Neurocomputing 181 (2016) 29–37

Contents lists available at ScienceDirect

Neurocomputing

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

Learning for an aesthetic model for estimating the traffic state in the traffic video

Xingmin Shi *, Zhenyu Shan, Na Zhao

Intelligent Transportation and Information Security Lab, Hangzhou Normal University, Hangzhou, Zhejiang, China

ARTICLE INFO

Article history:
Received 13 March 2015
Received in revised form
30 July 2015
Accepted 27 August 2015
Available online 17 November 2015

Keywords:
Aesthetic model
Traffic state
Random forest
Perceptual feature

ABSTRACT

With the increasing number of vehicles running on the urban roads, the traffic jam becomes much more serious. Properly estimating the traffic jam level from traffic videos is essential for the department of transportation management and drivers. Currently, for estimating the traffic state on videos, most solutions are built on evaluating traffic flow by counting the running vehicles per time unit or detecting their moving speed. However, the main challenge of these solutions is on the vehicle tracking method, in which the vehicles are necessary to be effectively and integrally segmented from the scenes. The solutions should tradeoff the accuracy of the estimation results and the efficiency of the method. In this paper, we propose a learning-based aesthetic model to estimate the traffic state on videos. The model uses multiple video-based perceptual features about traffic state to train the random forest classifier with the labeled data, and estimates traffic state by data classification. The evaluation experiments are conducted on a testing image set, and the results show that the traffic state estimation accuracy of the proposed model is higher than 98% and the efficiency performance is achieved in real-time.

1. Introduction

Owing to the rapid increment of vehicles in recent years, urban traffic jam has become a serious problem for the department of transportation management and drivers. Many researchers are pursuing research on effective methods of traffic state estimation to deliver accurate and real-time traffic information about the traffic jam. This traffic information is essential for traffic management, moving control, and driving guides, which are main research directions in intelligent transportation systems (ITS).

In the past few decades, a lot of research has been conducted on traffic state estimation. Most of them are built on traffic flow estimation by counting the running vehicles per time unit or detecting their moving speed. The traffic flow can be evaluated by the buried inductive loop detector [1], such as Sydney Coordinated Adaptive Traffic System (SCATS). In this system, the detector can obtain stimuli when a vehicle passes by the loop. However, the loop device buried under the road surface is not difficult to be broken and inconvenient to be maintained. Recently, another type of detector, the surveillance camera, has been widely used to evaluate traffic flow [2]. The video data provide more information than the inductive loop detector because the surveillance camera can capture more features. In video-based methods, besides the counting number of vehicles, we also use the estimated speed to evaluate the traffic flow [3]. However, the speed estimation greatly depends on the vehicle features extraction, which is seriously affected by changed conditions in videos. Moreover, multi-source traffic data, such as loop detector data, GPS data, and video data, are fused to improve the accuracy of traffic state estimation [4].

The general video-based methods exploit the segmentation of moving vehicles, either by frame-differencing or background subtraction. Unfortunately, both methods highly depend on the effectiveness and efficiency of the vehicle segmentation method to get accurate estimation results [5]. The methods often fail to work when the illumination conditions or environmental conditions change greatly. Moreover, the vehicle segmentation is difficult to be implemented due to the influence of moving objects or environmental objects, such as the pedestrians and blowing leaves. In order to overcome the limitation of general video-based methods, we propose to explore the traffic scene from a perceptual view, and build a learning-based aesthetic model with specific features, which are extracted from the videos for classifying the traffic state effectively. The vehicle corners and their spatial distribution features are used to construct the feature vector in this work, and the random forest classifier is utilized for perception training and traffic state classification. In summary, the main contributions of

* Corresponding author.
E-mail address: davidtrans@sina.com (X. Shi).

http://dx.doi.org/10.1016/j.neucom.2015.08.099
0925-2312/

this work include: (1) the vehicle corners and their spatial distribution feature are extracted from a perceptual aspect; (2) the proposed learning-based aesthetic model can effectively estimate the traffic state without vehicle segmenting and tracking process; and (3) the implementation of the proposed method is efficient and can be used in real-time applications.

The rest of this paper is organized as follows. Section 2 presents a review on the relevant methods proposed in latest literatures on aesthetic evaluation and traffic flow estimation. Section 3 fully illustrates the modules of the proposed method. The experimental results and analysis are elucidated in Section 4. Finally, the conclusion is drawn and future work is outlined in Section 5.

2. Related work

As aesthetic evaluation and video-based traffic state estimation are two main aspects related to our work, their related work is reviewed as follows.

2.1. Aesthetic approaches for image evaluation

Aesthetic evaluation is widely used in estimating the aesthetic quality, color harmony, beauty or the professional level of a photo or an image. To measure the aesthetic quality of an image, a bags-of-color-patterns method was proposed in [6]. In this work, the aesthetic quality of a photo is classified by taking color harmony into consideration. The color harmony model is built on computing the harmony score of local regions of a photograph instead of utilizing the global statistical information of colors. Generic image descriptor is another method utilized in [7] to evaluate the aesthetic quality of images. In this method, generic content-based local features, such as bag-of-visual-words (BOV), fishing vectors (FV), and GIST descriptors, are introduced due to that local-level patch-based information can be aggregated into an integrated global representation of images.

Fusion of various features can achieve better performance in aesthetic assessment. The effectivity of aesthetic assessment with various features including low-level features, mid-level semantic features, and type descriptors are evaluated in [8]. In this method, the training dataset is built by collecting free images from the public image sharing website, such as Flickr and DP Challenge, without any manual annotation. Comparing with the utilization of large amount of features for better understanding the aesthetics of an image, researchers select seven features into their top-down approach and improve the accuracy significantly [9]. In [10], the work extracts the relationship between the visual textures and the aesthetic perception properties, and proposes a layered prediction model to predict the aesthetic content for a given texture, which is defined as a vector consisting of computational features including low-level texture and color features. For evaluating the aesthetical features and classify the images, the Cellet, which connect spatially adjacent cells within the same pyramid level, can be utilized to represent the object [11]. The Graphlet is proposed to describe the object-level and spatial-level cues and forms the effective saliency descriptor [12], and the experimental results show that the active graphlet path is more indicative for photo aesthetics [13]. Moreover, structural cues are discovered and exploited in [14] and a new summarization technique is proposed in [15], in which the method can enforce video stability and preserves well-aesthetic frames.

Size has more or less impact on aesthetic evaluation. In the work [16], the researchers investigated the influence of size for aesthetic perception and found that the resolution and the physical dimensions can affect the appreciation of viewers. A series of regression models are proposed for predicting the aesthetic level of an image for a given size. Furthermore, the essential features related to the size-dependent property of image aesthetics are fully analyzed. In addition, machine learning methods are utilized to deal with the aesthetic quality estimation of an image [17]. The visual contents are extracted from the images and the support vector machine and tree classifiers are used for classification. They seek to explore the relationship between the perception feeling of a person and the low level content. Moreover, some universal metrics are adopted to evaluate the aesthetics in the game of chess [18].

2.2. Video-based traffic state estimation

For estimating the traffic state in traffic videos, virtual loop based methods and vehicle tracking methods are extensively used to count the vehicles. Virtual loops are defined or designated before dealing with the object tracking. When a moving object enters the region of virtual loops and exit the region, the counter will plus one, thus making the vehicle counting possible. The user-defined virtual loops are utilized to detect and count vehicles in [19]. In this work, the foreground mask is produced by Gaussian Mixture Model (GMM) and Motion Energy Images (MEI) method. The particle grouping is utilized in sub-sampling video frames according to their spatial and temporal coherence, as well as motion coherence. Such particles are clustered with the k-means algorithm, and their motion patterns and spatial information are taken into consideration in the clustering. Vehicle tracking is performed on these clusters corresponding to vehicles computed by evaluating the similarity of color histograms.

An extended Kalman filter is introduced into the real-time freeway traffic state estimation in [20]. This work is intended to pursue a general solution of real-time adaptive traffic state estimation in freeway networks. The solution is based on the stochastic macroscopic traffic flow model with extended Kalman filtering. The model parameters and the traffic flow variables are jointly estimated. As a result, the prior calibration is not necessary, and the solution is adaptive to various scenarios which can trigger the incident alarms. A complete system is proposed by Henri Nicolas et al. to analyze the behavior of vehicles in [21]. For improving the results, the scene characteristics and predefined traffic rules are employed in this work. The solution includes three steps. The first step is scene modeling, in which the scene structure and the traffic rules are automatically obtained. The second step is tracking multiple objects whose trajectories are evaluated. The final step is to evaluate the behavior of vehicles. This method can efficiently detect and estimate the behavior of vehicles. Predefined rules and the geometry constraints make it more complex and inconvenient in application. In order to estimate the speed of traffic flow, the road spaced occupancy and the traffic state, a traffic flow detection tool is implemented in [22].

3. Proposed method

3.1. Overview

The proposed method mainly consists of four modules, which are data acquisition, initial setup, features extraction, and traffic state training and classification. The acquisition has the responsibility of acquiring the video data from traffic surveillance cameras. Initial setup mainly concerns the lane setting. In practice, the traffic flow in a specific direction should be independently estimated. Generally, the view of the camera covers several two-way lanes. As a result, the lane region should be designated before image analysis in order to mitigate the effect from the adjacent