Semantic Segmentation for Pedestrian Detection from Motion in Temporal Domain

Guo Cheng, Jiang Yu Zheng
Department of Computer and Information Science, Indiana University Purdue University Indianapolis, 723 W Michigan St, Indianapolis, IN 46202, USA

Abstract

Instead of detecting pedestrians from their appearances in 2-D spatial images, which is time-consuming. This paper detects pedestrians along with their motion-directions in a temporal way. By projecting a driving video to a 2-D temporal image called Motion Profile (MP), we can robustly distinguish pedestrian in motion against smooth background motion. To ensure non-redundant data processing of deep network on a compact motion profile further, a novel temporal-shift memory (TSM) model is developed to perform deep learning of sequential input in linear processing time. In experiments containing pedestrian motion from various sensors such as video and LiDAR, we achieve the detecting rate of pedestrians at 90% in near and mid-range on the road. With the super-fast speed and good accuracy, this method is promising for intelligent vehicles.

Keywords – autonomous driving, motion profile, temporal-to-spatial, semantic segmentation.

I. Motion Profile for Video Data Reduction

Semantic Segmentation

far mid near
distance motion profile

Fig.1 Diagram of the motion-based method. (top-left) Frames in driving video with three (colored) zones covering far, mid, and near depths. (middle) Motion Profiles from three zones in the video formed by averaging pixels vertically in each zone to a line, and then consecutively copying lines into a temporal image. These temporal images contain vehicle, object, and pedestrian trajectories, which will be semantic-segmented through a Temporal-Shift Memory. Results of pedestrian trajectories at far, mid, near depths displayed in three colors (and mixed color in between these depths).

II. Pedestrian Motion in MP

Fig.2 Motion Profile (MP) from driving video. (left) A video volume. (right) A MP image (colored) consists of consecutive pixel lines extracted in video frames within a fixed zone. If the frame of HD video is 720 pixels in height, the data in the MP are 1/720th of video.

III. Patterns of Pedestrian Motion in MP

Fig.4 MPs from different sensors display the same semantic meaning.

IV. Semantic Segmentation of Pedestrian Motion

For the path planning and autonomous driving on normal roads, we classify pixels on the latest lines in the road profile into three semantic classes: Pedestrian in motion, human standing-still, background.

V. Sequential Semantic Segmentation with TSM

Fig.5 Semantic segmentation architecture on MP (left) and output (right).

VI. Quantitative Results

Table I   Accuracy of Pedestrian trace at pixel level.

VII. Evaluation

Fig.9 Evaluation of semantic results in a crowd section of MP.

Table II   Accuracy of Pedestrian trace at frame level.

Contact the Corresponding Author:

Name: Guo Cheng
Email: guocheng@iu.edu