UNIVERSITEIT VAN AMSTERDAM

Vulnerable Road User Detection and Orientation Estimation for Context-Aware Automated Driving
F.B. Flohr

SUMMARY

Vulnerable Road User Detection and Orientation Estimation for Context-Aware Automated Driving
Fabian Berthold Flohr

This thesis addresses the detection, segmentation and orientation estimation of persons in visual data. While the possible application domains of the proposed methods are manifold, ranging from image editing over surveillance to robotics and the intelligent vehicle domain, the latter is in focus of this work. In particular, the work focuses on the role of the Vulnerable Road Users (VRU, e.g. pedestrians and cyclists), being among the most critical objects for the realization of self-driving vehicles. A human driver is able to efficiently detect and localize a VRU on the street. Furthermore, a human driver recognizes important context information of the VRU (e.g. awareness, intention) and the environment (e.g. infrastructure elements), helping him to anticipate the VRU behavior. From an automation perspective, it is desirable to imitate or even outperform the skills of a human driver with a machine system. Motivated by this, the aim of this work is to establish an accurate machine representation of the VRU by using image-based cues to support context-aware automated driving.

The first addressed problem is a reliable ***detection of the VRU***, being a crucial preliminary step for all subsequent modules. The detection of VRUs is especially challenging due to their wide variation in appearance, arising from

articulated pose, clothing, background and visibility conditions (time of day, weather). To cope with these challenges, a stereo-based superpixel representation (i.e. stixels) is applied for efficient proposal generation. The resulting proposals are used within a Deep Convolutional Neural Network architecture to gain a robust object detection. Results are discussed on a newly introduced dataset, being the first dataset of this size, focusing on the challenging detection of cyclists in urban areas. Even with a significantly reduced proposal count compared to commonly used 2D proposal methods, competitive detection results are gained.

Based on the robust detection, ***pixel-wise VRU segmentation*** is considered to facilitate higher-level, semantic scene analysis (e.g. body part labeling, pose estimation, activity analysis). Furthermore a pixel-wise segmentation has the potential to enhance the detection and localization performance in itself. The large variety of VRU appearances make the problem again challenging. On the other hand, focusing on a single object class makes it possible to introduce a fair amount of prior knowledge on how pedestrians appear in images. The proposed method combines generative shape models and multiple data cues within an iterative framework. In each iteration, shape and data cues are refined leading to an accurate segmentation after only a few iterations. Experiments on a public segmentation dataset suggest that the proposed method outperforms state-of-the-art. To analyse the benefit of using additional disparity cues for segmentation, a new pedestrian segmentation dataset has been introduced.

Looking at ***head and body orientation of a VRU*** supports a human driver to estimate the motion state and attention of a VRU within a fraction of a second. Therefore, a new method for joint probabilistic head and body orientation estimation has been created that handles faulty part detections, continuous orientation estimation, coupling of the body and head localization and orientation, and temporal integration. For both head and body parts, responses of a set of orientation-specific detectors are converted into a (continuous) probability density function. Head and body parts are estimated jointly to account for anatomical constraints. The joint single-frame orientation estimates are integrated over time by particle filtering. The experiments involve data from a vehicle-mounted stereo vision camera in a realistic traffic setting. It is shown that the proposed method reduces the mean absolute head and body orientation error significantly compared to simpler methods, resulting in stable orientation estimates which remain fairly constant up to a distance of 25 m.

Methods have been applied in ***realtime system integrations*** on-board of experimental vehicles and tested in complex, real-world traffic scenarios. Visual context cues are deployed to gain an improved VRU path prediction. In particular, head and body orientation estimates are used to anticipate the behavior of a pedestrian by modeling situational awareness within a context-based Switching Linear Dynamic System. System components are described and influences on the vehicle intervention strategy are pointed out for the pedestrian case. Based on real world test sequences it can be confirmed that the prediction horizon can be increased up to 1 s without increasing the false alarm rate. A demonstration design has been worked out to present the system in an understandable way.

Data annotation and management are indispensable components for the development of machine learning applications. Accurate and correct data annotation has a direct influence on the quality of machine learning results. Furthermore, a data management and provisioning process is needed for handling the large amount of data needed to train complex models. Two software tools are presented to gain an efficient data annotation and management process. The tools have been used for all data annotation tasks in this thesis and have been shared with partners of public research projects.

The thesis completes with a conclusion of the individual chapters and overall insights. Various findings are discussed in relation to each other, and directions for future work are put forward.