2012 15th International IEEE Conference on Intelligent Transportation Systems
Anchorage, Alaska, USA, September 16-19, 2012

Cooperative Multi Sensor Network for Traffic Safety Applications at Intersections

Michael Goldhammer, Elias Strigel, Daniel Meissner, Ulrich Brunsmann, Konrad Doll, Klaus Dietmayer

***Abstract*— To significantly reduce injury and fatal accidents smart intersections equipped with sensors and communication infrastructure have been proposed. In this publication a novel multi sensor network to perceive the intersection environment is presented. Based on an intensive analysis of accident scenarios in Germany the system was designed to address 75 % of all severe and lethal accidents. 14 laserscanners, 10 cameras, signal phase tapping and an I2V communication unit have been installed at a public intersection in Aschaffenburg, Germany. By using computer based field of view modelling the sensor positions are carefully selected to avoid occlusions. Thus, the infrastructure perception system provides a bird's eye view. Our experiments show that spatial and temporal alignment of sensor data is achieved. We also demonstrate that a part of the sensor network, a calibrated stereo system, allows 3D coordinates in the field of view region of the cameras to be determined with an accuracy of 30 mm.**

I. INTRODUCTION

One of the most challenging locations for drivers are intersections. Accident scenarios at intersections are complex due to the number and variety of road users, intersection layout, speed range and the different directions from which traffic may approach. In 14 European countries almost 61,000 fatal accidents occured at intersections in the years 1996 - 2004 representing 21 % of all traffic accident fatalities [1]. Slightly more than 20 % of all fatalities occurred at intersections year upon year during the last decade, the portion being almost constant [2]. An even higher percentage can be observed when examining injury accidents in 27 European countries: 43 % of them happened at intersections in 2004 [2]. Similar numbers can be obtained for the United States of America: 21 % of all fatal crashes occurred at intersections in 2009 [3]. Since serious accidents occurring in comparatively simple driving scenarios are increasingly avoided or at least mitigated by means of advanced driver assistance systems (ADAS) already available in the market place (ESP as a means to avoid driving accidents is an example), it is to be expected that the percentage of serious intersection accidents (which occur in more complex intersection accidents) will continuously rise in the future.

Within those complex scenarios, extensive traffic density highly restricts the field of view (FOV) of drivers and vehicle based perception systems. To cope with this restriction smart intersections equipped with sensors and communication infrastructure have been proposed. If the positions of the sensors are carefully selected almost all occlusions can be avoided.

Several intersection systems, which observe the environment of intersections for different purposes, have been proposed. The Cooperative Intersection Collision Avoidance Systems-Stop Sign Assist (CICAS-SSA) program [4] in the United States of America uses sensors at intersections to determine unsafe conditions. The intersection approaches are observed by radar sensors and the central area by laserscanners. In addition, an infrastructure mounted driver interface was designed to warn the driver against crossing vehicles. The system was not designed to detect pedestrians or observe complex driving situations. The researchers of SafeSpot, subproject INFRASENS [5], use cameras, laserscanners and RFID-systems to detect dangerous situations. But especially the sensors for the environment perception at intersections have a low mounting position. Thus, they are prone to occlusions. A solely laserscanner based intersection perception system was developed in Intersafe2 [6]. Moving objects in a selected part of an intersection are recognized and tracked with high mounted laserscanners. No cameras are installed to use the texture information.

II. KO-PER INTERSECTIONS

The major contribution of this work within the Ko-PER project [7] is a multi sensor network at the intersections using cooperative perception strategies. The sensors monitor the surroundings to generate a bird's eye view of the current scene and get an exhaustive and reliable model of all traffic participants including pedestrians, bicyclists and motorcycles. Cameras and laserscanners are mounted more than five meters above street level to determine the class and state information of all moving objects at the intersection. The cameras enable the intersection system to use texture information to improve classification and tracking results. This information combined with a highly accurate map as well as signal phase information is communicated to approaching vehicles using the simTD technology [8]. The vehicles can use this information to analyze the current situation and react appropriately.

In addition, the recognition of pedestrians' intention is a field of research in the Ko-PER project. Latest European accident analysis indicate an increase of pedestrian fatalities as a proportion of all fatalities [2]. The knowledge of the intention of the most agile road users should enable the

M. Goldhammer, U. Brunsmann, K. Doll are with Faculty of Engineering, University of Applied Sciences Aschaffenburg, Aschaffenburg, Germany michael.goldhammer@h-ab.de

E. Strigel, D. Meissner, K. Dietmayer are with the Institute of Measurement, Control, and Microtechnology, Ulm University, Ulm, Germany elias.strigel@uni-ulm.de, daniel.meissner@uni-ulm.de

system to predict the trajectory of pedestrians more reliably and to warn drivers early.

In total, the intersection system has been installed at three test intersections with different scope of equipment. Two of them are situated at non-public test areas for the analysis of special and repeatable scenarios and one at a public intersection in Aschaffenburg, Germany, allowing the evaluation of real world scenes. This publication focuses on the installation in Aschaffenburg, because it incorporates the amassed know-how of the other test intersections. Furthermore, the complex intersection layout requires accurate planning and realization of the sensor positions, the cabling as well as sensor timing and synchronization.

The publication is structured as follows: After the discussion of the requirements and analysis for the development of the infrastructure perception system in Section III, the installation of the system in Aschaffenburg is introduced. Along with the design of the sensor network in terms of sensor selection, sensor timing and synchronization as well as cabling, the implementation considering the intersection layout and environment is described in Section IV. Finally, the presently achieved measurement accuracy and temporal as well as spatial alignment of the sensor network is presented in Section V before we summarize the current status in Section VI.

III. REQUIREMENTS AND ANALYSIS

The requirements for the introduced sensor network are derived from an intensive analysis of accident scenarios which occur at intersections. This analysis is based on the GIDAS (German In-Depth Accident Study) database which is the largest in depth accident study in Germany [9]. To create the GIDAS database, 2,000 accidents per year since 1999 in the area of Hannover and Dresden have been analyzed. All in all 29 types of relevant intersection accidents according to the GIDAS classification were identified and combined to 5 scenario clusters. Those scenarios cover 74.8% of severe and lethal accidents. If an intersection traffic safety system is able to deal with those accidents, a large portion of them might be avoided. The five scenario clusters are:

1) A vehicle turns left and conflicts with a car or two-wheeler (bicycle or motorbike) approaching from left.
2) A vehicle turns left and conflicts with a car or two-wheeler approaching from right.
3) A vehicle turns left and conflicts with a crossing car, two-wheeler or pedestrian.
4) A vehicle turns right and conflicts with a bicycle or pedestrian crossing the street.
5) A vehicle conflicts with a car that is driving ahead and which wants to turn left but has to wait in the middle of the intersection.

Fig. 1 illustrates scenario cluster 1 and cluster 4.

The areas which are important to cover with the sensor system and as a consequence the positions of the sensors can be directly derived from the important accident scenarios described above. All scenario clusters can be mapped to three of the four intersection approaches. Therefore, the sensors

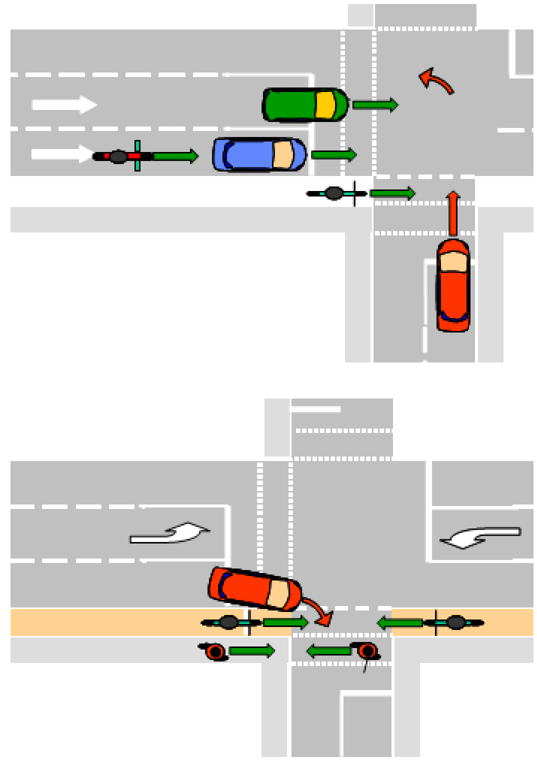


Fig. 1. Two examples of the five scenario clusters which cover 74.8% of severe and lethal accidents at intersections. The red car causes the conflicts.

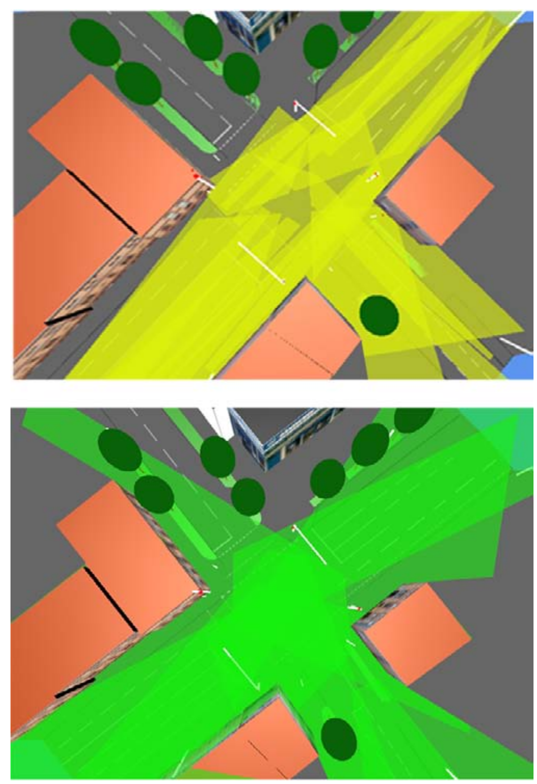


Fig. 2. Simulation of the observed areas by the laserscanner system (upper picture) and by the VGA camera system (lower picture). The FOV of each laserscanner is marked yellow, camera FOVs are marked green.

need to cover only them as well as the central intersection area.

Neither sensors mounted on cars nor drivers are able to resolve occlusions which mostly occur at intersections. One strategy is to mount the sensors several meters over the street level to enable a bird's eye view of the scene. To further minimize occlusions, the position and the FOV for each sensor has to be carefully selected. If one sensor is not able to detect what happens behind an object such as a big truck or a parking car, another sensor should cover that occluded area. To achieve a complete coverage of the important areas of the intersection a simulation environment was generated (see [10]) which enables the planning of sensor positions and orientations.

In order to achive a robust system we combine the advantages of complementary sensors. We use 3D range sensors to detect and track all traffic participants on the intersection. To improve classification of the objects detected by 3D range sensors, we utilize video sensors which provide texture information. Fig. 2 shows the simulation of the observed areas by the 3D range sensors and the camera system.

Fusing the data of the different sensors requires synchronized data acquisition. Computational power to process the data of all sensors in real time has to be provided. A broadcasting unit transmits the information about all dynamic objects, a static map of the intersection topology and the current state of the traffic lights to vehicles that can utilize this information. To ensure that all traffic participants and the intersection sensor network are using the same time base and the same coordinate system, global timestamps for the data have to be generated and all data have to be transformed into a global coordinate system.

The accuracy for determining the position of an object in the context of object detection needs to be about 30 cm to be able to distinguish the smallest traffic participants like children. For intention recognition of pedestrians the spatial resolution has to be in the range of centimeters to resolve small parts of a human body like feet or heads. The temporal resolution needs to be carefully chosen, such that movements can be detected very early. Experimentally we found that 20 ms is a reasonable value for this purpose.

IV. SYSTEM DESCRIPTION

To achieve real-world test conditions, a complex public intersection in an urban area is chosen to install the test setup. A main road with five lanes and a medium traffic volume of 22,000 to 23,000 vehicles per day is crossed by a smaller one with two and three lanes respectively (see Fig. 3).

At the intersection area, the main road has two straight-ahead lanes and a separate left-turn lane for each direction. The smaller roads have one lane per direction and a left-turn lane on one side. There are three crosswalks and a seperate bicycle lane along the main road, which are highly frequented by pedestrians and cyclists due to the location in direct proximity to an university. The intersection also features a trafficload dependent light signal system with dynamic cycle sequences and cycle times. The four corners of the intersection show different occlusions by roadside structures and parking cars, thus the road users' FOV in many common traffic scenarios are limited.

Fig. 3. Public intersection in Aschaffenburg (Germany) used for the test system (Picture is kindly provided by the Wuerzburg Institute for Traffic Sciences GmbH, http://www.wivw.de).

In advance to the installation of the multi sensor network setup, the FOV of all laserscanners and cameras were simulated for different positions and alignments using a 3D modeling software. The model was manually adjusted in order to achieve a spatial resolution in the range of 10 cm for neuralgic points as well as a sensor coverage up to 100 meters into the intersection approaches. Occlusions by stationary objects in the central area of the intersection were eliminated and robustness against occlusions by moving objects, e.g. trucks, was reached by using sensor mounting positions up to 12 meters above road level and overlapping FOV. In most cases the sensors and switch boxes necessary to connect the sensors could be attached to existing infrastructure around the intersection, like light and signal poles (see Fig. 4).

Within the setup 14 SICK LD-MRS research multilayer laserscanners (operating at 12,5 Hz) are used to acquire distance measurements from different perspectives. Thereby, four horizontally mounted scanners are used to cover the central intersection widespreaded (see Fig. 5), while eight vertically mounted scanners observe incoming lanes with high range into the intersection approaches. The calibration allows to transform all laserscanner data from the different sensor coordinate systems into one common intersection system, where it is merged into a close 3D reconstruction of the intersection area. In Fig. 5 the range data of the four laserscanners observing the center of the intersection Aschaffenburg are shown.

The optical sensor network consists of eight monochrome CCD cameras with VGA resolution (standard definition, SD), operating at 25 Hz and and two high definition (HD) cameras (2 megapixels, 50 Hz). Four SD cameras were equipped with wide-angle lenses for a comprehensive coverage of the central intersection area. Three SD cameras observe the incoming lanes and the last one provides a top view of a neuralgic intersection corner, where two crosswalks and a bicycle lane meet (see Fig. 6). This neuralgic area is also covered by the two HD cameras the optical axes of which form an angle of 90 degrees to cope with occlusions. The design especially adresses research on pedestrians' intention detection and provides a wide angle stereo system that allows for determination of 3D-coordinates via triangulation. It is designed for a spatial resolution better than 10 cm and a temporal resolution of 20 ms.

Fig. 4. Different mounting positions of sensor clusters. For most sensors and switch boxes existing infrastructure could be used as mounting position.

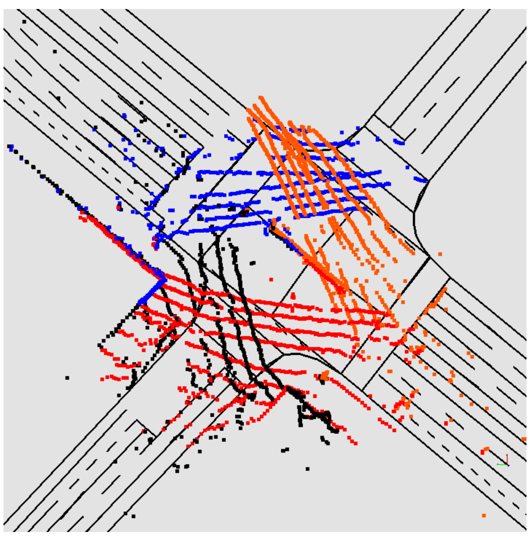


Fig. 5. Aligned range data (beam intersections with the floor) of the four horizontally mounted laserscanners depicted in different colors. Note, that these scanners cover the complete intersection area.

A port to the light signal systems allows to read out the current signal state and a separate stationary GPS reference system generates DGPS correction data. In order to ensure temporal association of the sensor data, a high-precision synchronisation module derives trigger pulses for the different sensor types from the GPS time. Simultaneously, dedicated UTC timestamps are sent to the data processing system, where they are associated to the sensor data. The transmission of the whole sensor data is done via Gigabit Ethernet using cable lengths up to 80 meters.

Bidirectional communication between the intersection system and vehicles (I2V and V2I) is realized by a roadside communication unit (RCU). The unit communicates via

Fig. 6. FOV of the cameras. Cameras 1 - 4 cover the central intersection area from different angles, cameras 5 - 7 cover three of the four approaches. The top view camera 8 and the HD cameras 9 and 10 focus at the neuralgic area of the intersection.

IEEE 802.11p standard (wireless access in vehicular environments) and is compatible to the infrastructure developed within simTD and INTERSAFE-2. Fig. 7 shows an overview of the system architecture.

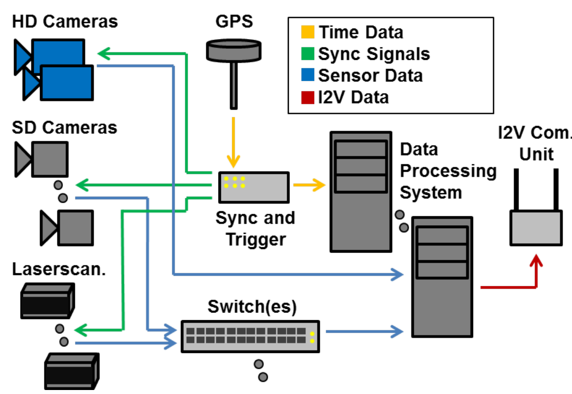


Fig. 7. Schematic overview of the system architecture. Trigger signals are generated from GPS time data and routed to the different sensors. Simultaneously, timestamps are sent to the data processing system and reassociated to the sensor data. After processing, the resulting information can be sent to vehicles via roadside communication unit.

V. EXPERIMENTAL RESULTS

For an evaluation of the spatial and temporal alignment of the sensor subsystems, movements of a test person within the central intersection area were recorded. The person was instructed to move quickly and in alternating directions. Via the laserscanner system the center of the pedestrain's point cloud was calculated and tracked, while the camera systems used the triangulated center of a ball-label above the test persons head. The trajectories were transformed into the common intersection coordinate system. Fig. 8 shows the comparison.

The results allow for an unique data association. We refer the remaining deviations in space which we observe to the method used. Neither the position of the head nor the

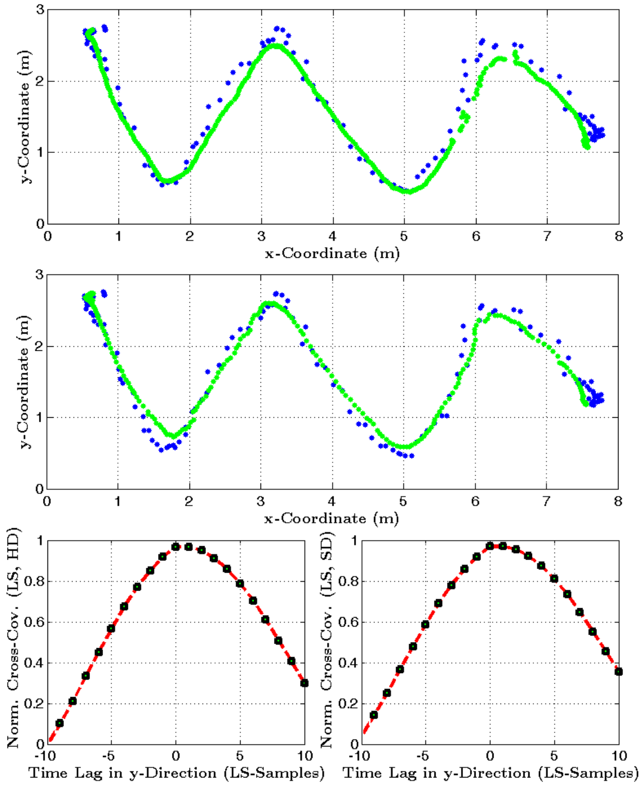
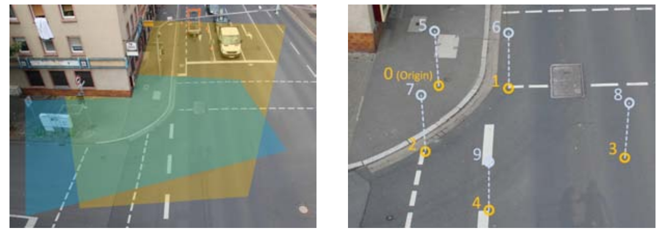


Fig. 8. Upper graph: Cut-out of the evaluated test person's trajectory, projected onto the horizontal xy-plane. Blue points represent the positions of the center of the laserscanner point cloud, green points represent the HD camera-evaluated position of the ball-label. Middle graph: Blue points represent the center of the laserscanner point cloud, green points the SD camera evaluated position. Lower graph: Cross-covariance of the trajectories evaluated by HD cameras (left), SD cameras (right), and the laserscanners (LS) versus time lag in y-direction.

position of the tracked ball have to exactly match with the center of the laserscanner point cloud of the whole body. In total, a standard deviation of 17 cm in x- as well as in y-direction was determined. This is well below the extensions of the detected person. The correct temporal association of the data was tested by an evaluation of the trajectorie's cross-covariance. The peak of the cross-covariance shows a temporal alignment between the laserscanner and both camera subsystems which is better than one sampling period of the laserscanners, i.e. approximately 80 ms.

The calibrated stereo subsystem of the two HD cameras provides a lateral resolution of 1 cm in the space region where the FOVs overlap (see Fig. 9a). For an evaluation of the absolute accuracy, five test points on ground level and five in two meters height, located inside the relevant area of the intersection, were defined (see Fig. 9b). The coordinates of the lower points were determined by the local land surveying office. The upper point coordinates were derived from the lower points by means of a vertically aligned bar. The resulting point positions determined by triangulation using the stereo subsystem showed errors in absolute position ($\sqrt{\Delta x^2+\Delta y^2+\Delta z^2}$) of < 20 mm for the lower and < 30 mm for the upper points. The latency of the stereo frame acquisition, mainly caused by the transmission speed limited by GigE Vision, is in the range of 15 ms.



a *b*

Fig. 9. a) Coverage of the two HD cameras. The overlapping area is used for triangulation. b) Points used for evaluation of the HD camera's stereo calibration.

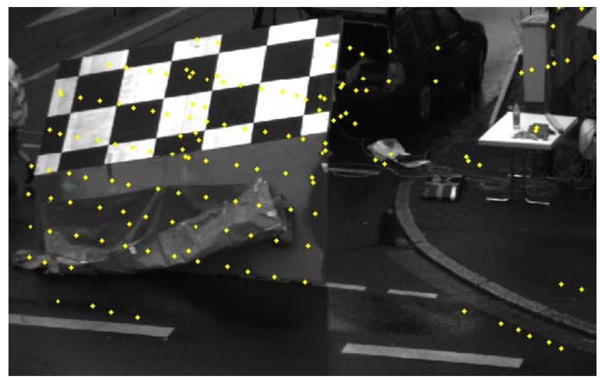


Fig. 10. Object used for cross calibration of a HD camera's picture data and projected 3D data of a single laserscanner.

A visual impression of the quality of the spatial and temporal cross-calibration of the laserscanner and camera subsystems is presented in Figures 10 and 11. Therefore, sequences with static obstacles and moving objects are used. In Fig. 10 the laserscanner data of the static calibration object projected into the image of a HD camera are shown. The image shows that the reflection points of the laser beams hitting the calibration object correspond to the respective pixel data in the camera image. Fig. 11 illustrates the projection of the laserscanner data of a real traffic scene including moving cars and pedestrians into images of two SD cameras. The spatial and temporal alignment is obvious from the matching of laserscanner- and pixel-data.

VI. CONCLUSIONS

In this paper we have presented a comprehensive sensor network operating at a complex public intersection of the Ko-PER-project including time associated data acquisition, data processing and broadcasting to traffic participants via I2V communication. The dedicated setup of complementary sensor systems enables the research on novel sensor fusion based algorithms on real life test data.

The system was adapted to the local surroundings of the intersection by 3D software simulation in order to gain a good balance between high resolution, occlusion resistance and large vision range far into the approaches. Additional information is provided by a DGPS system and a direct

Fig. 11. Projection of laserscanner-data into image of two SD cameras shows the spatial and temporal alignment of the multi-camera-system and the multi-laserscanner-system.

access to the existing traffic light control system. The comprehensive equipment also allows developers to find a minimized sensor sub-sample for their specific use cases, and thus paves the way towards widely available cost-effective infrastructure based systems.

Our future work comprises high-quality detection, classification and tracking of traffic participants in order to get a real time representation of the current traffic situation as virtual traffic map. Furthermore, pedestrian intentions, detected via high-resolution video data, could serve as an additional early indicator in order to avoid traffic incidents with vulnerable road users.

VII. ACKNOWLEDGMENTS

This work partially results from the joint project Ko-PER, which is part of the project initiative Ko-FAS, and has been funded by the German *Bundesministerium für Wirtschaft und Technologie* (Federal Department of Commerce and Technology) under grant number 19S9022B and 19S9022G. We greatfully acknowledge the BMW AG for the provided analysis of accident clusters from the GIDAS in depth accident statistics database.

REFERENCES

[1] European Road Safety Observatory, “Traffic safety basic facts 2006 junctions,” Tech. Rep., January 2007, safetyNet, Project co-financed by the European Commission.

[2] ——, “Traffic safety basic facts 2010 junctions,” Tech. Rep., 2010, safetyNet, Project co-financed by the European Commission.

[3] National Highway Traffic Safety Administration, “Fatality analysis reporting system (fars) encyclopedia,” NCSA Data Ressource Website, 2009. [Online]. Available: http://www-fars.nhtsa.dot.gov/Crashes/CrashesLocation.aspx

[4] Cooperative Intersection Collision Avoidance System, “The design of an optimal surveillance system for a cooperative collision avoidance system - stop sign assist,” CICAS-SSA Report 2, November 2008.

[5] SafeSpot, “Sp2-infrasens-sp infrastructure platform, final report: Results on test and validation,” SafeSpot, Tech. Rep. D2.5.2, 2010.

[6] Intersafe2, “Cooperative intersection safety,” FINAL REPORT D1.2, September 2011.

[7] [Online]. Available: http://www.ko-fas.de

[8] [Online]. Available: http://www.simtd.de

[9] [Online]. Available: http://www.gidas.org

[10] D. Meissner and K. Dietmayer, “Simulation and calibration of infrastructure based laser scanner networks at intersections,” in *Intelligent Vehicles Symposium (IV)*, 2010, pp. 670 –675.