SAE International

Data Fusion Strategies in Advanced Driver Assistance Systems

2010-01-2337
Published
10/19/2010

Michael Darms, Florian Foelster, Jochen Schmidt, Dominik Froehlich and Alfred Eckert
Continental AG

ABSTRACT

Data fusion plays a central role in more and more automotive applications, especially for driver assistance systems. On the one hand the process of data fusion combines data and information to estimate or predict states of observed objects. On the other hand data fusion introduces abstraction layers for data description and allows building more flexible and modular systems.

The data fusion process can be divided into a low-level processing (tracking and object discrimination) and a high level processing (situation assessment). High level processing becomes more and more the focus of current research as different assistance applications will be combined into one comprehensive assistance system. Different levels/strategies for data fusion can be distinguished: Fusion on raw data level, fusion on feature level and fusion on decision level. All fusion strategies can be found in current driver assistance implementations.

The paper gives an overview of the different fusion strategies and shows their application in current driver assistance systems. For low level processing a raw data fusion approach in a stereo video system is described, as an example for feature level fusion the fusion of radar and camera data for tracking is explained. As an example for a high level fusion algorithm an approach for a situation assessment based on multiple sensors is given. The paper describes practical realizations of these examples and points out their potential to further increase traffic safety with reasonably low cost for the overall system.

INTRODUCTION

Although the volume of traffic has noticeably increased within the last 15 years improvements in both driving and transport safety led to a cut by half of traffic fatalities in the EU. Beneath political and educational steps major improvement of active and passive vehicle safety technology led to this trend.

The European Charta for road safety [1] with the goal to cut in half the number of traffic fatalities by 2010 compared to the number of 2004 is driving forward this development. Signees of this charta support this goal in many ways. One of Continental's activities is the integral safety program ContiGuard® [2]. Many of the approaches within this integrated safety approach are based on data fusion techniques described in this paper.

Higher requirements are made on driver assistance systems that should alert the driver in case of dangerous situations or should assist him to solve these: safety related driver assistance systems. A typical representative is the application emergency brake assist [3], which integrates environment information into brake assistance. The quality and value of the information received by different single sensors and the combination through a data fusion approach determines the assistance concept and the increased safety level offered by the system.

The paper presents different data fusion strategies used in driver assistance systems which increase traffic safety. The presented approaches can be found in applications within Continentals safety activities, which show the relevance of the presented concepts in practice. In a first part a general overview over data fusion along with some basic definitions is given. The second part of the paper presents details and applications of the data fusion algorithms and shows their contribution to automotive safety applications.

DATA FUSION

According to Steinberg et al. the process of data fusion can be defined as follows: "Data fusion is the process of combining data or information to estimate or predict entity states." [4]. Examples for entities in the context of driver assistance applications can be objects in the environment like other observed vehicles or abstract objects like the pitch angle of the ego-vehicle.

The overall goal of data fusion applications is to combine data from different single sensors in a way, which combines the strengths and reduces the weaknesses of single sensors. Typically the following aspects are addressed in data fusion setups: Redundancy of information, complementarity of information, enhancement of timing aspects, and reduction of costs (see e.g. [5]). One of the main influence factors for costs is the architecture of the fusion system [6]. Besides that the architecture design strongly influences the possibility of cooperation of partners during the development process (encapsulation/coupling of components).

The Joint Directors of Laboratories (JDL) Data Fusion Model (see e.g. [4], [7]) distinguishes low level processing and high level processing of data. Low level data processing includes track estimation and object discrimination. Track estimation stands for the estimation of states as used for control loops (e.g. position, velocity, etc.). Object discrimination can be subdivided into detection and classification. Detection refers to all actions involved to decide whether an object is present or not, classification refers to algorithms allocating an object to a predefined class (e.g. vehicle, pedestrian, etc.). High level processing includes the situation assessment algorithms and the control of the data fusion process (e.g. allocating resources).

Data Fusion can occur on different abstraction levels of information. Typically raw data fusion, feature level fusion and decision level fusion are distinguished [7]. In a raw data fusion approach minimally preprocessed data is combined. Information can be pixels of imagers for example. One major advantage of this approach is the availability of the complete information of the single sensors. This way the fusion algorithm can be optimized to use all available information. On the negative side the resulting algorithms are rather inflexible as they are specialized for a particular sensor setup. Besides, large amounts of data have to be handled in the system (e.g. transmission of data to a centralized ECU). In a feature level fusion approach features are extracted from the raw data and then the resulting information is combined. This significantly reduces the communication bandwidth between sensors at the cost of an information reduction. The system can be better modularized and more easily be extended using well defined interfaces. In a decision level approach decisions made in earlier steps of the processing chain are combined to a fused decision. This approach can be used for object discrimination or during situation assessment. If the information underlying the decisions is independent an optimal result can be achieved while still maintaining a flexible system that can be extended with further information sources in the future. Examples of the different fusion strategies are presented in the following sections.

RAW DATA FUSION: STEREO CAMERA

In this section a stereo camera system is presented as an example of raw data fusion originating from different sensors, in this case two camera modules that are rigidly attached to each other. The raw data used here are pixels provided by two imagers. Stereo cameras become more and more interesting for safety related driver assistance systems, as they are capable of providing distance and height of arbitrary objects directly by fusion of two images. A monocular camera in contrast can provide depth from a single image only using models or specific assumptions about objects in the real world. This characteristic and the accuracy of the distance measurements for ranges up to about 70m for a typical setup in a vehicle make stereo cameras especially attractive for assistance scenarios at low speeds or when short reaction times at close range are required, e.g. in city traffic, traffic jams or for pedestrian protection. On top of stereo-based functions, the sensor system can provide the complete functionality of a monocular camera system, like traffic sign recognition or lane detection. Other sensor technologies that provide direct distance measurements like radar or lidar lack this option.

The computing steps required to generate depth from a stereo image pair are shown in Figure 1. Special care has to be taken regarding image capturing: In order to generate valid depth information, both imagers have to be synchronized such that sensor elements corresponding to the same objects in the world are exposed at the same time, as moving objects will lead to inaccurate depth estimation otherwise. This is particularly important as most state-of-the-art CMOS imagers nowadays use a rolling shutter, where imager rows are exposed one after another rather than simultaneously [16].

Starting from the images, two pre-processing steps are usually required: Conversion from raw pixel data to gray values, and rectification of the stereo image pair, which facilitates depth computation.

Depending on the color filter array (CFA) used, a conversion step will be necessary to convert the raw pixels to gray values (gray conversion). This step can be omitted if no CFA is used at all. Popular choices for CFAs are Red-Clear or Bayer pattern, which allows for full color reconstruction by applying red, green, and blue filters on the sensor elements. As gray conversion depends highly on the CFA, we refer the

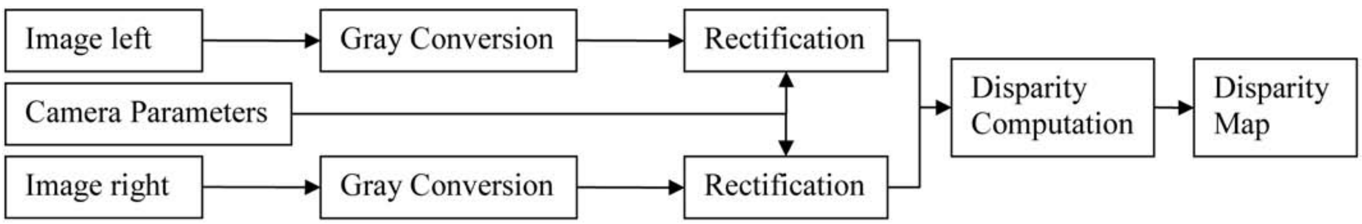

Figure 1. Stereo system overview: Raw images are converted to gray scale and rectified using intrinsic as well as extrinsic camera parameters. Rectified images are provided to the disparity computation module to facilitate correspondence search. The output is a disparity map containing depth information for each pixel.

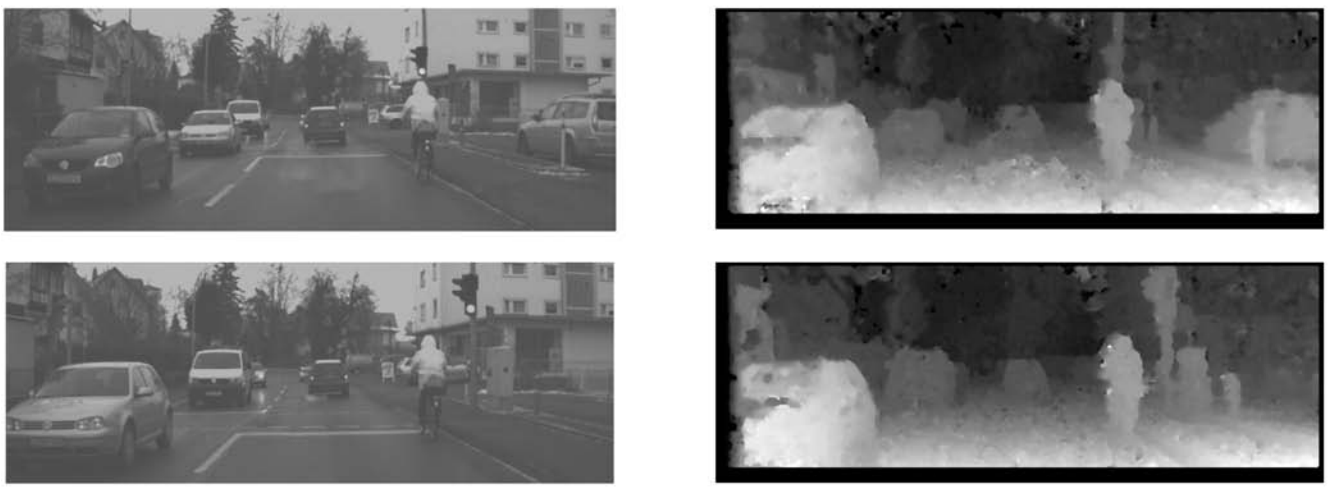

Figure 2. Disparity Map. Left: Images of traffic scenes captured in Lindau (Germany). Right: Disparity Images: The darker the pixel, the smaller the disparity (larger the distance).

reader to the literature. Details on Bayer to gray conversion can be found, e.g., in [8].

In order to compute depth information from the stereo image pair, an algorithm is required that matches object features (e.g., corners) in the left image to features in the right image, thus establishing correspondences between them. Using these correspondences and the stereo camera parameters (intrinsic as well as extrinsic ones), depth can be computed for each feature directly by triangulation (see [9] for details). This would lead to a sparse depth map of the captured scene, which is usually not sufficient for automotive applications.

Mainly, the goal of rectification is to simplify correspondence search. From computer vision it is known that stereo images are related by the epipolar constraint [9], which, amongst other things, states that corresponding points lie on epipolar lines, which all intersect in the epipole. Therefore, it is sufficient to perform correspondence search along these lines as opposed to searching the complete image. However, as these lines may have arbitrary slope, pixel interpolation is required for each search. Rectification fixes this by transforming the images in a way such that in the resulting rectified images corresponding points lie on the same scanline. Depending on the camera parameters, rectification is normally a projective transformation [9]. The rectification step is still applied for each image separately, leading to a rectified image pair; as rectification requires known calibration parameters, there is a dependency between the two images.

Disparity map computation is the step where information from both cameras is combined. The term disparity refers to the distance between x-coordinates of corresponding points in left (x_l) and right (x_r) image measured in pixels, given by $d = x_r - x_l$. The disparity d is directly related to depth $z = f \cdot b/d$, where F is the focal length and b the baseline (distance between optical centers) of the two cameras. For each pixel in one image (or both, if required), the disparity map contains the disparity - and therefore depth - for this coordinate. An example of a disparity map with disparities coded as gray values is shown in Figure 2.

Many different algorithms for disparity map computation have been proposed; they differ in density of the resulting map, occlusion as well as sharp discontinuity handling, and robustness against illumination changes. The methods can be classified in local, global, or semi-global ones: A local approach using block-matching by exploiting the advantages of a similarity accumulator resulting in a very efficient computation scheme for consistent dense disparity maps is presented in [10]. Local approaches are fast compared to global ones, on the expense of computational power required for the latter, while problems occur for homogeneous surfaces, which cannot matched properly in a local neighborhood. A more recent very successful algorithm called semi-global matching is presented in [11]. It combines pixel-wise matching with a 2D global smoothness constraint, resulting in accurate disparity maps that are highly dense, and well suited for automotive applications.

From this point on depth information is available for each pixel and can be processed further. Depending on the application this may be segmentation of objects based, depth computation for already existing objects, or a fusion step (e.g., with radar data), for instance based on a grid map that can be updated over time.

FEATURE LEVEL FUSION: RADAR AND CAMERA FOR VEHICLE TRACKING

As an example for a feature level fusion approach the fusion of data from radar with two independent fields of views and a monocular video sensor is presented. While the radar sensor combines by default already far range with near range measurements for known full speed adaptive cruise control (ACC) applications, the sensor fusion with an automotive camera further enhances the assistance functionality. This setup is chosen due to the assumption that radar sensors have appropriate measurement capabilities for longitudinal control functions, but still shortcomings in the lateral quantity recognition. One reason is the point target assumption of typical radar measurements [12]. In combination with moving radar reflection centers on real vehicle shapes this might lead to an inaccuracy of the position and velocity measurements. While the longitudinal velocity needed for ACC applications could be measured based on the radar Doppler measurement, the lateral velocity is more or less calculated only based from the position derivation.

To enhance the ACC functionality in cut-in and cut-out situations for faster pick-up or target release, the lateral velocity estimation has to be improved besides the position measurement itself. Since the used Kalman filter theory assumes zero-mean Gaussian distributed white noise process, improving the lateral velocity means avoiding possible systematic offsets, decreasing the signals variance and to reduce to latency of the estimated signal compared to the ground truth. A stronger low-pass filter for example would reduce the estimated variance, but will increase the latency of the signal in high dynamic situations. If a classification based mono video is used to enhance the tracking performance of the ACC system, no direct lateral velocity measurement is available, whereas the lateral position measurement has a higher accuracy and less systematic errors caused by moving radar reflections of the extended surface of the observed target vehicle.

Feature level fusion of multi sensor data sources is a well established technique to combine data from different sources to be fed into a single assessment function. Even if the theory was not originally designed for automotive applications the techniques of data association and object tracking described in the known radar literature (see e.g. [13]) are meanwhile also 'state of the art' in automotive radar sensors. Besides object tracking over time the Kalman filter could also be used to combine data of different sensor sources in an optimal way. The setup of combining radar with a mono video sensor will result in the fusion of more or less orthogonal measurements. This has to be distinguished from redundant sensor fusion systems where a second sensor is used to monitor and validate the data of the primary sensor. The stand-alone capability of both sensors to estimate the full position including lateral and longitudinal quantities is used for the essential data association of the two individual sensors. Joint probabilistic data association techniques (JPDA) are used for the final association decision needed to feed data into the Kalman filter processing stage [13].

Besides pure object tracking based feature fusion, the results of the camera based pattern analysis are further used to enhance the object classification. Furthermore, the result of the image based vehicle detector is used to support the obstacle identification processing to distinguish between relevant and non relevant objects from the function's point of view (fusion on decision level, see also [15]).

To verify the desired tracking enhancement, measurements with known ground truth data - based on an Automotive Dynamic Motion Analyser (ADMA) reference system [14] - were taken for dedicated single target situations. Based on these measurements the lateral positioning error decreased to one fifth and the latency of the lateral velocity was reduced by one half in high dynamic lateral oscillation maneuvers. Consequently, the expectations concerning the improvement of the pure tracking filter have been fulfilled. Since the pure enhancement of absolute figures compared to a ground truth measurement has no dedicated advantage for the driver an ADAS system, also the resulting function has to be evaluated.

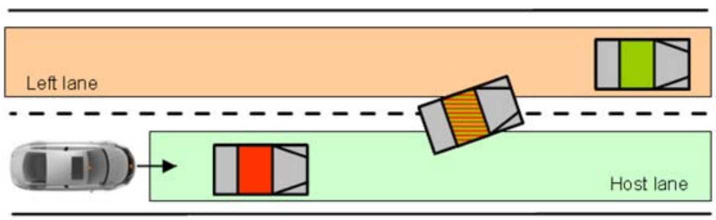

Figure 3. Lane change scenario.

After the tracking module has estimated the object position the ACC situation analysis has to assign lane information to all possible relevant objects by comparing the objects positions with the estimated lanes. Finally the function relevant object could be selected from the set of host lane objects. The earlier a lane change (Figure 3) can be detected the better the function will perform from driver point of view. In Figure 4 measurements of an exemplary situation are given. The upper part of the figure shows the distance to the ACC relevant object. In 17m distance the system picked up a car as target vehicle. This car accelerates and starts with a lane change to the left adjacent lane after approx. eight seconds which is finished three seconds later. The dotted green line indicates the target range over time diagram of the radar only system until the system, releases the target at 11s (point A in the diagram). The solid blue line gives the same information for the fusion system including radar camera feature level fusion inside a single object tracker. It could be easily seen that the target release of the fusion system is faster by 0.7s (point B).

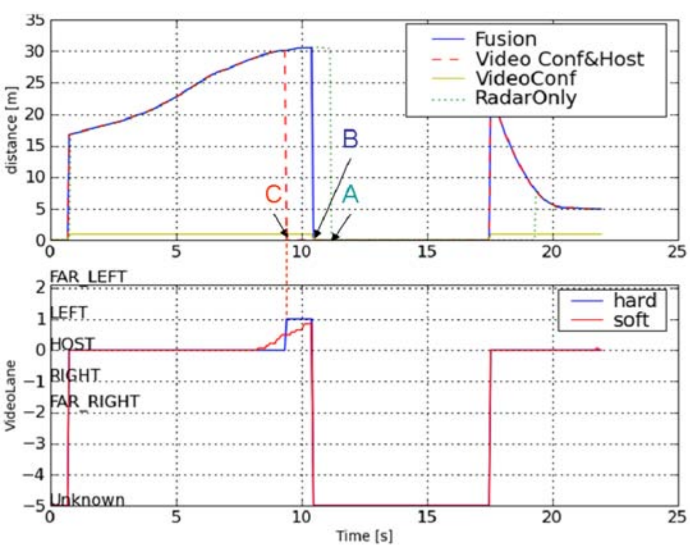

Figure 4. ACC target release for different fusion levels.

Since the target selection is always based on a comparison of object and lane positions, it will further help the function to also use the video based lane detection to update the driving tube estimation. Otherwise, the better quality of the target's position contradicts the lower accuracy of the lane estimation.

For the described measurement example, the utilization of video lane data leads to another second of earlier target release (point C). Especially in city traffic scenarios, where the detection range of the camera system is sufficient, this statement is valid. As shown in the lower part of the figure, the lane change was issued as soon as the middle of the vehicle passed the lane marking. Combing the above shown results we could state that the usage of video processing information shows the expected performance increase to the ACC function in case of lateral maneuvers like lane change scenarios.

DECISION LEVEL FUSION: FUSION OF SITUATION ASSESSMENT HYPOTHESIS

As example for a decision level fusion the fusion of different environment hypothesis is presented in this section. For an advanced driver assistance system application the behaviour of one or more target objects in the current traffic scenario has to be analyzed in a situation assessment step. In this step hypotheses about the current and predicted behaviour of the considered objects relative to the subject vehicle are generated. The necessary hypotheses depend on the requirements of the particular function to be realized. Examples of rather basic hypothesis types - which handle just one individual object - are run-up moving, run-up braking, passing, cut-in, following, ACC, and collision. Functions such as collision mitigation systems may demand for further, more complex hypothesis types in more complex scenarios. This includes considering multiple objects and their relationships, such as convoy scenarios or cut-in prediction. Situation hypotheses can be based on object lists and information on the subject vehicle and driver behaviour.

In a multisensor system the situation hypotheses generation can be performed by each single sensors without using information from other sensors (completely decentralized architecture). The individual hypotheses are then consolidated into a common list. For this, similar to fusion approaches on lower levels, situation hypotheses computed by different interpretation algorithms (sensors) are associated and combined into a result hypothesis. The association is based on the hypothesis type and the kinematics of the considered objects (e.g. distances, velocities). Hypotheses of the same type are associated if they have a matching type and object kinematics. The equality of the hypothesis types is not strictly required however. For instance, a passing hypothesis computed on a radar sensor may be associated with a cut-in detected by a camera (see Figure 5), assuming that the lateral performance of the camera is more reliable than that of the used radar. Similar considerations can be made for other hypothesis types as well. The association confidence can be reflected in a hypothesis probability.

Figure 5. Association of Pass and Cut-in.

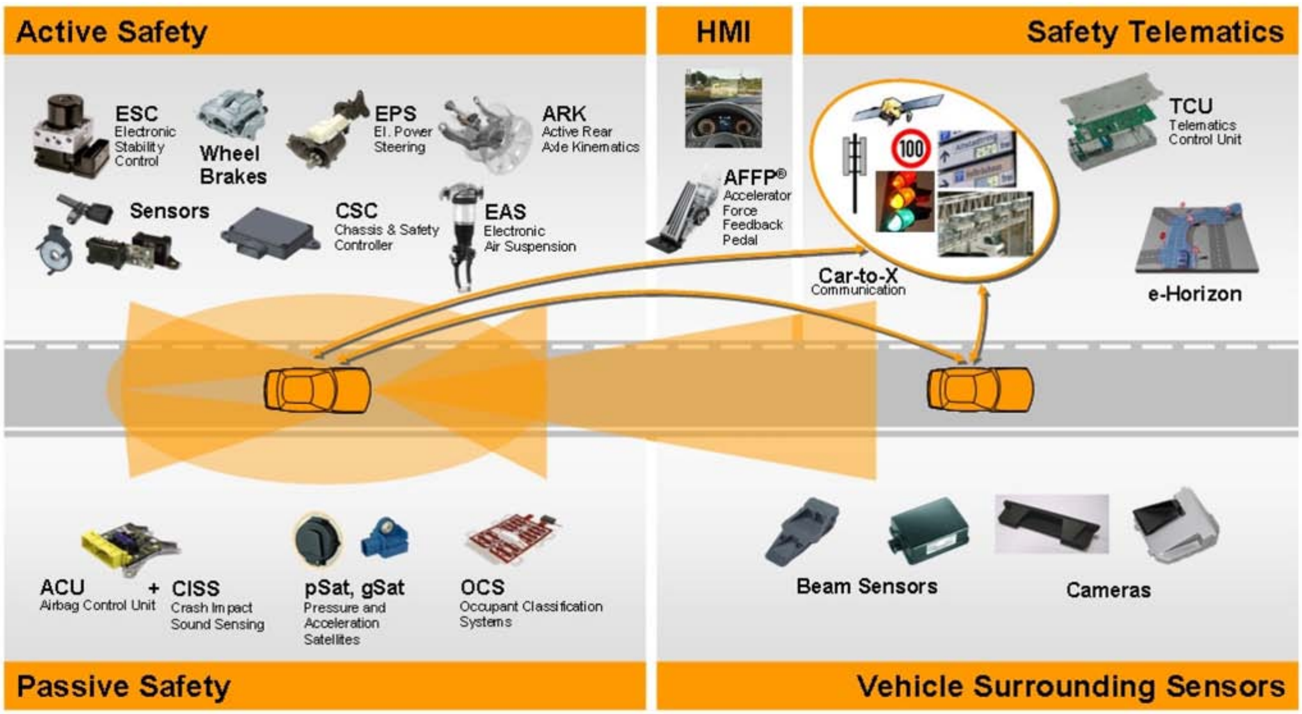

Figure 6. Five cornerstones of comprehensive vehicle safety [2].

This approach requires that all supported hypothesis types can be derived on the involved sensors. An advantage of this approach is the added validation check because in addition to object fusion the situation must be interpreted similarly by different sensors. Thus the robustness of the application is increased.

Each hypothesis type is computed for each candidate object. Then, depending on the particular function, a number of hypotheses are selected for further processing. The selection is based on hypothesis priority and kinematics. For instance, for a collision mitigation system the priority is derived from the potential accident hazard, severity, and probability. The number of hypotheses selected for further processing is small (typ. 3-6) in comparison to the number of objects. This again reduces the computational effort of the fusion in comparison to lower levels. One disadvantage of the fusion of situation hypotheses is that the situation analysis must be computed multiple times. Also this approach implies a particular design/architectural decision for the situation interpretation algorithms.

SUMMARY/CONCLUSIONS

Further development in traffic safety must include - beneath the individual domains active and passive safety - the interconnection of the different domains. Especially the integration of vehicle environment specific information and the man-machine interface will lead to an increase in safety and comfort (see Figure 6).

The development and integration occurs more and more on different platforms and levels of automotive guidance, stabilization, lane guidance and the navigation level including driver assistance. Driver assistance systems should relieve the driver of some of his driving challenges. The herewith achieved reduced driving loads increase the driving comfort and therefore have an impact on the road safety.

The article shows the advantages of sensor fusion algorithms for such driver assistance systems. The approaches presented are used within an integrated safety framework and are already deployed in series applications or are close to it. This shows that there is no single optimal fusion approach; all different fusion strategies contribute to an increase of safety and comfort of the overall system and have to be chosen adequately for the particular problem.

The different approaches lead to different fusion architectures, which have an impact on the cost structure and flexibility of the overall system. As described the raw data fusion concept of the stereo camera strongly influences the hardware design (e.g. timing of the involved imagers, selection of color filter array, etc.). More flexible is the feature level approach, which is best suited for orthogonal information. As described in the paper this is achieved using a radar and a camera sensor. In addition to that the architecture is flexible enough, to include other sensors, as single sensor modules are well encapsulated. The fusion on decision level introduces another layer of abstraction. As described the approach can be used to validate inputs of lower layers and increases the robustness of the overall system in complex scenarios.

REFERENCES

1. European Charta for Road Safety; www.erscharter.eu, 2003.

2. ContiGuard®; Press Release Continental AG 28.02.2008 "Continental extends his ContiGuard Safety System".

3. Baumann, K-H.; Scharneburg, R. and Justen, R., "The Vision of a Comprehensive Safety Concept," SAE Technical Paper 2001-06-0252, 2001, doi:10.4271/2001-06-0252.

4. Steinberg, A.; Bowman, C.: Revisions to the JDL Data Fusion Model. In Hall, D. L.; Llinas, J. (Editors): Handbook of multisensor data fusion. Boca Raton [u.a.]: CRC Press, ISBN 0-8493-2379-7, 2001.

5. Lou, R.C.; Kay, M.K.: Multisensor Integration and Fusion in Intelligent Systems, in: Autonomous Mobile Robots Volume 1, IEEE Computer Society Press, Los Alami-tos, California, 1991.

6. Hall, D. L.: Handbook of multisensor data fusion. Boca Raton [u.a.]: CRC Press, The electrical engineering applied signal processing series, ISBN 0-8493-2379-7, 2001.

7. Klein, L. A.: Sensor and data fusion concepts and applications, 2nd Ed., Bellingham, Wash: SPIE, ISBN 0-8194-3231-8, 1999.

8. Chang, L. and Tan, Y.-P.: Hybrid Color Filter Array Demosaicking for Effective Artifact Suppression. Journal of Electronic Imaging 15(1): doi:10.1117/1.2183325, 2006.

9. Hartley, R. and Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, 2nd edition, 2003.

10. Schmidt, J.; Niemann, H. and Vogt, S.: Dense Disparity Maps in Real-Time with an Application to Augmented Reality. In Proceedings Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002), pages 225-230, Orlando, FL USA, 2002.

11. Hirschmüller, H.:. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Volume 2, pages 807-814, 2005.

12. Foelster, F. and Rohling, H.: "Observation of lateral moving traffic with an automotive radar" presented at International Radar Symposium IRS 2006, Poland, May 24-26, 2006.

13. Blackman, S. and Popoli, R: "Modern Tracking Systems", Artech House, Bosten, ISBN 0-471-18407-1, 1998.

14. GeneSys Elektronik GmbH, "ADMA: Automotive Dynamic Motion Analyzer", http://www.genesys-adma.de", Feb. 2010.

15. Dintzer, P.; Randler, M.; Stephan, T.; Reichmann, T.; Henle, L. and Bauer, M.: Klassifikation stationärer Hindernisse für die Anwendung in Collision Mitigation Funktionen auf Basis eines hochauflösenden Radarsensors, presented at FAS2009, Germany, September 28-30, 2009.

16. Bradley, D.; Atcheson, B.; Ihrke, I. and Heidrich, W.: Synchronization and Rolling Shutter Compensation for Consumer Video Camera Arrays. Conference on Computer Vision and Pattern Recognition Workshops, Miami, FL USA, pages 1-8, 2009.

CONTACT INFORMATION

Dr.-Ing. Michael Darms
Advanced Engineering
Advanced Driver Assistance Systems
Continental, Kemptener Str. 99, D-88131 Lindau
Michael.Darms@continental-corporation.com

The Engineering Meetings Board has approved this paper for publication. It has successfully completed SAE's peer review process under the supervision of the session organizer. This process requires a minimum of three (3) reviews by industry experts.

ISSN 0148-7191

doi:10.4271/2010-01-2337

SAE Customer Service:
Tel: 877-606-7323 (inside USA and Canada)
Tel: 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE Web Address: http://www.sae.org
Printed in USA

SAE International™