

Comparison of Different Tracking Approaches on Pre-Fused Data for Automotive Perception System

László Lindenmaier

Budapest University of Technology and Economics, Faculty of
Transportation Engineering and Vehicle Engineering,
Department of Automotive Technologies
Budapest, Hungary
lindenmaier94@gmail.com

Viktor Tihanyi

Budapest University of Technology and Economics, Faculty of
Transportation Engineering and Vehicle Engineering,
Department of Automotive Technologies
Budapest, Hungary
viktor.tihanyi@gjt.bme.hu

Abstract—The development of autonomous vehicles requires different sensor types for constructing an accurate local environment model. The evolution of sensor technology is challenging the development of sensor data fusion. Although, the sensors are getting more robust and reliable. However, because of the development of sensors the data fusion should be able to handle diversified sensor clusters. Thus, the high-level sensor fusion algorithms, like track to track method, are getting more important. This paper introduces different tracking approaches, which are intended to support a track-to-track fusion method of an autonomous vehicle in highway situations.

Keywords—Autonomous vehicles, Automotive perception, Sensor Fusion, Track-to-track fusion, Tracking algorithm, Data association, JIPDA

I. INTRODUCTION

In the respective of autonomous vehicles, a robust and accurate environment model is mandatory including the surrounding dynamic objects. To construct this environment model usually different sensors are used such as radars, lidars and cameras. Different sensor types have different advantages and disadvantages. For instance, radars can measure the distance of objects accurately and it is independent of weather conditions, but the azimuth of objects could not be measured precisely by them. Camera based detectors are not just able to assign the objects to lanes, but they can classify them as well. Sensor fusion algorithm is intended to build a common environment model which is more robust than using either sensors alone.

Because of the development of sensor technology, the sensor clusters of autonomous vehicles are changing. Hence, the sensor fusion should be able to follow the change by a sensor-independent architecture. The level of fusion algorithm [4] refers to the abstraction level of input data. The higher the level of the fusion is, the more modular it is. But the lower the fusion level is, the more robust the environment model is. The track-to-track fusion method is a high-level fusion which ensures a sensor-independent architecture by pre-defined input interface.

Since, the track-to-track fusion is a higher-level fusion than the traditional algorithms [5] [6], this paper is intended to introduce tracking algorithms that can increase the robustness of the environment perception. In the topic of multi-object tracking numerous works are available [7] [8] [9] [10]. Although, in the track-to-track fusion method all the input objects are tracked. However, after the data association and attribute fusion, the pre-fused object list, the so-called object hypotheses should also be tracked. This step is intended to obtain the continuity of output objects. This paper is comparing different tracking methods for final object tracking to construct a robust environment model in highways.

II. TRACK-TO-TRACK FUSION

In the current work the data of a radar and two monovision camera detectors in the frontal area (see Fig. 1.) of ego vehicle are fused.

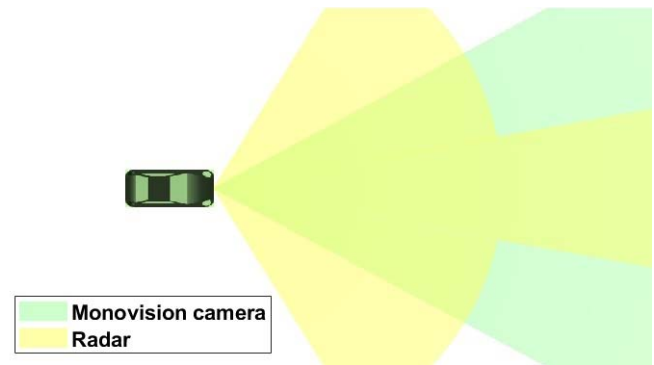


Fig. 1. FoVs of fused sensors

The track-to-track fusion method, which is applied on sensor data, is introduced in [1], but other works [2] [3] are also available in this topic. This fusion architecture has a big advantage compared to the traditional target level fusion, besides that it is modular because of the defined input interface. In the traditional target level fusion the input data are compared to the final object list (Fig. 2.), which leads to cardinality problem (the number of objects is uncertain). Since, in track-to-track fusion the data association is performed between the input tracks, a fused object hypothesis list is available after the attribute fusion (Fig. 3.). Thus, the cardinality problem could be resolved easier.

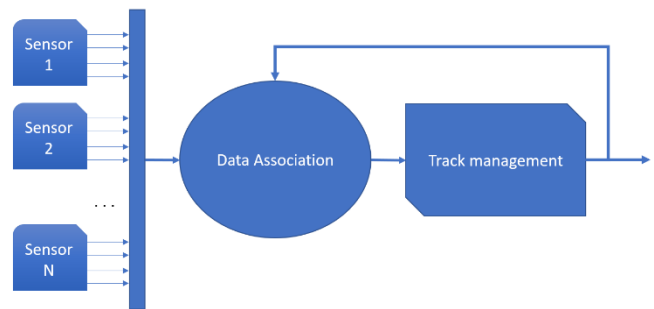


Fig. 2. Target level fusion architecture

Although, the input data consist of tracked objects, because of the clustering method a final tracking is needed. This tracking is intended to ensure the continuous object perception and the continuity of track IDs.

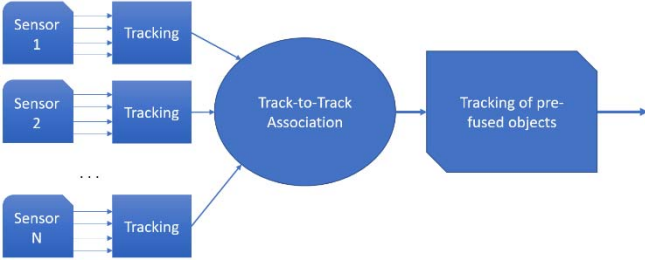


Fig. 3. Track-to-track fusion architecture

III. TRACKING ALGORITHMS

Tracking is intended to obtain the information that an object has been already perceived and to estimate its state recursively. In general, tracking algorithms consist of four main steps: the data association, the state update, object prediction and object management (see Fig. 4.). Different approaches exist for state update but the most common is Bayesian state estimation (e.g. Kalman filter) which also includes the object prediction. In case of single object, association is not necessary, because it is obvious that to which object the measurement belongs. However, multi-object tracking is more complex, because data association must be executed to define which detection (measurement) should update the state of an object. The object management is intended to handle the appearing (object birth) and disappearing (delete object from object list) objects.

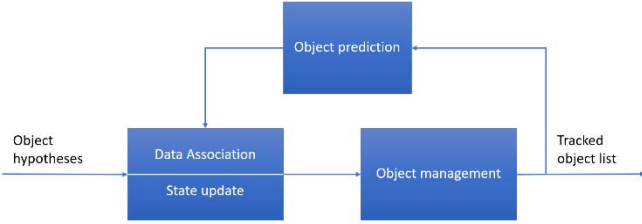


Fig. 4. Architecture of multi-object tracking

Since, all the input objects are already tracked and filtered by Kalman filter, on the pre-fused data additional filtering should not be carried out. Thus, this paper focuses on the data association and object management to ensure the continuous and robust object perception.

IV. DATA ASSOCIATION

The data association is intended to assign the sensor measurements to output objects. If the assignment was bijective, the association could be represented by a binary matrix. The rule of this association matrix is given in (1).

$$A(i, j) = \begin{cases} 1, & z_i \leftrightarrow x_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $z_i \leftrightarrow x_j$ denotes the assignment between z_i measurement and x_j object. Let N and M denote the number of objects and measurements, then $A \in \{0, 1\}^{M, N}$.

In some other approaches, such as in JIPDA (Joint Integrated Probability Data Association), the association is not bijective but probabilistic. Equation (2) demonstrates the probabilistic association matrix rule.

$$A(i, j) = p(z_i \leftrightarrow x_j) \quad (2)$$

A. Point-to-point association

This approach is the simplest one. It consists of two steps: construction of distance matrix and to solve the association problem. In this approach the objects are represented as a point object with state x which consists of different state variables (e.g. longitudinal, lateral position and velocity, etc.). The distance matrix could be filled by the Euclidian-distance, or either by Mahalanobis-distance of two objects. If the error covariance matrix of the object's state and the observation noise covariance matrix of measurement were available, the distance matrix would be constructed by (3).

$$D(i, j) = \sqrt{(z_i - \hat{z}_j)^T S_{ij}^{-1} (z_i - \hat{z}_j)} \quad (3)$$

where S_{ij} denotes the innovation covariance matrix between the z_i measurement and x_j objects, while $\hat{z}_j$ is the object's state transformed to the measurement space. In the current work the measurement vector contains the longitudinal and lateral position of objects. After construction of distance matrix, the association problem could be solved by two methods: LNN (Local Nearest Neighbor) and GNN (Global Nearest Neighbor) method. Both methods minimize the distance, but the LNN does it locally while GNN globally. The LNN method is based on Greedy algorithm. It is fast but provides a suboptimal solution. While GNN method could be performed by Hungarian method [14]. Although, in the current work the LNN method is used. However, in both methods a distance threshold should be defined to avoid false associations. In this approach the distance threshold is a key parameter. To define the threshold, the camera characteristic was investigated, because it has higher distance uncertainty than the radar. The camera characteristic, which gives the connection between the longitudinal distance in ego vehicle coordinate system and the pixel row index of camera image, is hyperbolic in a pixel column. In this research the middle column of the image was chosen to work with, because it is a good estimation for those areas of the image where object could appear in highway circumstances. By the derivation of the characteristic it can be proven that assuming constant uncertainty in pixel coordinate, the uncertainty is second order polynomial in vehicle coordinate system over the distance of object. Thus, the distance threshold is defined by a second order polynomial over the distance of the object, which is described by (4).

$$d_t(d(x_i)) = c_2 d(x_i)^2 + c_1 d(x_i) + c_0 \quad (4)$$

The second and first order coefficients of the polynomial are computed based on polynomial fitting. Assuming a constant uncertainty in pixel coordinate, the difference between the values of the hyperbola, whose abscissa difference equals to pixel uncertainty, is computed. This difference is computed along the rows of the image, that have a longitudinal distance in vehicle coordinate system. Then the polynomial fitting could be performed by Least Square Method. The c_0 is a user-defined parameter which is set to 3 m.

Furthermore, to avoid sudden lane change, a lateral threshold is also given, which is set to 1.8 m. To apply this lateral threshold the deviation of the lateral position should be modified which is described by (5).

$$\sigma_y = \frac{d_{t,lat}}{d_t} \sigma_x \quad (5)$$

where σ_y and σ_x denote the lateral and longitudinal deviation and $d_{t,lat}$ represents the given lateral threshold. Then the threshold area (see Fig. 5.) of an object, where a measurement could be assigned to it, is elliptic.

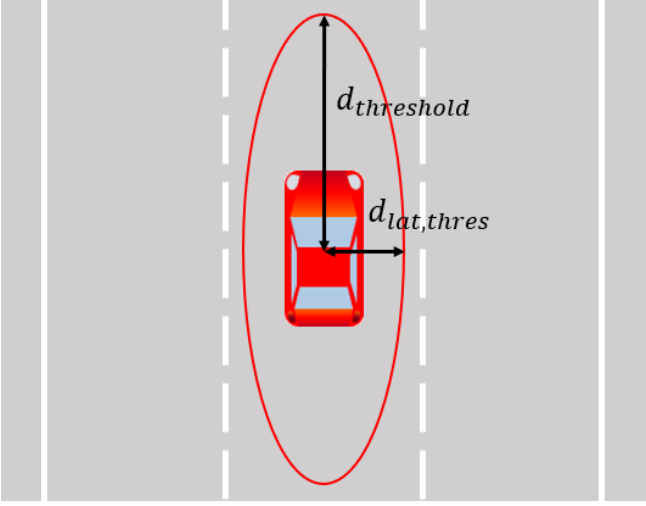


Fig. 5. Threshold area of an object

B. Extended 2D association

This approach is combining the point-to-point approach and a two-dimensional association method. One of the problems with the previous approach is that the zero-order coefficient of the polynomial is user-defined. It can occur false associations in the near zone of ego vehicle. Thus, two association matrices are constructed and the 2D based extends the previous approach. In 2D based association the objects are modelled as two dimensional and the distance matrix becomes intersection matrix. The intersection matrix is filled with the modified intersection area (see Fig. 6.) of detected object and tracked object.

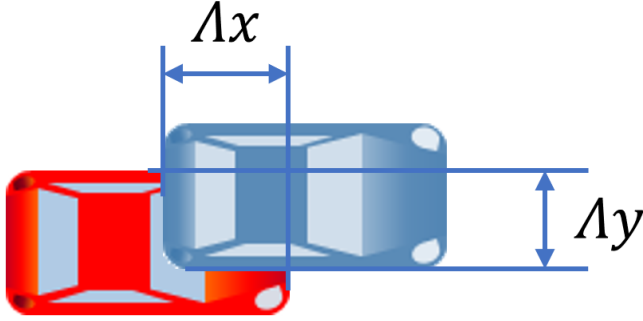


Fig. 6. Intersection area of two objects

The $A_I(x_i, z_j)$ intersection area is computed by (6).

$$A_I(x_i, z_j) = \Lambda x \Lambda y \quad (6)$$

The modification is justifiable, because the assignment with smaller lateral distance should be preferred to the same area value with lower longitudinal distance. Thus, the modified intersection area is computed by (7).

$$A_{I,m}(x_i, z_j) = A_I(x_i, z_j) + (1 - a_x)\Lambda x^2 + a_x\Lambda y^2 \quad (7)$$

where a_x denotes the ratio of longitudinal deviation and the sum of the longitudinal and lateral deviations. However, the definition of modified intersection area as in (7) would result positive area in case of zero intersection as well. Thus, the right modified intersection area is given by (8).

$$\hat{A}_{I,m}(x_i, z_j) = \begin{cases} A_{I,m}(x_i, z_j), & A_I(x_i, z_j) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Furthermore, the size of the objects considers also the uncertainty of the position, but this step is going to be detailed later. The combination of two method is based on the confirmation of point-to-point approach. If an object was not assigned to either measurements based on the point-to-point approach, but it was assigned to a measurement according to 2D approach, in the merged association matrix the assignment would be confirmed.

V. OBJECT MANAGEMENT

The object management is intended to handle the birth and disappearance of objects. This step is mandatory for a robust environment model, because it could ensure the continuous object perception and filter some of the false objects. In this paper three approaches are introduced and compared.

A. Frame based approach

Since, one of the goals of object management is to provide the continuity of objects, it must be ensured that the object is still perceived even in case of misdetection. Thus, object prediction is carried out and an object, which has not been assigned to either measurements, is predicted for a defined number of frames. The number of predicting frames should be modified depending on the object state, but it is not in focus of current work. The prediction is performed by a simple CV (Constant Velocity) model. The measurement that do not participate in assignment assumed to be new born objects. To avoid some of the false positive objects, the new born objects are masked for a couple of frames. If a new born object was assigned to a measurement consequently during these frames it initiates a valid, tracked object. The number of masking frames is set to 4.

B. Frame based extended with occupancy birth model

The main problem of the previous approach is that it considers perfect pre-fusion. However, sometimes the track-to-track fusion occurs false associations, thus more than one track is generated by an object. If in tracking module the data association was bijective (just one measurement is assigned to an object), the previous approach would not be able to filter the false association. In this approach the p_B probability of object birth is introduced. It is computed based on the occupancy grid model of environment and on distances of the other objects. The occupancy grid is constructed by the convolution of the 2D normal distribution of object position and the extension of object. The object extension is modelled by a rectangle with w width and L length. Then the probability of occupancy by an object in a point is computed by (9).

$$p_{occ,i}(x, y) = \mathcal{N}(\cdot, (\hat{x}_i, \hat{y}_i), \mathbf{P}_i^{xy}) * rect_{w_i, L_i}(x, y) \quad (9)$$

In (9) $(\hat{x}_i, \hat{y}_i)$ denotes the position vector of object and $\mathbf{P}_i^{xy}$ represents the part of the error covariance matrix that refers to the position. The occupancy grid is created by discretizing the environment model into small cells and the convolution is computed numerically in the middle of the cells for each confirmed and predicted object. The occupancy grid of the whole object list based on the occupancy grid of an object is computed by (10).

$$p_{occ}(x, y) = \min(1, \sum_{i=1}^N p_{occ,i}(x, y)) \quad (10)$$

Then the intersection probability of the object list occupancy grid and of the potential new born object occupancy must be computed by a simple multiplication of the two occupancy grids and applying the mode of the density function as in:

$$p_i(z_j|x_1, \dots, x_N) = \max(p_{occ}(x, y), p_{occ,j}(x, y)) \quad (11)$$

This value refers to the probability that the z_j object hypothesis is overlapping with at least one of the current objects in the environment model. However, in the process of modelling the probability of birth, the distance of the objects is also considered. The probability that the object hypothesis represents the same object is given by (12).

$$p_A(z_j|x_i) = \exp(-\frac{1}{2}d_{MH}(z_j, \hat{z}_i)^2) \quad (12)$$

Finally, the probability of object birth is computed by (13).

$$p_B(z_j|x_1, \dots, x_N) = (1 - p_I) \prod_{i=1}^N (1 - p_{A,j,i}) \quad (13)$$

This probability is investigated during the masking frames. If the probability was greater than a defined threshold during the masking period, the measurement would initiate a new object. The probability threshold is set 0.85.

Although, this method avoids the birth of an object that would overlap an existing, tracked object. However, it can occur, false negative object. If an object was predicted and an object hypothesis did not initiate a new object because of the occupancy grid, it would occur false negative for a couple of frames. To avoid this, as it has been already mentioned, that the size of the object is modified in the data association module. The convolution was computed for different object sizes and deviations. The difference between the mean value and the abscissa of $(1 - p_{B,t})$ value is investigated, which refers to half of the corresponding dimension. Based on numerical results two cases are defined. If the deviation of the position was lower than the corresponding dimension, the dimension should be increased by the deviation in both directions. If the deviation was much greater than the dimension, the dimension would not increase linearly with the deviation. Since this is just a numerical result, the size extension is computed simply by (14) and (15).

$$w_e = \min(6w, w + 2\sigma_y) \quad (14)$$

$$L_e = \min(6L, L + 2\sigma_x) \quad (15)$$

where w_e and L_e denote the extended dimensions, w and L are the original width and length, while σ_x and σ_y represent the longitudinal and lateral deviations. The size extension based on this approach helps avoiding false negatives.

C. Frame based extended with distance birth model

Although, the previous method solves the problem of overlapping objects due to false association. However, the convolution in most cases requires much computational capacity. This approach is intended to substitute the previous method by keeping the advantage of it but avoiding the great computational capacity requirement. Thus, a method is needed that can fuse the overlap probability and the assignment probability. To reach this, the distance between two objects is reinterpreted. First, four cases are defined (see Fig. 7.). Then longitudinal and the lateral distances are corrected as in (16) and (17).

$$\hat{d}_x(x_i, z_j) = d_x - L \quad (16)$$

$$\hat{d}_y(x_i, z_j) = d_y - \frac{w_i + w_j}{2} \quad (17)$$

where d_x and d_y are the longitudinal and lateral distances of the point objects, L is the length of the object behind, while w_i and w_j denote the width of the objects.

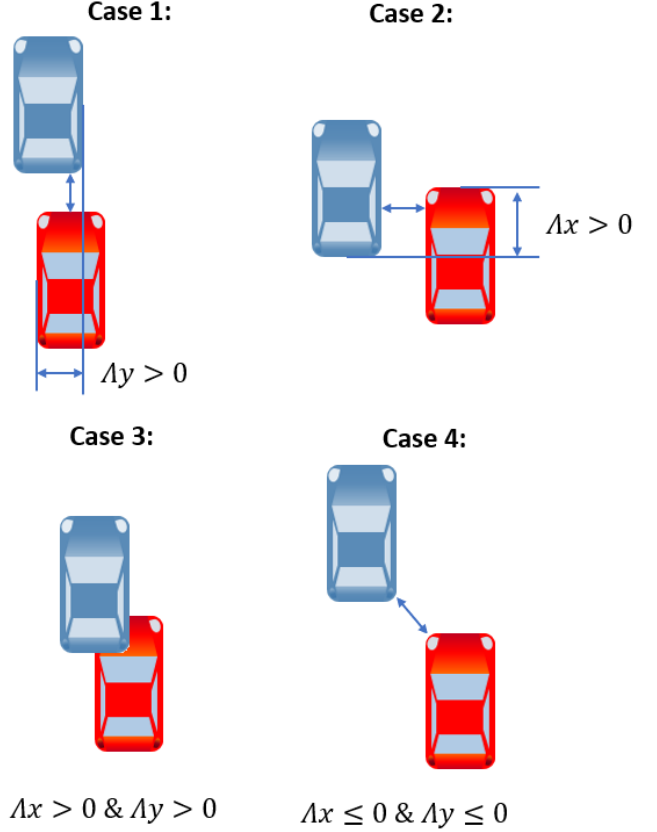


Fig. 7. Cases of distance interpretation

Note, that in this work the point representation of an object locates in middle of the rear side! Then the distances in different cases could be computed as in:

$$d_1(x_i, z_j) = \frac{\hat{d}_x(x_i, z_j)}{r\sigma_x + (1-r)\sigma_y} \quad (18)$$

$$d_2(x_i, z_j) = \frac{\hat{d}_y(x_i, z_j)}{r\sigma_x + (1-r)\sigma_y} \quad (19)$$

$$d_3(x_i, z_j) = 0 \quad (20)$$

$$d_4(x_i, z_j) = \sqrt{\left(\frac{\hat{d}_x(x_i, z_j)}{\sigma_x}\right)^2 + \left(\frac{\hat{d}_y(x_i, z_j)}{\sigma_y}\right)^2} \quad (21)$$

where σ_x and σ_y denote the resultant deviation of longitudinal and lateral positions, which equal to the sum of the deviations of object and object hypothesis. In (18) and (19) r is the ratio of the angle between the objects and of the right angle (Fig. 8.). After choosing the right distance computation method based on the current case, the conditional probability of object birth depending on one object is given by (22).

$$p_B(z_j|x_i) = 1 - \exp(-\frac{1}{2}d(x_i, z_j)^2) \quad (22)$$

Finally, the probability of object birth based on the whole object list could be computed as in:

$$p_B(z_j|x_1, \dots, x_N) = \prod_{i=1}^N p_B(z_j|x_i) \quad (23)$$

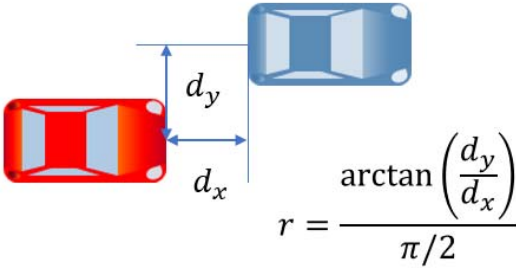


Fig. 8. Angle ratio between objects

One can recognize that this method fuses the occupancy grid and the distance of point objects without computing convolution. In addition, because of it is based on distance, the birth probability threshold is more scalable. This probability threshold is set to 0.5.

Since, this approach of birth probability is different from the previous approach, the increase method of dimensions is different as well. Because this approach is only distance based, the size extension is simpler. Based on the probability threshold one can compute a distance threshold by (24).

$$d_t = \sqrt{-2 \ln(1 - p_{B,t})} \quad (24)$$

Since distance threshold is computed based on Mahalanobis distance, the dimensions of object are increased based the distance threshold multiplied by the corresponding deviation. Then the dimensions are increased in both directions (see Fig. 9.)

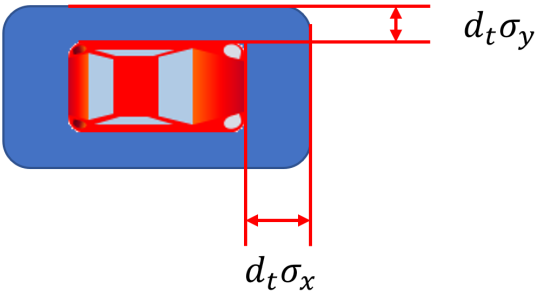


Fig. 9. Size extension

VI. RESULTS

The previously presented tracking approaches were evaluated and compared on a measurement sequence constructed by 1500 frames with 30 FPS (50 second). The measurement was logged on Hungarian highway called M0. A ground truth is created by annotating the video and transforming annotated objects into ego vehicle coordinate system. To evaluate the results data association is required between the reference objects and tracked objects. For this the Extended 2D association was applied with the same settings, since the ground truth is generated based on a video annotation. Then a reference object would assumed to be false negative, if none of the fused object was assigned to it. Similarly, a fused object would be evaluated as a false positive, if a reference object could not be assigned to it. If more than one fused object was associated to a reference

object, it would be registered as a false association, which is a key output of the evaluation. The results of the different approaches are included in Table I. and Table II, Table III. All off the possible combinations was tried to be able to compare them. Based on the data included in Table I., one can say that the result corresponds to the expectations in the respective of false associations. Applying Frame based object management occurs about twice as much false associations than using any of the birth models. The reason of this is the failure of the track-to-track pre-fusion because of erroneous detection by the camera. Sometimes the camera-based detector perceives the object, but the bounding box of the detection does not fit to the object, which can occur a false association in the track-to-track fusion. Without a birth model two tracks would be generated by one object (see Fig. 10.). Comparing the Occupancy based, and Distance based birth models, it can be said that the Distance based approach keeps the advantage of the Occupancy based without computing 2D convolution. Furthermore, because of the better scalability of birth probability threshold, it was able to resolve all false associations with Extended 2D data association (see Fig. 11.). Comparing the Extended 2D data association with Point-to-Point approach one can recognize that the Extended approach can also reduce the number of false associations in the near zone of the ego vehicle. In the respective of false negatives, the Occupancy based method occurs a bit more false negative objects because of the size extension estimation. However, with Distance based approach this problem is not occurred. The different tracking approaches does not influence significantly the number of false positives based on the Table III. In summary, it can be said that the performance of the Distance based object management with Extended 2D association is the best compared to other approaches.

TABLE I. NUMBER OF FALSE ASSOCIATIONS

False Associations	Object management		
	Frame based	Occupancy based	Distance based
Data Associations			
Point to-Point	49	27	5
Extended 2D	46	22	0

TABLE II. NUMBER OF FALSE NEGATIVES

False Negatives	Object management		
	Frame based	Occupancy based	Distance based
Data Associations			
Point to-Point	316	325	325
Extended 2D	310	319	310

TABLE III. NUMBER OF FALSE POSITIVES

False Positives	Object management		
	Frame based	Occupancy based	Distance based
Data Associations			
Point to-Point	250	248	245
Extended 2D	250	244	240

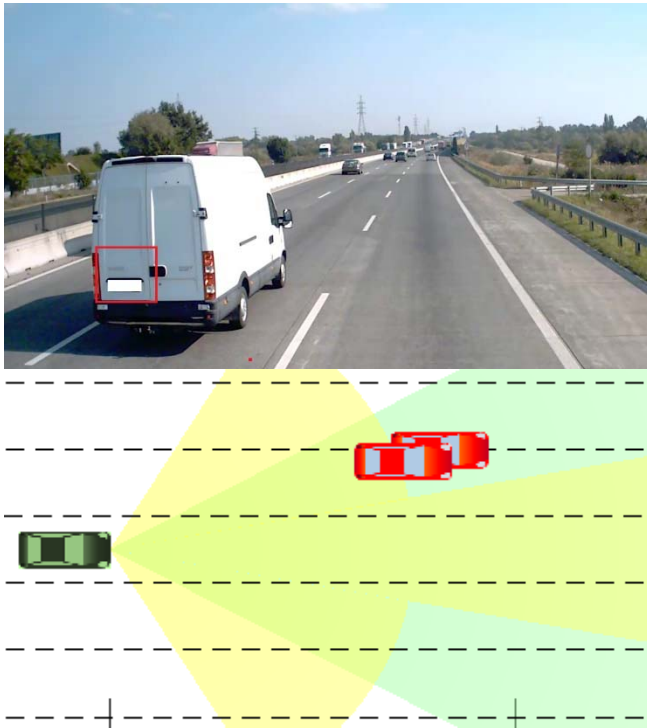


Fig. 10. False association without birth model due to camera-based detector



Fig. 11. False association resolved by Distance based birth model

ACKNOWLEDGMENT

The project has been supported by the European Union, co-financed by the European Social Fund. EFOP-3.6.2-16-2017-00002.

REFERENCES

- [1] A. Houenou, P. Bonnifait, V. Cherfaoui and J. F. Boissou "A Track-To-Track Association Method for Automotive Perception System" IEEE Intelligent Vehicles Symposium Spain, October 2012.
- [2] C. Blanc, L. Trassoudaine, Y. Le Guilloux and R. Moreira "Track to track fusion method applied to road obstacle detection" Campus des C zeaux Division D France, July 2004
- [3] A. M. Aziz „Fuzzy track-to-track association and track fusion approach in distributed multisensor multitarget multiple-attribute environment" Electrical Eng. Dpt., Military Technical College, Cairo, Egypt, 2006.
- [4] C. Otto "Fusion of Data from Heterogeneous Sensors with Distributed Fields of View and Situation Evaluation for Advanced Driver Assistance Systems", KIT Scientific Publishing 2013, pp. 15-20.
- [5] R. Omar Chavez-Garcia, T. D. Vu, J. Burler, O Aycard "Frontal Object Perception using radar and mono-vision" University of Grenoble France, 2012
- [6] C. Otto "Fusion of Data from Heterogeneous Sensors with Distributed Fields of View and Situation Evaluation for Advanced Driver Assistance Systems", KIT Scientific Publishing 2013, pp. 27-49.
- [7] Y. Bar-Shalom, M. K. Kalandros, L. Trailovic, and L. Y. Pao. "Tutorial on multisensor management and fusion algorithms for target tracking". Proc. of the 2004 American Control, pages 4737–4751, 2004
- [8] Y. Bar-Shalom and X. R. Li. "Multitarget multisensor tracking: principles and techniques". IEEE AES Systems Magazine, pages 41 – 44, Feb. 1996.
- [9] S. S. Blackman and R. Popoli. "Design and analysis of modern tracking systems" ArtechHouse, Incorporated, 1999.
- [10] Musicki, D., and R. Evans. "Joint Integrated Probabilistic Data Association: JIPDA." *IEEE transactions on Aerospace and Electronic Systems* . Vol. 40, Number 3, 2004, pp 1093-1099.
- [11] Y. Bar-Shalom and T. E. Fortmann, Tracking and Data Association, 179th ed. Academic Press, Inc., Mathematics in Science and Engineering, 1988.
- [12] M. M hlisch, Filtersynthese zur simultanen Minimierung von Existenz-, Assoziations-, und Zustandsunsicherheiten in der Fahrzeugumfelderfassung mit heterogenen Sensordaten. PhD Thesis, Universit t Ulm, Institut f r Mess-, Regel- und Mikrotechnik, 2009.
- [13] M. Munz, Generisches Sensorfusionsframework zur gleichzeitigen Zustandsund Existenzsch tzung f r die Fahrzeugumfeldererkennung. PhD Thesis, Universit t Ulm, Institut f r Mess-, Regel- und Mikrotechnik, 2011.
- [14] A. Frank, "On Kuhn's Hungarian method – A tribute from Hungary /Egerv ry," Research report, Research Group on Combinatorial Optimization, October2004.