ARTICLE IN PRESS
JID: INS [m3Gsc;May 4, 2017;0:0]
Information Sciences 000 (2017) 1–16
Contents lists available at ScienceDirect
Information Sciences
journal homepage: www.elsevier.com/locate/ins

Hybrid conditional random field based camera-LIDAR fusion for road detection

Liang Xiao[a], Ruili Wang[b], Bin Dai[a,*], Yuqiang Fang[a], Daxue Liu[a], Tao Wu[a]

[a] College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha, Hunan, China
[b] Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 13 December 2016
Revised 23 April 2017
Accepted 28 April 2017
Available online xxx

Keywords:
Sensor fusion
Road detection
Condition random field
Boosted decision tree

ABSTRACT

Road detection is one of the key challenges for autonomous vehicles. Two kinds of sensors are commonly used for road detection: cameras and LIDARs. However, each of them suffers from some inherent drawbacks. Thus, sensor fusion is commonly used to combine the merits of these two kinds of sensors. Nevertheless, current sensor fusion methods are dominated by either cameras or LIDARs rather than making the best of both. In this paper, we extend the conditional random field (CRF) model and propose a novel hybrid CRF model to fuse the information from camera and LIDAR. After aligning the LIDAR points and pixels, we take the labels (either road or background) of the pixels and LIDAR points as random variables and infer the labels via minimization of a hybrid energy function. Boosted decision tree classifiers are learned to predict the unary potentials of both the pixels and LIDAR points. The pairwise potentials in the hybrid model encode (i) the contextual consistency in the image, (ii) the contextual consistency in the point cloud, and (iii) the cross-modal consistency between the aligned pixels and LIDAR points. This model integrates the information from the two sensors in a probabilistic way and makes good use of both sensors. The hybrid CRF model can be optimized efficiently with graph cuts to get road areas. Extensive experiments have been conducted on the KITTI-ROAD benchmark dataset and the experimental results show that the proposed method outperforms the current methods.

1. Introduction

Road detection is a fundamental research topic in autonomous vehicles and has been studied for decades [11]. For autonomous vehicles, stable and accurate road detection is a prerequisite. As there are different kinds of roads, such as highways, urban roads and country roads, with different features, the approaches to detect them are different. In well painted highways, road detection can be replaced by lane detection, which is considered to be much easier. However, it is much more challenging to detect normal urban roads for many reasons such as the variations in road materials from segment to segment, the similarities of textures and heights between the road areas and non-road areas, the changes of illumination and weather and so on.

* Corresponding author.
E-mail addresses: xiaoliang@nudt.edu.cn (L. Xiao), r.wang@massey.ac.nz (R. Wang), bindai.cs@gmail.com (B. Dai), fangyuqiang@nudt.edu.cn (Y. Fang), liudaxue78@gmail.com (D. Liu), wutao@nudt.edu.cn (T. Wu).

http://dx.doi.org/10.1016/j.ins.2017.04.048

To achieve accurate and stable road detection, many algorithms based on different kinds of sensors have been developed. The most commonly used sensors are monocular cameras [4,15] and LIght Detection And Rangings (LIDARs) [8] which can acquire different kinds of information for road detection. Monocular vision captures the perspective projection of the scene and then the dense colors and textures can be used to group the pixels or super-pixels into road and background areas. However, monocular vision often suffers from the changes of illumination and weather, and it cannot capture accurate 3D information. Compared to vision, a LIDAR is an active sensor which works independently of the ambient light and it can measure the distances to objects accurately. However, in the point clouds captured by a LIDAR, neither color nor texture information is available and the points are rather sparse.

To overcome the inherent drawbacks and combine the merits of different kinds of sensors, multi-modal sensor fusion has been widely used [31,33,43,50,53]. For road detection, several camera-LIDAR fusion methods have been proposed. However, most of them are dominated by either cameras or LIDARs and fail to fully exploit the advantages of both sensors. For example, in [43], after projecting the LIDAR point cloud onto the image, the feature used for obstacle classification is dominated by the height information of the LIDAR points while the pixel information is ignored. In [21], the information from the images and LIDAR point clouds is utilized separately in a stage-wise fashion. The LIDAR point clouds are only used for ground seed extraction, while the following road detection and segmentation are dominated by the image. In [54], fusion is performed on the feature and region levels, resulting in a coarse level fusion. All these methods fail to fuse the image and LIDAR in fine granularity and through a joint model. This work aims to fill this gap. Aside from the multi-modal information, another kind of information that is crucial for improving the performance is the contextual information in each modality. Considering the strength of conditional random fields (CRF) in modeling contextual information [44], we extend CRF to a multi-modal setting and propose a novel hybrid-CRF-based camera-LIDAR fusion method to improve the performance of road detection. By formulating the road detection as a binary labeling problem, the labels (either road or background) of the pixels and LIDAR points are taken as random variables and a hybrid CRF model is built to solve the multi-modal labeling problem. The proposed method utilizes the learned boosted decision tree classifiers to derive the unary potentials of the pixels and LIDAR points. The neighboring smooth prior of the pixels and LIDAR points, together with the consistency constraint between the aligned LIDAR points and pixels are modeled via the pairwise potentials. This model integrates the information from the two sensors probabilistically and the information from both sensors are well exploited. The hybrid CRF model can be optimized efficiently by graph cuts [23] to get road areas. Experiments conducted on the KITTI-ROAD benchmark dataset [14] demonstrate that the proposed hybrid CRF model is effective in fusing multi-modal information and the results of road detection are better compared to that of the current existing methods.

The main contributions of this paper include: (i) A novel hybrid CRF model is proposed to fuse the image and LIDAR point cloud, in which the contextual consistency of the image and LIDAR point cloud, together with the constraint of cross-modal consistency is jointly modeled probabilistically, and (ii) the proposed sensor fusion framework is applied to urban road detection and our method achieves good performance on the KITTI-ROAD benchmark dataset [14]. The results of our method on the UM subset rank first on the leaderboard [1] apart from the deep-learning-based ones, which usually rely on models pre-trained on extra data for initialization and modern GPUs for fast computing.

The rest of this paper is organized as follows. Section 2 reviews the work on road detection. Section 3 shows how the LIDAR points and the images are registered. In Section 4, we first introduce the CRF-based labeling framework, then we provide the detailed information about the proposed hybrid CRF model. The training of the pixel and LIDAR point classifiers is described in Section 5, along with the feature extraction. The experimental results tested on the KITTI-ROAD benchmark dataset are given in Section 6. Finally, conclusions and directions for the future work are listed in Section 7.

2. Related work

As a fundamental problem in developing autonomous vehicles, road detection has been extensively studied. Various road detection systems have been developed based on different kinds of sensors as well as fusion of some sensor types.

The most frequently used sensor for road detection is the monocular camera [20]. Monocular-vision-based road detection is usually formulated as a classification problem, i.e., classifying each pixel or super-pixel into either road or background. Many kinds of machine learning methods have been applied to road detection, such as mixture of Gaussian [10], support vector machines [2], extreme learning machines [30,55], neural networks [42], boosting [15] and structured random forest [51]. In recent years, many new feature learning methods have been applied to road detection such as slow feature analysis [15], sparse coding and dictionary learning [28,29,32,52], convolutional neural network [3,35] and deep deconvolutional network [37]. Classification-based methods classify each unit independently and do not take the contextual interaction into consideration. Therefore, the prediction may be noisy. To solve this problem, conditional random fields (CRF) [19,41,45,50] are widely used to model the contextual interaction. Generally, CRF-based methods are supposed to get better performance than simple classification-based methods. However, when the image quality is badly affected by illumination or weather conditions, these methods may also get poor results.

LIDAR is another widely used sensor in autonomous vehicles. Various LIDAR-based road detection algorithms have been proposed and they can be roughly categorized into two groups: regression-based and classification-based algorithms. Based on the assumption of the continuity of the road area, the regression-based algorithms utilize one dimensional curve fitting [8,21] or two dimensional surface fitting [5,12] to segment the road. The classification-based algorithms extract features of the points or grid cells and then classify them based on certain intuitive rules or learning methods, such as elevation map