On Occlusion-Handling for People Detection Fusion in Multi-camera Networks

Anselm Haselhoff[1], Lars Hoehmann[1], Christian Nunn[2], Mirko Meuter[2], and Anton Kummert[1]

[1] Communication Theory, University of Wuppertal, D-42119 Wuppertal, Germany
{haselhoff,hoehmann,kummert}@uni-wuppertal.de
[2] Delphi Electronics & Safety, D-42119 Wuppertal, Germany
{christian.nunn,mirko.meuter}@delphi.com

Abstract. In this paper a system for people detection by means of Track-To-Track fusion of multiple cameras is presented. The main contribution of this paper is the evaluation of the fusion algorithm based on real image data. Before the fusion of the tracks an occlusion handling resolves implausible assignments.

For the evaluation two test vehicles are equipped with LANCOM® wireless access points, cameras, inertial measurement units (IMU) and IMU enhanced GPS receivers. The people are detected by using an AdaBoost algorithm and the results are tracked with a Kalman filter. The tracked results are transmitted to the opponent vehicle in appropriate coordinates. The final stage of the system consists of the fusion and occlusion handling.

Keywords: Object Detection, Sensor Data Fusion, Occlusion Handling, Camera Networks.

1 Introduction

Surveillance camera systems as well as driver assistance systems become more and more important. The used technologies are image processing, machine learning, sensor fusion, and communication. Thus the boundary between surveillance systems and driver assistance systems disappears. By means of Car2Infrastructure communciation both research fields can be connected. While in the automotive field systems like forward collision warning and communication units are already available, the next logical step is to connect these individual systems and generate an enriched view of the environment. Therefore a sensor data fusion can be applied. When dealing with object detection from different views, like in a multi-vehicle or camera network scenario, it is important to incorporate occlusion information. If occlusion handling is neglected, incorrect track-assignment can lead to strong errors of the object position.

In this work a system for fusing object detections in multi-camera networks is presented [2]. In addition the topic of occlusion handling is treated. It is not important for the system whether the cameras included in the network are static

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 113–119, 2011.

or mounted in a vehicle. The remainder of the paper proceeds as follows. It is started with the description of the overall system and the test vehicles in section 2. Afterwards the components of the fusion system are described separately, including the track assignment, occlusion handling and the fusion of people detections. Finally, the evaluation and the conclusions are presented in sections 4.

2 System Overview

For demonstration the system setup consists of two test vehicles equipped with LANCOM® wireless access points, cameras, inertial measurement unit, GPS, and a regular PC. The monochrome camera is mounted at the position of the rear-view mirror and is connected to the PC. The vehicle bus enables the access to the inertial measurement unit and GPS. The GPS is used to generate timestamps for the data that is subject to transmission. The GPS unit (AsteRxi system) delivers a position accuracy of a around $2cm$ and a heading accuracy of $1°$. The position information is obtained relative to one vehicle that is defined to be the dedicated master. Finally, the LANCOM® unit is responsible for the data transmission.

Fig. 1 illustrates the system that is used for multi-camera people detection and fusion. First, the image is scanned via an AdaBoost detection algorithm. The detection results are then tracked using a Kalman filter that is working in image coordinates. The tracked detections including the uncertainties are then transformed to appropriate coordinates that can be used for the fusion. The transformation of the uncertainties is implemented using the scaled unscented transformation (SUT) [4]. The calibration component delivers the needed camera position by means of the GPS unit. The transformed detections, including their uncertainties, are then transmitted to the other cameras as well as the camera position. The data is then synchronized using the GPS timestamps and passed to the track-assignment module. Afterwards it is checked for occlusion. Corresponding tracks are finally fused by means of a Track-To-Track fusion algorithm.

3 Fusion of People Detections

Each node in the multi-camera network sends the tracked detection results to the other nodes and receives the tracked detection results of the opponents. The detections are described by their position and their uncertainties in the world coordinate system (WCOS). The state estimates of each track are denoted by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, whereas the fused state is $\boldsymbol{\mu}$. The according covariance matrices are denoted by $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}$. The fusion is subdivided into two components, namely the track assignment and the Track-To-Track fusion.

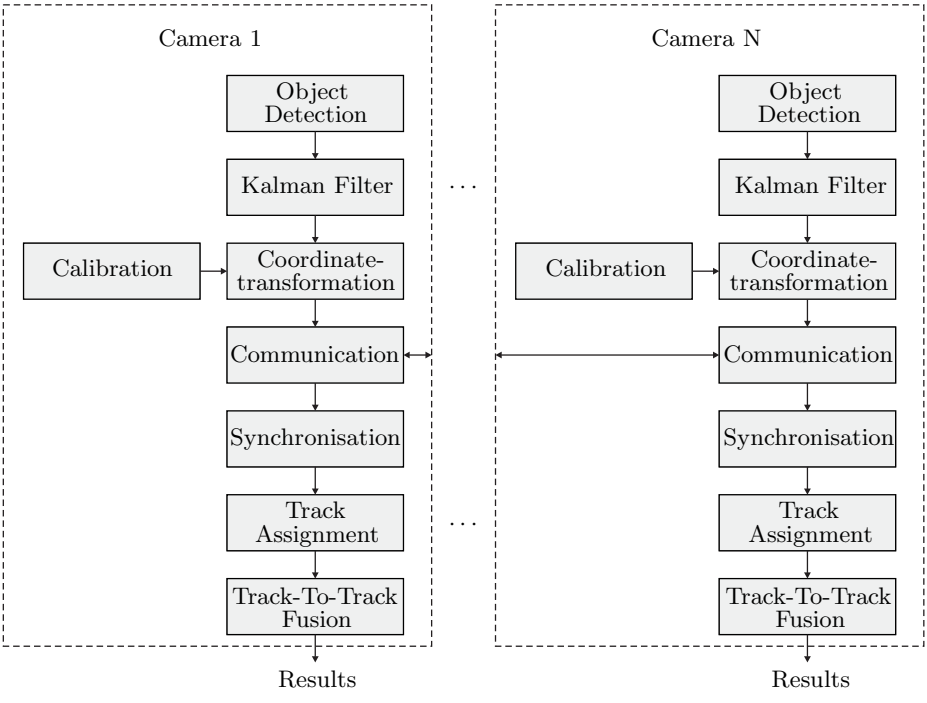


Fig. 1. System for Object Detection Fusion

3.1 Track Assignment and Occlusion Handling

The track assignment is based on a measurement Δ that defines the distance between each track

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \tag{1}$$

$$\Delta = \tilde{\boldsymbol{\mu}}^T \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \tilde{\boldsymbol{\mu}}. \tag{2}$$

Tracks are only fused if the track distance falls below a certain threshold and the tracks must fall into the field of view of both cameras. If these conditions are met the Hungarian method [5] is applied to perform the actual track assignment. After the initial track assignment it is checked for occlusion, to prevent the fusion of implausible assignments.

In Fig. 2 a typical occlusion scenario is shown. There are three ground truth objects, where from V_2 object 2 is occluded by object 1 and from V_1 object 3 is occluded by object 2 (see Fig. 2a). Fig. 2b illustrates a wrong track assignment, where object 3 and object 2 are labeled to belong to the same ground truth object. To resolve this incorrect assignment the projections of the bounding boxes are used (Fig. 2c). The idea is, that if the base point of an object projects into the polygon of any bounding box projection, occlusion has occured.

It is appropriate to fuse two objects that occlude each other, but if in addition one of the objects is occluded by another object, the fusion should be avoided. A simple algorithmic describtion is given in Algorithm 1.

Algorithm 1. Occlusion handling

O_i^1, O_j^2 objects from both vehicles ($V1$ and $V2$) that are subject to fusion
L_i^1, L_j^2 list occluding objects, determined by bounding box projections
if ($L_i^1 = \emptyset$ OR $L_i^1 = \{O_j^2\}$) AND ($L_j^2 = \emptyset$ OR $L_j^2 = \{O_i^1\}$) **then**
  ALLOW FOR FUSION
else
  AVOID FUSION
end if

3.2 People Detection Fusion

Once the track assignment is completed, the tracks are subject to fusion. The choice of the fusion method depends on the data at hand and to what extent the data or sensors are correlated. A comparison of fusion methods on simulated data is given in [3], a survey is given in [6] and a general treatment on data fusion can be found in [1]. Three typical methods for the fusion are *Covariance Fusion* (CF), *Covariance Intersection* (CI), and *Covariance Union* (CU). For our application the CF outperforms the other algorithms [2].

The *Covariance Fusion* (CF) can be defined by the following equations

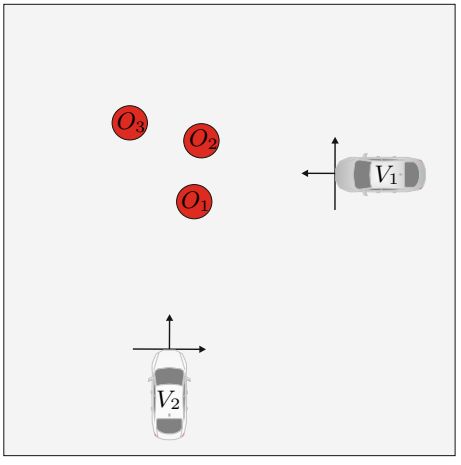
$$\mathbf{C} = \mathbf{C}_1 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \mathbf{C}_2 \tag{3}$$

$$\boldsymbol{\mu} = \mathbf{C}_2 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \boldsymbol{\mu}_1 + \mathbf{C}_1 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \boldsymbol{\mu}_2. \tag{4}$$

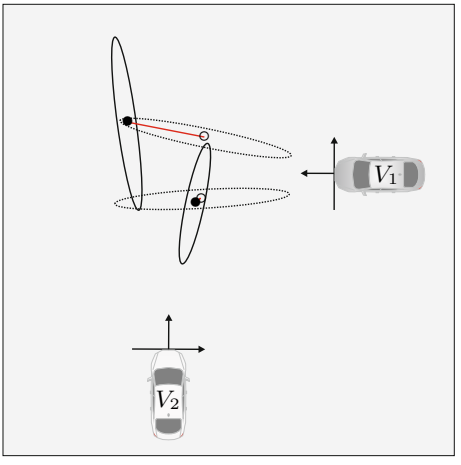
4 Evaluation

The evalution of the fusion method is performed on real data (see Fig. 3) with ground truth information in world coordinates. The system is tested on various video-sequences and different scenarios. The used scenarios are based on typical cross-way situations. In the first scenario both vehicles are approaching the object with an angle difference of 90° and in the second scenario the vehicles are placed at an angle difference of 180°. The distance of the vehicles to the objects at the starting position goes up to around $50m$.

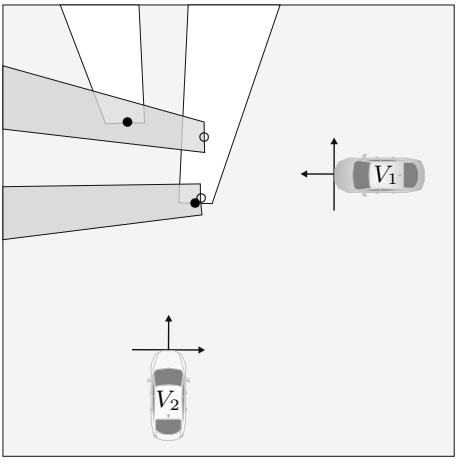
An example of the detection results with and without the occlusion handling is presented in Fig. 3. It becomes clear that the fused detections in Fig. 3a and 3b do not correspond to the ground truth. In contrast if the fusion is avoided for inconsistent tracks the results presented in Fig. 3c and 3d are obtained. These results reflect the ground truth data. To demonstrate the advantage of the fusion, first the detection results of the individual vehicles are evaluated, where a root mean square error (RMSE) of around 4 meters is obtained. The RMSE is calculated using the distance of the ground truth objects to the detections



(a) Ground truth data of three objects. Each camera can perceive two objects. Object 2 is occluded by object 1 and object 3 by object 2.

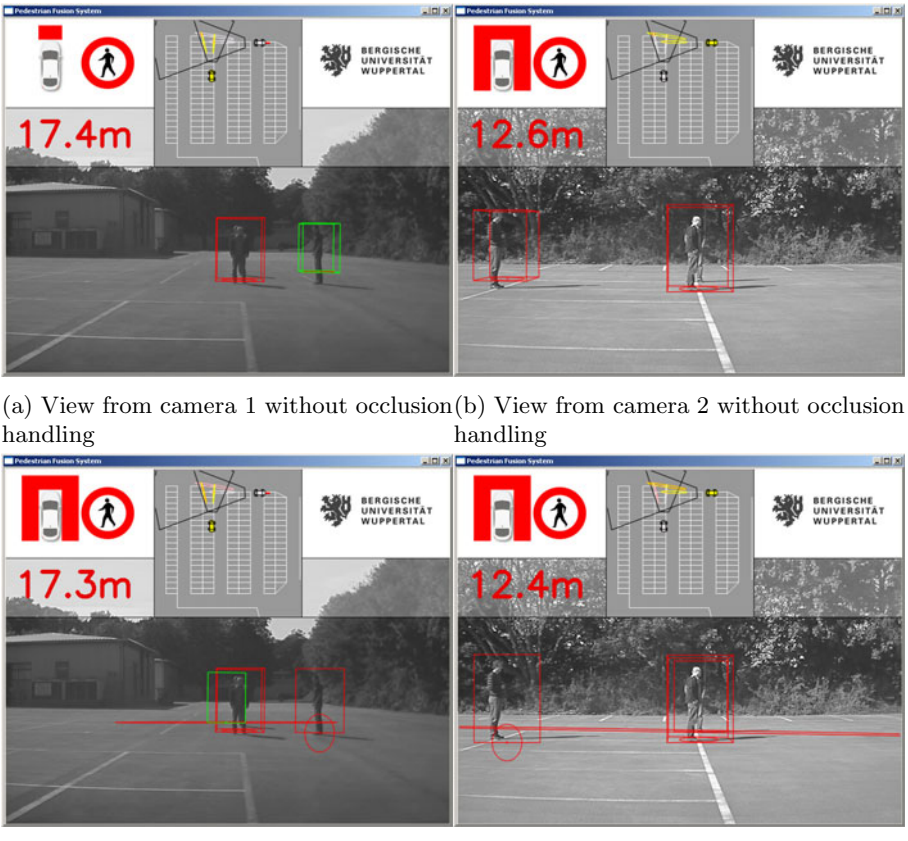


(b) Detected objects including the uncertainties. The track assignment is visualized by the red line. In this case a wrong track assigment has occured.



(c) Bounding box projections onto the ground plane can be used to resolve the wrong track assignment

Fig. 2. Example where the standard track assignment fails



(a) View from camera 1 without occlusion handling

(b) View from camera 2 without occlusion handling

(c) View from camera 1 using occlusion handling

(d) View from camera 2 using occlusion handling

Fig. 3. Fusion results of two cameras mounted in vehicles. Fig. 3a and 3b belong to the same timestamp, where a fusion leads to wrong results. The ego vehicle is gray and the remote vehicle is yellow. The trajectory of the vehicles is denoted by red dots. Fig. 3c and 3d belong to the same timestamp, where the occlusion handling is applied.

in world coordinates. As expected the lateral position (with reference to the vehicle coordinate system) of the objects can be measured precisely, whereas the depth information is inaccurate. The results in Table 1 reveal that the fusion dramatically improves the overall precision. The RMSE of d_w (distance of ground truth object and prediction) gets improved.

Table 1. RMSE: Fusion of two cameras

method	x_w [m]	y_w [m]	d_w [m]
scenario 1			
CF	0.35	0.55	0.69
scenario 2			
CF	0.15	1.00	1.02

5 Conclusions

A promising system for fusing detection results from multiple cameras was presented. The performance results underline the practicability of the approach. The CF method is appropriate for our application since the correlation of the data is very small, e.g. correlation coefficient of around 0.3. The occlusion handling avoids the fusion of objects that would result in erroneous ground truth objects.

References

1. Bar-Shalom, Y., Blair, W.D.: Multitarget-Multisensor Tracking: Applications and Advances. Artech House Inc., Norwood (2000)
2. Haselhoff, A., Hoehmann, L., Kummert, A., Nunn, C., Meuter, M., Mueller-Schneiders, S.: Multi-camera pedestrian detection by means of track-to-track fusion and car2car communication. In: VISAPP (to be published) (2011)
3. Matzka, S., Altendorfer, R.: A comparison of track-to-track fusion algorithms for automotive sensor fusion. In: Proc. of International Conference on Multisensor and Integration for Intelligent Systems, pp. 189–194. IEEE, Los Alamitos (2008)
4. Merwe, R.V.D., Wan, E.: Sigma-point kalman filters for probabilistic inference in dynamic state-space models. In: Proceedings of the Workshop on Advances in Machine Learning (2003)
5. Munkres, J.: Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5(1), 32–38 (1957)
6. Smith, D., Singh, S.: Approaches to multisensor data fusion in target tracking: A survey. IEEE Transactions on Knowledge and Data Engineering 18(12), 1696–1710 (2006)