Neurocomputing 143 (2014) 14–26

Contents lists available at ScienceDirect

Neurocomputing

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

Image esthetic assessment using both hand-crafting and semantic features

Lihua Guo [a], Yangchao Xiong [a], Qinghua Huang [a,*], Xuelong Li [b]

[a] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China
[b] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China

A R T I C L E I N F O

Article history:
Received 7 January 2014
Received in revised form
13 June 2014
Accepted 14 June 2014
Communicated by Tao Mei
Available online 24 June 2014

Keywords:
Image esthetic assessment
Clarity region detection
Hand-crafting features extraction
Semantic features extraction

A B S T R A C T

Automatically assessing the visual esthetics of images is of great interest in high-level vision research and has drawn much attention in recent years. Traditional methods heavily depend on the performance of subject region extraction. This paper proposes to use semantic features in the esthetic assessment system because they can implicitly represent the image topic and be helpful if the subject region extraction fails. Accordingly, a framework combining the hand-crafting features with semantic features is proposed to evaluate image esthetic quality. The experimental results show that the semantic features can improve the performance of image esthetic assessment.

1. Introduction

With the popularization of digital cameras and the rapid development of the Internet, the number of images that can be accessed is growing explosively. Automatic assessment of image quality that is consistent with human's perception has become more and more important with the increasing needs of professionals and home users. Image quality assessment (IQA) [19,20,24] score images using multi-kernel learning [26] and sparse coding [25], and focuses on the image esthetics. In the field of image esthetic assessment, there are various criteria (e.g. sharpness, composition, lighting balance, topic emphasis, and even the special photographic skills) for evaluating the image esthetics in photography. It is widely agreed that there are many rules of thumb regarding what generally makes an image appealing. For example, people would prefer a sharp and clear image to a blur image; balanced lighting with proper contrast is considered to be better than dim lighting; an image with a clear topic is deemed as more pleasing than the one with unnecessarily distracting background. These generally accepted rules make it possible to develop algorithms to automatically select images, which are more likely to be esthetically appealing.

1.1. Related work

Image esthetic quality assessment can be divided into two directions: subjective and objective esthetic assessment. The subjective method adopts human perception to assess image esthetic quality, whereas the objective method models a series of criteria to simulate human perception. Compared with the subjective image esthetic assessment, the objective method is simple, low cost, and easy to implement. But its main problem lies on the features to appropriately represent human perception.

In image classification, features are important, and even determine the final performance of classification system [11,12,21–23,29]. Recently, a lot of researchers proposed a number of features to solve the image esthetic assessment [1–6,9,14–16]. Tong et al. [1] used boosting to combine 846 global low-level features for the classification of professional and amateurish images. The features (e.g. color histograms, wavelets, and DCT moments) which are widely used in image retrieval applications were employed for IQA. Ke et al. [2] replaced the low-level features with high-level features (e.g. spatial distribution of edges, color distribution, etc.). Their method with a much smaller number of features outperforms than that using low-level features. Lu et al. [18] used the statistics of contourlet coefficients to indicate the variation of image quality. The adopted features [2,3,18] came from basic photographic technique.

In many cases human beings perceive subject areas from the background, and the subject areas can draw the most attention to human eyes. Therefore, some previous works [3,5,6,14] separated

* Corresponding author.
E-mail address: qhhuang@scut.edu.cn (Q. Huang).

http://dx.doi.org/10.1016/j.neucom.2014.06.029
0925-2312/

an image into subject and background regions, and extracted features from subject and background regions separately. Datta et al. [3] divided an image into 3×3 blocks and assumed the central block to be the subject area. However, such assumption is not valid in lots of images. In fact, it is generally known that photographers deliberately avoid putting subjects right at the center of image.

The saliency map and the clarity detector are two popular kinds of subject area extraction methods. A saliency detector is a natural choice to extract the region of interest. For example, Wong et al. [14] used the saliency map to define the subject areas. They assumed that the saliency map had higher brightness and contrast than other regions. After extracting the subject areas, their method [14] extracted a series of features based on the relationship between the salient and background regions. However, the saliency map often failed to correctly segment the subject region. For instance, if a certain part of the subject area has very high brightness and contrast, other parts will be ignored by the saliency detector. The clarity detector [6] uses local statistics of image pixels to tell the difference between clear and blurred regions, and extracted subject regions using a clarity region detection method and computed both the global and regional features. Unfortunately, the clarity-based subject region extraction still needs to improve [7,27].

Although recent techniques like clarity computation and saliency computation can separate the subject area from the entire image, hand-crafting features after subject detection cannot include all kinds of photographic techniques which are implicit contained in semantic information of the image. Marchesotti et al. [15] used generic image features to replace the hand-crafting features for assessing esthetic quality. Moreover, professional photographers may adopt various photographic techniques, and may have different esthetic criteria in mind when taking different types of images (e.g. landscape versus portrait). Tang et al. [5] proposed a method that segments an image and extracts visual features in different ways according to the categorization of image contents. They firstly divided the images into different categories based on the image content and extracted different

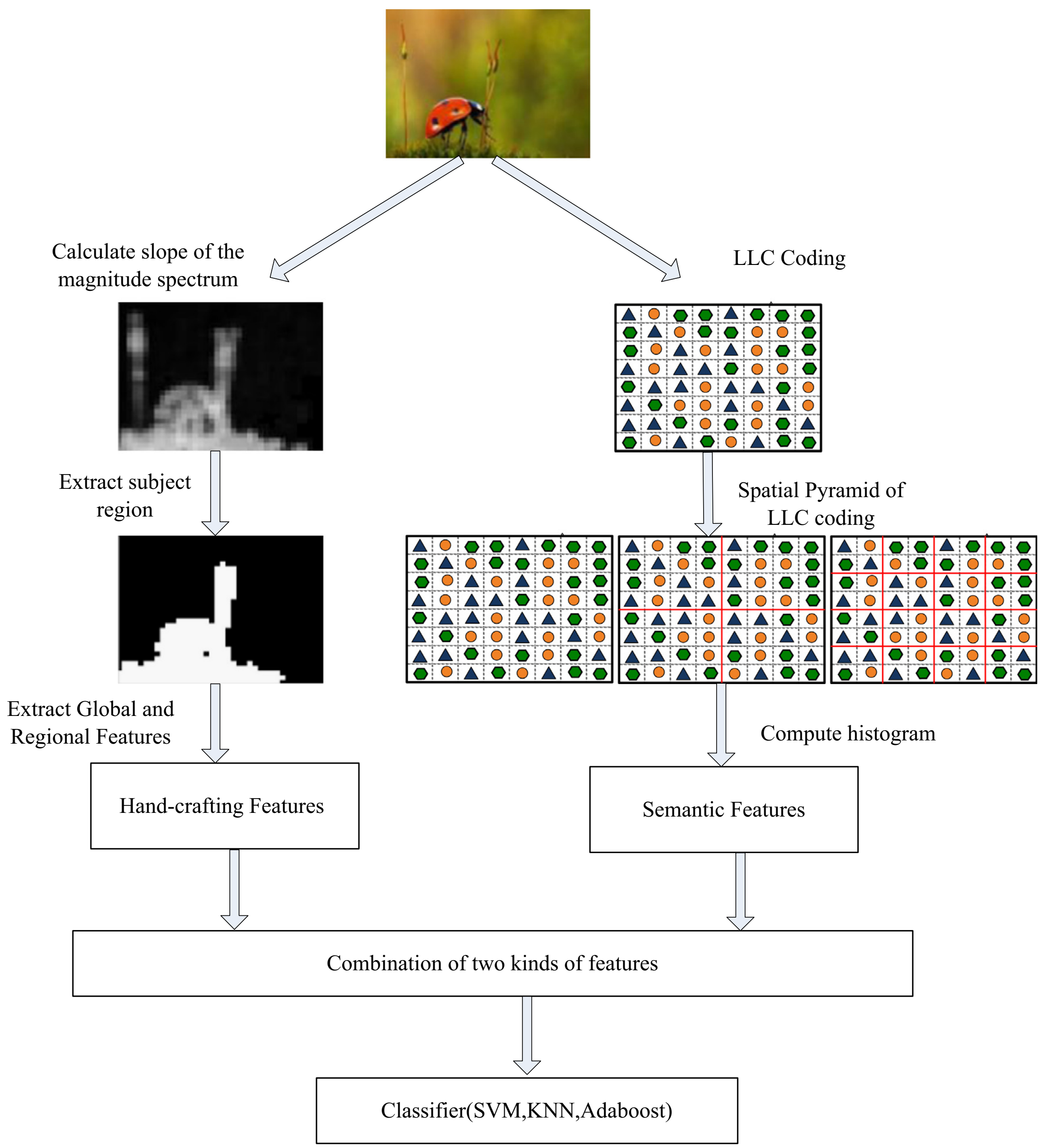

Fig. 1. Flow chart of our system.