Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Early and Multi Level Fusion for Reliable Automotive Safety Systems

Ullrich Scheunert, Philipp Lindner, Eric Richter, Thomas Tatschke, Dominik Schestauber, Erich Fuchs, Gerd Wanielik

Abstract— **The fusion of data from different sensorial sources is today the most promising method to increase robustness and reliability of environmental perception. The project ProFusion2 pushes the sensor data fusion for automotive applications in the field of driver assistance systems. ProFusion2 was created to enhance fusion techniques and algorithms beyond the current state-of-the-art. It is a horizontal subproject in the Integrated Project PReVENT (funded by the EC). The paper presents two approaches concerning the detection of vehicles in road environments. An Early Fusion and a Multi Level Fusion processing strategy are described. The common framework for the representation of the environment model and the representation of perception results is introduced. The key feature of this framework is the storing and representation of all data involved in one perception memory in a common data structure and the holistic accessibility.**

I. Introduction

DRIVEN by the growing demands for road safety systems more and more development activities in the area of intelligent vehicles are focusing on active and/or passive safety applications to build a virtual "safety belt" around the host vehicle in order to warn or respectively protect the passengers as well as vulnerable road users (like pedestrians, cyclists, etc.) in case of dangerous situations or accidents.

Many driver assistance and safety systems ([1],[2],[3]) are addressing this topic on the basis of different sensor configurations, different areas of perception and different data processing methods. Thereby the central challenge for this kind of systems is the accurate perception of the ego vehicle's surrounding with a high reliability and measurement precision. In the majority of cases this task is addressed with data fusion (e.g. multi sensor data fusion of complementary or redundant information, low-level data up to track-based fusion approaches, centralized or distributed information processing, etc.) of several car mounted sensor devices. As the safety-related applications on top require a perception performance of an unprecedented degree of reliability since the erroneous application of an emergency action could be quite counterproductive in terms of road safety improvement and driver acceptance, we propose new fusion methodologies, which go beyond the state-of-the-art track-based fusion.

It is well known that feature extraction processes utilizing a single sensor are inherently sensitive to disturbances, as demonstrated in [4]. The effect of such disturbances can be ameliorated however, by considering a combined processing. Indeed [5] notes that these approaches often provide significantly higher reliability.

Therefore it is our aim to present (extensions to) fusion systems that on the one hand focus on a early interaction within the pre-processing process and a common data interpretation of low-level sensor data (referred to as *early fusion*). On the other hand a fusion scheme is presented that combines the fusion of data on different abstraction layers (signal level, feature level, track level) with back-loop strategies (due to the fact that data components, which belong to one logical object, are scattered and distributed over multiple levels), which will later be referred to as *multi-level fusion*.

Both approaches strongly rely on the management of environmental data. Therefore we describe in this paper our definition of an automotive-oriented environment description and propose a general environmental data structure to store all relevant kind of data (in different abstraction layers) from the process of perception up to the application demands. Furthermore this environmental description can be applied for storing models of the environmental objects (previous knowledge) as well as measured and processed data (perception) for ADAS applications and provides an appropriate data structure for according sensor, object and situation refinement data.

Beginning with a motivation for the proposed environment description chapter II is reflecting the urgent need and the advances of a common and general environmental data structure. The perception memory – a holistic representation of data and knowledge within any fusion process is described. Chapter III introduces into the Multi Level Fusion for data fusion based vehicle detection in the project SASPENCE. The second data fusion approach performing Early Fusion in the project COMPOSE is described in Chapter IV. Chapter V is summarizing and concluding the paper.

Ullrich Scheunert, Philipp Lindner, Eric Richter, Gerd Wanielik are with the Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: {ullrich.scheunert; philipp.lindner; eric.richter; gerd.wanielik}@etit.tu-chemnitz.de).

Thomas Tatschke, Dominik Schestauber and Erich Fuchs are with the Institute for Software Systems in Technical Applications of Computer Science, FORWISS Passau, 94032 Passau, Germany (e-mail: {tatschke, schestau, fuchse}@forwiss.uni-passau.de)

II. MOTIVATION

Automatic perception is not adequately performed with a linear processing chain. The process has to have access to all relevant information about a scene in the real world including the measurements and all intermediate results of the processing. It must be able to revert to any information at any time of the processing. In classical and already known processing schemes the input data of any algorithms are normally lost after the processing has finished.

To meet increased demands in quality and reliability of the results of signal processing new methods have to be developed. These new algorithms are built up to be able to improve the overall processing result by optimizing all intermediate steps from a holistic point of view.

A. The Perception Memory Object

All types of data fusion algorithms and processes deal with an individual kind of data. Every algorithm has certain information as input and results in improved information at the same or different level of abstraction. To store all possible kind of information, perception processes producing a general data structure are needed. Therefore, the Perception Memory Object (PMO) is introduced, which mainly consists of links or references to other PMOs and the perception data itself. Links can represent hierarchical structures like maps by using *Component/ Component of* relations as well as flat structures, e.g. time lines (*Next/ Previous* relations).

B. The Perception Memory

To host and manage the interconnected PMOs a superior instance is required - the Perception Memory, which is a central information and administration unit. Several perception processes can connect to it and retrieve perception data of other processes respectively share their own data.

Hence, the Perception Memory is a holistic representation of data and knowledge of real world scenarios. As a concept for automated perception a structure is taken into account mainly consisting of several Perception Processes and a Perception Memory (see Figure 1).

C. The Extended 4D Environment Model

The Extended 4D Environment Model is also intended to be a holistic representation of the previous knowledge about real world scenarios in this case. In particular this covers feature models representing structure and shape of entities and digital maps as a wide area environment model. Besides that the environment model can also store information about logical relations between different objects and components of objects.

To address the tasks of processing and representing data on several levels as well as model knowledge Perception Memory Objects are used to represent the Extended 4D Environment Model inside the Perception Memory.

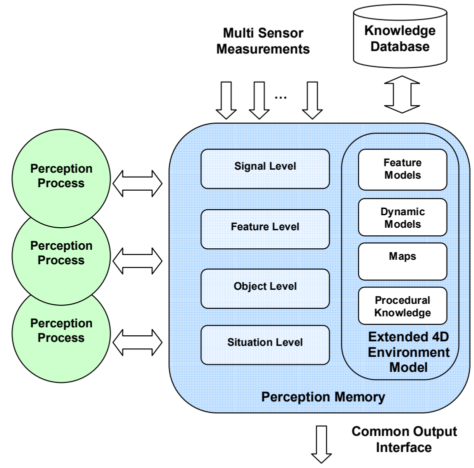

Figure 1: Perception Memory Concept

III. MULTI LEVEL FUSION

In the Multi Level Fusion approach data components which belong to one logical and physical object are scattered and distributed over multiple levels and evidently fused on several different levels too (e.g. signal level, feature level and track level). The chosen level of processing and information fusion is dependent on the object dealt with and the model used. Therefore a certain hierarchical fusion strategy can be defined for any object. This ensures that the tracking of the object is built on inputs from tracked features, untracked features as well as from signal level. By the use of back loops between the single levels, the multi level fusion approach allows adapting the sensor data processing. The multi level fusion management organizes e.g. the use of feature models to define which data from which level should be used to maintain an object track.

Sensor data processing on adaptive chosen levels allows the fusion strategy to be dependent on the actual sensor data and the observation situation of an object.

A. Multi Level Fusion in SASPENCE

The aim of the European PReVENT project SASPENCE is to "develop and evaluate an innovative system able to perform the reliable and comfortable Safe Speed and Safe Distance concept, which helps the driver to avoid dangerous situations." [15]. In this context sensor fusion of image data from visible light grayscale camera and radar data is used to reconstruct the entire road and obstacle situation.

B. Combination of Radar and Image Data

To detect the object of interest – vehicles – information of radar and image sensors are used and combined to yield a more reliable detection result. First the radar data – detections with information about strength, range and angle of the signal – is transformed to the vehicle coordinate system. Doing this the information can be projected to the image space of the camera. With this data we get a region of interest in the grayscale image to which all image processing steps are applied to (see white rectangle in Figure 3).

The radar based region of interest in the image is processed with a local orientation coding (LOC) operator to find distinctive image features like edges and line like structures. The Hough Transform is used afterwards to detect horizontal structures in the data created by the LOC operator. These features were chosen because passenger cars and trucks have dominating horizontal features when they are seen from the back. The characteristic structures are caused by the common design and can be found at the majority of these vehicles. The regions of interest in this context are the intersection of roof, car body and rear window, the license plate and the bottom of the car. These features can be found even under changing environmental and poor visibility conditions (see Figure 2).

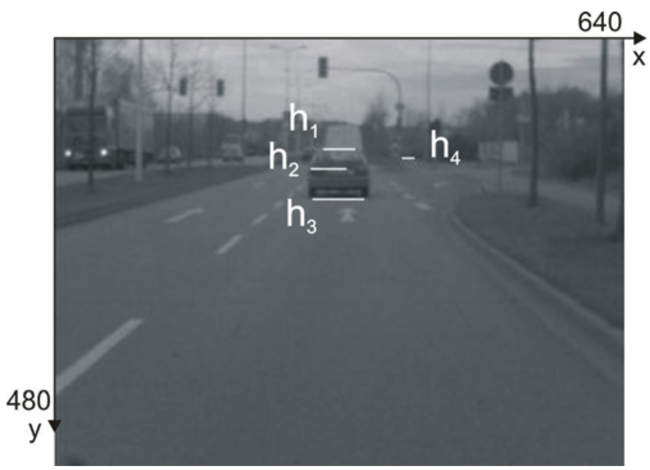

Figure 2: Edge detections in the car region

An accumulation of horizontal lines in the radar based image ROI is treated as strong evidence for the existence of an object with the features of interest, namely a vehicle. The horizontal structures are now combined to create a refined region of interest (blue rectangle in Figure 3) and an object hypothesis. This is in addition to the radar data which can support the image processing with coarse x-y information only due to the physical limitations of this kind of sensors.

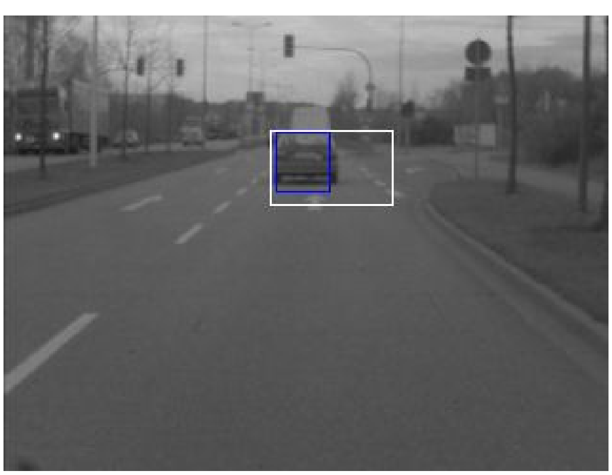

Figure 3: Radar ROI (white) with image-feature based ROI (blue)

C. Combination of features and their evaluation

To continue the approach to treat the detections (horizontal lines) as evidence for the object of interest we are combining the single lines to higher order primitives, e.g. to the bounding rectangle of the combined horizontal lines.

For doing this a hierarchical detection and classification procedure is used to find appropriate higher order structures. The step for finding at first appropriate horizontal line segments and then assigning additional segments to the first horizontal one is shown in Figure 4. To evaluate the line segments for the detection a potential function (evidence function) in Figure 5 is given. For more detailed information about Multi Level Fusion with fuzzy operators and evaluation techniques see [14].

Each line segment is evaluated by assigning a membership-value μ^W to it which expresses the assignment to the class "good horizontal line" in terms of the width of the line (in relation to the object distance measured by the radar sensor). This is done for each line in the image. The combination of the lines and their assigned membership value is done by a modified hamacher operator (see [14]). This results in a combined membership value, e.g. $\mu^{(0)}$ in Figure 4. This is the membership value for the combination of lines; the same can be done with the evaluation of features for the resulting rectangle. A possible attribute for these higher level structures is e.g. the width/height ratio of the rectangle which refers directly to the characteristic physical extent of a passenger car.

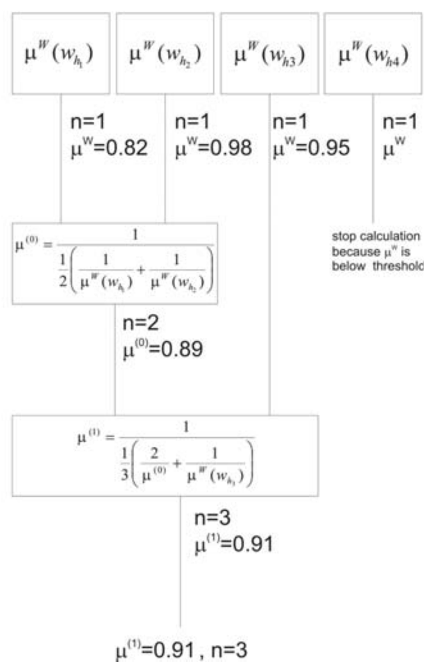

Figure 4: Classification and confidence measures for object recognition example with the modified hamacher operator to calculate $\mu^{(0)}$ and $\mu^{(1)}$ (for line numbers see also Figure 2)

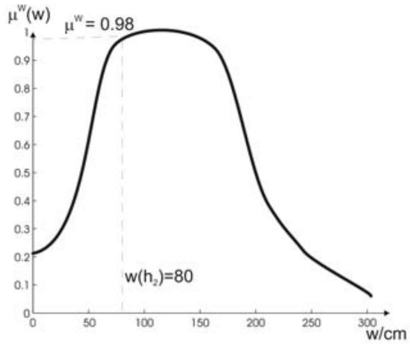

Figure 5: Potential function for the classification of the line segments in the example (here line h_2)

IV. Early Fusion

The main embarrassments of present-day sensor data fusion concepts basically arise from the lack of making (enough) use of redundant sensor information. Having this in mind the fusion methodology subsequently referred to as early fusion like others, aims at remedying these specific deficiencies by taking advantage of the synergetic effect of redundant multi sensor perception on a lower level in combination with a single, sufficiently rich and complete model of the observable environment during the perception instantiation process. In doing so this approach combines pieces of information already on a sooner level compared to state-of-the-art track-based fusion systems. Aim of this method is to interpret unbiased feature input data from different sensors as a whole, using modeled entities of the vehicle's surrounding and to explain all available sensor data with help of these models.

A. Vehicle detection extension for early fusion in PF2

An early fusion prototype, based on radar, laser scanner and FIR camera data, realizing such kind of fusion system has been developed in collaboration with BMW in the COMPOSE sub-project (cf. [6]). Within the ProFusion2 sub-project one aim is the extension of this prototype by a vision system and respective pre-processing to advance the object detection facilities of the original early fusion perception. Thereby particular attention should be paid on the resources of the vision-based algorithms as the early fusion prototype including our extension should still meet the real-time capabilities of the respective BMW demonstrator vehicle (cf. [6]).

This subsequent contribution focuses on the integration of vision-based vehicle detection algorithms into the methodology of early fusion.

B. Generating an object location hypothesis (ROI)

As vehicles can occur at almost any location in the input image at various scales and since it would be very time-consuming to scan the whole camera image for objects, it is necessary to somehow restrict the object detection process to certain regions of interest (ROI), where the detection of a vehicle is likely. Methods to reduce the number of possible object locations include pre-processing the image or taking advantage of other sensors. Examples of the former are edge or line segment detection [7] or symmetry detection [8]. Sensors that can be used to find possible object locations include laser scanners [9], a second camera for stereo vision [10] and radar measurements.

After drawing a comparison of a range of methods for restricting the image domain for the object detection process, we decided to use a combination of edge segment detection and radar measurement projection to narrow the locations in the image due to resource limitations (cf. Figure 6:).

Figure 6: Radar based regions of interest with detected vehicle

C. Vehicle Detection via a classifier cascade

Different types of classifiers, mostly support vector machines and artificial neural networks, have been successfully employed for the task of object detection. The type of classifier we used for vehicle detection described here are boosted ensembles of weak classifiers arranged in a classifier cascade (cf. [11]). The reason for this choice is twofold: This type of classifier can be evaluated very quickly (in combination with Haar wavelet functions as features and an intermediate data-structure as described in [12]) and thus is suitable for an application in a soft real-time environment or with limited resources like ours. Furthermore it also has shown good results for object detection.

For our detection system the classifier cascades were trained on 1062 grayscale images of rear views of cars together with 916 images of street scenes that did not contain vehicles as negative samples. The cascades were trained with a 50% maximum false-positive rate per stage. That is to say no more features were added to a stage, when it eliminated more than 50% of the negative samples in the training set for this stage. Weight trimming on samples was performed the way that samples for one round comprised 95% of the total weight mass.

At each classification step the part of the ROI that is within the search window is presented to the classifier to determine, whether there is a vehicle at this position in the ROI or not.

For all subsequent scans, the search window is scaled by a factor $\sigma > 1$ until its size covers the whole ROI. We found $\sigma = 1.1$ to be a satisfying compromise between speed and detection performance.

In our case the classifier was trained with 30 x 24 pixel images, so the search window was at least this size large. However, since the ROI is not much larger than any vehicle it can potentially contain, the initial size of the search window was such that its width was at least 60% of the ROIs width. This increases the detection process considerably and reduces the number of false positives.

D. Clustering and Filtering of Detections

The process of scanning the ROI results in several rectangular detections grouped around the actual vehicle. This is because in addition to those search window positions that contain the complete vehicle, those positions that are a little off and contain a large part but not the whole vehicle might also result in a positive classification. On the other hand, if only a single rectangle is detected in a certain area of the ROI without any overlapping neighbor rectangles, it is highly likely to be a false detection.

In order to avoid further uncertainty it is therefore necessary to group detected rectangles and remove those without enough neighbors. For any pair of rectangles $r_1(x_1, y_1, w_1, h_1,)$ and $r_2(x_2, y_2, w_2, h_2,)$ in the set R of all rectangles the equivalence $=_{\alpha\delta}$ is defined as

$$\begin{aligned} r_1 =_{\alpha\delta} r_2 \Leftrightarrow & (x_2 \le x_1 + \delta w_1) \wedge (x_2 \ge x_1 - \delta w_1) \wedge \\ & (y_2 \le y_1 + \delta h_1) \wedge (y_2 \ge y_1 - \delta h_1) \wedge \\ & (w_2 \le \alpha w_1) \wedge (\alpha w_2 \ge w_1) \end{aligned} \quad (1)$$

with parameters $\alpha = 1.2$ and $\delta = 0.2$. From the resulting partitioning $R_{/=_{\alpha\delta}}$ all partitions P with less then n members are eliminated: $R'_{/=_{\alpha\delta}} = \left(P \in R_{/=_{\alpha\delta}} \middle| n \le |P|\right)$. The minimum number of neighboring rectangles we used was $n = 3$.

Most of the false detections of the classifier are due to other, non-vehicle objects in the scene. One attempt that was made to remove some of these false positives was using a symmetry model to filter all detections. To determine whether a rectangular sub-region of an image, which has been classified as a vehicle, does indeed contain a vehicle, it is divided horizontally and vertically into three strips. That is, given a sub-region of width w and height h, this results in three horizontal strips of size $w \times (h/3)$ and three vertical strips of size $(w/3) \times h$.

Let r_1 and r_2 be two of these non-overlapping, rectangular sub-regions of the image. Let $S(r_1)$ and $S(r_2)$ be the sum of the pixel values in r_1 and r_2 respectively. Based on [16] we define the symmetric and asymmetric part as

$$f_{sym}(r_1, r_2) = S(r_1) + S(r_2) \quad (2)$$

$$f_{asym}(r_1, r_2) = |S(r_1) - S(r_2)|, \quad (3)$$

then the measure of symmetry between r1 and r2 is

$$F(r_1, r_2) = \frac{2 f_{sym}(r_1, r_2)}{f_{sym}(r_1, r_2) + f_{asym}(r_1, r_2)} - 1 \quad (4)$$

or $F(r_1, r_2) = 1$ iff $f_{sym}(r_1, r_2) + f_{asym}(r_1, r_2) = 0$. For an image sub-region to be considered as vehicle, at least two horizontal strips had to have a symmetry measure above 0.75 and no more than one vertical strip must have a symmetry measure above 0.85.

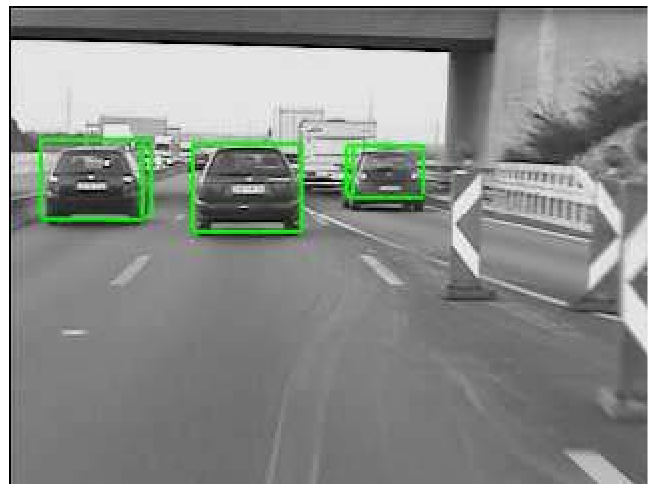

Figure 7: Vision-based detected vehicles in a highway scenario

E. Integration into the early fusion system

Based on the measurements of the vehicle detection method presented so far and of the other mounted sensors new object hypotheses are generated if the respective measurements could not be associated to already existing objects. This process mainly consists of a spatial measurement aggregation in combination with an object model fitting (cf. [6]). The respective measurements of the vehicle detection algorithm consist of a position, a width and height (and in case of a radar supported ROI generation a velocity) component of the detected vehicle, which are used to decide on a new vehicle hypothesis.

In case of an already existing object (hypothesis) a predicted measurement for the vehicle detection system is derived from the respective object model (in accordance to [6]). This process thereby considers besides the position, width, height and velocity aspects also visibility constraints and the orientation of the object in the respective sensor domain. These measurements are matched next via data association algorithms (e.g. GNN, bipartite matching) and the residua are passed to the filter to obtain an updated version of the object's state vector.

F. Performance evaluation of vehicle detection

For the performance evaluation of the actual presented vehicle detection and tracking system prerecorded traffic scenes where annotated with model descriptions of the vehicles that appear in each scene. Then it was recorded how many of the vehicles that appear in each frame of the video sequence were actually detected by the system. Additionally a “safety zone” was defined that covered the area in front of the car, where the detection of preceding vehicles was essential regarding traffic safety issues. All performance measurements therefore include two values, one for detection within this “critical zone” and one global value.

Two statistical figures were calculated from each experimental setup, the detection and the false-positive rate. Let F be the number of frames of a recording, let D be the total number of correct detections in all frames, let E be the total number of false detections in all frames, and let V be the average number of vehicles present in one frame. Then the detection rate calculates $D/(F \cdot V)$ and the false-positive rate is E/F. Results for the presented vehicle detection and tracking (cf. Figure 7) are given in TABLE 1. Details on the generation of reference data for the performance evaluation can be found in [13].

TABLE 1
EVALUATION OF THE VEHICLE DETECTION

Assessment Area	City scenario Detect. rate	City scenario F.P. Rate	Highway scenario Detect. Rate	Highway scenario F.P. Rate
Critical Zone	90.9	0.9	97.6	8.8
Global	84.4	9.8	97.7	9.4

Timing measurements on a Intel Pentium4 (2.8 GHz) system indicate that the whole detection process takes less than 6ms per frame on average which meets our requirements concerning the resources limitation of the early fusion demonstrator vehicle.

V. CONCLUSION

The paper introduces into the common architectural model of the project ProFusion2 consisting mainly of the Perception Memory and the Perception Memory Objects. It presents two data fusion approaches that can be assigned to the fields of Early and Multi Level Fusion. Results for the detection of vehicles in road environments are presented based on the collaboration of ProFusion2 with the PReVENT projects SASPENCE and COMPOSE.

ACKNOWLEDGMENT

The paper describes a part of the results obtained in the ProFusion2 project, which is a subproject of the PReVENT Integrated Project, an automotive initiative co-funded by the European Commission under the 6th Framework Program. Hence ProFusion2 addresses research work of common interest related to sensors and sensor data fusion including a modular architecture for sensor data fusion. The data screenshots (Figure 6 and Figure 7) shown base on the BMW / FORWISS COMPOSE fusion system from the BMW demonstrator and are by courtesy of COMPOSE. The presented vehicle detection system is integrated into the COMPOSE fusion system and uses its graphical interface.

REFERENCES

[1] Scheunert, Ullrich; Cramer, Heiko; Polychronopoulos, Aris; Amditis, Angelos; Wanielik, Gerd; Andreone, Luisa: Multi sensor data fusion for object detection: challenges and benefit. Journal Ingegneria Automotoristica, September, 2002.
[2] Fardi, Basel; Scheunert, Ullrich; Cramer, Heiko; Wanielik, Gerd: Multi Modal Detection and Parameter-based Tracking of Road Borders with a Laser Scanner. IV 2003, IEEE Intelligent Vehicles Symposium, Columbus, Ohio, USA, June 9-11, 2003.
[3] Scheunert, Ullrich; Cramer, Heiko; Wanielik, Gerd: Precise vehicle localization using multiple sensors and natural landmarks. Proceedings of the Seventh International Conference on Information Fusion, pp. 649-656, The 7th International Conference on Information Fusion, Stockholm, Sweden, June 28 to July 1, 2004.
[4] J. B. Romine and E. W. Kamen: Modelling and fusion of radar and imaging sensor data for target tracking. Opt. Eng., vol. 35, no. 3, pp. 659–673, 1996.
[5] S. Majumder, S. Scheding, and H. Durrant-Whyte: Multisensor data fusion for underwater navigation. Robotics and Autonomous Systems, vol. 35, pp. 97–108, 2001.
[6] L. Walchshäusl, R. Lindl, K. Vogel and T. Tatschke: Detection of road users in fused sensor data streams for collision mitigation. In 10th International Forum on Advanced Microsystems for Automotive Applications, pp. 307-314, Berlin, Germany, April 2006.
[7] C. Hilario, J. M. Collado, J. M. Armingol, A. de la Escalera, J. Marques, N. de la Blanca and P. Pina: Multi-resolution image analysis for vehicle detection. In Iberian conference on pattern recognition and image analysis, Vol. 2, Estoril, Portugal, June 2005.
[8] C. Jiangwei, J. Lisheng, G. Lie et al: Study on method of detecting preceding vehicle based on monocular camera. In Proceedings of the IEEE Intelligent Vehilces Symposium 2004, pp. 750-755, Parma, Italy, June 2004.
[9] N. Kaempchen, M. Buehler and K. Dietmayer: Feature-level fusion for free-form object tracking using laserscanner and video. In Proceedings of the IEEE Intelligent Vehicles Symposium 2005, pp. 453-458, Las Vegas, U.S.A., June 2005.
[10] A. Bensrhair, M. Bertozzi, A. Broggi et al: A cooperative approach to vision-based vehicle detection. In Proceedings of the IEEE Intelligent Transportation Systems Congress 2001, pp. 207-212, Oakland, U.S.A., August 2001.
[11] Y. Freund and R. Schapire: A short introduction to boosting. In Journal of Japanese Society for Artificial Intelligence, 14(5):711-780, September 1999.
[12] P. Viola and M. Jones: Rapid object detection using a boosted cascade of simple features. In Proceedings of the Conference on Computer Vision and Pattern Recognition, 2001.
[13] T. Tatschke, F.-J. Färber, E. Fuchs, L. Walchshäusl, R. Lindl: Semi-autonomous reference data generation for perception performance evaluation. Submitted to the 10th International Conference on Information Fusion, Quebec, Canada, July 2007.
[14] Scheunert, Ullrich; Lindner, Philipp; Cramer, Heiko; Multi Level Fusion with Fuzzy Operators using Confidence, The 9th International Conference on Information Fusion, Florence, Italy 10 - 13 July 2006
[15] http://www.prevent-ip.org/en/prevent_subprojects/safe_speed_and_safe_following/saspence
[16] T. Liu, N. Zheng, L. Zhao and H. Cheng: Learning based symmetric features selection for vehicle detection. In Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 124-129, Las Vegas, U.S.A., June 2005.