

Feature-Based Probabilistic Data Association (FBPDA) for Visual Multi-Target Detection and Tracking under Occlusions and Split and Merge Effects

Michael Grinberg

Florian Ohr

Jürgen Beyerer

Abstract—Uncertainties in the sensor data such as measurement noise, false detections caused by clutter, as well as merged, split, incomplete or missed detections due to a sensor malfunction or occlusions (both due to the limited sensor field of view and objects in the scene) make multi-target tracking a very complicated task. Thus one of the big challenges is track management and correct data association between detections and tracks. In this contribution we present an algorithm for visual detection and tracking of multiple extended targets under occlusions and split and merge effects. Unlike most of the state-of-the-art approaches we utilize low-level information integrating it in a unified approach based on a threshold-free probabilistic conception. The introduced scheme makes it possible to utilize information about composition of the measurements gained through tracking of dedicated feature points in the image and resolves data association ambiguities in a soft decision using a globally optimal probabilistic data association approach. Beside existence evolution consideration we also exploit the spatial and temporal relationship between stably tracked points and tracked objects, which along with observability analysis, allows us for reconstruction of compatible measurements and thus correct track update even in cases of splits, merges and partial occlusions of the tracked targets.

Index Terms—multitarget tracking (MTT), data association, split and merge handling, stereo video, environment perception, lateral vehicle perception, vehicle side monitoring, side-looking cameras, 6D vision, occlusion handling

I. INTRODUCTION

Most Bayesian target tracking methods, such as the Kalman filter [1], are based on filtering of successive sensor measurements. This requires robust detection of the tracked targets as well as correct data association. However, in many applications both target detection and data association are very complex tasks due to uncertainties in the data. In addition to the standard problem of noisy (i.e. inaccurate) measurements, one has to cope with incomplete or missed detections, false detections due to clutter, and wrong associations between tracks and detections. Another common effect is the incidence of multiple detections evoked by one target (split) or the fusion of multiple targets to one detection (merge). Similar effects may emerge due to partial or complete occlusion of the targets. For an implementation of a robust tracking system an appropriate data association method which can effectively handle all of the above-mentioned effects is therefore essential. In this contribution

we propose a data association scheme based on tracking of dedicated feature points for enabling of a robust multi-target object detection and tracking under occlusion and split and merge effects. Our method allows to compute affiliation probability of each point to the currently tracked objects. This knowledge is utilized in the data association step for a weighted combination of measurements obtained from point clouds to a single statistically most probable measurement which is then used for the object's state innovation. Additionally, a scheme for explicit observability modeling and propagation is proposed.

The paper is structured as follows: Section II summarizes related work, Section III deals with the observability modeling, in Section IV, an overview of the application and the overall object tracking framework is given, Sections V-VIII give a detailed description of each step of the proposed Feature-Based Probabilistic Data Association Algorithm, in Section IX, the experimental setup and results are presented. Finally, a short summary is given in Section X.

II. RELATED WORK

Most of the video-based object tracking applications utilize background estimation methods for extraction of moving image blobs, which are then tracked using methods such as the mean shift or CAMSHIFT algorithms [2], [3], [4]. Availability of the depth information (e.g. from stereo processing) makes it possible to track objects in three-dimensional space using estimates of their centers of gravity (CoG). In both types of applications though, possible confusion of neighboring objects may cause tracker failure. Further major challenges are occlusions and false detections (clutter). In particular, applications which can not utilize background estimation (e.g. video-based object tracking performed from a moving platform) suffer from these effects. One of the major problems hereby is that the tracker presupposes object existence and observability and does not account for their uncertainty. This problem was studied thoroughly in the radar tracking literature [5], [6], [7], [8].

In the past few decades several algorithms for multi-target tracking have been proposed. Probabilistic Data Association (PDA) [5], [6] as well as Joint Probabilistic Data Association (JPDA) [9], [10] proposed by Bar-Shalom et al. aim at utilizing all sensor measurements taking into account statistical distribution of errors and clutter. Mušicki et al. proposed in [7], [11] an approach that allows for representation of the uncertainty of the target existence, and for integration of

The authors are with the Fraunhofer Institute for Information and Data Processing (Fraunhofer IITB), Fraunhofer Str. 1, 76131 Karlsruhe, Germany, <http://www.iitb.fraunhofer.de/> {michael.grinberg, florian.ohr, juergen.beyerer}@iitb.fraunhofer.de

track initiations into the PDA framework (Integrated Probability Data Association, IPDA). The combination of JPDA and IPDA resulted in the Joint Integrated Data Association (JIPDA) algorithm [8] which was shown to be effective also for multi-sensor data fusion [12], [13].

All mentioned algorithms were developed to cope with distant target tracking applications using radar, LIDAR or sonar sensors. Thus they were designed under the assumptions that a target may be represented by its center of gravity, that a target may cause at most one detection, and that a detection may be evoked by at most one target.

When tracking extended targets, one has to deal with split and merge effects which are not handled by standard approaches. Those effects were tackled in research on visual tracking. Kumar et al. [14] proposed a method which copes with split and merge effects by evaluating the associations between the combined detections and tracks, and choosing the ones of maximal probability based on distance between contours of the resulting image blobs. However, even with identified split or merge events it is not possible to perform an appropriate update step. Instead, prediction is used for further processing until a resolution can be made. Genovesio et al. propose in [15] to create *virtual measurements* that always allow for an appropriate update. Thereby the most probable association is used. For splitting of joint measurements evoked by several tracks, they propose to use the k-means algorithm, which may become problematic when merged objects have unequal dimensions. Furthermore, gating failures arising when in case of a merge the CoG of the joint measurement does not lie within the gate of one or more corresponding tracks lead to measurement rejections as well as to missing or wrong splits as shown in Fig. 1.

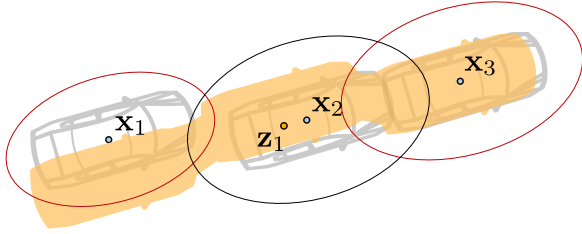


Fig. 1. Example of measurement rejection due to gating failure in the case of merged targets: CoG of the joint measurement z_1 does not lie within the gates of the tracks x_1 and x_3 .

III. MODELING THE TRACK'S OBSERVABILITY

Another important point is object observability. Mostly it is assumed to always be given. Occlusions if any are modeled and handled as measurement noise (varying object dimensions) or as a sensor malfunction (missing detections). Direct coupling (or equalization) of observability and existence leads to the existence-based track termination after several tracking cycles. For coping with this problem we propose an observability treatment scheme utilizing a grid-based object representation with occupancy and occlusion modeling for each cell. In addition to the two Markov chains that are used

for modeling target's dynamic state and existence evolution we consider the observability probability as a third Markov chain. This allows for a dedicated observability and occlusion handling and thus decoupling of existence and observability. Besides, we utilize knowledge gained by tracking dedicated feature points in the image for an appropriate formation of the measurements from initial point clouds for efficient handling of partial occlusions as well as split and merge effects. Additionally, in our contribution we combine the advantages of the described approaches. Measurements are combined in a statistically optimal way to account for association uncertainty. Global optimality of association between multiple detections and multiple targets is achieved by creating composite measurements for track updates in the context of maximization of the global association probability.

In the standard IPDA and JIPDA algorithm, two cross-coupled Markov chains are used for target state propagation (cf. Fig. 2). One Markov chain propagates the Gaussian

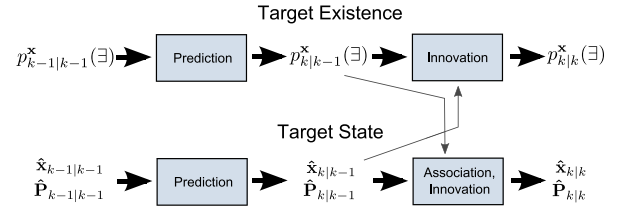


Fig. 2. Two cross-coupled Markov chains used in the IPDA algorithm

probability density function (pdf) of the target's dynamic state which is parameterized by its state estimate \hat{x} and the error covariance matrix \hat{P} . The second Markov chain models the probability mass function (pmf) of the target's existence.

In [11], two types of Markov chains are proposed for the target existence modeling. The first type has two states: one of the target being existent at time step k (denoted in the following as Ξ) and the second one of the target being non-existent at time step k (denoted as \nexists). The second type of existence Markov chain has three states: target is existent and observable (denoted in the following as \supset), target is existent but not observable (denoted as $\not\supset$) and target is non-existent, i.e. it is a phantom object (denoted as P) as depicted in Fig. 3(a). The Markov chain of the first type can be derived from the Markov chain of the second type by bundling the two first states as shown in Fig. 3(b). It is also possible to bundle the two last states (i.e. to model occluded targets as not existent, cf. Fig 3(c)) as done in [13]. This leads to the existence-based track termination of occluded objects.

As in [8], [16], [12], [13], most researchers are content with considering a Markov Chain with two states. In our approach, we explicitly consider target's observability. Contrary to other approaches, we model it as a separate Markov chain (cf. Fig. 3(d), Fig. 4) hereby decoupling object existence and observability. For this aim we exploit a grid based object representation which allows us to build appearance and occlusion maps for each object and thus explicitly compute occlusion degree and observability probability for each tracked object. Further details on this procedure are

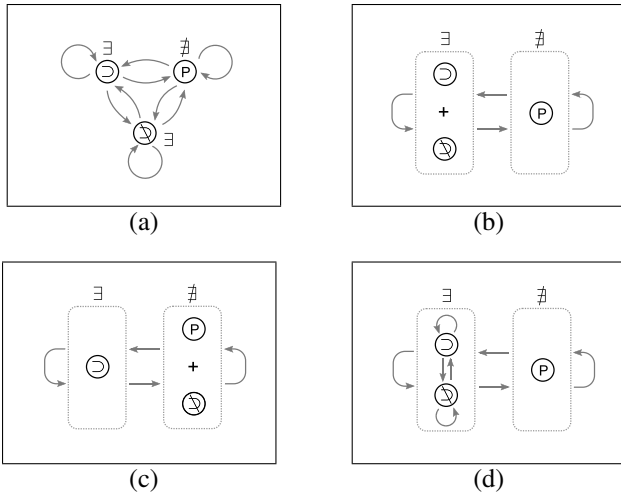


Fig. 3. Markov chain based modeling of object existence and observability. The three possible track states \supset , \emptyset and P (object existent and observable, object existent but not observable and object non-existent (phantom track)) can be bundled in different ways. The two umbrella states are visualized with \exists and $\#$.

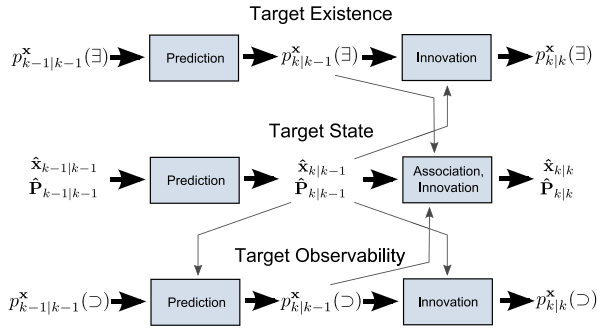


Fig. 4. Three cross-coupled Markov chains used in the FBPDA

given in Section V-C.

IV. OVERALL SYSTEM DESCRIPTION

The video-based object tracking framework which is taken as a basis for this work has been built at Fraunhofer IITB in the course of the EU funded project APROSYS [17]. The goal was detection of imminent side collisions to enable timely activation of novel occupant protection systems [18]. The sensor system under consideration here is a side-looking stereo video camera [19]. Ego-motion estimation and compensation is done using vehicle odometry data which was shown to deliver sufficient accuracy even for the structure-from-motion task [20].

Many useful constraints and assumptions being utilized in the applications developed for the front view do not apply to the side-looking perspective. Other than in the majority of such applications, neither symmetrical appearance nor restrictions about possible object locations and motion profiles (e.g. relative to the tracked lane markings) can be assumed. As Fig. 5 illustrates, many different object orientations have to be taken into account. Combined with a large variety of object types, this makes it extremely challenging to detect and track objects based on their appearance in real time. For



Fig. 5. Typical traffic scene observed by a side looking camera

the realization of a side-looking video system we thus had to choose a generic approach to achieve a robust and reliable object detection. Object hypotheses are therefore generated from the depth maps which are the result of stereo processing. After extracting a depth map, road location is estimated and then points belonging to the road are eliminated. Spatial clustering of the remaining three-dimensional points delivers initial point clouds which are used as an input in the data association step.

Due to the noisy depth extraction process, combined with well known problems of stereo vision systems such as difficulties of depth estimation in homogeneous image regions and gross depth errors in regions with regular patterns, the results of clustering may vary from frame to frame leading to split and merged object measurements. One of the possibilities of coping with this problem on the lowest level is the utilization of optical flow for feature tracking in the image and using of the resulting 2-D or 3-D motion patterns for hypotheses generation [21], [22], [23]. This however, can not be applied effectively for handling of merged stationary objects or objects with similar motion profiles. Yet, for data association this additional low-level information is very useful and may be utilized for successful resolution of the data association problem and for stabilization of tracking. In our algorithm we estimate the probabilities of the affiliation of the tracked points to the tracked objects and use them for both data association and object tracking. For detection and tracking of the feature points we use a 3-D point tracking system similar to the one proposed in [23]. Up to 3000 feature points are tracked simultaneously in 3-D space using Kalman Filters. Their six-dimensional state vectors $[x, y, z, v_x, v_y, v_z]^T$ are estimated from their image coordinates and displacement (optic flow) as well as stereo depth measurements. The measurement vector is thus $[u, v, d]^T$ with image coordinates (u, v) and feature depth d .

After elimination of the ground points, remaining points are clustered to point clouds which give detections for object tracking. For performing the association between point clouds and tracks and for track state propagation we propose to use the Feature-Based Probabilistic Data Association algorithm (FBPDA) which is described in the following sections. FBPDA consists of four major steps: time forward prediction, association between detections and tracks, reconstruction of composite measurements from associated point clouds and

innovation. These four steps are described in Sections V, VI, VII and VIII respectively. For effective noise handling as well as for handling of split and merged point clouds FBPDA provides affiliation probabilities of each point to the current tracks. These affiliation probabilities are exploited in the course of updating tracks' position, dynamics and dimensions.

Object parameters are estimated using an Extended Kalman Filter. Internally, objects are modeled as cuboids with a centroid (x, y, z) , dimensions (l, w, h) and geometrical orientation ϕ , motion orientation φ , speed v , acceleration a and yaw rate $\dot{\varphi}$. This corresponds to the Constant Yaw Rate Model with Acceleration, which has proved to deliver the best performance in the case of a side-looking system [24]. Due to the fact that often only a part of an object is visible to the sensors, orientation of the object's motion may differ from the estimated geometrical orientation. Object's geometric orientation and dimensions are updated taking into account the occlusion information as described in Section VII.

Figure 6 illustrates principal architecture of the implemented framework.

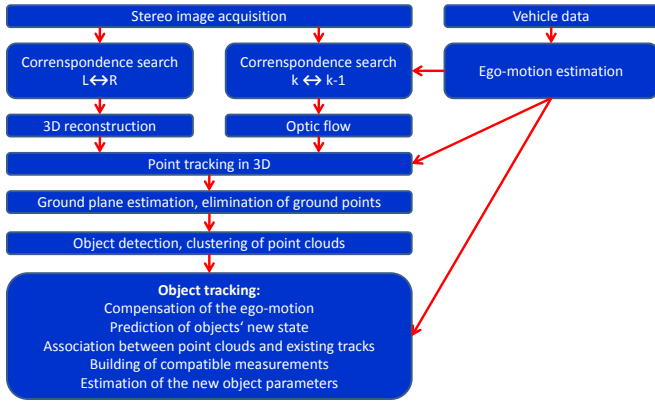


Fig. 6. Overall framework for visual object detection and tracking

V. TIME FORWARD PREDICTION

At each time step, after the ego-motion compensation, we start by predicting the new track attributes. This includes prediction of the track's dynamic state as well as prediction of its existence and observability probabilities.

A. State prediction

The FBPDA state prediction of a track is identical to the common Extended Kalman Filter prediction.

B. Existence prediction

The time forward prediction equations for the existence and non-existence probabilities $p_{k|k-1}^{\mathbf{x}}(\exists)$ and $p_{k|k-1}^{\mathbf{x}}(\nexists)$ of a track \mathbf{x} are:

$$\begin{aligned} p_{k|k-1}^{\mathbf{x}}(\exists) &= p^{\mathbf{x}}(\exists \rightarrow \exists) \cdot p_{k-1|k-1}^{\mathbf{x}}(\exists) \\ &\quad + p^{\mathbf{x}}(\nexists \rightarrow \exists) \cdot p_{k-1|k-1}^{\mathbf{x}}(\nexists) \\ p_{k|k-1}^{\mathbf{x}}(\nexists) &= 1 - p_{k|k-1}^{\mathbf{x}}(\exists) \end{aligned} \quad (1)$$

with $p_{k-1|k-1}^{\mathbf{x}}(\exists)$ being the a-posteriori existence probability from the last frame and $p^{\mathbf{x}}(\exists \rightarrow \exists)$ and $p^{\mathbf{x}}(\nexists \rightarrow \exists)$ denoting the persistence and the birth probabilities of a track \mathbf{x} . The last two factors are used for modeling the spatial distribution of the target birth and death probabilities. This makes it possible to account e.g. for the fact that at the borders of the field of view and at far distances the birth probability is higher than right in front of the sensors.

C. Observability prediction

For the computation of the observability probability and for the reconstruction of occluded measurements we use a grid-based 3-D representation of the targets. For each track \mathbf{x}_i , we define a 3-D grid with the origin at its centroid. The orientation of the grid is aligned with the track's orientation. Using this representation of predicted objects it is possible to calculate their appearance masks in the camera image $M_A^{\mathbf{x}_i}(u, v) \in \{0, 1\}$ by projecting the occupied grid cells into the image. The Appearance Probability Mask $M_{p(A)}^{\mathbf{x}_i}(u, v)$ for each track is given by

$$M_{p(A)}^{\mathbf{x}_i}(u, v) = p(A^{\mathbf{x}_i}(u, v)) = M_A^{\mathbf{x}_i}(u, v) \cdot p_{k|k-1}^{\mathbf{x}_i}(\exists) \quad (2)$$

where $A^{\mathbf{x}_i}(u, v)$ is the event of the track \mathbf{x}_i appearing at the image position (u, v) and $p_{k|k-1}^{\mathbf{x}_i}(\exists)$ is the predicted existence probability of the track \mathbf{x}_i . By overlaying the appearance probability masks of all the objects lying in front of the object \mathbf{x}_i we get the occlusion probability map for the respective object in the new frame. The occlusion probability $p^{\mathbf{x}_i}(\nexists, u, v)$ at each pixel (u, v) is calculated as

$$p^{\mathbf{x}_i}(\nexists, u, v) = p\left(\bigcup_{\mathbf{x}_r \in X_O^i} A^{\mathbf{x}_r}(u, v)\right) \quad (3)$$

with X_O^i being set of the objects lying at the pixel position (u, v) in front of the object \mathbf{x}_i . After the occlusion probability map for an object is built, we can estimate the occlusion probability $p^{\mathbf{x}_i}(\nexists, c)$ of the object's grid cells. This is done by projecting the cell centers into the image and grabbing the corresponding value of the occlusion probability map. Fig. 7 illustrates the grid-based determination of the occlusion probability. Based on these probabilities we can calculate the observability transition probabilities $p^{\mathbf{x}_i}(\exists \rightarrow \exists)$, $p^{\mathbf{x}_i}(\exists \rightarrow \nexists)$, $p^{\mathbf{x}_i}(\nexists \rightarrow \exists)$ and $p^{\mathbf{x}_i}(\nexists \rightarrow \nexists)$ of a track and predict its observability probability and occlusion probability in the new frame:

$$p_{k|k-1}^{\mathbf{x}}(\exists) = p^{\mathbf{x}}(\exists \rightarrow \exists) \cdot p_{k-1|k-1}^{\mathbf{x}}(\exists) \quad (4)$$

$$+ p^{\mathbf{x}}(\nexists \rightarrow \exists) \cdot p_{k-1|k-1}^{\mathbf{x}}(\nexists) \quad (5)$$

$$p_{k|k-1}^{\mathbf{x}}(\nexists) = 1 - p_{k|k-1}^{\mathbf{x}}(\exists) \quad (6)$$

VI. DATA ASSOCIATION BETWEEN DETECTIONS AND TRACKS

Given multiple active tracks and multiple detections there are often several assignment possibilities being more or less probable. Unlike other methods such as GNN, FBPDA does not choose one of these hypotheses for the innovation of a track, but considers all assignment possibilities in a soft decision. The complexity of the algorithm would grow

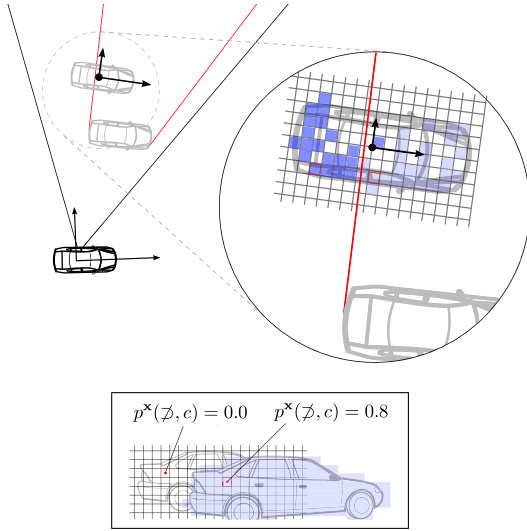


Fig. 7. Grid based estimation of object's occlusion probability. The occupied cells of the half-occluded car are marked in the top view in different shades of blue dependent on their occlusion probability. Occlusion probability is calculated from the occlusion map which is shown in the lower part of the Figure. The occluding car has existence probability of $p_k(\Xi) = 0.8$

exponentially with the number of targets and point clouds. This is a NP-hard problem [25]. For reduction of the total computational costs we introduce track clustering as a pre-processing step using a gating scheme similar to the one proposed in [8]. In the following, the disjunct track subsets can be processed independent from each other.

A. Assignment hypothesis definition

Assignment hypotheses between point clouds and tracks are modeled as a bipartite graph with edges $e : (\mathbf{x} \in X \mapsto \mathbf{z} \in Z)$ between elements of the set of targets X and elements of the set of detections Z . We define the set $X = \{X', \mathbf{x}_B, \odot\}$ as the aggregation $X' = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of the n current tracks plus two special elements \mathbf{x}_B and \odot representing a so far not known object and a clutter source and the set $Z = \{Z', \mathbf{z}_\emptyset, \mathbf{z}_\mathcal{D}, \mathbf{z}_\mathcal{I}\}$ as the aggregation $Z' = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m\}$ of the m current point clouds plus three special elements \mathbf{z}_\emptyset , $\mathbf{z}_\mathcal{D}$ and $\mathbf{z}_\mathcal{I}$. The element \mathbf{z}_\emptyset stands for erroneously missed detection caused by sensor failure, $\mathbf{z}_\mathcal{D}$ represents the correct absence of a detection because of occlusion of the target and $\mathbf{z}_\mathcal{I}$ the correct absence of a detection because of the target non-existence. Assignments between two special elements are prohibited. A valid assignment hypothesis can thus contain six types of edges e :

- $\mathbf{x}_i \mapsto \mathbf{z}_j$: assumption that point cloud \mathbf{z}_j has been caused by the track \mathbf{x}_i
- $\mathbf{x}_B \mapsto \mathbf{z}_j$: point cloud \mathbf{z}_j has been caused by a so far not known object
- $\odot \mapsto \mathbf{z}_j$: assumption that point cloud \mathbf{z}_j has been caused by clutter
- $\mathbf{x}_i \mapsto \mathbf{z}_\mathcal{D}$: track \mathbf{x}_i did not cause a point cloud because it is not observable

- $\mathbf{x}_i \mapsto \mathbf{z}_\mathcal{I}$: track \mathbf{x}_i did not cause a point cloud because it does not exist
- $\mathbf{x}_i \mapsto \mathbf{z}_\emptyset$: track \mathbf{x}_i did not cause a point cloud because of a sensing error

Fig. 8 illustrates four valid assignment hypotheses for the case of $n = 2$ tracks and $m = 2$ point clouds.

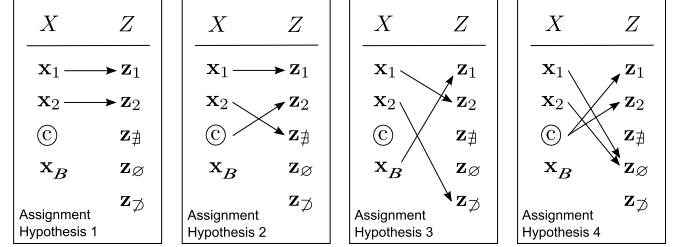


Fig. 8. Examples of valid assignment hypotheses

B. Calculation of the assignment hypothesis probabilities

The probabilities of the six edge types are calculated in analogy to [13]:

$$p(e = (\mathbf{x}_i \mapsto \mathbf{z}_j)) = \quad (7)$$

$$p_{k|k-1}^{\mathbf{x}_i}(\Xi) \cdot p_{k|k-1}^{\mathbf{x}_i}(\supset) \cdot p^{\mathbf{z}_j}(TP|\mathbf{x}_i) \cdot (1 - p^{\mathbf{z}_j}(FP))$$

$$p(e = (\mathbf{x}_i \mapsto \mathbf{z}_\mathcal{D})) = \quad (8)$$

$$p_{k|k-1}^{\mathbf{x}_i}(\Xi) \cdot (1 - p_{k|k-1}^{\mathbf{x}_i}(\supset))$$

$$p(e = (\mathbf{x}_i \mapsto \mathbf{z}_\emptyset)) = \quad (9)$$

$$p_{k|k-1}^{\mathbf{x}_i}(\Xi) \cdot p_{k|k-1}^{\mathbf{x}_i}(\supset) \cdot p^{\mathbf{x}_i}(FN)$$

$$p(e = (\mathbf{x}_i \mapsto \mathbf{z}_\mathcal{I})) = \quad (10)$$

$$(1 - p_{k|k-1}^{\mathbf{x}_i}(\Xi)) \cdot p_{k|k-1}^{\mathbf{x}_i}(\supset) \cdot (1 - p^{\mathbf{x}_i}(FN))$$

$$p(e = (\mathbf{x}_B \mapsto \mathbf{z}_j)) = \quad (11)$$

$$(1 - \sum_i p^{\mathbf{z}_j}(TP|\mathbf{x}_i)) \cdot (1 - p^{\mathbf{z}_j}(FP))$$

$$p(e = (\odot \mapsto \mathbf{z}_j)) = \quad (12)$$

$$(1 - \sum_i p^{\mathbf{z}_j}(TP|\mathbf{x}_i)) \cdot p^{\mathbf{z}_j}(FP)$$

with $p^{\mathbf{z}_j}(FP)$ being the probability that a clutter-caused point cloud appears at the position of \mathbf{z}_j , $p^{\mathbf{x}_i}(FN)$ the false negative probability for the track \mathbf{x}_i and $p^{\mathbf{z}_j}(TP|\mathbf{x}_i)$ being the likelihood function obtained from the Kalman filter.

With (7)-(12) it is now possible to calculate the probability of an assignment hypothesis $E = \{e_1, \dots, e_q\}$ by:

$$p(E) = \prod_{a=1}^q p(e_a) \quad (13)$$

C. Calculation of the association probabilities

The association probabilities $\beta_{\mathbf{z}_j}^{\mathbf{x}_i}$ are calculated as the sum of all assignment hypothesis probabilities including the edge $e = (\mathbf{x}_i \mapsto \mathbf{z}_j)$ divided by the sum of the probabilities of all assignment hypotheses assuming track \mathbf{x}_i as existent:

$$\beta_{\mathbf{z}_j}^{\mathbf{x}_i} = \frac{\sum_{\{E|e=(\mathbf{x}_i \mapsto \mathbf{z}_j) \in E\}} p(E)}{\sum_{\{E|e=(\mathbf{x}_i \mapsto \mathbf{z}_j) \notin E\}} p(E)}. \quad (14)$$

$\beta_{z_0}^{x_i}$ is the probability of the event that no point cloud originated from track x_i . It is calculated analogously to (14).

VII. HANDLING OF SPLIT AND MERGE EFFECTS AND RECONSTRUCTION OF COMPATIBLE OBJECT MEASUREMENTS

For handling of the possibility of split and merge events in the course of data association, one option is to allow the assignment of a set of point clouds to one track or a set of tracks to one point cloud, respectively. This would make it possible to identify splits and merges but would not allow to make an appropriate update since there is a one-to-one association necessary for doing so. Another possibility would be to create virtual measurements from the original measurements by splitting and merging them using predicted states of the tracked objects as shown in Fig. 9. This approach

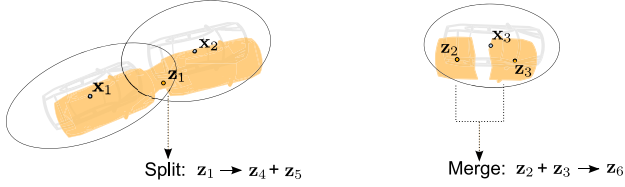


Fig. 9. Creation of virtual measurements for possible split and merge events

would maintain one-to-one matchings between objects and associated point clouds and would allow to update the track's state after creation of compatible state measurement. The disadvantages of the method are ambiguous partitioning of original measurements in case of merging targets and exploding number of feasible associations. Furthermore, when using a centroid-based tracking approach, in the case of split and merged targets the resulting measurement's CoG might not lie inside the gating ellipse of the targets' tracks as shown in Fig. 1. Another problematic case is the position update of a target which is partially occluded (due to either other objects in the scene or restricted field of view of the sensors). If target position is updated based on the CoG of the point cloud, it will cause a shift of the track's centroid and induce an impulse which will introduce bias to the estimated track's state.

To avoid such problems we reconstruct for each detection-to-track association the track's centroid and orientation using stably tracked points. We hereby utilize information about the affiliation of the tracked points to the tracks independently of the currently occurred splits and merges. The reconstructed centroids can then be used instead of the CoGs of the corresponding point clouds for the computation of the state measurement for each track.

A. Determination of the point-to-track affiliation probabilities

The affiliation probability $p(x_i \mapsto p_q)$ of a point p_q to a tracked object x_i is determined based on the association probability $\beta_z^{x_i}$ of the point cloud z containing p_q to that

object. For realization of a memory effect we filter the affiliation probabilities using a gain constant $g \in [0, 1]$:

$$p_k(x_i \mapsto p_q) = g \cdot \beta_z^{x_i} + (1 - g) \cdot p_{k-1}(x_i \mapsto p_q). \quad (15)$$

Initialization of a point is done with

$$p(\odot \mapsto p_q) = 1, \quad p(x_i \mapsto p_q) = 0 \quad \forall x_i. \quad (16)$$

B. Point cloud based reconstruction of the track's position and orientation

For the reconstruction of the track centroid p_O from an associated point cloud we first calculate the CoG p_{CoG} of the stably tracked points of this point cloud both in the current and previous frame. Hereby we are weighting the points' 3-D positions with their track affiliation probability (known from the previous frame). Having computed p_{CoG} in both current and previous frames, we can for each tracked point p_q reconstruct the vector $(\overrightarrow{p_{CoG}p_O})_q$ pointing from the CoG to the object centroid p_O in the current frame using knowledge about the relative orientation of this vector regarding the vector $\overrightarrow{p_{CoG}p_q}$ in the previous frame (cf. Fig. 10). Building a

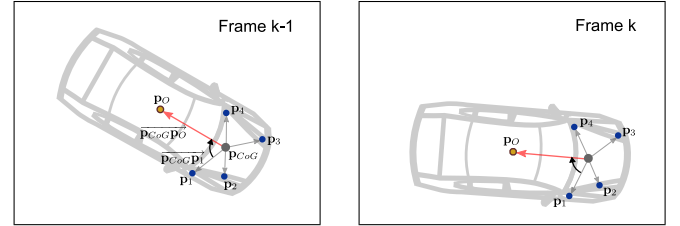


Fig. 10. Object position reconstruction based on spatial relationship to stably tracked points

weighted sum of the resulting vectors $(\overrightarrow{p_{CoG}p_O})_q$ according to the affiliation probability of the respective points p_q we get the new position of the track's centroid with respect to the considered point cloud. The track's new orientation can be obtained in the same way. Together those parameters form the reconstructed measurement of the point cloud. All these measurements weighted according to the association probabilities of the corresponding point clouds build a composed measurement which is then used for the innovation of the Kalman Filter responsible for the track's dynamics.

C. Grid-based reconstruction of the track's extent

The extent of a track is obtained in a similar way. We create one composite measurement from all point clouds associated to the track. This is done using the grid representation of the track which is aligned according to its new position and orientation gained through the innovation of the track's dynamics. The points of each associated point cloud are sorted into the grid. Thereby points of a point cloud contribute to the occupancy value of the grid cell according to their affiliation probability to the track and the association probability of the point cloud. For each cell c its current occupancy value o at time step k with respect to the track

\mathbf{x}_i is computed according to

$$o_k^{\mathbf{x}_i}(c) = \sum_{j=1}^m \left(\beta_{\mathbf{z}_j}^{\mathbf{x}_i} \cdot \sum_{q=1}^l p(\mathbf{x}_i \mapsto \mathbf{p}_q) \right) \quad (17)$$

with m being current number of point clouds and l being the number of points \mathbf{p}_q belonging to the point cloud \mathbf{z}_j and falling into the cell c . The grid is updated using this occupancy values. To avoid an update of the occupancy value of an occluded cell with 0, we filter the occupancy values of the cells using their occlusion probability. For this purpose we use the recursive filter depicted in Fig. 11. For each cell

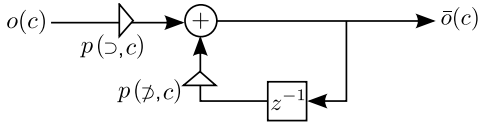


Fig. 11. Recursive filter used for filtering of cell occupancy probability

c its filtered occupancy value $\bar{o}_k(c)$ at time step k is given by

$$\bar{o}_k^{\mathbf{x}_i}(c) = p_k^{\mathbf{x}_i}(\triangleright, c) \cdot o_k^{\mathbf{x}_i}(c) + (p_k^{\mathbf{x}_i}(\not\triangleright, c)) \cdot \bar{o}_{k-1}^{\mathbf{x}_i}(c) \quad (18)$$

$$\text{with } p_k^{\mathbf{x}_i}(\triangleright, c) = 1 - p_k^{\mathbf{x}_i}(\not\triangleright, c). \quad (19)$$

Object's geometric orientation and dimensions are obtained using a RANSAC based estimation of the main visible object surface in the top view and fitting a rectangle with this orientation into the ground projection of the points. Together, these parameters form the resulting composite measurement which is then used for the innovation of the track's geometric orientation and its dimensions.

VIII. STATE, EXISTENCE AND OBSERVABILITY INNOVATION

In the innovation step, all three Markov chains are updated. The innovation of the target's physical attributes corresponds to the standard Extended Kalman Filter innovation. For maintaining the centroid position obtained from the grid based object extent computation as the reference point to be tracked, we switch to this point as a new reference point at the end of each frame. This prevents the object dimensions from jittering due to the one-sided changes of the visible object extent (e.g. in the case of a target entering the field of view). The a-posteriori probability of the track existence is calculated as the sum of the probabilities of all assignment hypotheses assuming track \mathbf{x}_i as existent divided by the sum of the probabilities of all possible hypotheses:

$$p_{k|k}^{\mathbf{x}_i}(\exists) = \frac{\sum_{\{E|e=(\mathbf{x}_i \mapsto \mathbf{z}_{\#}) \notin E\}} p(E)}{\sum_{\{E\}} p(E)}. \quad (20)$$

The observability update is done analogously to the existence innovation.

IX. EXPERIMENTAL RESULTS

The algorithm has been validated with both simulated and real data. For our simulations we defined several scenarios that caused problems for the standard approach such as splitting and merging point clouds, objects entering and leaving the field of view (FoV) and occlusion scenarios. These scenarios were simulated using a tool which generated point clouds with pre-defined parameters and behavior. Fig. 12 shows a very simple scenario of an overtaking car passing the FoV of the camera-carrying vehicle. Left column shows performance of the original filter. At the beginning, when the object is entering the FoV and "growing" in the image, the center of gravity of the visible points becomes slower causing erroneous underestimation of the target's velocity followed by a sudden acceleration when the entire object becomes visible. Splitting of the point cloud into two parts causes the track to decelerate rapidly and then to accelerate again as soon as the complete object is detected again. Finally, at the end of the sequence, while the object leaves the FoV of the cameras, the CoG of the visible points becomes slower again causing the track to decelerate and to partially remain in the FoV. Contrary to this behavior, FBPDA manages to always

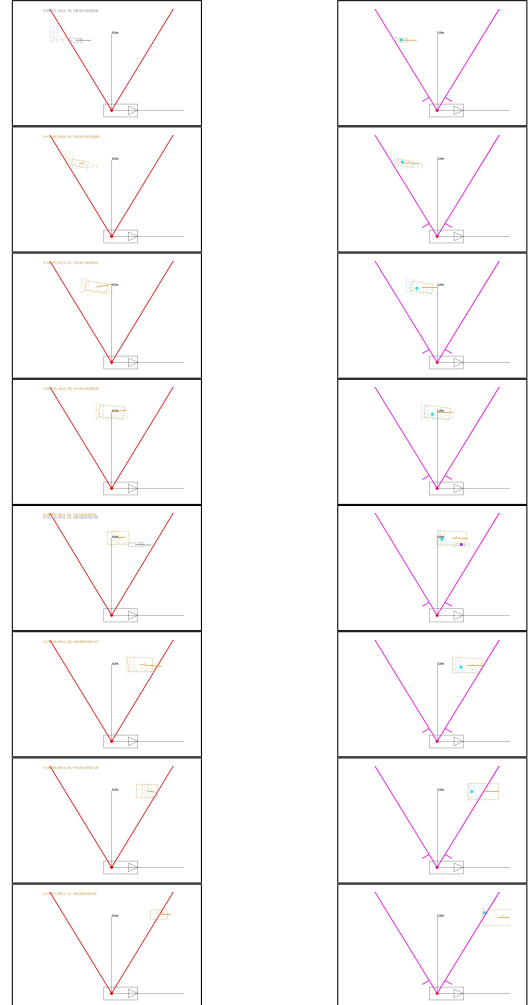


Fig. 12. Split scenario: FBPDA (right) vs. standard approach (left)

maintain the correct object velocity and even to correctly track it when it's almost completely outside of the field of view of the sensors (right column).

Similar effect can be observed in the case of an occlusion caused by another object in the scene. Fig. 13 shows a plot of the object's lateral position obtained by the original approach (red) and FBPDA (blue) vs. the ground truth (black). Compared to the standard association and tracking scheme which leads to considerable corruption of the position (and velocity) estimation and even termination and re-initialization of the track, FBPDA manages it to correctly update tracks' parameters through multiple merges, splits and occlusions.

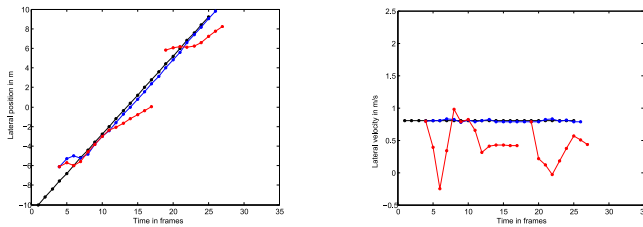


Fig. 13. Overtaking car scenario with occlusion, plots of the lateral position and velocity: original approach (red), FBPDA (blue) and ground truth (black).

X. CONCLUSION

In this contribution we have presented an algorithm for visual detection and tracking of multiple extended targets which is capable of coping with noisy, split, merged, incomplete and missed detections. The proposed Feature-Based Probabilistic Data Association approach resolves data association ambiguities in a soft threshold-free decision based not only on target state prediction but also on the existence and observability estimation modeled as two additional Markov Chains. This process is supported by a grid-based object representation which is used for the detailed occlusion analysis. For the correct estimation of the desired object parameters, low-level information about measurement composition is utilized which is gained through tracking dedicated feature points in the image and 3D space. Along with the occlusion analysis, spatial and temporal relationship between the set of stably tracked points and the object's centroid is exploited for the reconstruction of the desired object characteristics from the data even in case of detection errors due to limited field of view, occlusions, noise and sensor malfunction. For tracking applications that have to cope with the above-mentioned effects, our algorithm offers a much needed enhancement which has the potential to greatly increase detection and tracking performance and overall system robustness.

XI. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contribution of Dr. Dieter Willersinn, Prof. Mohan Trivedi, Dipl.-Inf. Michael Mai, and all members of the APROSYS Task Force at the Fraunhofer IITB. We would also like to thank Mr. Ron Tal and the anonymous reviewers for the valuable comments.

REFERENCES

- [1] E. R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, 1960.
- [2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Boston: Academic Press, 1990.
- [3] D. Comaniciu and V. Ramesh, "Mean shift and optimal prediction for efficient object tracking," in *Proceedings of International Conference on Image Processing*, vol. 3, 2000, pp. 70–73.
- [4] R. G. Bradski, "Computer vision face tracking as a component of a perceptual user interface," in *Workshop on Applications of Computer Vision*, 1998, pp. 214–219.
- [5] Y. Bar-Shalom and E. Tse, "Tracking in a Cluttered Environment with Probability Data Association," *Automatica*, vol. 11, pp. 451–460, 1975.
- [6] Y. Bar-Shalom, "Tracking Methods in a Multitarget Environment," in *IEEE Transactions on Automatic Control*, vol. 23, 1978, pp. 618–626.
- [7] D. Mušicki, R. Evans, and S. Stankovic, "Integrated Probabilistic Data Association (IPDA)," in *Proceedings of the 31st Conference on Decision and Control*, 1992.
- [8] D. Mušicki and R. Evans, "Joint Integrated Probabilistic Data Association - JIPDA," in *Proceedings of the 5th International Conference on Information Fusion*, 2002, pp. 1120–1125.
- [9] E. T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–183, 1983.
- [10] C. K. Chang and Y. Bar-Shalom, "Joint Probabilistic Data Association for Multitarget Tracking with Possibly Unresolved Measurements and Maneuvers," in *IEEE Transactions on Automatic Control*, vol. 29, 1984, pp. 585–594.
- [11] D. Mušicki, R. Evans, and S. Stankovic, "Integrated Probabilistic Data Association," in *IEEE Trans. on Automatic Control*, vol. 39, 1994.
- [12] M. Mählisch, W. Ritter, and K. Dietmayer, "De-cluttering with Integrated Probabilistic Data Association for Multisensor Multitarget ACC Vehicle Tracking," in *Proc. of Intelligent Vehicles Symposium*, 2007.
- [13] M. Mählisch, M. Szcot, O. Löhlein, M. Munz, and K. Dietmayer, "Simultaneous Processing of Multitarget State Measurements and Object Individual Sensory Existence Evidence with the Joint Integrated Probabilistic Data Association Filter," in *Proc. of the 5th International Workshop on Intelligent Transportation (WIT2008)*, 2008.
- [14] P. Kumar, S. Ranganath, K. Sengupta, and H. Weimin, "Cooperative Multitarget Tracking With Efficient Split and Merge Handling," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, 2006.
- [15] A. Genovesio and J.-C. Olivo-Marin, "Split and Merge Data Association Filter for Dense Multi-Target Tracking," in *Proceedings of 17th International Conference on Pattern Recognition*, 2004.
- [16] M. Mählisch, R. Schweiger, W. Ritter, and K. Dietmayer, "Sensorfusion Using Spatio-Temporal Aligned Video and Lidar for Improved Vehicle Detection," in *Proc. of Intelligent Vehicles Symposium*, 2006.
- [17] APROSYS - Advanced PROtection SYStems: An Integrative Project in the 6th Framework Programme, <http://www.aprosys.com>
- [18] J. Tandler, E. Zimmerman, V. Muntean, B. Seipel, T. Koch, D. Willersinn, M. Grinberg, C. Mayer, and M. Diez, "A new pre-crash system for side impact protection," *International Journal of Crashworthiness*, vol. 13, no. 6, pp. 679–692, 2008.
- [19] J. Tandler, D. Willersinn, and M. Grinberg, "Data Analysis of the APROSYS Side Pre-Crash Sensing System," in *Proceedings of Intelligent Vehicles Symposium*, 2007, pp. 926–930.
- [20] P. Woock, F. Pagel, M. Grinberg, and D. Willersinn, "Odometry-Based Structure from Motion," in *Proceedings of Intelligent Vehicles Symposium*, 2007, pp. 1112–1117.
- [21] T. Dang, C. Hoffmann, and C. Stiller, "Fusing Optical Flow and Stereo Disparity for Object Tracking," in *Proceedings of the IEEE V International Conference on Intelligent Transportation Systems*, 2002.
- [22] T. Dang and C. Hoffmann, "Fast Object Hypotheses Generation Using 3D Position and 3D Motion," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [23] U. Franke, C. Rabe, H. Badino, and S. Gehrig, "6D Vision: Fusion of Motion and Stereo for Robust Environment Perception," in *DAGM Symposium 2005*, ser. Pattern Recognition, 2005, pp. 216–223.
- [24] S. Rühl, M. Grinberg, and D. Willersinn, "Empirical evaluation of motion models for a side-looking driver assistance system," in *Proc. of the 5th Intl. Workshop on Intelligent Transportation*, 2008.
- [25] J. Collins and J. Uhlmann, "Efficient gating in data association with multivariate distributed states," in *Transactions on Aerospace and Electronic Systems*, vol. 28, 1992.

Year:
2009

Author(s):
Grinberg, M.; Ohr, F.; Beyerer, J.

Title:
Feature-based probabilistic data association (FBPDA) for visual multi-target detection and tracking under occlusions and split and merge effects

DOI:
10.1109/ITSC.2009.5309694 (<http://dx.doi.org/10.1109/ITSC.2009.5309694>)

© 2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Details:
Institute of Electrical and Electronics Engineers -IEEE-; IEEE Intelligent Transportation Systems Society -ITSS-:
12th International IEEE Conference on Intelligent Transportation Systems, ITSC 2009. Proceedings. Vol.1 : St. Louis, Missouri, USA, 4 - 7 October 2009
Piscataway, NJ: IEEE, 2009
ISBN: 978-1-4244-5520-1
ISBN: 978-1-4244-5519-5
ISBN: 978-1-4244-5518-8
ISBN: 978-1-4244-5521-8
pp.291-298