StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation

Dan Levi[1]
http://sites.google.com/site/danmlevi
Noa Garnett[1]
noagarnett@gmail.com
Ethan Fetaya[2]
ethan.fetaya@weizmann.ac.il

[1] Advanced Technical Center Israel
General Motors R&D
Herzlyia, Israel

[2] Weizmann Institute of Science
Rehovot , Israel

Obstacle detection is a fundamental technological enabler for autonomous driving and vehicle active safety applications. While dense laser scanners are best suitable for the task (e.g. Google's self driving car), camera-based systems, which are significantly less expensive, continue to improve. Stereo-based commercial solutions such as Daimler's "intelligent drive" are good at general obstacle detection while monocular-based systems such as Mobileye's are usually designed to detect specific categories of objects (cars, pedestrians, etc.). The problem of general obstacle detection remains a difficult task for monocular camera based systems. Such systems have clear advantages over stereo-based ones in terms of cost and packaging size.

Another related task commonly performed by camera-based systems is scene labeling, in which a label (e.g. road, car, sidewalk) is assigned to each pixel in the image. As a result full detection and segmentation of all the obstacles and of the road is obtained, but scene labeling is generally a difficult task. Instead, we propose in this paper to solve a more constrained task: detecting in each image column the image contact point (pixel) between the closest obstacle and the ground as depicted in Figure 1(Left). The idea is borrowed from the "Stixel-World" obstacle representation [1] in which the obstacle in each column is represented by a so called "Stixel", and our goal is to find the bottom pixel of each such "Stixel". Note that since we don't consider each non-road object (e.g. sidewalk, grass) as an obstacle, the task of road segmentation is different from obstacle detection. Notice also that free-space detection task is ambiguously used in the literature to describe the above mentioned obstacle detection task [1] and the road segmentation task [4].

Current "Stixel-based" methods [1] for general obstacle detection use stereo vision while our method is monocular-based. A different approach for monocular based obstacle detection relies the host vehicle motion and uses Structure-from-Motion (SfM) from sequences of frames in the video [3]. In contrast our method uses a single image as input and therefore operates also when the host vehicle is stationary. In addition, the SfM approach is orthogonal to ours and can therefore be later combined to improve performance. For the task of road segmentation, the common approach is to perform pixel or patch level [4]. In contrast, we propose to solve the problem using the same column-based regression approach as for obstacle detection. Our approach is novel in providing a unified framework for both the obstacle detection and road-segmentation tasks, and in using the first to facilitate the second in the training phase.

We propose solving the obstacle detection task using a two stage approach. In the first stage we divide the image into columns and solve the detection as a regression problem using a convolutional neural network, which we call "StixelNet". Figure 1(Right) shows an example network input and output. In the second stage we improve the results using interactions between neighboring columns by imposing smoothness constrains via a Conditional Random Field (CRF) over consecutive columns. The flowchart of the obstacle detection algorithm is presented in Figure 2(top). To train the network we introduce a new loss function based on a semi-discrete representation of the obstacle position probability. In this approach we model the probability $P(y)$ of the obstacle position as a piecewise-linear probability distribution.

The road segmentation is done is three stages. The first two, StixelNet (trained on the road segmentation task) followed by a CRF, are the same as in obstacle detection. The final stage performs a graph-cut segmentation on the image to achieve higher accuracy by enforcing road boundaries to coincide with image contours. The flowchart of the road segmentation algorithm is presented in Figure 2(bottom).

It is well known that having large quantities of labeled data is crucial for training deep CNNs. A major advantage of our unique task formulation is the ability to use laser-scanners, which are excellent at the given task, for labeling, thus eliminating the need for manual annotation. In addition, we further leverage this by fine-tuning StixelNet to the road segmentation task using a smaller amount of hand labeled data. Our experiments use the KITTI dataset [2]. On obstacle detection our approach is the state-of-the-art camera-based method even when compared to the stereo-based "Stixel" approach [1]. On the KITTI road segmentation challenge, our fine-tuned network, although not suitable to model all cases, is ranked second among all methods.

Figure 1: (Left) Obstacle detection example. (Right) Input to StixelNet and output example

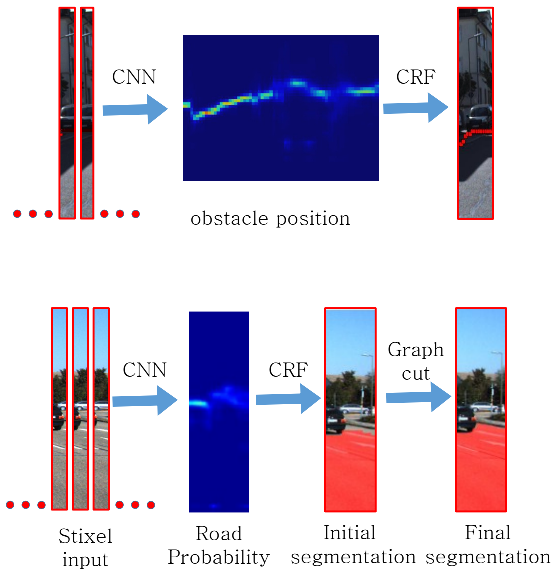

Figure 2: (Top) Obstacle detection algorithm flowchart. (Bottom) Road segmentation algorithm flowchart.

[1] Hernán Badino, Uwe Franke, and David Pfeiffer. The stixel world - a compact medium level representation of the 3d-world. In *Pattern Recognition*, 2009.

[2] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012.

[3] Qifa Ke and T. Kanade. Transforming camera geometry to a virtual downward-looking camera: robust ego-motion estimation and ground-layer detection. In *CVPR*, 2003.

[4] G. B. Vitor, A. C. Victorino, and J. V. Ferreira. A probabilistic distribution approach for the classification of urban roads in complex environments. In *Workshop on Modelling, Estimation, Perception and Control of All Terrain Mobile Robots on IEEE International Conference on Robotics and Automation (ICRA)*, 2014.