Environment-aware Optimization of Track-to-Track Fusion for Collective Perception

Georg Volk[1], Jörg Gamerdinger[1], Alexander von Bernuth[1], Sven Teufel[1] and Oliver Bringmann[1]

***Abstract*— Correct and complete perception is key for autonomous vehicles to plan safe maneuvers. Especially under harsh weather conditions the use of sensing capabilities from other road users via vehicle-to-everything (V2X) communication can contribute to more complete perception. However, information from other road users may contain additional uncertainties and lead to less accurate perception. Additionally, attackers may use the V2X channel to transmit malicious data.**

For building an accurate environmental model an autonomous vehicle needs as precise information as possible. To tackle the problems of additional uncertainties within collective perception we propose a methodology to check perceived information for its trustworthiness and validity. This is achieved by evaluating the perception capabilities of a holistic perception pipeline and checking collectively transmitted information for consistency. The proposed approach is evaluated under varying environmental conditions on a simulated highway scenario.

I. INTRODUCTION

Autonomous vehicles play a key role in future mobility. With the help of driverless cars it is possible to drastically increase the utilization of vehicles, hence reducing the overall costs for consumers. Furthermore, autonomous vehicles aim to improve the safety of transportation and make traveling more comfortable. To fulfill such promises an autonomous vehicle needs to have a complete and correct understanding of its environment to plan comfortable and safe maneuvers. However, road traffic developments, such as buildings at intersections, obstruct the vehicle-local field of view and make environment perception more challenging. Environmental influences like harsh weather conditions further reduce the local sensing range and perception capabilities [1]. A solution to cope with such difficult situations is the usage of collective perception (CP). It enhances local perception with additional sensor information from other vehicles or infrastructure elements transmitted by vehicle-to-everything (V2X) communication.

Information from foreign sources has to be correctly fused with the local dynamical model (LDM). Compared to vehicle-local sensor fusion, CP introduces additional difficulties such as bigger latencies, larger errors due to coordinate transformation, and larger or even unknown uncertainties. A common way to deal with CP is to transmit perceived object tracks from the LDM via V2X [2], [3]. This aids the compensation of transmission delays as it allows to align the data by predicting object positions with a given movement model. Afterwards a track-to-track fusion (T2TF) algorithm such as covariance intersection (CI) can be used to incorporate locally and collectively perceived information.

Additionally to the aforementioned problems the trustworthiness, validity, and confidence of the transmitted data is of great importance. The fewer tracks for a single object are transmitted, the higher the influence of an inaccurate state estimation from other cooperative vehicles can get. Hence, it is important to only fuse that information with the LDM which contributes to a more complete and accurate understanding of the vehicle environment. Furthermore, the influence of inaccurate information from another vehicle may also be a vulnerability which can be used by an attacker.

The goal of this work is to investigate different optimization strategies for the T2TF, which are shown using the example of CI. The strategies consider additional validity and consistency checks of collectively transmitted object tracking information to only fuse accurate and correct information. Besides the optimization of perception precision a reduction of vulnerability against attackers is achieved. Possible tracking-confidences and covariances will be pre-evaluated. Current environmental influences such as snow, fog, or rain in combination with the distance of the detected object are considered for a consistency check of confidence and covariance matrices transmitted from other vehicles. Invalid, inaccurate, or impossible state estimations will be neglected to improve the accuracy of the collectively perceived environmental model.

II. RELATED WORK

Collective perception increases the complexity of the assignment problem of measurements to tracks as well as the tracking and fusion of data. For tracking and fusion there exist two basic methodologies. The first is having a centralized tracking component which handles sensor data directly [4]. The second approach uses decentralized tracking components and fuses preprocessed sensor data which is available as tracks. For CP the second approach (T2TF) has the advantage that more information about object dynamics are present and V2X transmission latencies can be compensated [4].

For T2TF there exist different approaches solving the data fusion problem. CI was one of the first fusion method under unknown correlations and was introduced by Julier and Uhlmann [5].

The calculation of the global fused state $\hat{x}_g$ with covariance matrix (CM) P_g for two local trackers i and j with state vectors $\hat{x}_i, \hat{x}_j$ and corresponding covariance matrices P_i, P_j

[1]University of Tübingen, Faculty of Science, Department of Computer Science, Embedded Systems Group {georg.volk, joerg.gamerdinger, alexander.von-bernuth, sven.teufel, oliver.bringmann} @uni-tuebingen.de

is defined by Julier and Uhlmann [5] as

$$\mathrm{P}_g^{-1} = \omega \mathrm{P}_i^{-1} + (1-\omega)\mathrm{P}_j^{-1} \tag{1}$$

$$\hat{\mathrm{x}}_g = \mathrm{P}_g(\omega \mathrm{P}_i^{-1}\hat{\mathrm{x}}_i + (1-\omega)\mathrm{P}_j^{-1}\hat{\mathrm{x}}_j) \tag{2}$$

$$\omega = \arg\min \det \mathrm{P}_g,\ \omega \in [0,1]. \tag{3}$$

For the calculation of the weight ω it is also possible to minimize the trace (tr) of P_g instead of the determinant (det). The trace of a matrix corresponds to the sum of its diagonal values. For a matrix A with size $\mathrm{M} \times \mathrm{M}$, $\mathrm{tr}(A)$ is defined as

$$\mathrm{tr}(A) = \sum_{i=1}^{M} A(i,i). \tag{4}$$

Since minimizing Eq. 3 leads to high computational effort, some improvements for the approximation of ω have been introduced.

To improve the calculation of ω the "Fast Covariance Intersection" (FCI) by Niehsen [6] uses the traces of the CMs $\mathrm{P}_i, \mathrm{P}_j$. ω is defined as

$$\omega = \frac{\mathrm{tr}(\mathrm{P}_j)}{\mathrm{tr}(\mathrm{P}_i) + \mathrm{tr}(\mathrm{P}_j)}. \tag{5}$$

Besides the trace it is also possible to use Eq. 5 with the determinant of P_i and P_j to calculate ω.

For three or more tracks, multiple weights ω_j have to be calculated. For n tracks, the following constraint must hold

$$\sum_{j=1}^{n} \omega_j = 1.0. \tag{6}$$

As presented by Cong et al. [7], ω_j for up to n tracks, can be defined as

$$\omega_j = \frac{\epsilon_j}{\sum_{i=1}^{n} \epsilon_i}, j = 1, 2, \ldots, n \tag{7}$$

with

$$\epsilon_j = \frac{1}{\mathrm{tr}(\mathrm{P}_j^{-1})}. \tag{8}$$

The "Sequential Fast Covariance Intersection" (SFCI) by Cong et al. [7] allows a sequential fusion of data received from other sensors at different points in time. Cong et al. [7] also prove that SCFI leads to accurate and consistent results which are independent from the fusion order.

Fränken and Hüpper [8] introduced the "Improved Fast Covariance Intersection" (IFCI) which uses the information matrix (I) to define ω as

$$\omega = \frac{\det(\mathrm{I}_i + \mathrm{I}_j) - \det(\mathrm{I}_j) + \det(\mathrm{I}_i)}{2\det(\mathrm{I}_i + \mathrm{I}_j)}. \tag{9}$$

The information matrix corresponds to the inverse of the CM. For IFCI it is not possible to use the trace instead of the determinant [8].

As presented by Reinhardt et al. [9], CI is not necessarily optimal for three or more tracks which is shown by a counter-example.

Besides the CI, Uhlmann presented the covariance union (CU) fusion [10]. As explained by Castanedo [11], the CI addresses only the problem of correlated inputs but not inconsistent inputs. Two estimations are considered as inconsistent when the difference between the estimation is larger than the provided covariance [11]. To detect inconsistent inputs in CU, the mahalanobis distance can be used [12].

Another approach for T2TF with consideration of the cross covariance is presented by Tian and Bar-Shalom [13]. It is possible to use a recursive strategy to include the fused data from previous fusion iterations. If a strategy without previous information is used, fundamental equations from linear state estimation can be used. For this purpose all tracks must be from the same point in time, therefore a prediction to the time of fusion is performed. However, the use of cross covariance fusion does not provide the best global state estimation [4].

A centralized fusion concept is the "Information Matrix Fusion" (IMF) by Speyer [14]. IMF uses information from previous fusion iterations. For the fusion of the estimated state, the a priori and a posteriori state estimation of a local tracker must be known [4]. Nevertheless, IMF has shown to be less robust compared to CI and its optimized form IFCI regarding their fusion results [4].

In non-linear systems it is difficult to find the correlation between estimations. Hence, Noack el al. [15] presented a work in which the estimations are transformed into another state space with Gaussian densities such that correlation and covariance can be estimated which enables the usage of CI in non-linear systems.

The sample-based fusion by Steinbring et al. [16] uses a small set of samples, which is stored at each sender and be transmitted together with a state estimation. This enables a simple and correct reconstruction of the correlations of an estimation at a central fusion node.

An overview of different methods for data association, state estimation, and data fusion is presented by Castanedo [11]. A comparative study on T2TF methods is presented by Radtke et al. [17]. Five different algorithms are tested in an indoor localization setup, where the above-mentioned sample-based fusion [16] achieved the best results; CI on the other hand provides rather sub-optimal results. This can also be observed in the comparison of "Safe Fusion" for correlated inputs towards other T2TF methods as presented by Nygårds et al. [18]. In some cases the CI performs equal compared to other methods.

Since the performance of the CI is not sufficient for the requirements in automated driving and moreover is considered insecure towards attacks by forged data, we introduce optimization strategies to compensate these weaknesses.

III. PERCEPTION OPTIMIZATION

As seen in Sec. II multiple approaches to fuse data from different origins exist. The goal of this work is to improve robustness of collective perception against environmental influences and malicious attacks by forged data. Therefore, the robust but yet non-optimal CI data fusion shall be optimized such that only accurate and trustworthy data will be used to contribute to the collectively perceived environmental model. The validation of transmitted data will happen based on

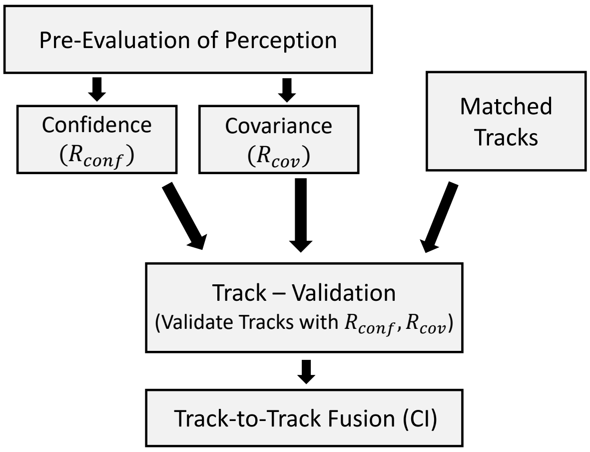


Fig. 1: Process overview from the pre-evaluation (see Sec. III-B) of a local perception over the validation of the received tracks in the track-validaton (see Sec. III-C) to the T2TF fusion.

a pre-evaluation of a local perception pipeline. With the pre-evaluation the capabilities of local perception systems are analyzed such that the T2TF algorithm is capable of evaluating the trustworthiness and validity of the collectively transmitted data before fusing it.

An overview of the complete pipeline of our suggested optimization is presented in Fig. 1. The pre-evaluation (see Sec. III-B) is performed to generate reference data $\mathrm{R}_{conf}/\mathrm{R}_{cov}$ for confidence and CMs. The reference data, paired with the matched tracks, is fed into a track-validation. The track-validaton performs the optimization strategies (see Sec. III-C) before the fusion is executed.

In the following, the used perception pipeline for local and collective perception and the pre-evaluation are described subsequently the suggested optimization strategies are presented.

A. Perception Pipeline

For simulation and investigation of CP the RESIST framework [19] with its improvements for environment-aware collective perception [1] and a sophisticated communication channel simulation and processing delays [20] is employed. The framework's main focus lays on realistic perception with simulation of environmental influences such as rain, snow, and fog [21]. Each cooperative vehicle within simulation performs local detection and tracking. Tracking is performed by a Kalman-filter [22] with a constant-velocity model [23]. Hence, track-state and CM consist of the 3D-position p (in m) and the 3D-velocity v (in $\mathrm{m/s}$). Hungarian matching [24] is used to associate new detections to present tracks. Locally perceived information is transmitted by V2X. For V2X channel simulation the corresponding channel load is computed based on the analytical model of IEEE 802.11p developed and validated by Sepulcre et al. [25]. Afterwards the collectively perceived objects are temporally aligned and transformed into the ego vehicles local coordinate system to perform a matching to the objects present in the ego vehicle's LDM. Finally a T2TF is performed to build up the collectively perceived environmental model.

B. Pre-Evaluation of Local Perception

The foundation of the proposed approach is the pre-evaluation of local perception capabilities. Hence, we assume that the sensor configuration and the used processing pipeline for local perception of cooperative vehicles is known, such that a sound analysis of its capabilities is possible. A cooperative vehicle v has a specific vehicle local perception system l_v. Furthermore, knowledge about the current environmental condition such as fog including its intensity defined as e is necessary.

The proposed approach investigates all perception systems l_v considering their perception accuracy in terms of covariance matrices of perceived objects and perception capabilities in terms of confidence of a track. The perception systems are evaluated considering different environmental conditions e and the perceived objects are clustered in distance bins d of size $s_{bin} = 5\,\mathrm{m}$. The bin size was chosen since it roughly corresponds to the length of a vehicle. Splitting the distances into sections is necessary because it is not possible to create reference information for every distance.

Additionally, adverse weather is considered for evaluation, as it has a significant influence on the general perception quality [1]; it decreases the confidence in detections while also increasing the variances in the CMs.

The results of a local perception are analyzed to get realistic and comparable confidences and CMs, to rate if a received track from a cooperative vehicle is plausible and can be considered as trustworthy and valid for the fusion. From these results two weather-related lookup-tables for the local perception capabilities $\mathrm{R}_{\mathrm{conf}}(l_v, e, d)$ and $\mathrm{R}_{\mathrm{cov}}(l_v, e, d)$ of each specific perception system l_v are built. For better readability the pre-evaluation $\mathrm{R}_{\mathrm{cov}}(l_v, e, d)$ for a specific distance bin and weather is abbreviated as $\mathrm{R}_{\mathrm{cov}}$ in the following. $\mathrm{R}_{\mathrm{conf}}(l_v, e, d)$ is similarly abbreviated as $\mathrm{R}_{\mathrm{conf}}$.

As track confidence the achieved recall [26] at distance bin d for l_v is used. To be classified as true positive detection for recall calcualtion, an Intersection over Union [27] greater 0.5 must be achieved.

An entry of the confidence matrix $\mathrm{R}_{\mathrm{conf}}$ at position (e, d) describes the pre-evaluated recall for a specific weather rate e and a specific distance bin d.

The above described splitting is also used for $\mathrm{R}_{\mathrm{cov}}$, which only describes the variances of the different track-state items because there exists no information about the correlation between the state-items. A set of CMs is produced for each environmental condition e, with each matrix representing a different distance bin d.

An exemplary reference CM $\mathrm{R_{cov}}$ is shown below:

$$\mathrm{R_{cov}} = \begin{pmatrix} \sigma^2_{p.x} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{p.y} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{p.z} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{v.x} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{v.y} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{v.z} \end{pmatrix}$$

As an evaluation baseline, a cloudy day without rain is used. A local perception with varying fog concentrations is performed to investigate the optimization strategies under various weather conditions. Therefore, fog is simulated with a drop diameter of $0.7\,\mu\mathrm{m}$ and densities from $0.01/\mu\mathrm{m}^3$ to $0.15/\mu\mathrm{m}^3$ with a step size of $0.01/\mu\mathrm{m}^3$.

Even with a comprehensive data set, it is possible that no objects were present for particular distances, and hence no reference data could be derived. As a result, we use a linear interpolation to fill in missing confidence and covariance values by interpolating between recall and variance values.

C. Optimization Strategies

Based on the pre-evaluation of Sec. III-B, our proposed method compares collectively received tracks to pre-evaluated matrices $\mathrm{R_{conf}}$ and $\mathrm{R_{cov}}$. This allows a sophisticated validation of the perceived data to improve the resilience of unoptimized data fusion methods against harsh environmental conditions. Two different optimization approaches will be investigated: a simple selection of tracks without using additional information and a more advanced filtering based on the pre-evaluation data. We will demonstrate our approach with the example of CI fusion described above.

Similar to the CU fusion, we consider the quality of a local track to decide whether it is fit for fusion. We examine the difference between the confidence and covariance from the local tracks to a large set of pre-evaluated confidence and covariance matrices for various weather conditions. The strategies mainly concern the optimization of precision, some of the strategies additionally avoid an influence of a possible attack.

*1) **Track Selection**:* As stated by Reinhardt et al. [9], the CI is not necessarily optimal for three or more tracks. For many tracks some information can deviate widely and therefore affect the fusion result negatively. Hence, one optimization strategy considers the reduction of tracks used for the fusion to two. Two ways are considered to select the tracks used for fusion, they are based on the confidence and the CM. A first strategy uses the two tracks with the highest confidence; the other strategy uses the two estimations with the minimal trace (see Eq. 4) of their CM. This optimization method can be used with or without a pre-evaluation of the given confidences respectively CMs. With a pre-evaluation, the given values are filtered after being plausibilized with the reference data. This variant is relatively simple and leads to a reduction in noise, since bad or noisy information is discarded if there are enough tracks present. If the number of cooperative vehicles is low, then the number of matched tracks for an object is also small. In this case the track selection does not affect the fusion since if less than three tracks are available, no selection is possible. Also this optimization strategy only addresses the optimization of quality but not malicious data. The track selection would consider a forged track with a high confidence or low covariance for the fusion.

*2) **Track Filtering**:* The second investigated optimization is a filtering of the received information based on a validation with the pre-evaluated reference data ($\mathrm{R_{conf}}$, $\mathrm{R_{cov}}$). This strategy covers both the fusion precision and the security aspect. Within this strategy we distinguish between the confidence based and CM based validation.

Each perception received from a cooperative vehicle has an assigned confidence, which describes the trustworthiness of the detection. As described in Sec. I, distant objects are perceived with less quality [1]. Hence, for higher distances a lower confidence is expected. If an attacker transmits a forged track with a high confidence to ensure that the information is used for the fusion, this can be validated by $\mathrm{R_{conf}}$. The pre-evaluation provides information about how likely a detection in a specific distance is; a received information with a distant detection and high confidence is unlikely. Hence, we can take the corresponding information from $\mathrm{R_{conf}}$ and verify the received information with a defined threshold. The threshold value here applies independent to the distance. Different confidence thresholds are investigated. If the confidence of the received track differs from the confidence specified in $\mathrm{R_{conf}}$ by more than the threshold, the track is discarded. This filtering enables an increase in fusion quality, because unlikely state estimations which probably cause an error will not be used for fusion. Furthermore, filtering implausible confidences could remove knowingly forged information in case of an attacker.

This approach can be transferred to the CMs of the received tracks. The pre-evaluation also provides reliable CMs in relation to distance and weather condition specified in $\mathrm{R_{cov}}$. To validate the covariance matrices, a comparison by the trace or elementwise by the main diagonal is possible.

The main diagonal of the CM represents the variances of the track-state items. Higher variances occur for inaccurate estimations such as for distant objects or objects partially concealed by other objects. Hence, these variances describe the accuracy of the corresponding state estimation. In order to achieve a fusion which is as accurate as possible, inaccurate estimations should not be considered. Based on the weight calculation of the CI (see Eq. 7) the influence of such inaccurate state estimations on the fused state is small but exists and thus should be removed to improve the overall fusion result.

A larger trace indicates a higher variance in estimation; thus an estimation with lower trace is preferred. If the trace of the received information exceeds the pre-evaluated trace by a given threshold, the received information is considered as inaccurate and thus will not be used for the fusion. Since some variance is valid, the threshold should not be chosen too small. Different trace thresholds

$$t_{\text{trace}} \in \{2.0, 3.0, 4.0, 5.0, 6.0\}$$

which apply for all distances are investigated.

Besides the trace, the elements of the main diagonal can be used elementwise for validation. Therefore, a threshold for each element must be defined. For an elementwise comparison a list of thresholds

$$t_{\text{elem}} = \{1.5, 0.8, 0.2, 2.0, 1.0, 0.3\}$$

is defined. Both t_{trace} and t_{elem} were defined based on the observations from the pre-evaluation. For the elementwise comparison a track s $(\hat{x}_s, \mathrm{P}_s)$ is discarded if for one element of the CM P_s deviates from the reference by more than the corresponding threshold. In combination with the filtering by the trace the following two conditions can be formulated:

$$\begin{gathered} \operatorname{tr}(\mathrm{P}_s) - \operatorname{tr}(\mathrm{R}_{\text{cov}}) > t_{\text{trace}}, \text{ or} \\ \mathrm{P}_s(i,i) - \mathrm{R}_{\text{cov}}(i,i) > t_{\text{elem}}(i) \text{ for } i = 1, 2, \ldots, 6 \end{gathered}$$

If one of the conditions applies, the track is considered inaccurate and discarded.

The advantage of the elementwise comparison towards the comparison by trace is the flexibility and consideration of deviation in more detail. This allows to fit the filtering exactly to the specified local perception system l_v from pre-evaluation. Also single deviating values can be discovered better.

In case of an attack on the communication channel, an attacker tries to influence the fusion result. Hence, a forged track would be sent alongside a significant low CM, because this would lead to a higher influence (see Eq. 7). As explained, the variances which represent the main diagonal of the CM are smaller the more accurate the state estimation is. In contrast to the above described optimization for inaccurate estimations, R_{cov} must not exceed the given track variances by more than the defined thresholds. If one of the two following constraints for the given CM P_t of a collectively perceived track and the corresponding matrix R_{cov} is evaluated as true, the track should be considered as invalid and must be discarded.

$$\begin{gathered} \operatorname{tr}(\mathrm{R}_{\text{cov}}) - \operatorname{tr}(\mathrm{P}_s) > t_{\text{trace}}, \text{ or} \\ \mathrm{R}_{\text{cov}}(i,i) - \mathrm{P}_s(i,i) > t_{\text{elem}}(i) \text{ for } i = 1, 2, \ldots, 6 \end{gathered}$$

IV. RESULTS

To show the advantages of the proposed optimizations for T2TF at the example of CI fusion, an investigation on a complex highway scenario with varying fog densities and number of cooperative vehicles is conducted. As a baseline, covariance intersection without any optimization is used. All received tracks are used for fusion; for the fusion with multiple weights the proposed weight calculation of Eq. 7 is applied. For the evaluation a 400 Frame VIRES VTD highway scenario with 36 vehicles in total is used from which 21 are cooperative vehicles equipped with sensors [1]. Three different rates of cooperative vehicles (1 vehicle (5.6 %), 5 vehicles (16.7 %), and 10 vehicles (30.6 %)) are used to compare the influence of different optimization strategies. These rates were chosen in such a way that both very low and relatively high rates, which will not be reached in the near future, are examined in order to be able to make future-proof statements. The average velocity is about $14\,\mathrm{m/s}$. The local perception including the tracking is performed as described in Sec. III. At first a theoretical evaluation of the approach on the example of a special use case will be shown. Afterwards the performance metrics “precision” and “recall” in combination with the variance of the state estimation will be used for evaluation.

A. Optimization Use-Case

The CI T2TF is vulnerable to inaccurate or forged information from an attacker (see Sec. I). This use case will cover an attacker scenario. However, the attacker scenario is also applicable to imprecise tracking information. The degree of influence of forged information is determined by the number of tracks, the track CMs, and the deviation of the forged information.

If an attacker sends a forged track with a low CM, this estimation is considered as good and is used for fusion. The use of such forged data can cause considerable deviations in fusion results. Filtering falsified information as invalid, based on the validation of the confidences and CMs of received state estimations could avoid an attacker’s effect. A forged track with a CM close to zero at $100\,\mathrm{m}$ distance perceived in fog with high density could be an example of such an infeasible state estimation.

The following example demonstrates CI’s vulnerability and the reduction of this vulnerability through our suggested methods. We assume two correct tracks with the states

$$\begin{aligned} \hat{\mathrm{x}}_{c_1} &= \{19.16, 2.37, 0.21, 9.34, -0.12, -0.09\} \\ \hat{\mathrm{x}}_{c_2} &= \{19.10, 2.39, 0.25, 9.30, -0.16, -0.03\} \end{aligned}$$

and the same arbitrary but valid CM P_{ct} with $\operatorname{tr}(\mathrm{P}_{ct}) = 7.77$.

The track state $\hat{\mathrm{x}}$ consists of position and velocity

$$\hat{\mathrm{x}} = \{p.x, p.y, p.z, v.x, v.y, v.z\}.$$

The fusion of these two states would lead to a correct fused state

$$\hat{\mathrm{x}}_{c_f} = \{19.13, 2.38, 0.23, 9.32, -0.14, -0.06\}.$$

Now we include a forged track

$$\hat{\mathrm{x}}_a = \{20.10, 3.37, 1.22, -25.30, -5.16, -1.03\}$$

with the CM P_a with trace $\operatorname{tr}(\mathrm{P}_a) = 0.77$ from an attacker which is about $50\,\mathrm{m}$ away.

The exact values of the CMs are less relevant in this case, more important is that the values of P_a are significantly smaller than in P_{ct} and close to zero. Since matching only takes position into account, a significant deviation of the velocity has no effect on matching. If this track is included into the fusion with the correct tracks $\hat{\mathrm{x}}_{c_1}$ and $\hat{\mathrm{x}}_{c_2}$, it would lead to an imprecise fused state

$$\hat{\mathrm{x}}_{w_f} = \{19.60, 2.87, 0.63, -7.95, -2.67, -0.53\}.$$

TABLE I: Overall recall [%] for a cloudy day, varying fog densities and different rates of cooperative vehicles

	Cloudy	Fog Density (1/μm^3) 0.01	0.07	0.13
5.6 % Coop. Vehicles				
Baseline	14.1	**12.8**	**5.1**	**1.8**
2TracksConf	**15.0**	**12.8**	**5.1**	**1.8**
2TracksCov	**15.0**	**12.8**	**5.1**	**1.8**
FilterTrace4	12.0	9.8	2.6	0.9
FilterTrace5	13.8	11.0	3.6	1.8
FilterElement	2.8	2.0	0.1	0.2
FilterConf0.2	3.2	**12.8**	**5.1**	**1.8**
16.7 % Coop. Vehicles				
Baseline	23.4	**22.4**	**15.6**	**6.9**
2TracksConf	**27.0**	21.6	**15.6**	6.8
2TracksCov	23.4	22.2	**15.6**	6.8
FilterTrace4	20.4	18.6	12.0	4.8
FilterTrace5	21.2	20.0	13.5	6.7
FilterElement	10.9	9.7	7.7	3.3
FilterConf0.2	14.9	22.4	**15.6**	**6.9**
30.6 % Coop. Vehicles				
Baseline	28.0	**28.3**	22.0	**10.9**
2TracksConf	28.7	27.6	**22.1**	**10.9**
2TracksCov	**29.9**	28.1	21.8	10.8
FilterTrace4	25.4	25.5	20.0	8.5
FilterTrace5	26.1	26.4	21.4	10.4
FilterElement	18.4	18.6	16.5	6.3
FilterConf0.2	22.9	**28.3**	22.0	**10.9**

TABLE II: Overall precision [%] for a cloudy day, varying fog densities and different rates of cooperative vehicles

	Cloudy	Fog Density (1/μm^3) 0.01	0.07	0.13
5.6 % Coop. Vehicles				
Baseline	71.8	72.0	79.9	50.0
2TracksConf	66.1	72.2	79.9	50.0
2TracksCov	66.6	72.0	79.9	50.0
FilterTrace4	70.3	75.1	74.4	71.9
FilterTrace5	73.1	71.8	75.8	50.0
FilterElement	**97.3**	**94.4**	**100.0**	**100.0**
FilterConf0.2	74.1	72.0	79.9	50.0
16.7 % Coop. Vehicles				
Baseline	61.2	58.9	67.4	67.7
2TracksConf	53.5	56.8	67.6	67.3
2TracksCov	61.0	58.4	67.6	67.3
FilterTrace4	62.4	60.3	73.2	77.2
FilterTrace5	61.9	58.4	71.2	70.7
FilterElement	**82.4**	**80.6**	**94.7**	**89.4**
FilterConf0.2	66.7	58.9	67.4	67.7
30.6 % Coop. Vehicles				
Baseline	56.5	54.1	57.4	68.8
2TracksConf	46.7	52.8	57.5	68.8
2TracksCov	48.7	53.7	56.9	68.3
FilterTrace4	57.4	58.1	65.1	77.2
FilterTrace5	58.0	55.6	64.6	72.1
FilterElement	**73.3**	**75.5**	**85.8**	**91.5**
FilterConf0.2	55.8	54.1	57.4	68.8

The estimation error vector

$$\hat{\mathrm{x}}_{\mathrm{error}} = \{0.47, 0.49, 0.40, -17.27, -2.53, -0.47\},$$

shows significant deviation to the correct state estimation $\hat{\mathrm{x}}_{c_f}$.

It could be observed that the received estimation seems to be invalid when comparing the CMs elementwise to the pre-evaluated data on the main diagonal with the difference threshold vector t_{elem} (see Sec. III-C.2). The trace of the corresponding reference CM $\mathrm{R}_{\mathrm{cov}}$ is $\mathrm{tr}(\mathrm{R}_{\mathrm{cov}}) = 6.05$. Also by difference of the traces with a threshold t_{trace} of less than 5.28, P_a is considered as invalid and thus $\hat{\mathrm{x}}_a$ will be ignored for the fusion.

It must be stated that the vulnerability of forged data cannot be completely avoided. The attacker information will be used for fusion if there is a reasonable CM, although the influence of such a false track state could be significantly decreased by the filtering of invalid tracks. Furthermore, this example only illustrates the attacker use case. For imprecise information the validation can be used in a similar way, with the difference of filtering imprecise data instead of data which can not be perceived that precisely given harsh environmental conditions.

B. Optimization Strategies

The presented optimization strategies in Sec. III-C are evaluated with the following parameterization:

- **Baseline** CI with all tracks
- **2TracksConf** Track selection by confidence
- **2TracksCov** Track selection by CM
- **FilterTrace4** Track filtering with $t_{\mathrm{trace}} = 4.0$
- **FilterTrace5** Track filtering with $t_{\mathrm{trace}} = 5.0$
- **FilterElement** Track filtering with t_{elem}
- **FilterConf0.2** Track filtering by confidence diff > 0.2

The results on recall and precision metrics are illustrated in Table I and II respectively. The baseline case represents the original CI as a reference. The two optimization alternatives for track selection and filtering are discussed in the following. Some exemplary results for the variances of the position estimation are presented in Fig. 2. Fig. 2a shows a baseline evaluation for a cloudy day with 10 cooperative vehicles, Fig. 2b and Fig. 2c show the evaluations of *2TracksCov* and *FilterTrace4* with the same conditions.

*1) **Track Selection**:* Two different track selection approaches were investigated. *2TracksConf* considers only the best 2 tracks per object based on their confidence. *2TracksCov* also considers the best two tracks but based on the CM. It can be seen that the track selection achieves almost similar results in recall and precision compared to the baseline. For a fog density of $0.07/\mu\mathrm{m}^3$ and more than $16.7\,\%$ cooperative vehicles recall as well as precision show slight increases. For track selection based on CM the results show similar recall and precision values as well. These results indicate that there are not more than two simultaneous perceptions for the same object in most situations. Hence, only considering the best two tracks makes one marginal difference compared to the baseline.

For *2TracksCov* at some distances a reduction in variance of position can be observed as presented in Fig. 2b. For a distance of $50\,\mathrm{m}$ the variance in x direction was reduced from $2.2\,\mathrm{m}$ to $1.7\,\mathrm{m}$ compared to the baseline case in Fig. 2a. A similar reduction from $6.5\,\mathrm{m}$ to $5.5\,\mathrm{m}$ can be observed for a distance of $125\,\mathrm{m}$.

*2) **Track Filtering**:* For track filtering three different approaches were investigated. Filtering based on the trace of the CM, elementwise filtering, and filtering based on the track confidence. For the trace-based approach two thresholds defined as *FilterTrace4* and *FilterTrace5* with a threshold of $t_{\mathrm{trace}} = 4.0$ and $t_{\mathrm{trace}} = 5.0$ were investigated. These thresholds were chosen as they were the most interesting from the defined t_{trace} in Sec. III-C. For $t_{\mathrm{trace}} = 4.0$ high decreases in recall for low number of cooperative vehicles have been observed. With more than $16.7\,\%$ cooperative vehicles recall only decreases slightly. However, precision could be increased drastically, especially for fog with high density. For a threshold of $t_{\mathrm{trace}} = 5.0$ the precision improvement compared to the baseline was partially lower on the one hand but on the other hand the decrease in recall was not that drastic compared to $t_{\mathrm{trace}} = 4.0$. Furthermore, an improvement in position estimation can be observed by looking at the variances shown in Fig. 2c. For a distance of $125\,\mathrm{m}$ the variance in x direction was reduced from $6.5\,\mathrm{m}$ to $5.0\,\mathrm{m}$ compared to the baseline case in Fig. 2a.

For elementwise filtering with t_{elem} defined as *FilterElement* the biggest improvements in precision were observed. However, this improvement in precision resulted in a severe drop of recall. Due to filtering inaccurate tracks, some objects were not detected at all. This resulted in a drop of recall on the one hand and a more accurate tracking represented by high precision rates on the other hand. If the accuracy of tracks is of greater importance compared to the number of correct detections, this optimization might be a viable solution. However, the influence on recall should not be neglected.

As last filtering-based approach the difference in confidence needs to be evaluated. Here, a threshold for the difference of 0.2 defined as *FilterConf0.2* was investigated. This threshold has been chosen as a deviation in tracking confidence by 0.2 is considered high compared to the tracking confidences found by our empirical studies on the highway scenario. With this approach the least difference compared to baseline could be observed for all fog densities. The results for cloudy showed a decrease in recall and improvements in precision for less than $30.6\,\%$ cooperative vehicles.

V. CONCLUSION & OUTLOOK

In this paper we presented different confidence and covariance based optimization strategies for track-to-track fusion which have been investigated on the example of covariance intersection. The suggested methods validate the received information from an arbitrary number of cooperative vehicles against local perception capabilities. This validation allows to reject and filter inaccurate estimations in order to improve the accuracy of the fused information and to discard invalid information which is received from a possible attacker. The presented approach has shown to reduce the recall on the one hand but improved the precision of data on the other hand as inaccurate information gets neglected.

The presented generic approach for validation of collectively perceived information can be used for resilience improvements not only for covariance intersection but other fusion methods as well. Therefore, other fusion algorithms shall be evaluated in the future. Furthermore, the validation of data is also a viable solution for filtering data before transmitting it. This may be a good solution for reducing the channel busy ratio when the communication channel is congested and will be investigated in the future. Moreover, the presented approach and the chosen parameters shall be evaluated on additional scenarios and environmental conditions. By combining the suggested optimization strategies additional improvements shall be investigated.

ACKNOWLEDGMENT

This work has been partially funded by the German Research Foundation (DFG) in the priority program 1835 under grant BR2321/5-2.

REFERENCES

[1] G. Volk, A. von Bemuth, and O. Bringmann, "Environment-aware Development of Robust Vision-based Cooperative Perception Systems," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. Paris, France: IEEE, Jun. 2019, pp. 126–133.

[2] A. Rauch, F. Klanner, R. Rasshofer, and K. Dietmayer, "Car2X-based perception in a high-level fusion architecture for cooperative perception systems," in *2012 IEEE Intelligent Vehicles Symposium*. Alcal de Henares , Madrid, Spain: IEEE, Jun. 2012, pp. 270–275.

[3] M. Aeberhard, S. Schlichtharle, N. Kaempchen, and T. Bertram, "Track-to-Track Fusion With Asynchronous Sensors Using Information Matrix Fusion for Surround Environment Perception," *IEEE Transactions on Intelligent Transportation Systems*, 2012, iSBN: 1524-9050.

[4] A. Rauch, "Entwicklung von Methoden für die fahrzeugübergreifende Umfelderfassung," Ph.D. dissertation, Universität Ulm, 2016.

[5] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068. International Society for Optics and Photonics, 1997, pp. 182–193.

[6] W. Niehsen, "Information Fusion based on Fast Covariance Intersection Filtering," in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997)*, vol. 2. IEEE, 2002, pp. 901–904.

[7] J. Cong, Y. Li, G. Qi, and A. Sheng, "An order insensitive sequential fast covariance intersection fusion algorithm," *Information Sciences*, vol. 367, pp. 28–40, 2016.

[8] D. Franken and A. Hupper, "Improved Fast Covariance Intersection for Distributed Data Fusion," in *2005 7th International Conference on Information Fusion*, vol. 1. IEEE, 2005, pp. 7–pp.

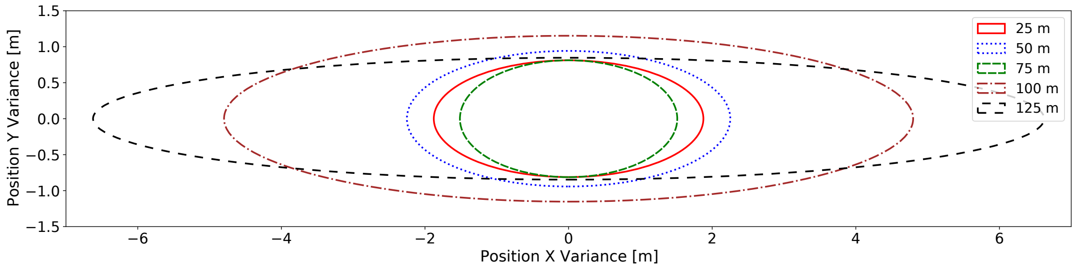
[9] M. Reinhardt, B. Noack, P. O. Arambel, and U. D. Hanebeck, "Minimum Covariance Bounds for the Fusion under Unknown Correlations," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1210–1214, 2015.

[10] J. K. Uhlmann, "Covariance consistency methods for fault-tolerant distributed data fusion," *Information Fusion*, vol. 4, no. 3, pp. 201–215, 2003.

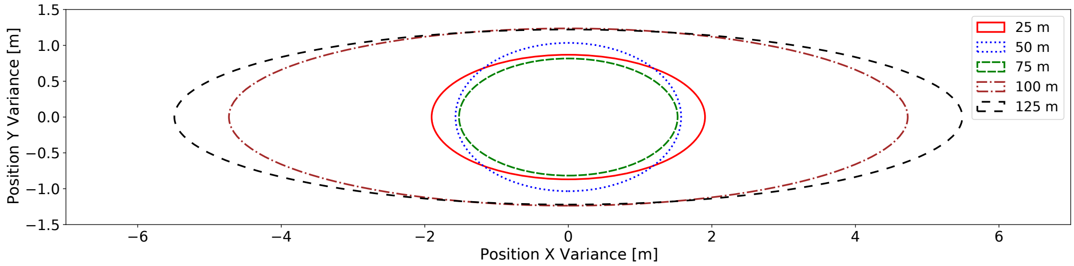
[11] F. Castanedo, "A Review of Data Fusion Techniques," *The scientific world journal*, vol. 2013, 2013.

[12] P. C. Mahalanobis, "On the generalized distance in statistics." National Institute of Science of India, 1936.

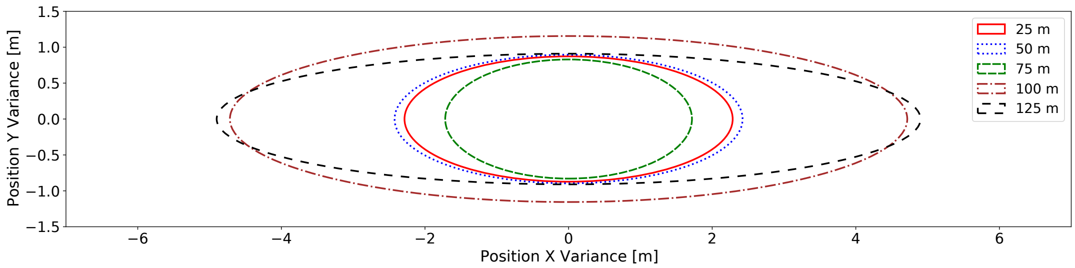
[13] X. Tian and Y. Bar-Shalom, "Exact algorithms for four track-to-track fusion configurations: All you wanted to know but were afraid to ask," in *2009 12th International Conference on Information Fusion*, 2009, pp. 537–544.



(a) *Baseline* evaluation for cloudy environment and 10 cooperative vehicles.



(b) *2TracksCov* evaluation for cloudy environment and 10 cooperative vehicles.



(c) *FilterTrace4* evaluation for cloudy environment and 10 cooperative vehicles.

Fig. 2: Variances of the fused position estimation (X and Y) in $[\mathrm{m}]$.

[14] J. L. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-gaussian control problem," *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pp. 1126–1131, 1978.

[15] B. Noack, M. Baum, and U. D. Hanebeck, "Covariance Intersection in Nonlinear Estimation based on Pseudo Gaussian Densities," in *14th International Conference on Information Fusion*. IEEE, 2011, pp. 1–8.

[16] J. Steinbring, B. Noack, M. Reinhardt, and U. D. Hanebeck, "Optimal Sample-based Fusion for Distributed State Estimation," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 1600–1607.

[17] S. Radtke, K. Li, B. Noack, and U. D. Hanebeck, "Comparative Study of Track-to-Track Fusion Methods for Cooperative Tracking with Bearings-only Measurements," in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. IEEE, 2019, pp. 236–241.

[18] J. Nygårds, V. Deleskog, and G. Hendeby, "Safe Fusion Compared to Established Distributed Fusion Methods," in *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2016, pp. 265–271.

[19] S. Mueller, D. Hospach, J. Gerlach, O. Bringmann, and W. Rosenstiel, "Framework for Varied Sensor Perception in Virtual Prototypes," in *MBMV*, 2015, pp. 145–154.

[20] G. Volk, Q. Delooz, F. A. Schiegg, A. Von Bernuth, A. Festag, and O. Bringmann, "Towards Realistic Evaluation of Collective Perception for Connected and Automated Driving," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1049–1056.

[21] A. v. Bernuth, G. Volk, and O. Bringmann, "Simulating Photo-realistic Snow and Fog on Existing Images for Enhanced CNN Training and Evaluation," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 41–46.

[22] G. Welch, G. Bishop *et al.*, "An introduction to the Kalman filter," 1995.

[23] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Transactions on aerospace and electronic systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[24] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[25] M. Sepulcre, M. Gonzalez-Martín, J. Gozalvez, and R. Molina-Masegosa, "Analytical models of the performance of IEEE 802.11p vehicle to vehicle communications," *ArXiv:2104.07923 [cs.NI]*, 2021, https://arxiv.org/abs/2104.07923.

[26] D. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[27] S. H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. D. Reid, and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," *CoRR*, vol. abs/1902.09630, 2019.