2011 IEEE Intelligent Vehicles Symposium (IV)
Baden-Baden, Germany, June 5-9, 2011

A generic video and radar data fusion system for improved target selection

Dennis Müller*, Josef Pauli†, Mirko Meuter*, Lali Ghosh* and Stefan Müller-Schneiders*

* Delphi Electronics & Safety
42119 Wuppertal, Germany
{dennis.mueller, mirko.meuter}@delphi.com
{stefan.mueller-schneiders, lali.ghosh}@delphi.com

† University Duisburg-Essen
Fakultät für Ingenieurwissenschaften
47048 Duisburg, Germany
josef.pauli@uni-due.de

***Abstract*— This paper presents an automotive video and radar data fusion framework that can be used as a preliminary stage of an automatic cruise control or collision mitigation by braking system. The fusion framework finds the optimal assignment of radar and camera target reports and provides improved state estimates for the fused targets. A sophisticated critical path selection is presented and used in the critical target selection module that aims to select the most relevant target. This module is capable of identifying targets that cut into the ego lane or cut out from the ego lane and incorporate that into the final target selection. The selected target is then compared to a state of the art algorithm within the radar sensor. Additional test drives were made to evaluate the performance of the new algorithm. Due to its low computational effort and the sensor independent design the presented algorithm is suitable to be used in the automotive embedded environment.**

I. Introduction

Many state of the art high and medium class vehicles are equipped with a radar that is responsible for the ACC (Automatic Cruise Control) functionality. This system constantly monitors the distance to the leading vehicle and compares this with the appropriated safety distance as well as the driver selected maximum speed. It then accelerates or decelerates the vehicle if necessary and possible. If however the distance to the leading vehicle becomes too small or a new vehicle suddenly appears in the ego lane (for example a short distance cut in from the right lane) these systems do not trigger a full braking. This is due to the low confidence a single sensor can achieve in these situations. If in addition to the radar a camera system becomes available a fusion scheme is thinkable that aims to confirm observed radar targets by the camera. The camera can also improve the target selection because the camera can actually observe the road geometry ahead and thus improve the driving path prediction. With such a system it is possible to develop a fusion application that will improve the common ACC module due to the improved target selection methods.

In the literature vision and radar data fusion methodologies are commonly divided into low level, intermediate level and high level fusion approaches. Low level fusion combines raw sensor data[1] from multiple sensors and produce new data with a similar low level of abstraction that is expected to be more informative. [1] present a novel filter framework based on the finite set theory to filter multiple independent sensor reports. This allows to delay the final decision about target existence and association and can yield better false alert rates and higher detection rates. However such approaches require a lot of processing power and need to have detailed information about the underlying sensors, thus making it hard to exchange the sensors after the system is developed. In [2] another low level fusion approach is presented that suffers from the same problem. Intermediate level fusion approaches require one sensor to produce highly abstract information, such as finally filtered target reports and use the raw signal from the second sensor to confirm or improve this information. [3] uses the raw camera image and image processing methods to distinguish solid radar observations such as guard rails from ghost targets that originate from multi-path reflections of the radar beam. [4], [5], [6] all use image based features to confirm the target existence and improve the target state and thus also require access to the camera raw data. [7] shows an example for high level fusion similar to the approach presented in this work. We will present such a fusion system and show how we combine the available information from both the radar and the camera system. We treat both sensors as independent thus no feedback channel between the two sensors is required. A minimal set of information is required to be delivered by the sensors. To keep the assortment of acceptable sensors small, only the current object position and the object velocity are required. In addition to that the sensor shall also report the state uncertainty in form of a covariance matrix. In general the radar has a higher range and range rate accuracy but the camera has superior accuracy in the lateral position and velocity component. By fusing the available information we can acquire a system that can reliably detect targets that cut into or out of the ego lane. We can thus feed these targets earlier to the subsequent ACC module which in turn can react faster on them. This can prevent a collision in the cutting in case and improve the drivers experience in the cutting out case. The remainder of this paper is structured as follows: In section II, we will discuss the problem of sensor synchronization, both in the spatial and time domain. We will then show how we tackle the track-to-track association problem and how we fuse the available information in section III. Section IV will discuss our critical path estimation and section V will then present our actual target selection scheme that is based on both the fused target state estimate and the selected critical path. Section VI will discuss our evaluation methodology while section VII will conclude this paper with a short summary and discussion of future steps.

[1] *raw* in this sense refers to sensor data that is as less processed and filtered as possible

II. Data Synchronization

The fusion system presented in this work was designed to be as independent from the sensors as possible. This should allow for a later exchange of one or both of the sensors in the final system with the least amount of work necessary. We only require both sensors to deliver tracked and confirmed target reports with a state estimate and a corresponding uncertainty. The state report should at least contain the position of the obstacle and the velocity. In our setup, the radar reports its targets in the polar coordinate domain and can only measure the range rate. From the radar internal tracking process an estimate about the lateral rate is also available. The camera reports its observed targets in the cartesian coordinate domain and can only measure the position of the obstacle but again a longitudinal and lateral velocity component is derived in the internal tracking process. The tracker is responsible for managing the target identities (e.g. the same target should be assigned with the same identifier tag as long as it is observable) and it is also responsible for delivering the corresponding state uncertainties in form of a covariance matrix. Our radar operates at 10 Hz but the camera system delivers a new report at 30 Hz. The sensors do not need to be synchronized. Since the fusion framework will be operated in an automotive environment we can also assume that some basic information about the vehicle motion will be available. To keep the system as generic as possible we will only require the current yaw rate and the current velocity of the vehicle to be available in the fusion framework.

We have arbitrarily chosen to operate the fusion framework at the same cycle as the camera, so whenever the camera reports a new environment report (a list of targets) the fusion framework is processed afterwards. The fusion framework maintains a list of all reported targets by both sensors. At every cycle the current target state is predicted using the recent vehicle motion reports received from the CAN bus. If a new sensor report became available since the last cycle this will be used as a starting point for the synchronization. If not, the most recent prediction will be used. Let $X_0 = (x, y, \dot{x}, \dot{y})^T$ denote the target state estimate in cartesian coordinates where x denotes the longitudinal distance, y denotes the lateral distance and $\dot{x}$ and $\dot{y}$ denote the longitudinal and lateral velocity components, respectively. It can then easily be shown that if we assume a constant yaw rate and velocity during the time interval T the state vector becomes

$$X_T = F \cdot X_0 + M \cdot v_0 \tag{1}$$

where

$$F = \begin{pmatrix} c & -s & cT & -sT \\ s & c & sT & cT \\ 0 & 0 & c & -s \\ 0 & 0 & s & c \end{pmatrix} \tag{2}$$

and

$$M = \frac{1}{\alpha} \begin{pmatrix} s & c-1 & sT & cT \\ 1-c & s & -cT & sT \\ 0 & 0 & s & c-1 \\ 0 & 0 & 1-c & s \end{pmatrix} \tag{3}$$

with $c = \cos(-\alpha T)$, $s = \sin(-\alpha T)$ and $v_0 = (v, 0, 0, 0)^T$ where v denotes the host car velocity in the time interval T and α denotes the yaw rate in the time interval. Note that F and M are functions of α. Expanding (1) around α_0 then yields

$$X_T \approx (F_0 + F'(\alpha - \alpha_0))X_0 + (M_0 + M'(\alpha - \alpha_0))v_0 \tag{4}$$

where F_0 and M_0 denote the matrices F and M evaluated at $\alpha = \alpha_0$ and F' and M' denote the derivatives of F and M with respect to α, respectively. It then follows that

$$E[X_T] = F_0 X_0 + M_0 v_0 \tag{5}$$

and (see [8], [9] for details)

$$\begin{aligned} V[X_T] &= \sigma_\alpha^2 F' \left(Q_X + X_0 X_0^T\right) F'^T + F_0 Q_X F_0^T \\ &+ \sigma_\alpha^2 M' \left(Q_v + v_0 v_0^T\right) M'^T + M_0 Q_v M_0^T \\ &+ \left(C + C^T\right) \end{aligned} \tag{6}$$

where Q_X denotes the variance estimate of X_0, σ_α^2 denotes the (assumed) variance of the yaw rate measurement and $Q_v = (\sigma_v, 0, 0, 0)^T(\sigma_v, 0, 0, 0)$ denotes the (assumed) variance of the velocity measurement. C denotes the covariance between the left and right summand in (4). It can readily be seen that

$$\begin{aligned} C &= E\left[(F'(\alpha - \alpha_0)X_0) \cdot (M'(\alpha - \alpha_0)v_0)^T\right] \\ &= \sigma_\alpha^2 F' X_0 v_0^T M'^T \end{aligned} \tag{7}$$

Note that in the limiting case $\alpha \mapsto 0$ equation (4) converges and (5) and (6) can be further simplified. Using (5) and (6) we can predict the current target state and uncertainty from the last available measurement into the current time frame by using the velocity and yaw rate reported from the host car. If more than one velocity/yaw rate pair was reported since the last fusion cycle, the above motion equations are applied iteratively for each set of motion parameters and time intervals.

III. Track-To-Track Association

At each fusion cycle a comprehensive list of all observed targets from both the radar and the camera are available. The target state reports are already synchronized into the current time frame given the host vehicle motion parameters and the corresponding state uncertainties are updated accordingly. One might argue that a possible solution to the track-to-track association and state estimation problem might be achieved by implementing a kalman filter based framework that utilizes the given sensor information as measurement inputs. However it is known (compare [10]) that due to the high temporal correlation of the *measurements* and a possible inter-sensor correlation such an approach in general underestimates the fused objects covariance. We have thus decided to use a global association optimization scheme combined with a covariance trace-minimization based data fusion technique (compare also [11]). With $X_i^{(c)}$ we denote the current (synchronized) state estimate of the $i-th$ camera target. $X_j^{(r)}$ denotes the state estimate of the $j-th$ radar target. With

$$\Delta X_{i,j} = X_i^{(c)} - X_j^{(r)} \tag{8}$$

we denote the state estimate difference. If both reports originate from the same target and the assumption that the

state errors are uncorrelated[2] it follows that

$$D = \Delta X_{i,j}^T \left(P_i^{(c)} + P_j^{(r)} \right)^{-1} \Delta X_{i,j} \tag{9}$$

is chi-square distributed with $N = 4$ degrees of freedom.

Using only the current state might introduce ambiguity or in general less confidence in the current association. Although all available information from one sensor is combined within the current target state and state uncertainty report in an optimal way due to the underlying tracking process, using only the current target state report does not always allow to distinguish ambiguous situations. For example in figure 1

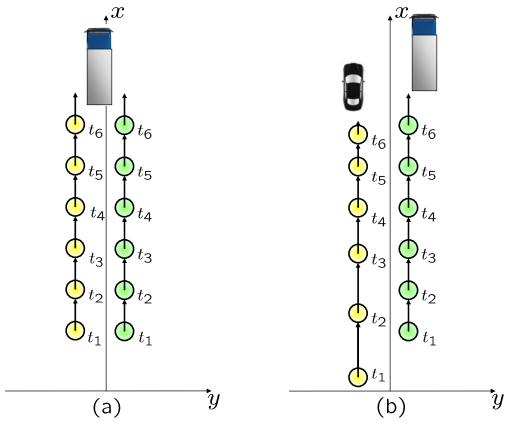

Fig. 1. Ambiguous association situations

two situations are drawn. The yellow dots represents a target state report from one sensors while the green dots represents target state reports from the other sensor. The black arrows indicate the velocity component of the target state report. In both shown situations six time instances for the reports are shown, each one labeled with $t_1, \ldots, t_6$. In subfigure (a) one would accept the hypothesis that both sensor reports originate from the same target because they behave in the same way. However in subfigure (b) one would clearly reject the hypothesis because the yellow target seems to decelerate while the green target continues to travel with constant speed thus the sensor report can not describe the same target. This situation is easy to distinguish if only the time instance t_1, t_2 or t_6 would be given. However in the time instance t_3, t_4 and t_5 the state report seems to be almost identical, both with respect to the position and velocity component. If only that state estimate would be taken into account it is likely that the system will associate these targets with each other which would then yield a wrong fused state estimate and give a false confirmation for the targets. It is thus advantageous to also take the history of recent measured states into account to calculate an association feasibility. Two consecutive sensor reports are highly correlated due to the underlying tracking process that is performed in the sensor. Unfortunately this correlation is not known and is not reported by the sensor in any way. The sum

$$D^* = \frac{1}{N} \sum_{i=1}^{N} D_i \tag{10}$$

denotes the average of all single feasibility measures obtained from equation (9) where the lower index i indicates the $i-th$ recent sensor report, thus D_1 is calculated using the most recent report, D_2 is obtained by using the sensor reports before D_1 and so on. N is the total number of available reports and thus the smaller number of reports available from the radar and the camera. To keep the computational requirements low the maximum number of stored reports is limited. D^* is not chi-square distributed due to the correlation between the summands. However D^* still has the desired effect that for sensor reports from the same target, D^* will be small since every single feasibility measure will be and for sensor reports from two distinct targets, D^* will be large because at least some of the feasibility measures will be. To find a global optimal association between radar and camera target we utilize the Hungarian method (see [13] and [14]). Target reports that were associated with each other can then be combined to yield a fused estimate about the true target state. For simplicity we omit the indices and assume that we are interested in the fused state estimate for the $i-th$ camera target $X^{(c)} = X_i^{(c)}$ and the $j-th$ radar target $X^{(r)} = X_j^{(r)}$. We are interested in a linear combination of both estimates

$$X^{(f)} = W^{(c)} \cdot X^{(c)} + W^{(r)} \cdot X^{(r)} \tag{11}$$

under the constraint

$$W^{(c)} + W^{(r)} = I \tag{12}$$

We aim to minimize the trace of the resulting covariance. The solution to that minimization problem is given by (compare [15])

$$W^{(c)} = P^{(r)} \cdot \left(P^{(c)} + P^{(r)} \right)^{-1} \tag{13}$$

$$W^{(r)} = P^{(c)} \cdot \left(P^{(c)} + P^{(r)} \right)^{-1} \tag{14}$$

with the final covariance

$$P^{(f)} = P^{(r)} \cdot \left(P^{(c)} + P^{(r)} \right)^{-1} \cdot P^{(c)} \tag{15}$$

IV. Critical Path Selection

We assume that the camera system is able to detect and track the visible road markings ahead and that at every time step the camera system reports a state vector describing the current road geometry with a parabola model. Even though more sophisticated geometry models might yield a more accurate description of the road geometry this choice was made so that one can easily exchange the camera sensor since every state of the art lane detection system should at least be able to deliver to this requirement. The state vector is given as

$$x^{(l)} = \begin{pmatrix} \gamma^{(l)} & \beta^{(l)} & d_l^{(l)} & d_r^{(l)} \end{pmatrix}^T \tag{16}$$

[2]Note that in general we can not make this assumption since both $X^{(r)}$ and $X^{(c)}$ are predicted using the same vehicle motion date obtained from the CAN Bus (compare equation (1)). However if we receive the sensor updates frequently enough the correlation introduced by this is small enough to be negligible. In case the cross-covariance should not be neglected, similar formulas can be derived that incorporate this correlation, see [12].

with corresponding covariance matrix $P^{(l)}$. $\gamma^{(l)}$ denotes the curvature, $\beta^{(l)}$ the heading and $d_l^{(l)}$ and $d_r^{(l)}$ the distance to the left and right lane marker respectively. The upper index (l) denotes that these parameters correspond to the model parameters of the lane detection and tracking system. We will also use the upper index (p) for parameters that describe the driving path estimation. From the underlying tracking process in the software model we can assume that $x^{(l)}$ is Gaussian distributed with covariance $P^{(l)}$. In addition to the current state and state covariance the camera system also reports a confidence level for both the left lane marker detection and the right lane marker detection.

The driving path is determined by only the vehicle ego motion, hence the current velocity and the current yaw rate. If the velocity and yaw rate would be constant the vehicle would drive on a circle with constant radius. The curvature of this circle is then given by

$$\gamma^{(p)} = \frac{\phi}{v} \tag{17}$$

where ϕ is the current yaw rate and v denotes the current velocity. The heading is always $\beta^{(p)} = 0$ and the left and right offset can only be assumed to be half the width of the driving path, thus it is assumed that the vehicle is centered in the driving path, hence

$$\left(d_l^{(p)}, d_r^{(p)}\right) = \left(-\frac{1}{2}, \frac{1}{2}\right) \cdot \text{Path Width} \tag{18}$$

The path width is assumed to be equal to the last reported lane width. If no valid lane width report is available a fixed width of 3.5 meters is selected.

If we assume the measured yaw rate ϕ and the measured velocity v to be also Gaussian distributed without bias and if we further assume that the velocity is significant above zero[3] it can be shown (see [16]) that the curvature distribution can be approximated by a Gaussian distributed with

$$E\left[\gamma^{(p)}\right] = \frac{\mu_\phi}{\mu_v} \tag{19}$$

and

$$V\left[\gamma^{(p)}\right] = \left(\frac{\mu_\phi \mu_v}{{\mu_v}^2 - {\sigma_v}^2}\right)^2 - \frac{{\mu_\phi}^2 - {\sigma_\phi}^2}{{\mu_v}^2 - {\sigma_v}^2} \tag{20}$$

We can further assume a standard deviation for the heading of 1° and a standard deviation for the offset of 0.75 meters. The correlation between all the parameters is assumed to be zero. With that, $P^{(p)}$ is known.

Adapting the measurement gate idea known from the kalman filter [12] we can decide whether the estimated driving path can be believed to be a noisy measurement sample of the lane or if the path is different from the lane. Therefor we calculate

$$D = \left(x^{(l)} - x^{(p)}\right)^T \left(P^{(l)} + P^{(p)}\right)^{-1} \left(x^{(l)} - x^{(p)}\right) \tag{21}$$

D is chi-square distributed with four degrees of freedom. The chi-square distribution thus provides a threshold for D that is used to decide whether we accept the hypothesis that the driving path is equal to the lane or if we have to reject that hypothesis. In every time step we obtain the current lane model estimate from the lane detection and tracking system, the corresponding covariance and confidence levels for both sides. If both sides have too less confidence the driving path model is favored in this particular time step. If the confidence level is high enough for at least one of the sides the kalman filter gate approach explained above is utilized to decide whether the driving path and the lane parabola are similar enough to be considered identical or not in this particular frame. A simple hysteresis gate is applied to the single frame decision to further smooth the overall system decision. Because the presented algorithm prefers the driving path in cases where both possible models differ the system is susceptible to ego motion sensor interference. If for example the yaw rate sensor is faulty the driving path will most likely be completely different from the lane. The system will then prefer the wrong driving path over the potentially correct lane estimate.

[3]to be more precise: We assume that $v > 6 \cdot \sigma_v$, where σ_v denotes the standard deviation of the velocity. This assumption ensures that the density function of v is almost zero for all negative velocities.

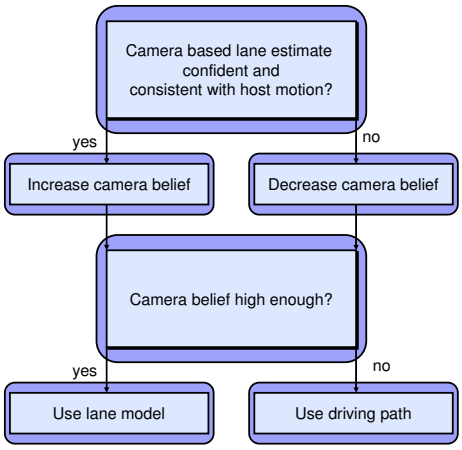

Fig. 2. Critical Path Selection Algorithm

V. Target Selection Scheme

A critical path was selected based on either the camera based lane estimate or the host motion deduced driving path. In both cases the critical path is modeled as a parabola together with uncertainty estimates given as the expected covariance matrix of the parabola parameters. If the critical path is selected as the lane estimate this covariance matrix is provided by the lane detection and tracking system. Otherwise the covariance matrix is estimated as described in the previous section. In addition to that the track-to-track association scheme described in section III provides state estimates containing a position estimate, a velocity estimate and again the covariance matrix between these states. For a vision based target and fused targets this state estimate and covariance is provided by the vision system or obtained from trace minimization fusion. The covariance estimate for pure radar based targets is provided by the underlying radar sensor.

Let $X^{(v)}$ and $P^{(v)}$ denote the state and covariance estimate for a vehicle target and $X^{(p)}$ and $P^{(p)}$ denote the state and covariance estimate for the critical path respectively. The combined state and covariance is then given as

$$X = \begin{pmatrix} X^{(v)} \\ X^{(p)} \end{pmatrix} \tag{22}$$

and

$$P = \begin{pmatrix} P^{(v)} & 0 \\ 0 & P^{(p)} \end{pmatrix} \tag{23}$$

Given the combined target/path state X and the corresponding covariance matrix P we can assume that the true target position and the true critical path state is normal distributed with the given mean and covariance. We can thus draw samples $T_1, \ldots, T_k$ from this distribution. Each sample $T_i \sim \mathcal{N}(X, P)$ is a possible description of the observed situation, called an *instance*. For each instance we can thus determine whether the target would be left of the ego lane, right of the ego lane or inside the ego lane. If it is the later it can not perform a cut in maneuver. However if it is either to the right or the left it is possible to calculate the *time to cut in*. Therefor the obstacle motion is modeled as

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t \begin{pmatrix} v_x \\ v_y \end{pmatrix} \tag{24}$$

where x_t denotes the longitudinal distance of the obstacle at time t, y_t denotes the lateral distance at time t and v_x and v_y denote the velocity in longitudinal and lateral direction respectively. thus we assume a constant velocity of the obstacle. The critical path is modeled as a parabola. We thus have

$$y = \frac{1}{2}\gamma x^2 + \beta x + \text{offset} \tag{25}$$

where γ is the curvature, β is the heading and offset is either the offset to the left border of the critical path if the target is believed to be left to the critical path or the offset to right border of the critical path otherwise. Substituting 24 in 25 and solving for t then yields the *time to cutin*. Note that one usually gets two solutions and that they can be negative. However it is easy to see that the smallest positive solution is the relevant *time to cutin*.

Drawing many samples from the assumed target state distribution would thus yield a fairly good approximation to the actual distribution of the *time to cutin*. This is however not possible because it would take just too much processing power to do so for all observed targets. A modification was thus developed that allows obtaining a close approximation to the distribution without drawing too much samples each time. For that we draw a small number of samples at a time and calculate the *time to cutin* for each sample as described above. A list of evaluated samples is maintained. For a new target this list is empty. Once samples are drawn the resulting cutin time is added to this list together with an initial weight of 1. So in each time step a reasonable number of samples is added to this list. Once a new frame becomes available it is also reasonable to assume[4] that the *time to cutin* of the obstacle is reduced by ΔT, the time difference since the last camera image. We thus process all stored samples in the list and reduce the stored time to cutin estimate by ΔT. In addition to that the weight is also reduced. This reflects the fact that an old sample drawn some time ago no longer accurately represents the current situation and is thus less viable. There is a maximum number of samples in the list and the oldest sample will be removed from the list if the capacity is maxed out. If the above assumption holds, the distribution of the *time to cutin* can be fairly accurate described by this list. If not, the reconstructed density becomes distorted and the following detection criterion will not trigger a false alert. A *cut in maneuver* is detected if we have collected enough evidence that the actual *time to cutin* is less than a threshold T_{cutin}. To validate that we loop through the list and count the weights of each sample below the threshold and sum them up. We do the same for each sample that is above the threshold. If we have more weight below the threshold than above it this is considered as a *cut in maneuver*. It is obvious that this is equivalent to the question whether the median of the distribution is less than T_{cutin}. Because we had to reject some samples earlier[5] we also demand that the sum of weights below the threshold exceeds a fixed limit. This prevents outliers from triggering the *cut in maneuver* detection.

A *cut out maneuver* is detected in the same way as a *cut in maneuver*. Each target/path state instance that represents an in ego lane target gives rise to a *time to cutout* sample. The point of intersection with the critical path is evaluated for both the left and the right border. Again in general there exists two solutions for each side and they can be negative. And again it is easy to see that the smallest positive solution is the relevant *time to cutout*. This sample is added to a second list of evaluated instances and again a *cut out maneuver* is detected when there is more mass below the time to cut out threshold than above it and this mass exc eeds the same threshold.

As already described above we drew 100 samples from the combined target state X and covariance matrix P in each frame. For each sample we can decide whether for this sample the car is inside the ego lane, left to it or right to it. For a target that is currently considered out of the ego lane at least 55% of the drawn samples must indicate that it has moved into the ego lane for it to be considered as in ego lane. On the other hand if it is considered as in ego lane and less than 45% of the drawn samples indicated that it is still in the ego lane it is marked as outside of the ego lane.

A target is then valid for the target selection if it is either believed to be in the ego lane and not cutting out or within the neighboring lanes but cutting in. We also require the target to be within a certain, velocity dependent range. We only accept radar targets that are either reported as moving (their radar based range rate exceed a certain threshold) or are camera confirmed (the target could be associated with a camera target). This includes stationary targets that have never been observed as moving by the radar but have found an association partner by the camera. We can thus trigger a full force braking if the vehicle approaches an already full stopped traffic jam end. Out of all valid target the closest is selected. Again, a small hysteresis is used to smooth the

[4]see our comment in section VII regarding this assumption

[5]Instances which indicate no cutin or instances in which the target was inside the ego lane were rejected.

TABLE I
SITUATION BASED EVALUATION RESULTS

Situation	Radar Winner	Fusion Winner	Total
Cut in	3	27	30
Cut out	4	22	26
New moving	1	5	6
New stationary	0	2	2
Other	1	0	1
Total	9	56	65 (2.4%)

target selection output and prevent flickering of the module output.

VI. EXPERIMENTAL RESULTS

Throughout this work a production ready radar system has been utilized. That radar system comes with an internal target selection scheme that is based purely on the driving path derived from the vehicle motion and the pure radar based target state estimate. The radar target selection is then feed into, for example, an ACC software module that actually controls the vehicles engine and brakes based on that target selection scheme. The target selection software in the radar is fully evaluated and validated through hundred thousands of kilometers. In the scope of this paper we utilize this target selection as ground truth in the following sense:

Whenever the fusion framework selects the same target as the radar we accept this selection as correct. If however the fusion framework selects a different target than the radar, for example because it recognizes a cut in maneuver from a closer target, this situation is recorded and manually analyzed. With this setup but disabled actuators exhaustive test drives were performed on public roads. These test drives took place on five different days and on each day roughly three hours of driving were performed, resulting in approximately 15 hours of driving total. Out of that 65 video sequences with 20 seconds each (roughly 22 minutes, $\approx 2.4\%$) have been recorded by the system. Table I shows the result of this manual evaluation. The first column describes the situation in which a video sequence was recorded. The *New moving* and *New stationary* refers to a situation where a never before seen moving or stationary target appears in the ego lane, for example when approaching the end of a traffic jam. The column *Radar Winner* and *Fusion Winner* count how often the target selection of the radar or the presented fusion scheme was superior. Since only 2.4% of the total time driven were recorded it can be concluded that in unambiguous situations both algorithms come to the same conclusion and select the same target, indicating that the fusion process does not increase the false alert rate by selecting wrong targets. However in the aforementioned situations (cut in or cut out from other vehicles) our algorithms shows clearly superior performance compared to the radar only based approach as can be seen from the above tabular.

VII. SUMMARY AND FUTURE STEPS

In this paper we have presented our generic data fusion framework that is capable of fusing target reports obtained from a radar system and a camera system. The required information from both sensor was kept as low as possible to allow for an easy exchange of one or both sensors. Only the position and velocity of each observed target is required. An improved path selection and the fused target state estimates were used to identify targets that perform a cut in or cut out maneuver. This information was then used in the critical target selection scheme. The final target selection scheme was then compared to a state of the art target selection available within the radar system.

Future steps include a more sophisticated evaluation of the target selection, especially with enabled actuators to actually experience the fusion based target selection. The cut in maneuver detection is based on the assumption that the *time to cut in* decreases by ΔT in the time interval ΔT. This assumption is frequently violated and in these situations the obtained density estimate from the weighted particle list is distorted. Even though the presented algorithm is capable of detecting cut in maneuvers early enough a more sophisticated model is sought that better describes the dynamics of a cut in or cut out maneuver.

ACKNOWLEDGMENT

This work results from the joint project Ko-PER, which is part of the project initiative Ko-FAS, and has been partially funded by the German BMWi (Federal Department of Commerce and Technology) under grant number 19S9022D.

REFERENCES

[1] M. Mählisch. *Filter Synthesis for Simultaneous Minimization of Detection, Association, and State Uncertainties in Automotive Environment Perception with Heterogeneous Sensor Data*. PhD thesis (in german), University of Ulm, 2009.

[2] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier. Fade: a vehicle detection and tracking system featuring monocular color vision and radar data fusion. *IEEE Intelligent Vehicles Symposium Proceedings*, 2002.

[3] A. Sole, O Mano, G.P. Stein, H. Kumon, J. Tamatsu, and A. Shashua. Solid or not solid: vision for radar target validation. *IEEE Intelligent Vehicles Symposium Proceedings*, June 2004.

[4] A. Kuehnle. Symmetry-based vehicle location for ahs. *SPIE - Transportation Sensors and Controls: Collision Avoidance, Trafic Management, and ITS Proceedings*, 1998.

[5] T. Zielke, M. Brauckmann, and W. von Seelen. Intensity and edge-based symmetry detection with an application to car-following. *CVGIP: Image Understanding*, 1993.

[6] L. Bombini, P. Cerri, P. Medici, and G. Alessandretti. Radar-vision fusion for vehicle detection. *3rd International Workshop on Intelligent Transportation (WIT 2006)*.

[7] U. Hofmann, A. Rieder, and E.D. Dickmanns. Radar and vision data fusion for hybrid adaptive cruise control on high-ways. *Intern. Conference on Computer Vision Systems*, 2001.

[8] L. A. Goodman. On the exact variance of products. *Journal of the American Statistical Association*, 1960.

[9] L. A. Goodman. The variance of the product of k random variables. *Journal of the American Statistical Association*, 1962.

[10] S. Matzka and R. Altendorfer. A comparison of track-to-track fusion algorithms for automotive sensor fusion. In *Proc. of International Conference on Multisensor and Integration for Intelligent Systems , 2008 IEEE*, pages 189–194, aug. 2008.

[11] S. A.H. McMullen D. L. Hall. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Boston, London, 2004.

[12] Y. Bar-Shalom and X. Li. *Multitarget-Multisensor-Tracking: Principles and Techniques*. YBS, 1995.

[13] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly, vol. 2*, 1955.

[14] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 1957.

[15] Smith, Randall, and Cheeseman. On the representation and estimation of spatial uncertainty. *Int. J. Rob. Res.*, 5(4):56–68, 1986.

[16] J. Hayya, D. Armstrong, and N. Gressis. A note on the ratio of two normally distributed variables. Technical report, 1975.