Robust multiple hypothesis based enumeration in complex traffic scenarios

Levente Lipták, László Kundra, Dr. Kristóf Csorba

The measurement of the traffic load on the road network is essential for calibrating the traffic lights and designing future developments. One of the first published solutions is dated back to 1937, in which a strip through the road was used to count vehicles, with an electronic device attached at the end. Nowadays this work is still done by so-called enumerators, human traffic counters. This is due to the very strict precision requirements, and the handling of the complicated traffic situations. Our aim is to ease their work with a high accuracy vehicle counter software[1][2]. The videos are produced by cameras placed on street furniture, and processed with Open Computer Vision Library (OpenCV) providing statistics of the vehicle types and their motion path. The application has to handle varying weather conditions, the continuously changing lights, and shadows. The cameras are deployed in new junctions every day, lead to different viewpoints at different junctions, and diverse traffic situations. The ideal camera position would be high above the center of the junction, in top view, avoiding any overlapping. Their positions are limited to the reachable locations, which are usually not ideal for image processing, and lead to overlapped vehicles. As the industry steel employs enumerators in many cases, the recorded videos have often low quality which is sufficient for humans, but contain lots of noise for automatic processing. Because of this our system has to be prepared for very low quality input as well.

The videos are processed with multiple methods to provide the best results. Detections of slow or stopping vehicles are poor using only motion based methods due to the background model contamination effect merging vehicle images into the background model. However, this is a simple task for methods based on easily recognizable unique shapes (Maximally Stable Extremal Regions) or corner points (Shi-Tomasi Corner Detector). In the case of a fast vehicle their performances are reversed. Fusing together information extracted from the image simultaneously is not a simple task, because detections are described by different data types, and also have dissimilar noise to be eliminated. Though the produced set of information is complex, it is necessary to define the criteria of similarity to be able to pair identical objects on consecutive frames. Furthermore the system should be able to handle a situation containing overlapping vehicles moving on the same or on different paths.

To reach the highest accuracy, multiple hypotheses are created in the more complicated situations, increasing the probability that they include the one representing the correct situation. More of them could be useful, if some parts of the movement curves of multiple vehicles are joined. However, some of them are useless, and they have to be filtered out in possession of more information. Erroneous hypotheses could be in the result set too (for example pedestrians and wind-blown objects), which also have to be filtered. This paper is presenting the conclusion of 2 years of research, aiming at main architecture, hypothesis building, and filtering of the duplicates and improbable hypotheses based on a physical model.

References

[1] V. Kastrinaki, M. Zervakis, K.Kalaitzakis: A survey of video processing techniques for traffic applications, In Image and Vision Computing, Vol. 21, Elsevier, 2003. pp. 359-381.

[2] Norbert Buch, Sergio A. Velastin, James Orwell: A Review of Computer Vision Techniques for the Analysis of Urban Traffic IEEE Transactions on Intelligent Transportation Systems (Volume:12 , Issue: 3), 2011, pp 920 - 939