

Performance Evaluation of a Track to Track Sensor Fusion Algorithm

Lóránt Kovács¹, László Lindenmaier², Huba Németh², Viktor Tihanyi², Ákos Zarándy^{1,3}

¹Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary

²Budapest University of Technology and Economics, Department of Automotive Technologies

³Computer and Automation Research Institute of the Hungarian Academy of Sciences (MTA-SZTAKI),
Computational Optical Sensing and Processing Laboratory

Abstract—The development of autonomous vehicles requires algorithms responsible for constructing the model of the local surroundings of the vehicles from sensor measurements. This task is also called sensor fusion. The performance of these methods should be compared to each other, using a common metric, in order to select the proper algorithmic approach for a given task, the optimal sensor set or sensor placement. This paper describes the evaluation metric and its application on a track to track sensor fusion algorithm. The current application focuses on highway scenarios, the system in focus uses only the forward looking sensor information and uses the CLEAR-MOT metric.

I. INTRODUCTION

The autonomous vehicle's environment is monitored by a set of sensors of multiple sensor types. The algorithm responsible for creating a common model from the sensors' multiple detections is the so called *sensor fusion*. The performance evaluation of the different sensor fusion methods on a pre-defined data set is required to check the capabilities of each algorithm on real scenarios, to select the suitable method for a given task. This paper summarizes the evaluation results of a track to track sensor fusion algorithm written in [1].

II. DATA COLLECTION

A vehicle (*ego vehicle*) was equipped with several sensors (cameras, radars, lidars), in preparation for developing autonomous driving functionality focusing on highway application. For the algorithmic evaluation of the selected fusion algorithm [1], only the forward-looking cameras and radars are used.

Three different forward-looking automotive radars (*Radar X,Y,Z*) are used, which give track level information on the detected objects in the sensors' field of view. The camera footage was processed offline with three different object detection algorithms (*Detector A,B,C*) for detecting objects (vehicles) and object classes (vehicle types) frame by frame.

Using the sensor calibration and configuration parameters, the collected measurement data is transformed to a common coordinate system, which is fixed in middle of the front of the ego vehicle.

The data collector vehicle was driven on motorways in different traffic scenarios. The radar and camera recordings were collected, accompanied by the vehicle components' sensor data, using an onboard data logging system.

III. ALGORITHM

The automotive radar manufacturers usually do not provide public interface to the raw detection information of their radar sensors. With the sensors used for this study, only the detected objects and their properties: track, position, speed, and the object's class and size were available. This constraint sets the level of fusion at *track-to-track sensor fusion*. Methods to process raw radar reflections were introduced in [2], [3]. However, in this paper, we used a semi processed output of the sensors.

The implemented algorithm follows the approach of Houenou et al., described in [1]. Their solution is split into two main parts.

The first, *sensor layer* is responsible for handling and processing the sensor specific data, including synchronization [4], re-sampling and even tracking. The output of this layer is universal: all the sensor dependent parameters are removed. From here, the measurement data are independent of their originating sensor, sharing the same properties.

The second, *fusion layer* works from these sensor independent data packages at every time stamp. Sensor tracks are associated and merged using their track-to-track association algorithm. The generated meta-tracks are predicted, resulting system tracks for every object as the output of the fusion algorithm.

The result of the fusion algorithm follows the same sensor independent data structure as the output of the sensor layer.

IV. REFERENCE DATA SET

The evaluation requires a reference data set, the ground truth. This might contain automatic, or manual annotations (position, bounding box, class) on the objects.

1) *Sensor dependent ground truth*: In this case, each sensor type has its own ground truth. The *radar ground truth* can be created in an experiment, where both the surrounding cars and the ego vehicle are equipped with a differential GPS. Knowing the reference vehicles' size and their location, the radar ground truth can be calculated for every measurement time stamp. The *camera ground truth* requires manual annotation on a selected camera's video footage. When the measurement sensor set is extended with lidars, the preprocessing of their signals, could be used for ground truth generation or validation by detecting

the reference vehicles, estimating both their position and object class.

2) *Video annotation based ground truth*: An easier approach is when only selected video segments are annotated from the measurement database [5]. Having the camera placement and distortion parameters, the annotations from the camera image are transformed to the real word coordinates, but the possible calibration and transformation errors could make the transformed data (now in the real world domain) less accurate. On the other hand, this approach holds the possibility to extend the ground truth database without making new measurements, just by annotating additional videos from previous recordings. The annotation accuracy can be increased by independent re-annotation of the ground truth.

In this current paper, this later, video annotation based ground truth is used. The annotated vehicle bounding boxes were transformed from the image plane to the vehicle coordinates, establishing the capability to evaluate the algorithm performance both in the image plane and in the world ground plane.

V. EVALUATION

The motivation for numerical description of the fusion algorithm's performance is to have the ability to compare algorithms, their results in different measurement scenarios. The evaluation outcome also helps to choose the proper sensor setup and algorithm parameters.

The evaluation of the fusion algorithm output requires a comparison metric. In this paper, 2D object overlap based metric was used in the image plane, and 2D object distance based metric was applied on the real world ground plane.

A. Frame wise object level evaluation

For each video frame, a similarity measure matrix is calculated for both the image plane and the real word plane fusion result (F) and ground truth (GT) object positions.

On the image plane, the intersection over union (Jaccard index, Eq. 1) was calculated on the object bounding boxes, resulting the distance matrix (*DistIM*) in the image plane [6].

$$d(GT, F) = \frac{bbox(GT) \cap bbox(F)}{bbox(GT) \cup bbox(F)} \quad (1)$$

On the real word ground plane, the euclidean distance of the objects has been calculated separately in the x (looking forward from the ego vehicle) and y (left of the ego vehicle) directions (*DistRWX*, *DistRWY*). This yields the capability to set different matching tolerances in the different directions. This direction dependent threshold is needed, as the ground truth data transformed to the real word coordinates is less accurate in the x direction. To filter the distance matrices, the matching thresholds seen in Table I were applied.

To find optimal correspondences among the ground truth and the fusion result, that minimize the overall distance error, Munkres algorithm (Hungarian algorithm) [7] is used on the filtered distance matrices, resulting the association matrices (*AssocIM*, *AssocRWX*, *AssocRWY*). On the image plane, this

TABLE I
SIMILARITY MEASURES AND THRESHOLDS

Distance matrix	Distance metric	Applied threshold
Image plane	Jaccard index (Eq. 1)	0.5
Real world - X	Euclidean	15 (m)
Real world - Y		4 (m)

TABLE II
FRAME WISE EVALUATION PARAMETERS

Parameter name	Abbreviation	Description
Ground truth object count	GTcount	number of ground truth objects
Fusion object count	OBJcount	number of fusion result objects
True positive	TP	annotation on the object
False negative	FN	no object on the annotation
False positive	FP	no annotation on the object
True positive	TPrate	TP/ GTcount
False negative	FNrate	FN/ GTcount
False positive	FPrate	FP/ GTcount
Precision:	Prec	$\frac{TP}{TP+FP}$
Recall:	Rec	$\frac{TP}{TP+FN}$
F1 score:	F1Score	$2 \frac{Prec * Rec}{Prec + Rec}$

clearly defines the association, while on the ground plane the correspondence matrix (*AssocRWX*) is given by the logical *AND* of the *AssocRWX*, *AssocRWY* matrices:

$$AssocRWX = AssocRWX \wedge AssocRWY \quad (2)$$

In some cases, the association in the image plane and in the ground plane gave different number of ground truth to fusion object assignments. To make an advantage of this, a third, handcrafted association approach is created by "mixing" the real word and image association matrices (*AssocMIX*). This was done by applying logical *OR* on the *AssocIM* and *AssocRW*, and filtering this result matrix that only one assignment could be done in each row and column.

The parameters listed in Table II. have been calculated and accumulated frame by frame concurrently for the three association matrices. The results are discussed at VI-A.

B. Global trajectory evaluation

The frame wise evaluation described in the previous subsection does not take into account the temporal correspondence existing among the consecutive, frame wise "measurements". Handling the ground truth and fusion data as object trajectories establishes the temporal correspondence [8]. This approach in the image plane is similar to evaluating the performance of multi object tracking methods: the ground truth objects' path is compared to the fusion result objects' trajectory. This sets the [8] and the MOTChallenge: The Multiple Object Tracking Benchmark¹ as a guideline for this evaluation. Two sets of measures are used in this approach: the CLEAR metrics [9],

¹<https://motchallenge.net/>

and a set of track quality measures [10]. The MOTChallenge benchmark environment is available online.

The ground truth and the fusion result data were restructured to the MOTChallenge file format. At each frame, the bounding boxes of the ground truth and the fusion result are matched using the Hungarian algorithm.

Multiple Object Tracking Accuracy (MOTA): is the sum of the accumulated false negatives (FN), false positives (FP) and the number of object ID mismatches (IDSw) averaged by the total number of ground truth objects (GTcount) [9], [11]:

$$MOTA = 1 - \frac{FN + FP + IDSW}{GTcount} \quad (3)$$

In other words: "this is the accuracy of the tracker when it comes to keeping correct correspondences over time, estimating the number of objects, recovering tracks." [12]

The ground truth trajectories are classified as mostly tracked (MT) if it is successfully tracked for at least 80% of its life span, mostly lost (ML) if it is only recovered for less than 20% of its total length, and partly tracked (PT) in between.

"ID precision and ID recall shed light on tracking trade-offs, while the IDF1score allows ranking all trackers on a single scale that balances identification precision and recall through their harmonic mean." [13]

The results are discussed at VI-B. The list of the evaluated track quality measures:

- ID Precision (*IDPrec*): fraction of fusion objects, that are correctly identified
- ID F1 Score (*IDF1*): ratio of correctly identified objects over the average number of ground-truth and fusion results
- Recal (*Rec*): percentage of detected targets
- Precision (*Prec*): percentage of correctly detected targets
- False Alarm Rate (*FAR*): number of false alarms per frame (FP/GTcount)
- *TrackNum*: number of ground truth trajectories
- *MT*, *PT*, *ML*: number of mostly tracked, partially tracked and mostly lost trajectories
- False positives (FP)
- False negatives (*FN*): number of missed targets
- ID switches (*IDSw*): number of id switches
- Fragmentations (*Frag*): how many times a ground truth trajectory is untracked.
- Multi-object tracking accuracy (*MOTA*)

VI. RESULTS

The fusion algorithm[1] is implemented using Matlab and Simulink, and it is capable of processing prerecorded measurements. The algorithm was run offline several times, with different sensors enabled on two ground truth data containing 1500 annotated frames each.

A. Object level evaluation on the real world ground plane

The aggregated results of the object level evaluation of the sensor fusion algorithm in the real word plane is shown in Table III.

The *OBJcount* column shows the importance of having different sensor types in the fusion: if only cameras are used (1st and 2nd rows), the number of detections is not sufficient. This can be explained by the fact, that this evaluation is made on the real world plane, whereas the camera based detections are made on the image plane. A few pixel position offset on the image can move the object out of the assignment boundary in the real word.

Row 3 and 11 show, that by adding Radar X to the only radar based fusion, the *OBJcount* increases, but FP rate increases. This explains why the rest of the evaluated setups do not use Radar X. An opposite tendency can be seen in row 9 and 10: by adding Detector C to fusion, the FP rate drops, and the F1 score increases.

The best sensor setup with the highest F1 score on this ground truth is the row 3, followed by row 4 and 7: the Radar Y,Z give better performance without a detector added to their fusion. The underlying reason could be the possible error multiplication of the camera to real word transformation.

B. Track level evaluation in the image plane

The aggregated results of the track level evaluation of the sensor fusion algorithm in the image plane is shown in Table IV. Assuming that the best sensor setup can be selected based on the ID F1score and the MOTA value, row 3 and 7 have the top two values. By adding the detector A to radars Y and Z, the number of false negatives decreases, while false positive increases. The number of trajectory segments does not change significantly.

There is a significant difference on this ground truth in the measurements, where only detector B or C is present, and the ones where detector A is on duty.

VII. CONCLUSION

This paper evaluates the performance of a selected track to track sensor fusion algorithm with numerical results the on measurement data recorded on highway in traffic. Applying the metric on the selected sensor fusion algorithm helps to select the proper sensor set and algorithm parameters. The results give also the opportunity to compare the performance of different fusion algorithms on the ground truth dataset with the metric.

ACKNOWLEDGMENT

This research is supported by the *Foundation for the Hungarian Vehicle-technical higher education*. The support of the Széchenyi 2020 Program, of the Human Resource Development Operational Program, of the Program of Integrated Territorial Investments in Central-Hungary (project numbers: EFOP-3.6.2-16-2017-00013, 3.6.3-VEKOP-16-2017-00002 and EFOP-3.6.2-16-2017-00002), and of the European Structural and Investment funds are gratefully acknowledged.

TABLE III
OBJECT LEVEL COMPARISON OF THE FUSION RESULT IN RW PLANE FOR DIFFERENT SENSOR SETUP

		GTcount	OBJcount	TP	FN	FP	TPrate	FNrate	FPrate	Prec	Rec	FIScore
1	Detector A,B	7822	1582	807	7015	775	0,1	0,9	0,1	0,51	0,1	0,17
2	Detector A,B,C	7822	619	200	7622	419	0,03	0,97	0,05	0,32	0,03	0,05
3	Radar Y,Z	7822	6732	4067	3755	2665	0,52	0,48	0,34	0,6	0,52	0,56
4	Radar Y,Z + Det C	7822	6088	3796	4026	2292	0,49	0,51	0,29	0,62	0,49	0,55
5	Radar Y,Z + Det B	7822	6395	3260	4562	3135	0,42	0,58	0,4	0,51	0,42	0,46
6	Radar Y,Z + Det B,C	7822	5358	2668	5154	2690	0,34	0,66	0,34	0,5	0,34	0,4
7	Radar Y,Z + Det A	7822	6858	3834	3988	3024	0,49	0,51	0,39	0,56	0,49	0,52
8	Radar Y,Z + Det A,C	7822	6282	3458	4364	2824	0,44	0,56	0,36	0,55	0,44	0,49
9	Radar Y,Z + Det A,B	7822	6484	3235	4587	3249	0,41	0,59	0,42	0,5	0,41	0,45
10	Radar Y,Z + Det A,B,C	7822	6016	3285	4537	2731	0,42	0,58	0,35	0,55	0,42	0,47
11	Radar X,Y,Z	7822	6992	3801	4021	3191	0,49	0,51	0,41	0,54	0,49	0,51
12	Radar X,Y,Z + Det A,C	7822	6387	3257	4565	3130	0,42	0,58	0,4	0,51	0,42	0,46
13	All	7822	5977	2939	4883	3038	0,38	0,62	0,39	0,49	0,38	0,43

TABLE IV
TRACK LEVEL COMPARISON OF THE FUSION RESULT IN IM FOR DIFFERENT SENSOR SETUP

		IDF1	IDPrec	IDRec	Rec	Prec	FAR	Track Count	MT	PT	ML	FP	FN	IDS w	Frag	MOTA
1	Detector A,B	18,95	56,32	11,39	20,23	100	0	24	1	11	12	0	6240	3	10	20,19
2	Detector A,B,C	9,05	61,71	4,88	7,91	100	0	24	1	2	21	0	7203	0	1	7,91
3	Radar Y,Z	67,16	72,59	62,48	79,35	92,2	0,03	24	12	9	3	525	1615	10	31	72,51
4	Radar Y,Z + Det C	54,03	61,73	48,04	72,72	93,43	0,02	24	11	9	4	400	2134	21	32	67,34
5	Radar Y,Z + Det B	53,72	59,72	48,82	74,98	91,71	0,03	24	11	8	5	530	1957	42	49	67,67
6	Radar Y,Z + Det B,C	54,17	66,63	45,64	63,53	92,74	0,02	24	5	14	5	389	2853	29	40	58,18
7	Radar Y,Z + Det A	65,97	70,6	61,9	79,8	91,02	0,03	24	12	9	3	616	1580	21	32	71,66
8	Radar Y,Z + Det A,C	59,78	67,11	53,9	74,48	92,74	0,03	24	10	11	3	456	1996	25	34	68,33
9	Radar Y,Z + Det A,B	55,68	61,43	50,92	75,77	91,41	0,03	24	9	12	3	557	1895	38	40	68,17
10	Radar Y,Z + Det A,B,C	57,18	65,76	50,58	71,45	92,9	0,02	24	7	12	5	427	2233	37	41	65,52
11	Radar X,Y,Z	60,32	63,9	57,12	79,21	88,62	0,04	24	15	6	3	796	1626	31	45	68,64
12	Radar X,Y,Z + Det A,C	50,76	56,46	46,1	74,62	91,39	0,03	24	11	9	4	550	1985	40	39	67,08
13	All	52,15	60,2	46	70,08	91,72	0,03	24	4	16	4	495	2340	55	50	63,05

REFERENCES

- [1] A. Houenou, P. Bonnifait, V. Cherfaoui, and J. F. Boissou, "A track-to-track association method for automotive perception systems," *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 704–710, 2012.
- [2] C. Knill, A. Scheel, and K. Dietmayer, "A direct scattering model for tracking vehicles with high-resolution radars," in *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2016-Augus. IEEE, jun 2016, pp. 298–303.
- [3] A. Scheel, C. Knill, S. Reuter, and K. Dietmayer, "Multi-sensor multi-object tracking of vehicles using high-resolution radars," in *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2016-Augus, 2016, pp. 558–565.
- [4] N. Kaempchen and K. Dietmayer, "Data synchronization strategies for multi-sensor fusion," *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, vol. 85, no. 1, pp. 1–9, 2003.
- [5] H. Cho, Y.-W. Seo, B. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, may 2014, pp. 1836–1843.
- [6] A. Milan, K. Schindler, and S. Roth, "Challenges of ground truth evaluation of multi-target tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 21 Suppl 2, no. iii. IEEE, jun 2013, pp. 735–742.
- [7] F. Bourgeois and J.-C. Lassalle, "An extension of the Munkres algorithm for the assignment problem to rectangular matrices," *Communications of the ACM*, vol. 14, no. 12, pp. 802–804, dec 1971.
- [8] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A Benchmark for Multi-Object Tracking," pp. 1–12, mar 2016. [Online]. Available: <http://arxiv.org/abs/1603.00831>
- [9] R. Stiefelhagen, K. Bernardin, R. Bowers, R. T. Rose, M. Michel, and J. Garofolo, "The CLEAR 2007 Evaluation," in *Multimodal Technologies for Perception of Humans*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 4625 LNCS, pp. 3–34.
- [10] Bo Wu and R. Nevatia, "Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*, vol. 1, no. c. IEEE, 2006, pp. 951–958.
- [11] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 319–336, 2009.
- [12] K. Bernardin, "CLEAR 2007 Evaluation Plan – 3D Person Tracking Task," pp. 1–17, 2007.
- [13] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance Measures and a Data Set for Multi-target, Multi-camera Tracking," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9914 LNCS, no. c, pp. 17–35.