, followed by normalization to obtain the DCNN score for user input test images. The image similarity retrieval method adopts the work of Tian et al. [22].

Assume $U = \{I_{u1}, I_{u2}, ..., I_{um}\}$ is the set of images selected by the user as preferences, where I_u represents a user-preferred image. Features are extracted from the entire training set to obtain a model Γ. For a given image I_u, its visual features are extracted to retrieve similar images from Γ. For retrieval purposes, Equation (3) is used to explore images adjacent to user preference images in the joint visual space.

Ψ is the adjacent visual space of user-preferred images. For each user preference image, a dynamic strategy is

The preference function model is shown in Equation (4):

$$D(\cdot) \cdot A(\cdot) \cdot DAS(\cdot) \cdot DAS(\cdot)$$

1.2 Aesthetic Scoring Algorithm

This paper employs widely recognized representative aesthetic rules in the aesthetic analysis phase, such as image color, simplicity, contrast, and composition [14].

1.2.1 Color To evaluate image color harmony, this paper uses adjacent colors (subordination color) (Figure 3a), cooperate colors (Figure 3b), and complementary colors (Figure 3c). Conventional color analysis methods are adopted: first converting color images from RGB format to HSV (hue, saturation, value) format, then analyzing and computing the image color distribution histogram, and finally converting the HSV format back to RGB format for display. After obtaining the image color distribution histogram, thresholds are set based on the proportion of each color in the distribution. If a single color occupies a high proportion, the image uses adjacent colors; if the image contains two colors with high proportions simultaneously, it can be determined as either cooperate colors or complementary colors. The color wheel shown in Figure 3 [Figure 3:

see original paper] can determine which category the image belongs to. Using the color wheel to measure image color composition can accurately determine whether the color matching is reasonable, providing convenience for users to select images with reasonable color coordination.

1.2.3 Clarity In image quality evaluation, clarity is one of the important factors for judging image quality. This paper adopts the method proposed by E. Mavridaki et al. [14] to detect image clarity, using Equation (1) to extract clarity features, obtaining a 635-dimensional clarity feature vector for the test image. Subsequently, similarity classification is performed with the clarity feature model of the entire dataset on SVM.

$$Q = \frac{1}{M \times N} \sum_{x} \sum_{y} \sqrt{(I(x,y) - I(x+1,y))^2 + (I(x,y) - I(x,y+1))^2}$$

where $I(x,y)$ corresponds to the grayscale value or brightness component at position (x,y), and $M \times N$ is the total number of pixels in the image.

1.2.4 Composition Reasonable composition can better express image content, making the subject prominent and the form novel and unique. Researchers have proposed many image composition principles, such as rule of thirds, golden ratio composition, symmetrical composition, and diagonal composition. This paper mainly employs three composition rules: "center composition,""rule of thirds composition," and "golden ratio composition." To detect the "center composition rule," this paper uses the objectness measure proposed by Alexe et al. [24] to detect the subject position in the input image, then calculates the Euclidean distance and angle between the subject position and the image center. For detecting "rule of thirds composition" and "golden ratio composition," similar calculation methods are used to estimate the Euclidean distance and the angle between the subject position and the image center. The Euclidean distance and angle calculations are shown in Equation (2):

$$d = \sqrt{(CenX - ObjCenX)^2 + (CenY - ObjCenY)^2}$$

where $CenX, CenY$ represent the center coordinates of the image, and $ObjCenX, ObjCenY$ represent the center coordinates of the subject region.

2 Experiments

2.1 Dataset

The experiments in this paper use the CUHKPQ public dataset collected and organized by Luo et al. [13]. The CUHKPQ dataset contains 17,673 images, each provided by amateur photographers to professional photography websites with manually labeled ground truth quality. The entire dataset is divided into seven categories: 'people,' 'plants,' 'night,' 'static,' 'architecture,' 'animals,' etc. Additionally, each image is rated (high or low quality) by ten independent viewers. Considering user subjective preferences, this paper assumes that comparisons within the same category are more detailed. Therefore, experiments are conducted on individual categories and the entire dataset to verify this assumption. During experiments, the dataset is randomly divided into two parts, with one part used as the training set and the other as the test set.

2.2 DCNN Experiments

As shown in Figure 2, this paper first requires users to randomly select m images they like, then uses DCNN to extract deep features from these images, using these m images to represent user preferences in images. Meanwhile, DCNN is used to extract features from the selected training dataset, and the following described image retrieval method is employed to obtain a user preference model.

Assume $U = \{I_{u1}, I_{u2}, ..., I_{um}\}$ is the set of user-selected preference images, where I_u represents a user-preferred image. Features are extracted from the entire training set to obtain a model Γ. For a given image I_u, its visual features are extracted to retrieve similar images from Γ. For retrieval purposes, Equation (3) is used to explore images adjacent to user preference images in the joint visual space.

$$S(I_i) = \{I_i, I \in \Gamma, I \in \Psi\}$$

where Ψ is the adjacent visual space of user-preferred images. For each user preference image, a dynamic strategy is adopted to select m retrieved images from Γ, and these images are concatenated into retrieval results $S(I_i)$. Once the entire retrieval result is obtained, it can be used to learn a user preference function. Finally, when a user inputs an image to the preference function, n recommended images can be obtained (where $n = 1, 2, 3, ...$; users can set the value of n themselves). The preference function model is shown in Figure 5 [Figure 5: see original paper].

2.3 Aesthetic Experiments

Inputting user-selected preference images m into the aesthetic rule calculation model yields corresponding binary classification features for aesthetic rules. Finally, using an SVM classifier, an aesthetic rule scoring model for the input m images is obtained. The algorithm flowchart is shown in Figure 6 [Figure 6: see original paper].

User satisfaction results are shown in Table 1 . Through data analysis of user satisfaction and actual user feedback during experiments, recommending 5 images each time (the specific quantity can be set by users) yields the highest user satisfaction without causing multiple selections or visual fatigue.

2.4 Comprehensive Experiments

Finally, the preference model generated by DCNN and the traditional aesthetic scoring model are weighted and cross-validated using Equations (5) and (6) to obtain the final user preference recommendation model.

$$DAS = \alpha \cdot T + \beta \cdot D$$

$$\alpha + \beta = 1$$

where DAS is the final score for recommended images, used to recommend images to users in descending order. α is the user weight for DCNN scores, and β is the user weight for aesthetic rule scores.

Table 2 shows the average precision (AP) for each category of images with different recommendation quantities. Users select their preferred images from each category as initial input, which serves as an expression of user preferences. After similarity retrieval as configured in this paper, a training data sequence is obtained. Images liked by users are labeled as 1, and disliked images as 0, to calculate the AP value for each category. Table 2 shows that when recommending 5 images to users, the AP values for animal (73.5%), people (80.2%), and architecture (81.7%) categories are the highest. When recommending 10 images, static (80.4%) achieves the highest value. When recommending 15 images, landscape (80.8%) is highest. When recommending 20 images, night scene (80.7%) has the highest AP value. To improve recommendation accuracy, when users select images from a single category, the preference recommendation model recommends the quantity of images with high AP values. To more intuitively observe differences in AP values across different recommendation quantities, the data from Table 2 is plotted as a line chart in Figure 7 [Figure 7: see original paper].

3 Experimental Results Analysis

In the experiments, 53 participants aged 19-56 were selected for image recommendation experiments, and their feedback data was collected. The definition of satisfaction degree is shown in Equation (7), and results from 5 random users are displayed in Figure 5 [Figure 5: see original paper].

$$\text{User Satisfaction Degree} = \frac{\text{Number of photos liked by user}}{\text{Number of photos recommended at one time}}$$

To verify the effectiveness of the proposed preference recommendation algorithm, MAP values were calculated and compared with current work, as shown in Table 3 . ZDE et al. [8] used Galois lattices and emergent semantic information to determine final recommendation lists by finding similarities between relevant users, achieving a MAP value of 67.23% when recommending 20 images. Liu et al. [9] used sparse topic modeling of image content and classical probabilistic matrix factorization to obtain useful image information, achieving MAP values of 62.32% when recommending 5 images, 56.56% when recommending 10 images, and 54.62% when recommending 20 images. Li et al. [20] used different methods to calculate a MAP value of 75.0% when recommending 5 images. The proposed preference recommendation algorithm achieves a MAP value of 76.9% when recommending 5 images, which is higher than [8,9,20]. The MAP value when recommending 10 images (73.3%) is also significantly higher than Liu et al. [9]' s result (56.56%). When recommending 20 images, the MAP value (74.8%) is substantially higher than ZDE [8]'s (67.23%) and Liu [9]'s (54.62%). Comparison with previous work Su [25] also shows improvement in MAP values. Comparing MAP values across different recommendation quantities reveals that the MAP value is highest (76.9%) when the recommendation quantity is 5, making it the optimal recommendation quantity.

4 Conclusion

This paper proposes a user-preference image recommendation algorithm using deep convolutional neural networks (DCNN), support vector machines (SVM), and traditional aesthetic rule calculation methods. The algorithm first uses DCNN to extract deep features from user input images, thereby learning user preferences in deep features. Second, it calculates and learns user aesthetic preferences for input images using aesthetic rules. Finally, the two are weighted and cross-combined to obtain the final user preference model. When users input an image into this preference model, satisfactory preference images can be recommended. Since different people have different levels of visual fatigue, users have different requirements for the number of recommended results. Therefore, this paper allows users to determine the number of recommended images. Experiments demonstrate that the proposed user-preference image recommendation algorithm can recommend satisfactory images to users.

References

[1] Barrilero M, Uribe S, Alduan M, et al. In-network content based image recommendation system for Content-aware Networks [C]// Proc of Computer Communications Workshops. 2011: 115-120.

[2] Asyan Hamidi, Abdul Ghiti Hamidi. A comparative study of text-based image retrieval and content-based image retrieval techniques [J]. Journal of Capital Normal University: Natural Science Edition, 2012, 33 (4): 6-9.

[3] Miyazawa Y, Yamamoto Y, Kawabe T. Context-Aware recommendation system using content based image re-trieval with dynamic context considered [C]// Proc of International Conference on Signal-Image Technology & Internet-Based Systems. Washington DC: IEEE Computer Society, 2013: 779-783.

[4] Zhou Mingquan, Yan Guohua, Wei Na. Content-based image retrieval technology [M]. Beijing: Tsinghua University Press, 2007.

[5] Li Youfeng. Research on image retrieval related algorithms based on color and texture features [D]. Chengdu: University of Electronic Science and Technology of China, 2009.

[6] Rodriguez L, Diago L, Hagiwara I. Color recommendation system combining design concepts with interactive customers preference modeling from context changes [C]// Proc of IEEE Congress on Evolutionary Computation. IEEE, 2010: 1-8.

[7] Bo Qirong, Peng Jinye A Novel Interactive Image Recommendation System [J]. Computer Engineering & Applications, 2011, 2: 248-251.

[8] Zomahoun D E, Yetongnon K. EMERGSEM: Emergent Semantic and Recommendation System for Image Retrieval [C]// Proc of the 10th International Conference on Signal-Image Technology and Internet-Based Systems. Washington DC: IEEE Computer Society, 2014: 256-263.

[9] Liu Lei. A sparse image recommendation model using content and user preference information [C]// Proc of IEEE//WIC//ACM International Conference on Web Intelligence. IEEE, 2017: 232-239.

[10] Zhang, Jing, Yang Ying, Tian Qi, et al. Personalized social image recommendation method based on user-image-tag model [J]. IEEE Trans on Multimedia, 2017, 19 (11): 2439-2449.

[11] Tong Hanghang, Li Mingjing, Zhang Hongjiang, et al. Classification of digital photos taken by photographers or home users [C]// Pacific Rim Conference on Advances in Multimedia Information Processing. Springer-Verlag, 2004: 198-205.

[12] Ritendra D, Dhiraj J, Li Jia, et al. Studying aesthetics in photographic images using a computational approach [C]// Proc of European Conference on Computer Vision. Springer-Verlag, 2006: 288-301.

[13] Luo Wei, Wang Xiaogang, and Tang Xiaoou. Content-based photo quality assessment [C]// Proc of International Conference on Computer Vision. IEEE, 2012: 2206-2213.

[14] Mavridaki E, Mezaris V. A comprehensive aesthetic quality assessment method for natural images using basic rules of photography [C]// Proc of IEEE International Conference on Image Processing. 2015: 887–891.

[15] Dhar S, Ordonez V, Berg T L. High level describable attributes for predicting aesthetics and interestingness [C]// Computer Vision and Pattern Recognition. 2011: 1657-1664.

[16] Lu Xin, Zhe Lin, Jin Hailin, et al. RAPID: Rating Pictorial Aesthetics using Deep Learning [J]. IEEE Trans on Multimedia, 2015, 17 (11): 2021-2034.

[17] Zeng Chun, Xing Chunxiao, Zhou Lizhu. Overview of personalized service technology [J]. Journal of Software, 2002, 13 (10): 1952-1961.

[18] Wang Yeqing, Li Yi, and Porikli Fatih. Finetuning Convolutional Neural Networks for visual aesthetics [C]// Proc of International Conference on Pattern Recognition. IEEE, 2017: 3554-3559.

[19] Liu Jianguo, Zhou Tao, Wang Binghong. Research Progress of Personalized Recommendation System [J]. Advances in Natural Science, 2009, 19 (1): 1-15.

[20] Li Yuncheng, Mei Tao, Cong Yang, et al. User-curated image collections: Modeling and recommendation [C]// Proc of IEEE International Conference on Big Data. IEEE, 2015: 591-600.

[21] Krizhevsky Alex, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097-1105.

[22] Tian Xinmei, Dong Zhe, Yang Kuiyuan, et al. Query Dependent Aesthetic Model With Deep Learning for Photo Quality Assessment [J]. IEEE Trans on Multimedia, 2015, 17 (11): 2035-2048.

[23] Mclaren K. XIII: The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula [J]. Coloration Technology, 2010, 92 (9): 338-341.

[24] Alexe B, Deselaers T, Ferrari V. Measuring the objectness of image windows [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2012, 34 (11): 2189-2202.

[25] Su Shimei, Wang Meng, Xu Yongbo. User-specific sorting method in photographic pictures [J//OL]. Application Research of Computers, 2019, 36 (2).

[26] Yang Kuiyuan, Hua Xiansheng, Wang Meng, et al. Tag tagging: towards more descriptive keywords of image content [J]. IEEE Trans on Multimedia, 2011, 13 (4): 662-673.

[27] Wang Weining, Wang Li, Zhao Mingquan, et al. Image aesthetic classification based on parallel deep convolutional neural network [J]. Journal of Automation, 2016, 42 (6): 904-914.

Note: Figure translations are in progress. See original paper for figures.