Abstract—In this contribution we present a concept for improvement of object tracking in applications that suffer from severe detection errors such as incomplete, merged, split, missing and clutter-based detections due to noisy data, sensory and algorithmic restrictions and occlusions. It is based on utilization of low-level information that is gained through tracking dedicated feature points with known relationship to the tracked objects. The proposed Feature-Based Probabilistic Data Association and Tracking Algorithm (FBPDA) can be applied not only in the field of driver assistance systems but also in surveillance applications and further video-based object tracking applications. The main requirement is the possibility to robustly track dedicated feature points in the image (and in 3D space). For this aim, both correlation-based techniques (optic flow) and correspondence-based techniques using e.g. SIFT [1] or SURF [2] features can be used.

Index Terms—environment perception, feature-based, multitarget tracking, track-before-detect, detection-by-tracking, probabilistic data association, point tracking, 6D-Vision, split and merge handling, occlusion handling, stereo video

I. MOTIVATION

In general object tracking can be divided into two tasks: the first one is object (re-)identification and the second one is object state estimation. While dynamic state estimation in case of unambiguous data association is a quite well solved task, data association and object (re-)identification in the case of a-priori unknown number of objects in the scene and possibly missing, incomplete, merged, split or clutter-based detections are still an active research field. Uncertainties in the measurement process, in object state modeling and in measurement-to-track association are the main cause for severe tracking errors. In case of extended targets this is even worse since object observability represents another uncertainty source.

Fig. 1. Data association problem in case of split detections (truck in the background). In case of an object entering the field of view its visible extent grows causing wrong position and velocity estimation (car in the foreground).

Fig. 1 illustrates some reasons and effects of incorrect data association using the example of a car-mounted side-looking stereo camera developed for a side pre-crash detection system. In the frame corresponding to the upper image, clustering of the 3-dimensional point clouds (visualized by means of colored vertical lines to the bottom) delivered three detections that were correctly associated to the visible...
object parts. In the course of the sequence, lacking feature points in the homogeneous area on the truck door and on the occluded rear part hindered remaining points from being clustered together resulting in three independent detections. Only one of these detections (the middle one) could be associated with the track, for the other two detections new tracks had to be initiated. In general, such "splits" (as well as "merges") lead to erroneous estimation of the track position and velocity and to track losses. A similar effect can be expected in the case of the white car in the foreground. In the course of the sequence it is entering the field of view (FoV) and thus "growing" in the image. It is obvious that a centroid-based filtering of such a "growing" track leads to erroneous position and velocity estimation.

II. FBPDA - MAIN IDEAS

For coping with the above-mentioned effects we developed the Feature-Based Probabilistic Data Association and Tracking Algorithm (FBPDA). Its main ideas are:

- World model with a mechanism for propagation of uncertainties about object existence, dynamics, and observability.
- Detailed object extent representation for refined occlusion modeling and propagation.
- Detection-by-tracking instead of tracking-by-detection paradigm. Measurements are derived from detections utilizing low-level information gained by tracking dedicated feature points.
- Affiliation of the tracked feature points to a certain object is modeled by means of affiliation probability which is updated each frame.
- Different possible track configurations are weighted according to the probabilities of the events leading to those configurations.

In the following sections detailed description of these ideas is given.

III. MODELING OF UNCERTAINITIES

FBPDA is based on the following assumptions: Since the number of objects in the scene is a-priori unknown and the sensor system may sometimes get split, merged, missing or clutter-based detections, the existence of detected targets may not be assumed as given but should be modeled as a probability mass function (pmf) or as a degree of belief. This existence probability should be taken into account when resolving association problem. On the other hand, a possibility that an object is occluded and therefore can not be detected should also be taken into consideration when performing data association. Thus it also should be modeled by means of a pmf. Since dynamics and existence of occluding objects are also subjects to uncertainty, the three uncertainties (existence, observability and dynamic state) have to be propagated in mutual dependency. This can be done by means of three cross-coupled Markov chains as shown in Fig. 2. Details on their derivation can be found in [3] and [4].

![Fig. 2. Three cross-coupled Markov chains for propagation of object state and uncertainties related to the existence (denoted as $\exists$), dynamics and observability (denoted as $\supset$). The middle Markov chain represents propagation of the estimated object dynamic state $\hat{x}$ and its covariance $\hat{P}$, the upper Markov chain – propagation of the existence probability $p^x(\exists)$ and the lower Markov chain – propagation of the observability probability $p^x(\supset)$.](image)

The a-priori spatial probability density function (pdf) for birth, death and persistence of a track as well as the pdf for missing or clutter-based detections can be derived from detector performance metrics as shown in [5].

IV. DETAILED OCCLUSION MODELING

For correct interpretation of incomplete detections (e.g. in case of partial occlusions) it is not sufficient to estimate the overall object observability probability. Hence, object visibility modeling has to be refined. In order to model effects of occlusion, object extent in space and in the image has to be defined in detail. This can be done in different ways. The simplest and at the same time very generic method for representing object extent is a 3D occupancy grid as proposed in [3]. A top view of such a grid is shown in Fig. 3 (a). For each grid cell its occupancy degree and thus its potential for causing an occlusion can be estimated and propagated from frame to frame. Projecting occupied cells into the image delivers a coarse image-based representation of the tracked objects. Under incorporation of the corresponding existence probabilities, this allows for building of an occlusion probability map for each tracked object. An example of such
occlusion probability map is shown in Fig. 3 (b). After having estimated possible occlusions, occupancy degree of occluded cells may be filtered in order to prevent "shrinking" of partially occluded objects. Details on this process can be found in [3] and [4].

The main advantage of the grid-based object representation is its simple algorithmic description and scalability. But it also contains several drawbacks. One of them is the coarse quantization of the object volume and the necessity of the occupancy-based occlusion estimation. The second drawback is the fact that the information about coherent image regions and their relation to the tracked objects gets lost. This can be avoided with a more customized object representation. One possibility for doing that is a model-based object representation, which requires a model-based object detection and in many cases is not possible. Another possibility is representation in terms of linked image regions that are segmented according to their intensity, depth and motion as shown in Fig. 12(b). Such regions can be re-identified and tracked not only by means of optic flow but also by means of robust descriptors. This might help to overcome difficulties emerging in case of failing optic-flow-based feature point tracking (e.g. in case of temporary occlusions).

V. SOLVING ASSOCIATION AMBIGUITIES

As mentioned in Section I, one of the major problems of the multi-target tracking algorithms are data association ambiguities in case of split, merged and incomplete detections which results in erroneous object state estimation. Contrary to other approaches such as PDA [6], JPDA [7] and JIPDA [8], we propose solving these ambiguities by means of a novel feature-based data association and tracking scheme which is based on the detection-by-tracking principle. The key point for solving the tracking problem is the feature-based track state innovation which is visualized in Fig. 5. Instead of "blind" update of the track state by using associated detections as shown in Fig. 4, we first reconstruct "correct" measurements utilizing for this aim information gained from the tracking of dedicated feature points. Details on this procedure are given in Section VI.

![Fig. 3. Grid-based object representation with resulting occlusion map for the object in the background (green car)](image)

![Fig. 4. Wrong data association in a classical tracking-by-detection approach in case of a clutter-based cluster merge.](image)

![Fig. 5. Basic idea of the FBPDA: utilization of the information about point-to-track affiliation for "reconstruction" of appropriate measurements for the tracking (detection-by-tracking).](image)
Fig. 6. Object position reconstruction based on spatial relationship of the track reference point $r_O$ to the stably tracked points.

Fig. 6. The center of gravity (CoG) $p_{CoG}$ of the object points that could be tracked from the previous into the current frame is calculated in both frames ($k-1$ and $k$). Knowing its geometrical relationship to the object reference point $r_O$ in the frame $k-1$ (old track centroid), this point can be reconstructed in the new frame ($r_{O_{old}}$). Innovation of object dynamics is thus done with this, correct reference point and not with the new centroid $r_{O_{new}}$. Only after having performed the innovation, we may set $r_{O_{new}}$ as the new reference point.

It’s worth to mention that tracked feature points contribute to the estimation according to their estimated affiliation to the corresponding track.

The same procedure is used for the object extent as shown in Fig. 7. Since an object might not always evoke a measurement (e.g. in case of occlusions or sensor failures), its predicted parameters are also incorporated (weighted with corresponding probabilities).

VII. TRACK MANAGEMENT

Track management models events of the tracks’ birth, persistence and death and makes use of the global evidence for these events. The corresponding probabilities are derived from the data association step as described in [3], [4] and serve as a basis for initialization, confirmation and termination of tracks. For each measurement among others a hypothesis is considered for being originated from a new not yet known track. The birth probability takes into account the probability of the measurement being evoked by neither clutter nor other already tracked objects. In the following frames this hypothesis should be confirmed in order to establish a new track. In opposite case it will be rejected. Similarly, for each track the track death event is considered. Its probability accounts for the case that the track gets associated no measurements not because of occlusion or sensor malfunction but because the corresponding object simply does not exist (this also incorporates the case of objects that have left the field of view of the sensors and are not of interest for the tracking system any more). In general, this track management scheme implements a Track-before-Detect scheme well known in the radar community [9]. Fig. 8 gives an example for life cycles of an object-based and a clutter-based tracks.

Fig. 8. Probability-based track initiation, confirmation and termination. Track $x_1$ is caused by a real target. It is confirmed soon after appearing in the field of view of the sensors and is terminated after leaving it completely. Track $x_2$ is caused by clutter and thus is terminated before being confirmed.

VIII. PROTOTYPICAL IMPLEMENTATION

The prototypical implementation of the proposed algorithm was done as an enhancement to the video-based object tracking system built at Fraunhofer IOSB in the course of the EU funded project APROSYS [10]. The goal there was detection of
imminent side collisions to enable timely activation of novel occupant protection systems [11]. It has been realized by means of a stereo camera system and a radar network. The sensor system under consideration here is a side-looking stereo video camera [12]. For the realization of a robust and reliable object detection, a generic approach had to be chosen. It should be able to detect different object classes with arbitrary position and orientation relative to the ego-vehicle. In contrast to many forward looking driver assistance systems applications, no a-priori information about symmetry, shadows, position relative to lanes and motion direction could be assumed. Thus, object hypotheses were generated merely from the range and motion data. Fig. 9 illustrates the principal architecture of the implemented video-based object detection and tracking framework.

Similar to the 6D-Vision system proposed by Franke et al. [13], up to 3000 feature points are tracked simultaneously in 3-D space using Kalman Filters. Their six-dimensional state vectors are estimated from the stereo depth measurements as well as from their displacement in the image between consecutive frames (optic flow). After estimation of the ground plane and elimination of the ground points, remaining point clouds are clustered and give detections for the subsequent object tracking. Object parameters are estimated using an Extended Kalman Filter. Internally, objects are modeled as cuboids with a centroid \((x, y, z)\), dimensions \((l, w, h)\), geometrical orientation \(\phi\), motion orientation \(\varphi\), geometrical orientation \(\varphi\), 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 [14]. 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. Ego-motion estimation and compensation are done using vehicle odometry data that has been shown to deliver sufficient accuracy even for the structure-from-motion task [15]. For handling of occlusions we have chosen a grid-based object representation with grid cell size of 20cm. Object’s geometric orientation and dimensions are updated taking into account the occlusion information as described in Section VI.

IX. EXPERIMENTAL RESULTS

The algorithm has been tested with both simulated and real data. Hereby we concentrated on scenarios that caused problems for the classical approach such as splitting and merging point clouds, objects entering and leaving the FoV and occlusion scenarios.

(a)  
(b)

Fig. 10. Frame from the test sequence: (a): original image. (b): depth map

Fig. 11 shows a comparison of our results with results achieved with the classical approach in a sequence where a pedestrian causes an occlusion of a car in the background leading to splits and merges of corresponding point clouds. The corresponding input video and depth images are shown in Fig. 10. While the classical approach loses the original track of the car and has to establish two new tracks instead (that in turn are lost as soon as the point clouds merge again) our algorithm manages to correctly update the car track throughout the sequence despite multiple splits, merges and occlusion. The visualization of the visible grid cells of both tracked objects is shown in Fig. 12 (a).

X. CONCLUSION AND FUTURE WORK

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 approach resolves data association ambiguities in a soft decision based not only on target state prediction but also on the existence and observability estimation modeled as two additional Markov Chains.
Along with the occlusion analysis, spatial and temporal relationship between the set of stably tracked points and the object’s reference point is exploited, allowing for the reconstruction of the desired object characteristics from the data even in case of detection errors due to limited FoV, occlusions, noise and sensor malfunction. For tracking applications that have to cope with the above-mentioned effects and allow tracking on the feature level, our algorithm offers a much-needed enhancement which has the potential to greatly increase detection and tracking performance and overall system robustness.

For a refined object extent modeling and occlusion handling we proposed to use a grid-based object representation. Our current work concentrates on the alternative object extent representation based on its appearance in the image. Introduction of an intermediate tracking level (segment tracking) which makes use of both image segmentation and feature tracking offers another powerful tool for improvement of the overall tracking performance.

REFERENCES