

IMAGE AND LASER SCANNER PROCESSING AS CONFIDENT CUES FOR OBJECT DETECTION IN DRIVING SITUATIONS

Álvaro Catalá-Prat^a, Frank Köster^a, Ralf Reulke^b

^a German Aerospace Center (DLR), Institute of Transportation Systems, Lilienthalplatz 7, 38108 Braunschweig, Germany - (alvaro.catalaprat, frank.koester) @dlr.de

^b Humboldt-Universität zu Berlin, Department of Computer Science, Unter den Linden 6, 10099 Berlin, Germany - reulke@informatik.hu-berlin.de

Commission V, WG V/5

KEY WORDS: Environment, Object, Detection, Laser scanning, Image, Texture, Segmentation, Fusion

ABSTRACT:

In the area of advanced driver assistance and automation systems knowledge about the vehicle environment is becoming more and more important in order to increase traffic safety. This paper is concerned with the detection and tracking of objects in the proximity of the ego-vehicle while driving on highways. For this purpose, a camera sensor and a laser scanner are used. The processed data of the sensors is then fused at object level in a competitive way. The paper focuses on the generation of object observations by applying the mentioned sensors. In the case of the camera system, an image processing method based on texture information is presented. The texture information is adaptively calculated in order to be independent of the lighting conditions. Taking into account knowledge about the image structure in driving situations, texture segments are classified and object observations are generated. In comparison to other methods, objects are detected independently of any features, model and movement assumptions. For object generation from laser scanner data, a method characterizing detected object contours by means of a shape indicator (long, corner, round, concave etc.) is proposed. Different to other works in this field, in the method presented here explicitly obtains the objects' optimal reference point and the observability of the objects' components. The experiments conducted both with simulated and real data show the plausibility of the methods to be used as cues for an object fusion system.

1. INTRODUCTION

In the area of advanced driver assistance and automation systems the focus is often set on driving safety, traffic efficiency, and comfort. For these purposes, knowledge about the vehicle environment is becoming more and more important, such as safety aspects of the road and road signs, about other vehicles and pedestrians, and about weather conditions. This paper focuses on object detection by means of two cues, a camera and a laser scanner, which are then fused in a further step.

Detection systems are often based on a single sensor. However, using multiple cues in a fusion system offers important advantages, such as increased detection scope, accuracy, reliability, availability and robustness. Since this increases the system's complexity, fusion mechanisms must be applied carefully. Fusion strategies are applied at different abstraction levels. Sensor and feature level fusion is usually done in a cooperative way (e.g. validation through a second sensor). On the other hand, object and tracking level fusion usually follows a competitive strategy. For this purpose, the data is first pre-processed independently in each sensor and then brought together at a common level. The lower level strategies often reach a better performance and availability, while those at a higher level are more reliable (in case of sensor failure), modular and expandable, as well as decoupled from the sensing systems.

In order to fuse at object level, every sensor cue must be prepared to provide object observations. These tasks are independent of each other and of the fusing system, which

increases reliability. Feedback loops are avoided, in order to reduce the negative memory effects of false positives (phantom objects) and outliers.

The aim of this paper is to present methods to process camera and laser scanner data as cues of an object detection system consisting of an object level fusion. Both sensors are integrated in the experimental vehicle ViewCar (www.dlr.de/ts), presented in Figure 1. The applied monocular camera ($f=3.5$ mm objective and $7.5 \mu\text{m}$ pixel size) is used in gray level mode. The applied laser scanner is a profiler with four parallel beams that are pre-processed in the sensor into one reflection point per direction (120° aperture, with angle resolution between 0.125° and 0.5°). The sensors are focused on the frontal area of the vehicle. Desirable properties of object detection cues are: availability (processing rate), confidence (high true positive rate, low false positive rate), (self-) validation, and accuracy estimation.



Figure 1. Experimental platform, the ViewCar, containing a laser scanner and a camera (among other sensors).

The paper is structured as follows. After the introduction, motivation and aims given in this section, the next section deals with related work in the area of object detection. Section 3 gives an overview of the applied system structure and fusion mechanisms, followed by the detailed description of the methods for camera images (Section 4) and laser scanner data (Section 5). In Section 6, the experiments conducted to test the methods are presented and discussed. Finally, Section 7 presents conclusions drawn from the experiments and gives outlook of possible future work.

2. RELATED WORK

Object detection and tracking is a major topic in the area of driver assistance and automation systems. While many systems have been based on a single sensor, an increasing trend to fusion based solutions can be observed. In this section, works related to the extraction of object observations from sensor data are presented. At this level, no further attention is paid whether the extracted observations are used in a single sensor tracking system or in combination with other sensors in a fusion system. While sensors such as stereo-vision, radar, and multi-beam lidar are also used by some systems, this paper focuses on monocular vision and laser scanner based systems.

A number of works use a camera as detecting sensor in driving situations. The methods found in the bibliography can be divided in the following groups. Model based methods are trained to detect target objects, e.g. by means of Haar wavelet features (Viola and Jones, 2001). In this group, objects looking differently than the trained models cannot be detected. A second group of methods is based on optical flow and the segmentation of image regions with similar movement, e.g. (Giachetti, 1998) and (Wohlfeil, 2008). For these systems, static objects and objects moving with the same speed as the ego-vehicle or close to the vanishing point are difficult to detect. As a third heterogeneous group, methods based on object features can be found, which often lean on a given object model (like cuboid or polyhedron). Common features are the object shadow (as a horizontal edge), vertical edges, contour lines, vehicle wheels, and symmetry of the vehicle's rear, as in (Benshair, 2001), (Wender, 2008), (Neumaier, 2007) and (Wohlfeil, 2008). These methods assume that the target objects have certain features and those are visible in the images, which is not always guaranteed. In this paper, a new method based on texture feature and the structure of the texture image is developed and tested. Texture features have the advantage that object regions stand out even though no clear edges, no clear movement, or no expected features might be visible. Thus, the method is appropriate to be used even with low quality images.

On the other hand, several references concerning object detection in driving situations applying a laser scanner can be found in the bibliography. Since the data provided by the sensor already represents 3D points, the two main issues of object detection consist of data segmentation and model fitting. Sensor data segmentation is usually done depending on the distance and relative velocity between consecutive reflection points of the scan – see (Dietmayer, 2001). The fitting to an object model is done in different ways. (Labayrade, 2005) reduces the object to a point for collision avoidance. (Dietmayer, 2001) and (Wender, 2008) consider the two extreme contour points and the closest reflection point to the sensor. Depending on these three points the object is then classified in O-shaped, I-shaped or L-shaped. In (Lindner, 2008), the rotating-calipers-algorithm

is applied to fit the object into a box. This and other authors also process sensor data by means of an occupancy grid algorithm. All these methods do not consider characterizing round or concave contours. In this paper, a new method to detect the object basic shape in a differentiated way is presented, based on the internal contour angles. Besides, the optimal object reference point and the observability of the object sides are calculated.

3. SYSTEM STRUCTURE

The methods presented in this paper are embedded in a multi-layer multi-sensor data fusion system, which has been introduced in (Catalá-Prat, 2008) and (Catalá-Prat, 2008b). In Figure 2, the modular structure of the complete fusion system is presented. The tasks are divided in different abstraction levels (sensor level, object level, and application level), which are explained in the following.

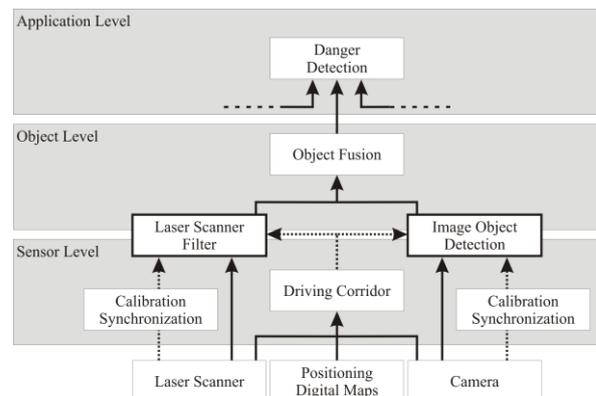


Figure 2. Modular system structure of the multi-sensor multi-level fusion system.

At sensor level, synchronization and calibration issues are treated. The latter is especially important since vehicle movements produce sensor data vibrations and loss of validity of alignment parameters. Additionally and to support the task of object detection, a method to detect the driving corridor has been developed. By means of driving corridor information, the object detection task can be filtered, and the amount of data can be strongly reduced. This means a gain in both processing speed and confidence. The developed method is based on the low level cooperative fusion of camera data (lane markers), positioning and information from digital maps (number of lanes of current road), and laser scanner data (static objects at road boundaries). Some details about this method have been presented in (Catalá-Prat, 2008).

At the application level, a danger recognition based on the statistical detection of atypical events has been developed. At this abstract level, the inclusion of other sources than object data for danger recognition, such as road information, is also possible. This module is explained in (Catalá-Prat, 2008b).

The presented system focuses on the detection and tracking of object information at object level. The observations extracted from camera and laser scanner data – explained in the following – are the input of the object fusion. In the fusion kernel, each object (object hypothesis) is tracked by a filter. In this work, an asynchronous information filter is applied, (Rao, 1993). This

filter can be understood as a reformulation of the well-known Kalman filter, and is appropriate for the estimation of object state based on observations coming from different sensors and with different data rates.

In each filter step, all object hypotheses are predicted to the new observation time, compared and associated to the incoming observations. With this data, the object hypotheses are then updated, and the object list is administrated.

All incoming observations as well as the fused object hypotheses are defined based on the same object model. This represents a reduction of the fusion costs. However, it also implies an increase of the pre-processing complexity of object detection (extraction of observations, including adaptation of uncertainty). The object model applied for object tracking and fusion consists of an orientated cuboid model with a constant velocity, see Figure 3. This model is very simple, which assures fast processing and converging filtering, but is limited against manoeuvring objects.

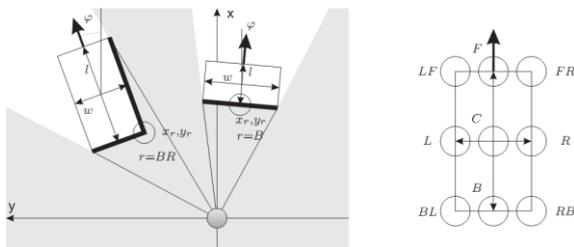


Figure 3. Object model used, including reference, width, length and heading angle. On the right, possible references of an object (L=left, F=front, R=right, B=back, C=centre).

In order to catch observability changes and object manoeuvres, as well as to reduce the influence of noise, outliers and split and merge effects in the observations, further mechanisms have been developed, which are only briefly introduced in this paper and will be extensively presented in future publications. These include the multi-hypotheses association, the variable object reference (Figure 3), the partial observability matrix (with corresponding filter adaptations), and strategies to duplicate and unify object hypotheses. In order to achieve a homogeneous representation, all objects are considered to be aligned to the ego-vehicle (width as lateral dimension, length as longitudinal dimension). Thus, their width and length might be exchanged after analyzing the object movement (in the object fusion).

The advantage of using a logical reference of the object is that complex 3-dimensional objects are represented by a point, and thus the tracking and fusion mechanisms stay computationally less time-consuming. In order to compare two objects (as in the data association), the reference of an object can be changed to any other reference, as long as the corresponding sides are observable. In this process the accuracy of the position must also be adapted by means of propagation of uncertainty.

4. IMAGE PROCESSING FOR OBJECT DETECTION

The first of the methods presented in this paper consists of the extraction of object observations from camera images. The proposed method is based on the extraction of the image texture, the segmentation of the image in regions with the same

texture value and the analysis of the structure of the segmented image. In order to reduce computation costs and the detection of noise objects, this method is only applied on the region of the image within the driving corridor (see Section 3).

Texture features are extracted in different ways in the bibliography: e.g. by means of statistical features (cooccurrence matrix), model based features, geometric features and others. Most of the methods are suboptimal under variable lighting conditions and in perspective images. In this work, the mean gray value in the neighbourhood is taken for every pixel as texture feature, which allows fast computation and shows good results. Due to the varying lighting conditions and traffic situations, the texture calculation must be done adaptively, as shown in Figure 4. Therefore the image region directly in front of the vehicle is assumed to belong to the road and used as reference to set the gray value step and offset (defining thus the asphalt colour).



Figure 4. Adaptive calculation of the texture image (reference area marked in white).

In a further step, the structure of the texture image is calculated in the form of a segment graph. Each segment with a uniform texture value is set as a node and all its neighbours are calculated and saved as connexions in the graph. This expensive algorithm has been optimized, for example by means of the driving corridor. See an example in Figure 5.

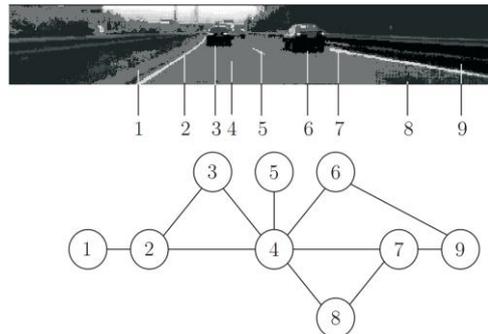


Figure 5. Example of a simplified segment graph to represent the structure of the texture image.

By means of the segment graph and the corresponding gray texture values, an analysis is proposed in order to classify each segment. Possible image segment classes are *road*, *object*, *lane marker*, and *spot*. A rule set is applied iteratively in order to exclude unfeasible classes for each texture segment. The rules are based on the typical structure of a driving situation, as for example: *A texture segment can only be of type object if it is in contact with a lighter road segment*; *A segment can only be of type road marker if it is in contact with a darker road segment*; and so on. Whenever a texture segment only has one possible class, this class is assigned to it. In order to initialize the analysis, the segment corresponding to the texture reference region (see Figure 4) is set as *road*.

After image segment classification, those segments classified as objects are used to generate object observations based on the model presented in Figure 3. Since split and merge effects can appear at the obtained object segments, only nearly horizontal sections of the segment's lower boundary are extracted (x_{i1} , y_{i1} , x_{i2} , y_{i2}). Thus, all objects are assumed to be observed from their closest side to the sensor (reference=B, heading= 0°). In future works, this extraction will be extended in order to consider object lateral sides and object heading.

Under the assumption of a flat road, the observations are transformed from image into road coordinates by means of an inverse resection.

$$\mathbf{p}_r = \mathbf{T}_i^r \cdot \mathbf{p}_i \quad (1)$$

The matrix \mathbf{T}_i^r represents the transformation between image and road coordinates. It consists of the inverse of \mathbf{T}_r^i , which corresponds to the well-known collinearity equations \mathbf{T}_w^i (from world to image coordinates) simplified with the z-component as follows:

$$\begin{pmatrix} x_i \\ y_i \\ w \end{pmatrix} = \mathbf{T}_w^i \cdot \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (2)$$

This equation is adapted to a road point \mathbf{p}_r , where $x_r = x_w$, $y_r = y_w$ and $z_r = z_w = \text{constant}$ (e.g. $z_r = 0$). The obtained matrix corresponds to \mathbf{T}_r^i :

$$\begin{pmatrix} x_i \\ y_i \\ w \end{pmatrix} = \begin{pmatrix} T_{w1,1}^i & T_{w1,2}^i & (z_r T_{w1,3}^i + T_{w1,4}^i) \\ T_{w2,1}^i & T_{w2,2}^i & (z_r T_{w2,3}^i + T_{w2,4}^i) \\ T_{w3,1}^i & T_{w3,2}^i & (z_r T_{w3,3}^i + T_{w3,4}^i) \end{pmatrix} \cdot \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} \quad (3)$$

In a final step, the obtained object observations are refined and validated via edge information, as can be seen in Figure 6. To do this, an object standard height is assumed and with it, a window is calculated in image coordinates (analogously to Equation (3)). If the amount of edges contained in the window reaches a threshold, the object is validated. Edges in direction to the vanishing point are discarded of this step, since they are often present in road structures (without objects).

Accompanying to the generation of object observations, the uncertainty in image coordinates, derived from edge information, is transformed into road coordinates by means of propagation of uncertainty and equation (1).

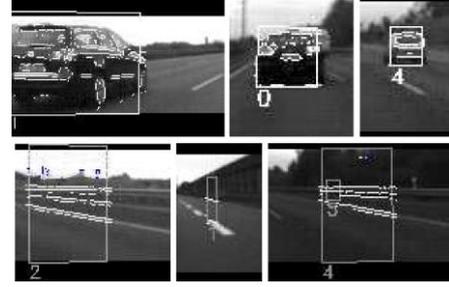


Figure 6. Examples of validated (above) and not validated object observations (below).

5. LASER SCANNER PROCESSING FOR OBJECT DETECTION

As a second cue of object observations a new method to process laser scanner reflection points is presented in this section. For the object segmentation, the in-sensor pre-processing is used. As a pre-filtering, only objects within the detected driving corridor are processed (see Section 3). The focus of this section is set on the fitting of the segmented reflection points (contours) to the object model presented in Figure 3. As indicator of the basic shape of an object, the contour point with minimal angle to the contour extreme points is selected. Further calculations are done in order to characterize the detected object with its dimensions, heading, reference and observability.

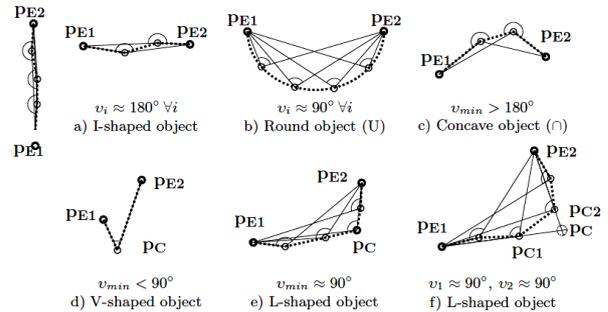


Figure 7. Decision of the basic shape of a contour depending on the minimal inner angle, v_{\min} .

The basic shape of an object is identified depending on the minimal angle between both contour extreme points (\mathbf{p}_{E1} and \mathbf{p}_{E2}) and an inner contour point (\mathbf{p}_C), $v_{\min} = \min\{\angle \mathbf{p}_{E1}, \mathbf{p}_C, \mathbf{p}_{E2}\}$, as represented in Figure 7. If this angle is close to 180° , then the object has an oblong shape (I-shape). If no clear minimum is present and all angles are close to 90° , it is a round (or slightly rounded) object (U-shape). If the minimum angle is larger than 180° , then a concave object is given (\cap -shape). If a clear minimum close to 90° is found, then the object has L-shape. In other cases, other shapes shall be considered.

In real data, the vertex in L-shaped objects is often not detected. Therefore, it is convenient to take the two contour points with minimal inner angle and to estimate the vertex point as the intersection point between the lines $\overline{\mathbf{p}_{E1}, \mathbf{p}_{C1}}$ and $\overline{\mathbf{p}_{E2}, \mathbf{p}_{C2}}$ (see Figure 7.f).

After having decided the basic shape of the object, its heading and size are calculated as shown in Figure 8. The size of an L-shaped object depends on the inner angle of its vertex. If it is less than 90° (Method 1), then the longest side is taken as the first dimension, and the distance from this side to the other extreme point is taken as the second one (see Figure 8.d). If the angle is greater than 90° , this would result in a too short estimation (Figure 8.c). In this case (Method 2), the vertex is first shifted on the longest side to the closest point to the other extreme (see Figure 8.e).

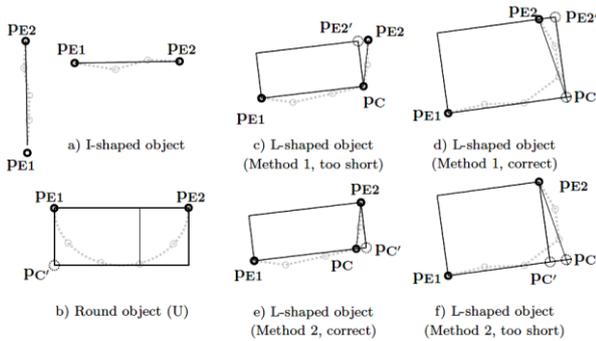


Figure 8. Calculation of the object size.

The calculated raw heading and size are then adapted to the model of Figure 3. To do this, a check of the basic shape, heading and position of the object is applied, and its width, length, heading and optimal reference point are calculated.

The presented method, as in most methods of the bibliography, is not able to deal with every contour unambiguously (see example in Figure 9). For this reason, the most probable object observations are generated and passed to the tracking and fusion module, where the ambiguity is solved with help of the object's movement.

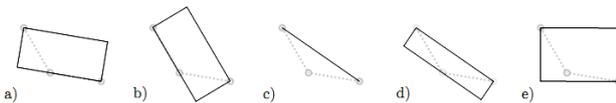


Figure 9. Presence of ambiguities in the object characterization from laser scanner data.

In a last step of this method, the observability of the detected object is analyzed and saved in the observability matrix. A component (length or width) is set to not observable in any of the following cases, which are represented in Figure 10:

- The contour extreme is close to the sensor range limit (Figure 10.a).
- The contour extreme is possibly beyond another object (Figure 10.d).
- Assuming that the object has double length (or width), the coverage angle from the sensor nearly does not change (Examples: not observable Figure 10.b, and e, but observable Figure 10.c).

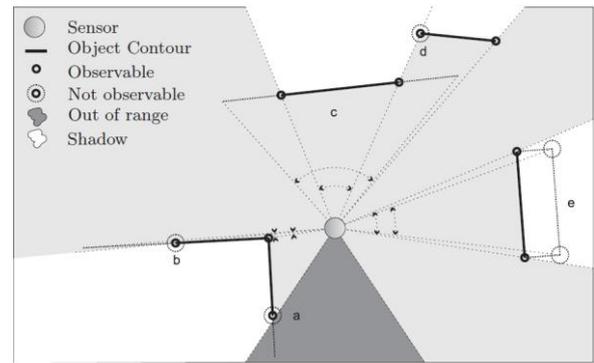


Figure 10. Different observability constellations of detected laser scanner objects (see cross-references in text).

Based on the accuracy of each reflection point provided by the sensor and on the extracted object properties, the object accuracy is calculated by means of propagation of uncertainty. An example of laser scanner observations over a sequence of real data is shown in Figure 11.

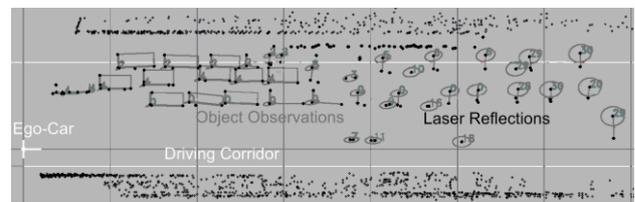


Figure 11. Example of laser scanner detections in a sequence of real driving data.

6. EXPERIMENTAL RESULTS

A set of experiments has been carried out in order to test and check the plausibility of the presented methods. On the one hand, the simulator environment of the Institute of Transportation Systems of the DLR (see www.dlr.de/ts) has been applied, which includes realistic traffic and sensor simulation (camera, laser scanner and others). In the simulator, traffic situations can be targeted and repeated in controlled conditions. In addition to the simulated experiments, real driving data collected with the ViewCar (see Figure 1) has also been used to check the plausibility of the methods. In both cases, the information about the driving corridor is assumed to be provided. This reduces the processing time and the false positive rate (due to noise in the road sides).

The experiments conducted include different highway scenarios of interest, such as vehicle following, overtaking and lane changing. Table 1 contains a summarized overview of the results. As a first criterion, the position accuracy is considered $\bar{\varepsilon}_x$ and $\bar{\varepsilon}_y$. In simulated drives, the mean absolute error (compared with the known ground truth) is given. In real drives the ground truth is unknown and therefore, the estimated error covariance from tracking is considered instead. Both measures have the same order of magnitude. As a second criterion, the detection's reliability (false negative and false positive rates*) is taken into account. The false positive rate in real driving data is very high due to the inclusion of side objects.

* $f.n. = \#f.n./\#t.p.$; $f.p. = \#f.p./\#t.p.$

	Camera (24 fps, range ~60 m)		Laser scanner (12.5 Hz, range ~150 m)	
	Sim (64 s)	Real (46 s)	Sim (64 s)	Real (46 s)
$\bar{\varepsilon}_x$ (m)	1.309	1.110	0.069	0.293
$\bar{\varepsilon}_y$ (m)	0.298	0.317	0.034	0.197
$f.n.$ (%)	8.207	8.867	2.990	7.564
$f.p.$ (%)	6.307	12.988	2.518	24.492

Table 1. Quantitative results in simulated and real driving situations. (see cross-references in text).

Based on Table 1, the following conclusions can be drawn. On the one hand, it is proven that both sensors are capable of detecting object observations with high accuracy and reliability. Thus, the sensors are appropriate to be used as cues for the object level fusion. On the other hand, the sensors used complement each another well. The laser scanner shows a higher accuracy, reliability and range, while the camera has a higher frame rate. Finally, since the results on simulated data and real data were similar, a validation of the results with simulated data is possible. Although no ground truth was available, a visual comparison shows qualitatively good results.

7. CONCLUSIONS AND OUTLOOK

In this paper an object detection system based on two cues has been presented. The methods have been introduced in the context of a multi-layer multi-sensor fusion system. The object fusion has been carried out at object level, for which object observations were extracted from each sensor.

The extraction of object observations from camera images has been done applying texture calculation and image structure extraction. Compared to other works in the bibliography, the method has the advantage of detecting objects even though these might not have clear edges, features or movement. The important thing is that the object has a different texture (or gray value) than its background. However the method still has some limitations, as in roads containing different asphalt regions, spots or dirt (e.g. fallen leaves). As in other works, shadows from other objects and bridges can also lead to wrong results.

In the second part of the paper a method to characterize objects detected by a laser scanner has been presented. With the proposed method, based on the inner angle of the contour points, a more detailed object shape characterization (including round and concave objects) than with conventional methods is reached. Additionally, observability checks are explicitly performed on the objects. A common limitation of the method as in other works is the presence of ambiguities, which have to be solved at the fusion level.

The results of the conducted experiments show the plausibility of the system. The detected object observations are appropriate to be used as the input of an object fusion system. Both sensors present strengths and weaknesses, which is ideal for a fusion system. All experiments have been run in offline mode without real time constraints. However, a set of code optimizations is planned for future work, with which we expect the system to provide high data rates.

In the future, the limitations of both object cues will be improved and the fusion mechanisms will be further optimized. The image processing will be extended to other object

perspectives. The complete system will be adapted and tested with other scenarios such as oncoming traffic and urban situations. Based on the detected objects the danger recognition methods will be further improved and evaluated.

REFERENCES

- Bensrhair, A., Bertozzi, M., Broggi, A., Miche, P., Mousset, S., and Toulminet, G., 2001. A cooperative approach to vision-based vehicle detection. In: *IEEE Intelligent Transportation Systems*. Oakland (CA), USA, pp. 209-214.
- Catalá-Prat, Á., Köster, F., and Reulke, R., 2008. Sensor Fusion for the Description of Driving Situations with the Data of the Test Vehicle ViewCar. In *Optische Technologien in der Fahrzeugtechnik. VDI-Bericht 2038*, Leonberg, Germany, pp. 127-141.
- Catalá-Prat, Á.; Köster, F., and Reulke, R., 2008b. Early Detection of Hazards in Driving Situations through Multi-Sensor Fusion. In: *FISITA World Automotive Congress*, Munich, Germany, Vol. 2, pp. 527-536.
- Dietmayer, K., Sparbert, J., and Streller, D., 2001. Model based Object Classification and Object Tracking in Traffic scenes from Range Images. In: *IEEE Intelligent Vehicles Symposium*. Tokyo, Japan, pp. 1-6.
- Giachetti, A., Campani, M., and Torre, V., 1998. The use of optical flow for road navigation. In: *IEEE Transactions on Robotics and Automation*, 14(1), pp. 34-48.
- Labayrade, R., Royere, C., and Aubert, D., 2005. A collision mitigation system using laser scanner and stereovision fusion and its assessment. In: *IEEE Intelligent Vehicles Symposium*, pp. 441-446.
- Lindner, P., Scheunert, U., and Richter, E., 2008: Multi Level Fusion for Environment Recognition. In: *PREVENT Fusion Forum e-Journal* (2), pp. 24-30.
- Neumaier, S., and Färber, G., 2007. Vision-Based 4D-Environmental Perception for Advanced Assistance Functions. In: *it - Information Technology*, 49(1), pp. 33-39.
- Rao, B., Durrant-Whyte, H., and Sheen, J., 1993. A fully decentralized multi-sensor system for tracking and surveillance. In: *The International Journal of Robotics Research*, 12(1), pp. 20-44.
- Viola, P., and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 511-518.
- Wender, S., and Dietmayer, K., 2008. 3D vehicle detection using a laser scanner and a video camera. In: *Intelligent Transportation Systems*, 2(2), pp. 105-112.
- Wohlfeil, J., and Reulke, R., 2008. Detection and Tracking of Vehicles with a Moving Camera. In: *Optische Technologien in der Fahrzeugtechnik. VDI-Bericht 2038*. Leonberg, Germany, pp. 181-195.