Information Fusion for Cooperative Vehicles

Karin Tischler, Christian Duchow and Britta Hummel

Institut für Mess- und Regelungstechnik, Universität Karlsruhe (TH)
D-76128 Karlsruhe, {tischler, duchow, hummel} "at" mrt.uka.de

Abstract: The cooperative perception by an inter-vehicle network promises a multitude of improvements for advanced driver assistance systems. First, for the simultaneous estimation of the vehicle ego pose and the road network infrastructure, we propose a local, highly model-based fusion architecture for digital map and video information. Second, to extend the vehicle's field of view, we include remote information concerning the state of each participating vehicle and its object detections. For the object-level fusion of these detections a centralized tracking process by Kalman filtering is employed. The algorithms are evaluated in simulated vehicle network scenarios which are based on real sensor data.

1 Introduction

The field of information fusion has a large amount of applications in driver assistance systems. Most systems require an extensive knowledge and understanding of the vehicle's environment which is complex and highly dynamic. The perception of an individual vehicle is limited by its sensor field of view and additional occlusions e. g. by other vehicles, the road curvature or buildings. Therefore, inter-vehicle data exchange promises great improvements extending the observable domain to inaccessible regions. Each vehicle acts as sensor and fuses its information with the perception of others to an enhanced environmental description. A precondition for the joint representation is a precise estimation of the vehicle pose. Fusing a digital map with video information, we additionally obtain a meaningful description of the topology and geometry of the relevant roads. The joint description is a prerequisite for cooperative driving with the intent of cooperative decision making and path planning. This may lead to a more efficient utilization of the road capacity.

2 Architecture for the cooperative network

For a cooperative perception, multiple vehicles have to be equipped with an inter-vehicle communication, the localization in a joint coordinate system and environmental sensors.

A reliable communication between multiple vehicles is an important precondition for the cooperation. Due to high velocities and changing communication partners, a dynamic

ad hoc network is necessary. For a joint environmental perception, the information of all other vehicles in communication distance shall be incorporated in the tracking of own data to get a more complete understanding. By a transfer of remote data to other vehicles, the network could be enhanced to areas beyond the own communication radius. In order to avoid perception loops in a complex network and hereby induce self-fulfilling prophecies, fused remote data is not retransferred at the moment. Based on the joint perception, some vehicles could decide to build a group for cooperative driving manoeuvres. In this case, a lossless communication has to be ensured. They have to compare and match their descriptions to get an identical interpretation of the traffic situation.
For a useful exchange of moving object data, the transmission latencies have to be less than the measurement rates. Sufficient technology standards are still missing, but Wireless LAN (IEEE 802.11b) already enables first tests on communication within multiple vehicles, partial with infrastructure support [1]. In our work, the communication between multiple vehicles is supposed to work properly without a noticeable delay.

Besides sensors to register its own state, a vehicle may be equipped with environmental sensors, e. g. video or radar, and use multiple fusion levels to perceive and describe its environment. In the exchange between different vehicles the data rate is limited. For the inter-vehicle fusion, we interpret the resulting object detections of each vehicle as input from distributed sensors. The transmitted data includes the detected objects and necessary sensor characteristics, like the current field of view and measurement uncertainties.

The fusion of perceptions from distributed sensors needs a spatiotemporal alignment in a joint frame. The common description should include the uncertainties the of vehicle localization and object measurements. In [3], we introduced the basic transformations for the representation of uncertain localization data in arbitrary Cartesian earth-fixed and vehicle coordinate systems. Choosing the vehicle coordinates for the joint representation leads to an accurate representation of own measurements whereas for remote object measurements the localization uncertainty of the ego vehicle is added to that of the remote vehicle. It is worth noting that, if each vehicle incorporates the data of other agents, the resulting descriptions will not be identical. Inconsistencies are inevitable due to different uncertainties, the delay in the processing and communication (obsolete data are discarded) and different tracking decisions.

The vehicle's state including position, orientation and velocity in a world frame is provided by a Global Positioning System (GPS). It is enhanced by a map matching and the fusion with video information presented in the next chapter. The GPS also serves as timer based on Universal Time Coordinates (UTC) to synchronize the computer clocks and therefore provide accurate time stamping for the detections.

3 Fusion of digital map and video information

Cooperative perception and driving requires an adequate description of the surrounding traffic infrastructure, e. g. roads and junctions, and a good position estimation with respect to this infrastructure. Furthermore, information regarding the topology and geometry of a junction in front of the vehicle is useful for the tracking of vehicles and other objects. For the simultaneous estimation of the road network geometry and the vehicle pose, we propose a fusion architecture for digital map and video information.

The problem is formulated as a set of hypothesis tests. A digital map is used to form *a priori* information regarding the general topology of the relevant junction. The subsequent, detailed topology estimation incorporates expert knowledge about allowed or likely junction configurations together with a set of possibly detected video features, as for example arrow markings or the number of lanes. The intermediate result is a set of road or junction topology hypotheses. Expert knowledge and the information of the digital map are then again used to deductively infer geometry hypotheses. The large amount of information in various steps of this process necessitates an abstract conceptual knowledge representation. Therefore, we have designed a *Roads&Junctions* domain ontology. The design of the topology estimator is part of our ongoing research.

The position estimation from GPS and map matching is refined by sampling the space of possible positions around the original position estimation. Each pair of a position hypothesis and geometry hypothesis constitutes a hypothesis test as follows: We select the feature of local orientation as test feature — it is possible to extend the approach to include more features. We calculate the areas in the image in which we expect the lane markings or the curb to be. These areas are computed from the geometry hypothesis and the position hypothesis in question. We also calculate the expected local orientation in the image. The observed local orientation is computed as in [2]. We only consider regions with a significantly high ratio of the two eigenvalues of the grey value gradient covariance matrix (cf. [4]), which additionally lie in the estimated ground plane.

We then formulate a probabilistic model for the difference between expected and observed orientations. If the image, the geometry hypothesis and the position hypothesis do not correspond, we model a uniform probability density distribution for the differences in orientation. If the image, the geometry hypothesis and the position hypothesis match, we expect the differences in orientation to be generally small. We choose a probability density distribution to reflect this expectation. From this probabilistic model, we derive measures to indicate which geometry hypothesis together with which position hypothesis matches the image best. The resulting position hypothesis and geometry hypothesis is a refinement of the GPS position estimation and an estimate of the junction in question. Details and initial results of this approach can be found in [2].

4 Fusion of detections on object level

To fuse object detections from multiple vehicles, we have chosen a centralized tracking algorithm where all measurements are transformed in the same frame first. With the tracking all objects are directly fused in a joint environmental description. Depending on the application, each vehicle may process its data in world or vehicle coordinates. For a local navigation decision, the vehicle frame is preferable: the relevant region in vehicle's front has the highest accuracy. For the cooperative driving we need a global representation which has the advantage that the data of all agents are transformed in the same way. This reduces the inconsistencies in the descriptions of multiple vehicles. In this work, we will use the world frame, but our methods may be applied to a vehicle-centered representation as well.

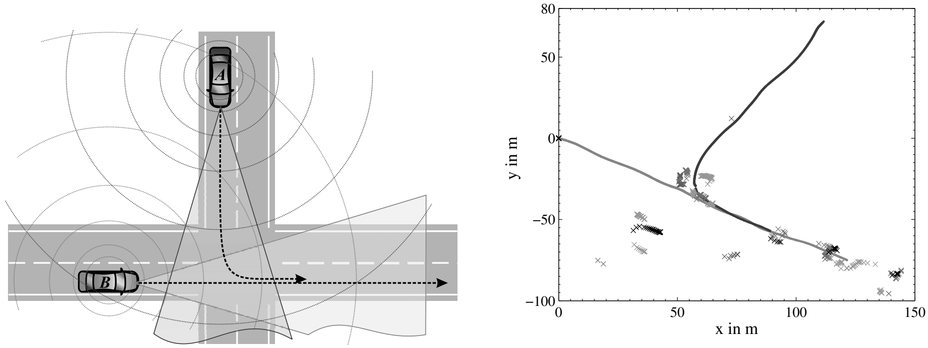
For the tracking we use multiple Kalman filters and the constant acceleration model that is considered to be sufficient for the evaluation of our system. Later adjustments, especially for a better handling of stationary objects, may be done e. g. by an interacting multiple model filter. The association of a measurement to the corresponding object prediction (with position and velocity) is done in a nearest neighbor decision according to the Mahalanobis distance as described in [3].

The measurements of different vehicles are supposed to be not synchronous and remote data get additional communication delay. Therefore, measurements will update the Kalman filters according to their acquisition time stamps. The filter deals with changing prediction intervals that are shorter than the measurement interval of each single sensor. Detections of the same object by multiple vehicles will be fused in the corresponding Kalman filter track. If an object is detected by multiple vehicles, the track will get more reliable. The system has to be completed with a plausibility check especially to notice inconsistent detections in overlapping measurement areas.

5 Simulation scenario and evaluation

Using real measurement data sets from our test vehicle, we create a traffic scenario as depicted in Fig. 1(a). The vehicles A and B are supposed to be able to communicate in an ideal way. The object measurements are required by the radar sensor of our experimental vehicle that detects single static objects as well. The simulated scenario serves as the basis for the tests of the various steps of our fusion approach.

In this particular scenario, A and B are approaching the junction on different roads, where B has the right of way and A wants to turn left. By data exchange, both vehicles communicate their ego state and detections. This kind of information could be used within an upcoming driver assistance system to adjust the deceleration of A earlier to give way to B at the junction and then turn left without stopping. Such single behavior according to coordinated environment information is a first step towards cooperative driving which means a coordination of the behavior. Additionally, A gets information about the oncoming traffic of B and vice versa by exchanging their radar-based object detections. Without distributed sensing these objects would have been beyond the current field of view. Thus, the sight of both vehicles is significantly enhanced.
Fig. 1(b) displays the result of the centralized tracking process that incorporates the vehicle measurements and all object detections in one global representation. For the sake of reliability, only tracks with a limited covariance are retained. After vehicle A turns on the road of B, the ego data of vehicle B is fused with the corresponding object detections from A's sensors, which yields a more reliable estimate.



(a) (b)

Figure 1: Fused position tracks of vehicles and object data (b) in the junction scenario (a).

6 Conclusions

With the presented fusion approach, we obtain a comprehensive description of complex traffic scenarios including knowledge about the traffic infrastructure from a digital map and video. The centralized fusion enables a tracking of information from multiple vehicles in a joint frame. In the selected traffic scenario, the tracking algorithm has shown the desired output of the cooperative perception.

Acknowledgement

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the scope of the Transregional Collaborative Research Centre on Cognitive Automobiles (SFB/Tr 28).

References

[1] B.-W. Chuang, J.-H. Tarng, J. Lin and C. Wang, *System Development and Performance Investigation of Mobile Ad-Hoc Networks in Vehicular Environments*, Proc. IEEE Intelligent Vehicles Symposium, Las Vegas, pp. 302–307, 2005.

[2] B. Hummel, Z. Yang and C. Duchow, *„Kreuzungsverstehen“ – ein wissensbasierter Ansatz*, Schwerpunktheft Fahrerassistenzsysteme, it - information technology, Jan. 2007 (in German).

[3] K. Tischler, *Sensor Data Fusion for Cooperative Perception with Multiple Vehicles*, Proc. Cognitive systems with Interactive Sensors, Paris, 2006.

[4] J. Zhang, H.-H. Nagel, *Texture-based segmentation of road images*, Proc. IEEE Intelligent Vehicles Symposium, pp. 260–265, 1994.