Neurocomputing 188 (2016) 50–62

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Computational aesthetics of photos quality assessment based on improved artificial neural network combined with an autoencoder technique

Yunlan Tan [a,b], Yimin Zhou [b], Guangyao Li [b], Anmin Huang [a,*]

[a] School of Electronics and Information Engineering, Jinggangshan University, Jiangxi, China
[b] College of Electronics and Information Engineering, Tongji University, Shanghai, China

ARTICLE INFO

Article history:
Received 27 December 2014
Received in revised form
4 April 2015
Accepted 5 April 2015
Available online 15 December 2015

Keywords:
Computational aesthetics
Quality assessment
Artificial neural network
Autoencoder
High aesthetics
Low aesthetics

ABSTRACT

Photograph aesthetical evaluation has been widely investigated in these decades. For fine-granularity aesthetic quality prediction, a novel aesthetics classifier based on improved artificial neural network combined with an Autoencoder technique is presented. First, we download large consumer photographic images from a well-known online photograph portal. Then, we extract 56 features normalized to 0–1 and train the networks with photographs of high and low ratings to test the quality of photos. Experimental results show that the accuracy of classification is above 86.67%, which is better than all state-of-the-art methods. Meanwhile, it is observed from experiments that the extracted features are consistent with the humans' visual perception systems.

1. Introduction

According to DPChallenge's statistics [1], photographers have uploaded over 587,000 professional photographs. This convenience brings us new industries like network sharing center and portals attracting people of the same interest, i.e. professional or amateur photographers around the world. Inevitably, works are compared and assessed by the insiders of the community. Thus it proposes a challenging problem to differentiate high quality photographs from low ones for several reasons. Firstly, visual data is very rich and ambiguous because aesthetic assessments are highly subjective. Secondly, when judging photographs, people are often confronted to personal tastes. Finally, if one might agree that low level degradations (e.g. out of focus image) are in general an indicator of poor quality, it is more difficult to find a consensus on higher level visual properties such as color harmonies, layout, lighting conditions, etc. With all these difficulties, one might even question the possibility to learn generic models encoding photographic preference. Therefore, computational aesthetics of photo quality assessment has been investigated through these decades [2].

Objective quality assessment refers to construction of algorithms that gauge the perceived quality of these human visual stimuli. Image features (e.g., low-level, high-level) are adopted by computer vision researchers for the purpose of image objective quality assessment [3] and many successful algorithms has been proposed for this purpose [4–8]. Sun et al. [9] try to design optimized visual features to mimic human perception on photo assessment. Luo et al. [4] propose a photo quality assessment method, which is first extracted the subject region from a photo, and then formulate a number of high level semantic features for photograph quality classification. Ke et al. [5] design a high level sematic features based the spatial distribution of edges, blur, and the histograms of low-level color properties such as brightness and hue. They test their features on a large and diverse dataset and the system is able to achieve a classification rate of 72%. Tong et al. [6] try to classify photos as professional ones or snapshots, but they use the Corel image database, which is too homogeneous to separate the two classes. In addition, they simply collect a large set of low level features from the image retrieval literature. Wang and Datta [7] are the first to realize the quantization of image features, including the brightness, color distribution, wavelet, region composition and depth of field, and apply SVM or linear regression to classify the high from the low quality photographs and achieved an accuracy of 70.12%. Luo et al. [4] produce better results than [5–7] with much less number of features, but it is still not good enough with a classification rate of 72% on a large dataset. The

* Corresponding author.
E-mail address: huanganmin@jgsu.edu.cn (A. Huang).

http://dx.doi.org/10.1016/j.neucom.2015.04.124
0925-2312/

main problem with existing methods is that they compute features from the whole image. This significantly limits the performance of the features since a good photo usually treats the foreground subject and the background very differently [4].

Professional photographers usually differentiate the subject of the photo from the background to highlight the topic of the photo. High quality photos generally satisfy three principles: a clear topic, gathering most attention on the subject, and removing objects that distract attention from the subject [10,11]. Photographers try to achieve this by skillfully manipulating the photo composition, lighting, and focus of the subject. Motivated by these principles, in this paper, we make the assumption that the salient regions of an image contain the subject and utilize the method of frequency-tuned salient region detection [12] to segment salient objects. Then we develop several highly effective quantitative metrics on subject. We address the limitation in existing work by employing a saliency approach to extract features associated with the subject.

Apart from image features extraction, optimization of image processing algorithms using perceptual measures of visual quality as objective functions remains relatively unexplored. Jiang et al. [13] propose a novel regression method, namely Diff-RankBoost based on RankBoost and support vector techniques. They predict coarse-granularity aesthetic categories over 450 real consumer photographic images. Wang et al. [14] perform a large scale studies on analyzing algorithm performance, namely the structural similarity index (SSIM). Marchesotti et al. [15] learn the linear SVMs with a hinge loss using the primal formulation and a Stochastic Gradient Descent (SGD) algorithm. They demonstrate a significant margin relying on the binary labels ("good" and "bad") provided by the authors. The ACQUINE [16] aesthetics value measurement system is a typical aesthetic evaluation and search engine developed by them. However, this method is impractical for large datasets due to the limit of SVM. Wu et al. [17] extend their SVM classification method to predict aesthetic adjectives rather than aesthetic scores. They introduce a probabilistic post-processing step that alleviate effects due to misleadingly labeled training data. Ke et al. [5] extend their work by adding visual features of images, but both methods remain the same drawbacks. Wong et al. [18] present a saliency-enhanced image classification method, but this method must be preprocessed to detect the saliency region. Moreover, all of the above researchers donot go further into the classification methods on the side of machine learning. Sidra [16] provide a classification method based on machine learning, but their dataset seem too small to produce a highly convincing result.

Although foregoing machine learning methods have been introduced to evaluate the aesthetic values of photographs by a great number of researchers, most of them are applied to a relatively small set of images and just attained a passable accuracy for classification. We formulate photo quality evaluation as a machine learning problem in which we map the characteristics of a human-rated photograph in terms of its underlying adherence to the rules of composition. Our method can be compared with the approach suggested in [7,18,19], where the authors apply a saliency map to estimate visual attention distribution in photographs. We complement the saliency information extracted from an image using a high-level semantic segmentation technique that can infer the geometric context of a scene [6,12,18]. With the help of the above methods, we extract aesthetic features that can be used to measure the deviation of a typical composition from ideal photographic rules of composition. These aesthetic features are subsequently used as input to learn the visual aesthetic model.

In this work, we present an image processing method related to features detection, training, classification and prediction by exploiting Datta's work [7]. In addition, we apply artificial neural network and Autoencoder [20], a deep learning method, which is firstly adopted for aesthetic value evaluation to Dattra's method to achieve a more accurate result. To this end, we do the following contributions in this paper: (1) An empirical study on extracting visual aesthetics on real-world images is performed, (2) A smooth mapping between user input visual attractiveness and high-level aesthetic features is brought, while those features are consistent with the humans visual perception systems, and finally (3) photographic quality assessment classifier under ANN+Autoencoder framework is constructed.

This paper is organized as follows. In Section 2, the related works are reviewed. In Section 3, the method of image feature extraction and the techniques of the improved artificial neural network and Autoencoder are illustrated. In Section 4, the experimental results are presented. In Section 5, several conclusive remarks are drawn.

2. The proposed method

We formulate photo quality evaluation as a machine learning problem in which we map the characteristics of a human-rated photograph in terms of its underlying adherence to the rules of composition. Unlike the above methods, we propose an esthetic value evaluation method which combine the common feature extraction method with the improved artificial neural network and Autoencoder. This is the first time to train Datta's dataset with an artificial neural network using Autoencoder to preprocess the input. The overview of the proposed method is illustrated as in Fig. 1. The whole process could be demonstrated as three phases and each takes a lot of work. The first phase is data collecting. We download tens of thousands of images from internet and exclude those evaluation records over 100 votes. The second phase is features extraction. We extract 56 features normalized to 0–1. The third phase is to build the Autoencoder, with which different amounts of encoded sizes is utilized. The last phase is training and classifying as much as possible combination of configurations. Our task is to find a smooth mapping function $R^{(56+L)} \to R^2$on dataset D, where $L=7, 14$, and 28. That is to say, the $56+L$-dim feature vector is mapped into 2-dim classification vector. The proposed method is configured to classify the photographs into high and low esthetic values. The experiments show that the proposed method achieves a considerably good result.

2.1. Features extraction

We address the limitation in existing work by employing a saliency approach to analyze and extract features associated with the subject. Like Wong et al. [18], we make the assumption that the salient regions of an image contain the subject and use the salient regions to represent the subject. Instead of applying the approach presented by Laurent Itti [21], we exploit an automatic method presented by Radhakrishna Achantay [12]. Altogether we extract 56 features of local and global features for every selected image. The feature set is well chosen but not absolute because our main goal is to build a strong classifier and regression model. The first selected features are referred as candidate ones for further refinery and are denoted as $F=\{f_i \mid 1 \le i \le 56\}$ that are described as follows.

Photography viewers and connoisseurs guide us to guess the certain aspects that are crucial in evaluating the esthetics measure. Those esthetics measures are believed to be embedded in numerical values of certain features. We treat each image crawled from the web as an individual and refine features from them. From the angle of viewers, we could imagine the color is not a decisive factor in evaluation of the esthetics measure since each color has its lovers, we cannot make some of them take advantageous of others. The chromatic properties of images are, as expected, what differentiate