Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Dominant vanishing point detection in the wild with application in composition analysis

Xiaodan Zhang, Xinbo Gao*, Wen Lu, Lihuo He, Qi Liu

State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 11 March 2018
Revised 8 May 2018
Accepted 23 May 2018
Available online xxx

Communicated by Jungong Han

Keywords:
Image composition analysis
Vanishing point detection
Deep learning
Image aesthetic analysis

ABSTRACT

The vanishing point provides a strong ability to infer the 3D structure of the scene. It finds great application in image composition analysis, lane detection, camera calibration and salience detection. Many methods have been proposed to predict the location of vanishing point. They are usually based on geometrical and structural features such as lines or contours. However, such methods suffer deteriorated accuracy due to the large number of outlier line segments in natural landscape images. In this paper, we propose a semantic-texture fusion network to detect the dominant vanishing point in the image. The proposed network includes two branches. The first branch is based on the Holistically-Nested Edge Detection Network which extracts textural features. The second branch aims to extract the semantic features. In order to boost the representational power of a network, we adopt the Squeeze-and-Excitation block to model the interdependencies between the semantic features and the textural features. Experimental results reveal a step forward against the state-of-the-art vanishing point detection methods in natural landscapes. Based on the detection results, we further demonstrate how the proposed model can be used to provide on-line guidance to amateur photographers.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Under the pin-hole camera model, a set of parallel lines in three-space will project to converging lines in the image plane. A common point of the intersection is known as the vanishing point (VP). The VP encodes the 3D orientation of the lines, and thus is essential to many tasks, *e.g.* automatic navigation [1], image calibration [2], 3D reconstruction and fixation prediction [3,4].

The technique of VP detection has been extensively studied in recent years [5–9]. Despite the extensive studies on VP detection, they mainly focus on man-made environment which includes large number of line segments. Besides, they also assume the three dominant vanishing directions in the scenes are mutually orthogonal. The Manhattan-world hypothesis is in general valid for simple urban scenes. However, for more complex scenes in natural images, there may not be more than one VP, which makes traditional VP detection methods inapplicable. Besides, the visible edges can be weak and not detectable via local photometric cues. Further, there are many noisy lines in the background, such as those on tree's branches. Nearly all pairs of lines will intersect in the image plane, even if they are not parallel in 3D, leading to many spurious candidates for vanishing points. In this scenario, it is difficult to pick out the VP without semantic information.

This paper explores the possibility of detecting dominant VP by deep learning model in the wild. The dominant VP is the VP (i) associates with the major geometric structures of the scene, and (ii) conveys a strong impression of 3D space of depth to the viewers [10,11]. VP is one of the most commonly used techniques in photo composition. As shown in Fig. 1, it is a powerful tool to illustrate depth. Photographers often use it to attract the viewer's attention. Nowadays, there is a demand to develop more intelligent functionalities for digital cameras, among which automatically helping to take wonderful images has recently begun to attract research attention. Solution to this problem is promising, in the sense that it can bring dramatic convenience for millions of amateur photographers. A best way to learn photography is to learn from the experienced peers. When we take photos, we may wonder what the experienced photographers would take in a similar situation. Inspired by the work of Zhou et al. [10], we tackle the problem by retrieving images which have similar point of view with similar semantic contents. The dominant VP encodes the angle of perception when a viewer sees a scene. Thus our VP detection network captures rich viewpoints information. We apply our network to an image retrieval application which aims to provide amateur users with on-site feedback about the composition of their photos.

* Corresponding author.
E-mail addresses: xdanzhang@stu.xidian.edu.cn (X. Zhang), xbgao@mail.xidian.edu.cn, xbgao.xidian@gmail.com (X. Gao), luwen@mail.xidian.edu.cn (W. Lu), lihuo.he@gmail.com (L. He), liuqi625@sina.com (Q. Liu).

https://doi.org/10.1016/j.neucom.2018.05.071
0925-2312/© 2018 Elsevier B.V. All rights reserved.

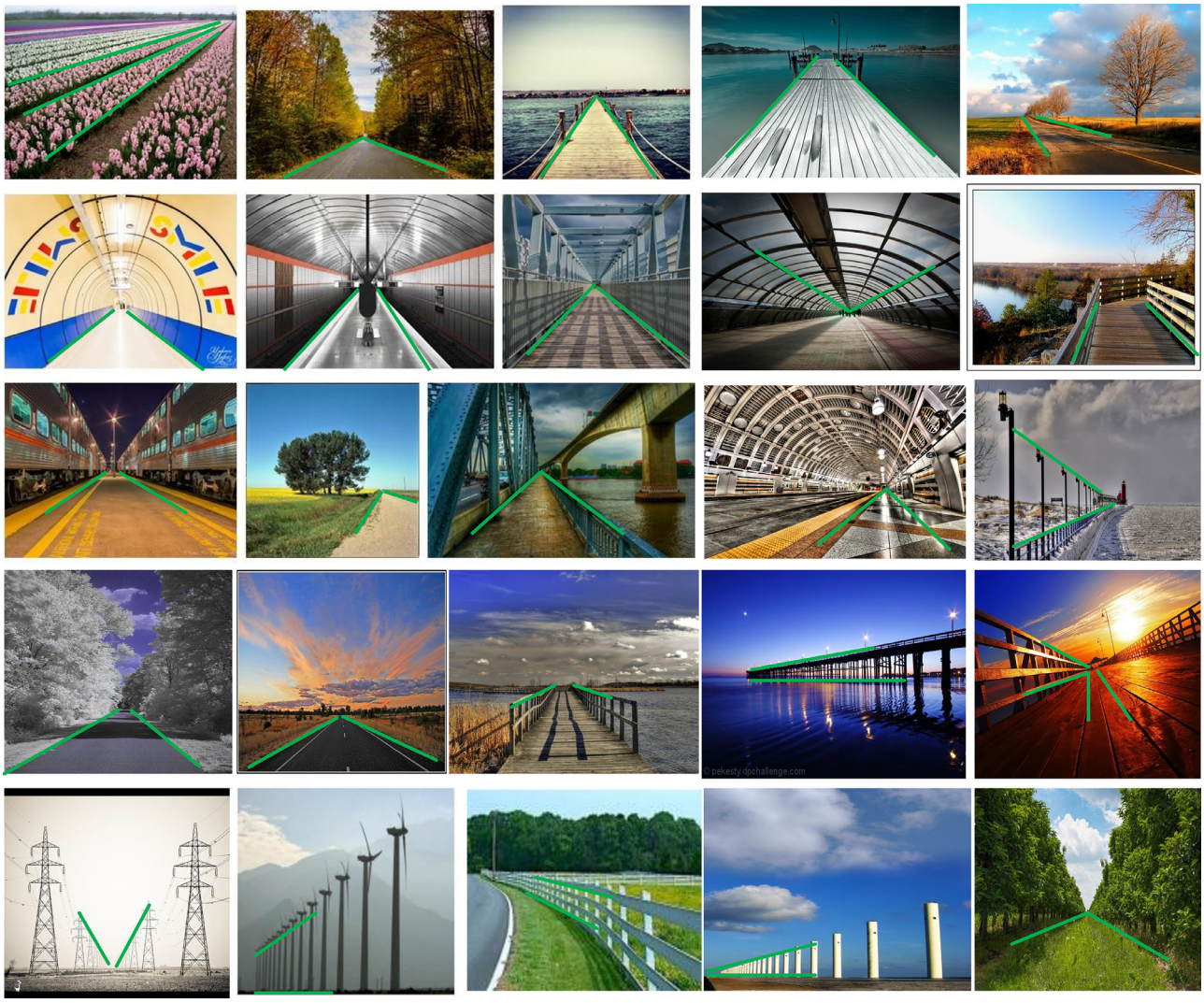

Fig. 1. The use of vanishing points in photography. Experienced photographers often use the vanishing points and the associated lines to draw the viewers to key elements. Manually labeled ground truth lines are marked in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this paper, we make the following three contributions.

- We propose a semantic-texture fusion network to the vanishing point detection in the wild. By combining the semantic and textural features, our method significantly outperforms state-of-the-art methods for natural scenes.
- In order to boost the representational power of a network, we adopt the Squeeze-and-Excitation block to model the interdependencies between the semantic features and the textural features.
- We demonstrate the application of our model by providing online guidance to amateur photographers.

The paper is organized as follows. We summarise the related work in Section 2. We describe the semantic-texture fusion network in Section 3. Experimental results and the application are introduced in Section 4. Conclusions are presented in Section 5.

2. Related work

The proposed method is closely related to two research topics: image composition analysis (2.1) and vanishing point detection (2.2).

2.1. Image composition analysis

Image composition modeling problem has been studied for decades. It finds great applications in image aesthetic assessment, emotion prediction, and image recommendation [12–21]. In early works, Tang et al. [12] propose to model the important techniques such as simplicity, color harmony, lighting, for image aesthetic assessment. Yeh et al. [15] model the rule of thirds and image simplicity based on the image segmentation results and salience information. Liu et al. [13] propose a compound operator of crop-and-retarget to modify the composition aesthetics of the image. Image composition also has strong relationship with emotion prediction. In [19], Zhao et al. model the composition principles, such as *balance, emphasis, harmony, variety, gradation,* and *movement* to predict the emotion. However, the above methods implement the simple photography composition guidelines from 2D rendering of visual elements. In this paper, we focus on the linear perspective effects which encode the 3D information of the scenes.

2.2. Vanishing point detection

The history of vanishing point detection method is extremely rich. Many previous work is based on the Manhattan world

Please cite this article as: X. Zhang et al., Dominant vanishing point detection in the wild with application in composition analysis, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.05.071

Download English Version:

https://daneshyari.com/en/article/6863645

Download Persian Version:

https://daneshyari.com/article/6863645

Daneshyari.com