Evidential Trustworthiness Estimation for Cooperative Perception

Antoine Lima[0000−0001−8543−1528], Véronique Cherfaoui[0000−0003−2064−9838] and Philippe Bonnifait[0000−0002−5842−1399]

Université de Technologie de Compiègne, CNRS UMR 7253, Heudiasyc, France

Abstract. Intelligent Vehicles can exchange their perception information using wireless technology in a cooperative and decentralized manner. This has the potential to extend the range of perception and thus improve anticipation for complex driving maneuvers and decision making. However, information received from other peers can be erroneous and has to be used carefully. In this paper, we present a method that allows each peer to assign a trust in the information received from other peers based on comparisons with its current knowledge of the world. We describe how this process is managed using the Dempster-Shafer theory. We also present how positive and negative evidence cues can be developed in this problem, in particular by using detectability grids. An experimental evaluation, carried out with real vehicles, is reported to show that this formalism behaves correctly.

Keywords: Cooperative Perception· Trust· Multi-Robot System· Belief Functions.

1 Introduction

In order to navigate safely, intelligent vehicles need to perceive their environment. Their on-board sensors, such as cameras or LiDARs, are generally sufficient for local navigation tasks but not for more complex maneuvers because of the limited range of the sensors and because there are occlusions in their Field Of Views (FOVs). For example, in Fig. 1, v_1 cannot see if a vehicle is coming from behind the building on its right and will thus have to either be cautious and engage slowly or break strongly once the vehicle becomes visible. Cooperative Perception (CP) aims at improving the navigation performance in such situations by taking advantage of perceptual information captured by others. Indeed, using upcoming wireless technologies, it is possible for vehicles and the infrastructure to exchange perceptual information with each other. By integrating this information to its own, one's knowledge of its surroundings can be extended further and behind obstructions. For example, in Fig. 1, v_3 could warn v_1 of its presence and v_2 could warn v_1 that a vehicle is present in front of it. However, although the authenticity of peers can be cryptographically guaranteed using public-private key pairs [5], security vulnerabilities or perception malfunctions (e.g. sensors failures) can still generate erroneous information that should not be incorporated to one's own.

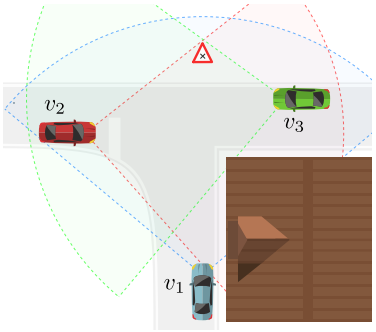


Fig. 1: Three cooperative vehicles at an intersection. v_1 cannot see v_3 because it is hidden by a building but v_2 and v_3 can see each other.

To prevent this, we propose an information processing and data fusion system that confronts the information received and the information from the embedded system to estimate the trustworthiness of the peers. This information can then be used in a cooperative tracker to attenuate or ignore information from untrustworthy peers. This process is done locally by each peer, without communicating its trust, by verifying that the received information matches with its knowledge of the world. For example, detecting objects at the same location creates trustworthiness while mismatching or illogical information creates untrustworthiness. In Fig. 1, because v_1 partially shares objects and FOV with v_2 and v_3, it will trust them and thus anticipate v_3 earlier.
After a review of related works in Section 2, we will introduce the problem at hand using a dense representation of the detectable or undetectable space from the point of view of different peers in Section 3. In Section 4, trustworthiness estimation will be formulated. Finally in Section 5, a simulation study and experimental results based on real data will be given and analyzed.

2 Related Works

The field of Cooperative Perception began with [7] demonstrating its potential for safety in intelligent transportation systems. Since then, the European Telecommunications Standards Institute (ETSI) standardized the Cooperative Perception Message (CPM) [6], composed of the sender position, sensor descriptions and a list of objects. It is used in many cooperative approaches, as studied in [4].
A part of the research effort is focused on preventing attacks as CP works on a public network. The most common form of attack prevention is misbehavior detection, as reviewed in [8]. In this paper, the authors list and classify numerous approaches as being standalone or shared, distributed or centralized and node or data centric. For example, in [12], errors are detected by comparing received positions with detections from embedded sensors. In [3], four levels of checks ranging from simple bound check to object comparison are used to emit

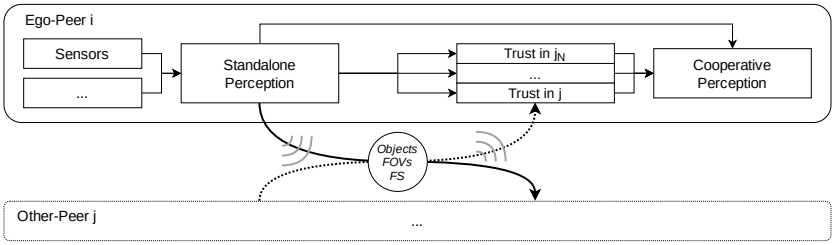
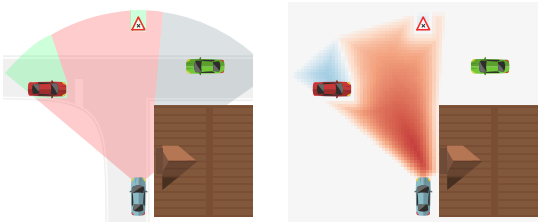


Fig. 2: Decentralized cooperative perception with trustworthiness estimation.

reports on misbehaving peers. When enough reports are received about a faulty peer, their certificate is revoked, excluding them from communication. In [2], a probability of trustworthiness is estimated for each pair of peers, by checking the consistency of their object lists and their respective detection probabilities across space. More recently, [11] compared occupied or free space in the form of grids and verified that detected objects matched with these grids. In [1], sensors estimate a probability of existence for each object. When fusing object from multiple sensors, they switch to an evidential representation and use a persistence probability to model the field of view of fused sensors. A trust parameter representing the sensor's information reliability is fixed for each sensors.
Our method can be seen as an unification of [3,2,11] where fault detection generates untrustworthiness and confirmation creates trustworthiness.

3 Problem Statement With Object Detectability

Consider a driving situation composed of N vehicles $v_1, v_2 \ldots$. Every cooperative vehicle perceives surrounding objects $o \in O$ and Free Space FS. Objects can be any kind of road user (e.g: pedestrians, cooperative or non-cooperative vehicles) or static features (e.g: traffic signs) whereas FS are areas explicitly characterized as being free. Objects and FS are broadcast to peers in the communication range, supposed to be further than the perception range. In addition, the Fields Of View (FOV) of the sensors are shared as per the CPM such that all exchanged information is vectorial in order to reduce the amount of data exchanged. Upon reception, each peer assesses the trustworthiness of the sender, as illustrated in Fig. 2.
As the problem at hand combines multiple point of views, a method to define which area was seen by the cooperating peers is needed. For this, we extend the detection probability of [2] with the capacity to state that there cannot be objects in the measured FS by using an evidential representation. In this representation, the ground plane is divided into cells of fixed size that contain a mass function ${}^i m^{\mathcal{D}}_{x,y}$ defined on $2^{\Omega^{\mathcal{D}}}$ where $\Omega^{\mathcal{D}} = \{D_i, \not{D}\}$. D_i represents that peer i can detect an object in the cell at position $[x, y]^T$ and $\not{D}$ that an object cannot be detected at that point. Conversely, the global grid is noted $m^{\mathcal{D}}$ and the detectability of an object is $m^{\mathcal{D}}_o = m^{\mathcal{D}}_{x_o,y_o}$.



(a) (b)

Fig. 3: Illustration of a detectability grid from the point of view of v_1. (a) Hidden area in grey, FS in red and detectable in green. (b) Resulting grid: non-detectability in red, detectability in blue and unknown in grey.

This grid (called detectability grid) is built with the process illustrated in Fig. 3. Outside the FOVs and behind buildings, the state is unknown. The FS characterized by local sensors (e.g. LiDAR points that hit the ground) is used to express the impossibility to detect objects within it. In space neither free nor unknown, an object is likely to be detected.
Object detectability is used in two different ways when receiving information. First, the sender's detectability grid is reconstructed to assess its objects coherency (e.g. there is no object in the FS or out of the FOV). Then, the receiver's detectability grid is constructed to verify the coherency of received objects and to only compare objects that are detectable by it and the sender.

4 Evidential Trustworthiness Estimation

Trustworthiness in the information sent by other peers is estimated by every peer individually as a mass function noted $m_j^{\mathcal{T}}$ about peer j. It is designed to be used in a subsequent cooperative fusion to ignore or discount objects originating from untrustworthy peers. As such, it is defined on $2^{\Omega^{\mathcal{T}}}$ with $\Omega^{\mathcal{T}} = \{T, \not{T}\}$ to express that information from j is trustworthy and can be integrated without hesitation or conversely not integrated at all. In the rest of this paper, trust mass functions are normalized and will be given in the following order: $T, \not{T}, \Omega^{\mathcal{T}}$.
Mass functions are particularly adapted to the problem at hand for several reasons. Firstly, similar to humans, trust evolves over time and can be forgotten when peers are not interacting anymore, which can be managed with discounting. Secondly, as new peers have an unknown degree of trustworthiness, choosing a wrong prior could lead to ignoring good information or including misleading one during transient phases. Finally, this provides more information for a subsequent cooperative tracker to make more or less cautious decisions when peers are only partly trustworthy (because their information contains both valid and invalid values).
Trustworthiness is sequentially estimated using an evidential network. Similar to state filtering, the current estimate at time t $m_j^{\mathcal{T}}(t|t)$ is derived as the combination of the previous estimate $m_j^{\mathcal{T}}(t-1|t-1)$ and new evidence about "Co-

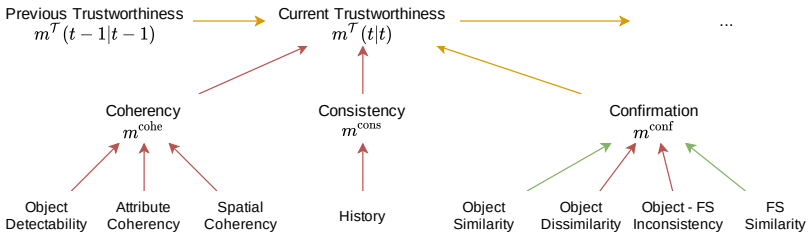


Fig. 4: Evidential network for trustworthiness computation at time t. Red arrows only convey untrustworthiness, green trustworthiness, and orange both.

herency", "Consistency" and "Confirmation" as illustrated in Fig. 4. They are respectively denoted m_j^{cohe}, m_j^{cons} and m_j^{conf}, defined on $2^{\Omega^{\mathcal{T}}}$ and group leaves of the network as described later on. Those leaves express simple and non-dogmatic constraints on either trustworthiness or non-trustworthiness based on different aspects 'of the received information.
In the combination process, every term is discounted by an associated factor that is not be explicitly noted here for the sake of clarity. Therefore, leaves always express some degree of belief on $m(\Omega^{\mathcal{T}})$. Dempster's rule $\oplus$ is well adapted in this case and is used for combination:

$$m_j^{\mathcal{T}}(t|t) = {}_{\Lambda_{\Delta t}} m_j^{\mathcal{T}}(t-1|t-1) \oplus m_j^{\text{cohe}} \oplus m_j^{\text{cons}} \oplus m_j^{\text{conf}} \tag{1}$$

where $\Lambda_{\Delta t}$ is a discounting factor that depends on the elapsed time Δt, moving an $\Lambda_{\Delta t}$-proportion of every focal set to the unknown [9].

4.1 Coherency

m^{cohe} models that the information contained in a message has to be coherent within itself. Multiple constraints are combined using Dempster's rule, three of them are given here as an example:

$$m_j^{\text{cohe}} = m_j^{\text{obd}} \oplus m_j^{\text{atc}} \oplus m_j^{\text{spc}} \tag{2}$$

m_j^{obd} expresses that objects cannot exist inside the FS or outside the perception range of the peer. For this, the detectability measure ${}^j m^{\mathcal{D}}$ of the sending peer j is used. For example, an object that is in the FS is by definition undetectable and its detectability will be low. Similarly objects outside the FOV are unknown and will have a low detectability. We use a constant $D^{\min}$ threshold to assign a mass on the untrustworthiness parametrized with a constant β^{pen} for such objects:

$$m_j^{\text{obd}} = \bigoplus_{\substack{o \in O_j \\ {}^j m_o^{\mathcal{D}}(D) < D^{\min}}} \begin{bmatrix} 0 & \beta^{\text{pen}} & 1-\beta^{\text{pen}} \end{bmatrix} \tag{3}$$

m_j^{atc} expresses that object attributes have to be likely. For example, the speed v_o of the object o has to be coherent with a normal behaviour which is handled

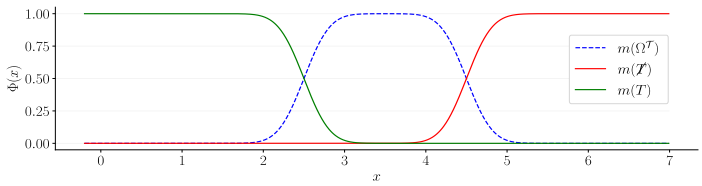


Fig. 5: Sigmoid function $\Phi(x)$ with parameters $\mu = 2$, $\sigma = 0.25$ and $\delta = 2$ producing an arbitrary mass function m.

by a scalar sigmoid function Φ illustrated in Fig. 5 with parameters μ^v, σ^v and δ^v chosen to reflect speed limits:

$$m_j^{\text{atc}} = \bigoplus_{o \in O_j} \Phi(v_o, \mu^v, \sigma^v, \delta^v) \tag{4}$$

with

$$\Phi(x, \mu, \sigma, \delta) = \left[m_1 = \frac{1 - \text{CDF}\left(\frac{x-\mu-2\sigma}{\sigma\sqrt{2}}\right)}{2} \; m_2 = \frac{1 + \text{CDF}\left(\frac{x-\mu-2\sigma-\delta}{\sigma\sqrt{2}}\right)}{2} \; 1 - m_1 - m_2 \right] \tag{5}$$

m_j^{spc} expresses that objects have to be spatially coherent. For example, cars should be close to the road network. Again a scalar metric is computed and used as input to a sigmoid function Φ. For instance, let C_j be the subset of O_j classified as cars and d_o the distance between a car o and the road. With adapted parameters μ^d, σ^d and δ^d, the mass is:

$$m_j^{\text{spc}} = \bigoplus_{o \in C_j} \Phi(d_o, \mu^d, \sigma^d, \delta^d) \tag{6}$$

Please note that other constraints can be added following the same formalism, such as modeling that object sizes or covariances should be of reasonable values.

4.2 Consistency

m^{cons} models that objects must follow coherent trajectories in time and not change their dynamics in an unpredictable way (by making improbable position jumps between two messages for instance). For this, previously received objects are predicted $O_j(t|t-1)$ and associated with newly received objects O_j using an assignment function noted A. The similarity function described in [13] is used to compare objects and is noted $m^{\mathcal{S}}_{a,b}$. It compares different characteristics of the objects and yields a mass function defined on $\Omega^{\mathcal{S}} = \{S, \not S\}$ to express that a and b can correspond to the same physical object or to two different ones:

$$\begin{aligned} m^{\mathcal{S}}_{Oj} &= \bigoplus_{o(t-1), o \in A(O_j(t|t-1), O_j)} m^{\mathcal{S}}_{o(t-1),o} \\ m_j^{\text{cons}} &= \begin{bmatrix} 0 & m^{\mathcal{S}}_{Oj}(\{\not S\}) & 1 - m^{\mathcal{S}}_{Oj}(\{\not S\}) \end{bmatrix} \end{aligned} \tag{7}$$

Thus m_j^{cons} expresses untrustworthiness when objects mismatch with their past but is vacuous otherwise.

As we have seen, the coherency and consistency constraints (m_j^{cohe} and m_j^{cons}) can only express untrustworthiness. Therefore, other criteria have to be used to allow trustworthiness to increase.

4.3 Confirmation through free space and objects

m^{conf} models that received objects and FS should match with the current knowledge of the world of the receiving peer. For this, the detectability $m^{\mathcal{D}}$ of the receiving and sending peers are used to represent that peers have different FOVs and that comparisons cannot be made on non-overlapping areas.

$$m_j^{\text{conf}} = m_j^{\text{osi}} \oplus m_j^{\text{odi}} \oplus m_j^{\text{ofi}} \oplus m_j^{\text{fsi}} \tag{8}$$

m_j^{osi} and m_j^{fsi} model that trustworthy information should match with the local one. The received FS is compared using the method of [11] in m_j^{fsi}. Received objects O_j are matched using an assignment function noted A and compared using the similarity function $m^{\mathcal{S}}$ defined in [13]. The local object detectability grid $m^{\mathcal{D}}$ is used as a discounting factor to only compare objects that are locally detectable:

$$\begin{aligned} m_{Oj}^{\text{osi}} &= \bigoplus_{o,o_j \in \mathcal{A}(O,O_j)} \left(1-m_{oj}^{\mathcal{D}}(D)\right) m_{o,o_j}^{\mathcal{S}} \\ m_j^{\text{osi}} &= \begin{bmatrix} m_{Oj}^{\text{osi}}(S) & 0 & 1-m_{Oj}^{\text{osi}}(S) \end{bmatrix} \end{aligned} \tag{9}$$

Conversely, m_j^{odi} models that received objects must not mis-match local ones O. For this, the j-detectability of objects not matched with the assignment function A is used:

$$\begin{aligned} m_{Oj}^{\text{odi}} &= \bigoplus_{o \notin \mathcal{A}(O,O_j)} \left(1-m_o^{\mathcal{D}}(D)\right)^j m_o^{\mathcal{D}} \\ m_j^{\text{odi}} &= \begin{bmatrix} 0 & m_{Oj}^{\text{odi}}(D) & 1-m_{Oj}^{\text{odi}}(D) \end{bmatrix} \end{aligned} \tag{10}$$

Similarly m_j^{ofi} models that the received objects O_j must not be inconsistent with the free space FS estimated locally:

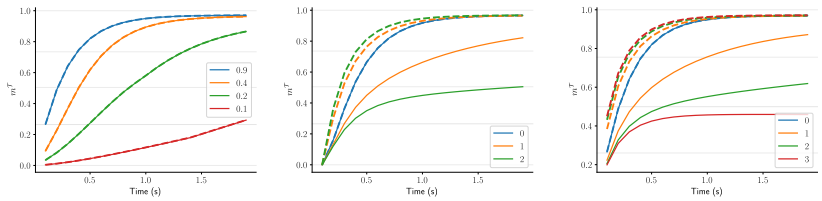
$$m_j^{\text{ofi}} = \bigoplus_{o_j \in O_j} \begin{bmatrix} 0 & m_{oj}^{\mathcal{D}}(\not{D}) & 1-m_{oj}^{\mathcal{D}}(\not{D}) \end{bmatrix} \tag{11}$$

5 Results

In order to illustrate and validate our approach, we implemented the equations detailed in Section 4, first in a simple situation in Section 5.1 then on real data in Section 5.2.

5.1 Simulation Study

In this section, trustworthiness estimates are obtained by running simulations implementing Fig. 2 on the situation of Fig. 1 for 2 seconds with varying parameters. Curves of Fig. 6 correspond to the trustworthiness a vehicle attributed



(a) (b) (c)

Fig. 6: Simulation results: Trustworthiness between two vehicles with different levels of object detectability in (a), number of ghost objects in (b) and number of objects with incoherent sizes in (c). Continuous lines are $m(T)$ and dashed lines are $m(T) + m(\not{T})$.

to another one under different conditions. Simulation results about a particular parameter are plotted on top of each other to compare its impact, while others remain unchanged and optimal.

One can see in Fig. 6a that the object detectability value plays a major role. When it is low, trustworthiness converges more slowly, which is a desired behavior. In Fig. 6b, the presence of objects that do not exist creates untrustworthiness while matched objects creates trustworthiness. The same can be seen in Fig. 6c, where erroneous sizes generate untrustworthiness.

5.2 Experimental Results

To validate our approach, it has been applied to real-world data using the same dataset as in [10]. In it, three vehicles v_1, v_2 and v_3 were driven in an busy roundabout with v_1 stopped at one of the roundabout entrance while v_2 and v_3 followed each other inside of it. In post-processing, LiDAR point clouds and RTK GNSS receivers have been processed to generate object lists and FSs. The different parameters (e.g. discounting factors) have been tuned on some preliminaries tests to get smooth trust variations. Fig. 7 shows the trust estimated by the three vehicles in each other over the course of 22 seconds. At the beginning, trustworthiness in the others is completely unknown. Exchanged information is faithful up to time $t = 12$ s when v_3 starts sending erroneous information (incoherent sizes, omitted and ghosts objects) then stops at time $t = 16$ s. In this case, v_3 is voluntary lying to v_1 and v_2 but its internal information remains correct. Note that v_2 and v_3 always share perceived areas, but only do with v_1 from $t = 6$ to $t = 12$ s. As a result, v_1 is uncertain in the trustworthiness of v_2 and v_3 and reciprocally when they do not share objects, while v_2 and v_3 trust each other rapidly. A transient phase can be observed when v_1 starts perceiving common areas with v_2 and v_3. At first, only untrustworthiness is expressed because small inconsistencies between objects at the boundary of their FOVs accumulate without enough positive information to counteract. Once enough FS and objects are shared, trustworthiness can increase.

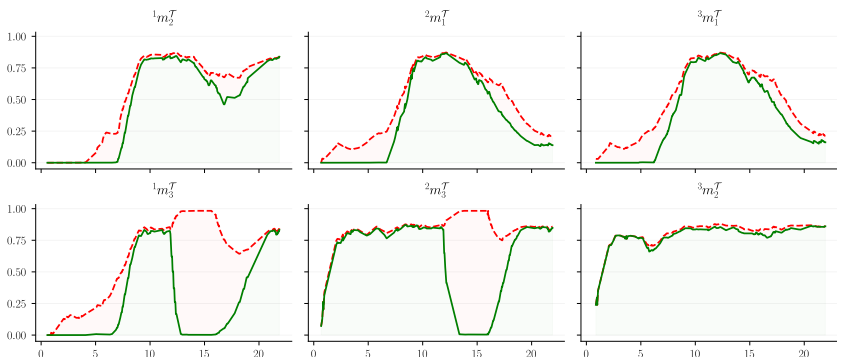


Fig. 7: Real data: Estimated trustworthiness of v_1, v_2 and v_3 in each other, where ${}^{i}m_{j}^{\mathcal{T}}$ denotes the trust of i in j. The green curve is $m(T)$ and the red curve $m(T) + m(\not{T})$.

When v_3 sends erroneous information, one can see on ${}^{1}m_{3}^{\mathcal{T}}$ and ${}^{2}m_{3}^{\mathcal{T}}$ that it is detected by v_1 and v_2, with untrustworthiness being rapidly estimated and maintained. Once v_3 stops sending erroneous information at $t = 16$ s, v_1 and v_2 increase their trustworthiness in it.
These results illustrate well that to have trustworthiness in another peer, overlapping and coherent objects and free space are necessary. In our opinion, the laws of physics of the real world are the most reliable way to induce trust. Finally, by comparing ${}^{2}m_{1}^{\mathcal{T}}$ with ${}^{3}m_{1}^{\mathcal{T}}$ and ${}^{1}m_{3}^{\mathcal{T}}$ with ${}^{2}m_{3}^{\mathcal{T}}$, one can see that the trustworthiness estimated about a particular peer differs from the other due to different points of view. This is another consequence of our choice to manage trust in a decentralized way.
In terms of computation performance, the trustworthiness estimation can take up to 500 ms per iteration with our current implementation in Python/C++. As such, it cannot run in real time as communications are at 10 Hz. However, this is not necessarily an issue as this process can be run asynchronously at lower frequencies.

6 Conclusion

In this paper, we have proposed a method to estimate trustworthiness in other peers in the context of decentralized cooperative perception. This formulation combines misbehavior detection techniques and positive confirmation thanks to mass functions to express trustworthiness, untrustworthiness or lack of information about another peer. The trust information that peers create in others is personal and never shared. Thus, two peers will have different trusts in a third peer. The convergence of the method has been illustrated in a simple simulation, then confirmed on a real-world situation. It has shown to react quickly to erroneous information to prevent its propagation. In future work, the interaction trust and object estimation will be studied. In addition, a cooperative dataset

with a ground truth on the existence of objects will be acquired to illustrate the effectiveness of this formulation and the impact of trustworthiness estimation on the non-propagation of erroneous information.

Acknowledgments This work was carried out within SIVALab, a shared laboratory between Renault and Heudiasyc (UTC/CNRS).

References

1. Aeberhard, M., Paul, S., Kaempchen, N., Bertram, T.: Object existence probability fusion using dempster-shafer theory in a high-level sensor data fusion architecture. In: IEEE Intelligent Vehicles Symposium (2011)
2. Allig, C., Leinmüller, T., Mittal, P., Wanielik, G.: Trustworthiness Estimation of Entities within Collective Perception. In: IEEE Vehicular Networking Conference (2019)
3. Ambrosin, M., Yang, L.L., Liu, X., Sastry, M.R., Alvarez, I.J.: Design of a Misbehavior Detection System for Objects Based Shared Perception V2X Applications. In: IEEE Intelligent Transportation Systems Conference (2019)
4. Caillot, A., Ouerghi, S., Vasseur, P., Boutteau, R., Dupuis, Y.: Survey on Cooperative Perception in an Automotive Context. IEEE Transactions on Intelligent Transportation Systems (2022)
5. Chowdhury, M., Gawande, A., Wang, L.: Secure Information Sharing among Autonomous Vehicles in NDN. In: International Conference on IoT Design and Implementation (2017)
6. ETSI: TR 103 562: Analysis of the Collective Perception Service (2019)
7. Günther, H.J., Mennenga, B., Trauer, O., Riebl, R., Wolf, L.: Realizing collective perception in a vehicle. In: IEEE Vehicular Networking Conference (2016)
8. van der Heijden, R.W., Dietzel, S., Leinmüller, T., Kargl, F.: Survey on Misbehavior Detection in Cooperative Intelligent Transportation Systems. IEEE Communications Surveys Tutorials (2019)
9. Kurdej, M., Cherfaoui, V.: Conservative, Proportional and Optimistic Contextual Discounting in the Belief Functions Theory. In: International Conference on Information Fusion (2013)
10. Lima, A., Bonnifait, P., Cherfaoui, V., Hage, J.A.: Data Fusion with Split Covariance Intersection for Cooperative Perception. In: IEEE International Intelligent Transportation Systems Conference (2021)
11. Liu, X., Yang, L., Alvarez, I., Sivanesan, K., Merwaday, A., Oboril, F., Buerkle, C., Sastry, M., Baltar, L.G.: MISO- V: Misbehavior Detection for Collective Perception Services in Vehicular Communications. In: IEEE Intelligent Vehicles Symposium (2021)
12. Obst, M., Hobert, L., Reisdorf, P.: Multi-sensor data fusion for checking plausibility of V2V communications by vision-based multiple-object tracking. In: IEEE Vehicular Networking Conference (2014)
13. Zoghby, N., Berge-Cherfaoui, V., Denœux, T.: Optimal object association from pairwise evidential mass functions. In: International Conference on Information Fusion (2013)