Neurocomputing 158 (2015) 174–183

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptive road detection via context-aware label transfer

Qi Wang [a,*], Jianwu Fang [b,c], Yuan Yuan [b]

[a] Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, PR 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, PR China
[c] University of Chinese Academy of Sciences, Beijing 100049, PR China

A R T I C L E I N F O

Article history:
Received 11 November 2014
Received in revised form
31 December 2014
Accepted 30 January 2015
Communicated by Qi Li
Available online 11 February 2015

Keywords:
Computer vision
Road detection
Depth map
Label transfer
Context-aware
MRF

A B S T R A C T

The vision ability is fundamentally important for a mobile robot. Many aspects have been investigated during the past few years, but there still remain questions to be answered. This work mainly focuses on the task of road detection, which is considered as the first step for a robot to become moveable. The proposed method combines the depth clue with traditional RGB information and is divided into three steps: depth recovery and superpixel generation, weakly supervised SVM classification and context-aware label transfer. The main contributions made in this paper are (1) Design a novel superpixel based context-aware descriptor by utilizing depth map. (2) Conduct label transfer in an efficient nearest neighbor search and a temporal MRF model. (3) Update the learned model adaptively with the changing scene. Experimental results on a publicly available dataset justify the effectiveness of the proposed method.

1. Introduction

To autonomously navigate a robot in an outdoor environment, the vision system should be capable of perceiving the surrounding world. For example, the robot needs to know who he is interacting with [1], which way he could follow [2], and where he should stop to conduct his mission [3]. Among these abilities, road detection [4] is the primary one for a robot to become moveable. For this purpose, the road detection task needs to provide a clue of the drivable road in an input image or video so that the intelligent system can plan its path. In this paper, we put our focus on the road detection problem, which is fundamentally important not only for a robot, but also for an autonomous vehicle with an advanced driver-assistance system [5], such as object tracking [6] and anomaly detection [7].

Since the road images may differ with each other greatly, the detection task is actually not an easy task. Take Fig. 1 as an example. The roads have different pavements and are laid on different places, leading to various color, texture, and shape appearance. Along with these factors, the lighting condition is another influential one, inducing shadows on the road surface. These complexities in together make a reliable road detection difficult. Motivated by this fact, in this paper we propose a robust road detection method based on depth fusion and label transfer in a video sequence. The stable depth clue ensures the robustness of the proposed model and the ever-updating mechanism makes the transferred labels accurate (Fig. 2).

1.1. Overview of the proposed method

Though many works have been proposed in the past few years, most of them focus on the individual image. In fact, video sequences are the most frequently confronted situation instead of single images. Therefore, we lay our attention on the video sequence in this paper. The task is to infer the road area in each frame given a camera recorded street scene. The proposed method in this paper is named as context-aware label transfer (CALT), which is divided into the following three steps:

- *Preprocessing*: To facilitate the processing, the input video frames are firstly segmented by SLIC [8] to get superpixels. The subsequent label transfer is based on the obtained superpixels. Since we want to utilize the depth clue in the framework, the depth map of each frame is also reconstructed according to a consistent depth recovery technique [9].
- *Context-aware label transfer*: The preprocessed sequence is tackled frame by frame in this step. For the first frame, its ground truth labels (road or non-road) are manually marked. For the subsequent frames, their labels are sequentially

* Corresponding author.
E-mail addresses: crabwq@nwpu.edu.cn (Q. Wang), fangjianwu@opt.ac.cn (J. Fang), yuany@opt.ac.cn (Y. Yuan).

http://dx.doi.org/10.1016/j.neucom.2015.01.054
0925-2312/

transferred according to the previous one, taking the obtained result of current frame as the updated ground truth. This operation is iteratively conducted until all the frames are treated.

- *Result demonstration*: After the above procedure, each superpixel in a frame is allocated a label. To get a consistent labeling map without noisy labels, a Winner-Take-All (WTA) smoothing is applied to rule out the isolated inaccurate labels. Then a geometric triangle constraint is employed to restrict the road area. The final results are then demonstrated overlaid in the original sequence.

1.2. Contributions

Although there are existing road detection methods by employing depth map and label transfer, the proposed one in this paper is distinguished with them in the following aspects, which also makes the main contributions of this paper.

- Design a novel superpixel based context-aware descriptor by utilizing depth map. Most existing methods only consider the color information of the obtained image or video. Several works employ the depth map, but the way they incorporating it into the frameworks is simple and straightforward [10,11]. In this paper, we segment the image into superpixels and capture its characteristic by simultaneously concatenating the color and depth features. This combination is effective because it leverages the superiorities of the color's distinctiveness and depth's robustness. Based on this characteristic, a context-aware descriptor is developed to represent the superpixel's relationship with the adaptive circular neighborhood, which further paves the way for the optimization of label transfer.
- Conduct label transfer in an efficient nearest neighbor search. Label transfer can reduce the inference problem of training sophisticated parametric models for an unknown image to the problem of matching it to an existing set of annotated images [12]. But in this process, accurate registration is a challenging task [13]. For a more precise correspondence, we choose to transfer the labels between superpixels with the most similar contextual clues in adjacent images, instead of searching for the best match in a big training set. Since the two examined images are similar to each other, we utilize a new dense pixel correspondence method [14] to register the near superpixels respectively in the adjacent frames, which effectively exploits the video temporal relations between frames.
- Update the learned model adaptively. Traditional offline methods learn the model only once in the beginning. This strategy leads to an expensive training stage with large amount of images. More importantly, it means even if the actual scene changes much from the training ones, there is no adaptability. Based on this consideration, the proposed model updates the parameters of the classifier frame by frame, yet in an efficient manner. Novel properties of specific labels are dynamically updated, which ensures that the model can handle the changing scene.

Fig. 1. Road images with different colors, textures, shapes and lightings.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the recovery of depth map. Section 4 describes the main part of the proposed method – context-aware label transfer. Section 5 gives the experimental results to justify the effectiveness of the proposed method. Conclusions are finally made in Section 6.

2. Related work

The techniques for road detection can be categorized according to the types of road images, which are structured ones (e.g., a road in urban street) and unstructured ones (e.g., a road in rural area) [15].

For the structured road detection, the captured images have clear road markings and the designed algorithms are based on these extracted markings [16]. Among the earliest attempts, Bertozzi and Broggi [17] assume that the road markings are visible. Based on this assumption, the stereo image pair is first mapped to another 2D space to remove the perspective effect. Then the left image is used to recognize the road markings and both the two images are further employed to detect the free areas ahead the vehicle. Wang and Frémont [18] first use sky removal to enhance the axis-calibration stability. Then the stereo vision based extension is applied to extract the line function and reconstruct the ground plane. But the stereo images are vulnerable to weather conditions such as rain, snow, fog, and darkness. The radar sensor, on the contrary, being an active sensor and operating at millimeter wavelengths, can provide an alternate image of the scenario in front of the vehicle. For example, Ma et al. [19] propose a Bayesian model to interpret the radar and optical road images. In this procedure, they incorporate the lane and pavement prior to guide the boundary detection. Feng et al. [20] design a system equipped with a 2D laser radar. By measuring the distance from the radar to the road surface, a rectangle-searching algorithm is implemented to find the road rectangle containing the most road points. Besides the multi-sensor approaches, many other ones mainly focus on the feature representation, which can avoid the inconvenience of sensor setups and the undesired radioactive

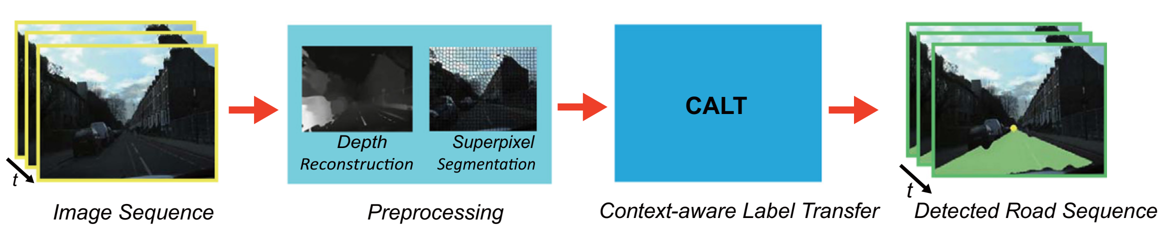

Fig. 2. General framework of the proposed method.