Improved data association using buffered pose adjustment for map-aided localization

Anthony Welte*, Philippe Xu*, Philippe Bonnifait*, Clément Zinoune†

Abstract— Maps provide an important source of information for autonomous vehicles. They can be used along with cameras and lidars to localize the vehicle. This requires the ability to correctly associate observations to features referenced in the map. The problem is all the more difficult than all observations are not necessarily referenced, and all map features might not be detectable with the embedded sensors.

This paper presents an adjustment technique than enables to increase the number of associations that can be made while limiting the chance of obtaining wrong associations. This is achieved by matching observations in batches in a buffer and matching them regularly. Periodically, the observation buffer is used to adjust the trajectory used to match observations. This is done without making any assumption on the association between observations and map features through a likelihood maximization process. The adjusted trajectory then provides the best associations that are used for real-time localization.

The method was tested with data recorded on public roads using an experimental vehicle. The results show that thanks to the trajectory adjustment step and the use of an observation buffer, the number of associations that can be made is increased. This also results in greater localization accuracy and consistency.

I. INTRODUCTION

Autonomous vehicle navigation highly depends on an accurate and reliable localization with respect to a representation of the driving environment which is stored in a map. With High-Definition (HD) maps which contain georeferenced features [12], [20] the localization system accuracy can be improved significantly [19], [10], [6]. Building such maps can be expensive but we believe that, in the near future, HD maps of the road network can be deployed on a large scale. Accurate mapping being itself a challenging task, the resulting map often contains errors or is incomplete [16] due to constant evolution of road networks. Such faulty data must be carefully taken into account.

Improvement of localization using accurate maps can only be done using sensors able to observe referenced landmarks. Smart cameras that measure distances to road markings and road edges (ground features) are being equipped on commercial vehicles for Advanced Driver-Assistance Systems (ADAS) such as lane keeping assistant systems. These sensors have already been used for localization and are particularly useful to improve the across track localization accuracy [19].

Another promising type of sensors is lidar. As lidars measure distances, they are especially interesting for localization when combined with maps. Lidars have not only been used

with dense maps [1], [22], [13] but also with sparse feature maps [5], [7]. From lidars, particular landmarks can be detected not only using the Cartesian information, but also thanks to the intensity data returned by most lidars. Indeed some features such as road signs are easily identified using lidar intensities as these features have been designed to be seen at night, thus return most of the laser light.

Using detected features for localization is not straightforward. Indeed, detections need to be matched to their corresponding feature in the map. This problem is especially important as a wrong association can have a significant effect on the localization solution which cannot be acceptable for safety-critical applications like autonomous driving.

Localization and mapping are often studied in a SLAM context for navigation in unknown environments. In this context, finding loop closures is essential to build the map. For autonomous vehicle navigation these are rare. The problem of finding correct association is nonetheless similar with SLAM. This work focuses on the matching problem of ground features and road sign measurements with georeferenced features in HD maps provided by a third party. The main contributions of the paper are the following. i) Our matching approach uses a smoothing step to improve localization estimates on a local window. ii) An optimal rigid adjustment of the local trajectory is then computed. Since the dimension of the adjustment parameter is small, it enables a real-time and tractable computation of a multi-hypothesis adjustment without approximation. This adjustment increases the number of good associations, leading to an improved localization. iii) We explain how a filtering scheme is implemented in parallel to provide to the vehicle a high-frequency pose information. iv) An experimental evaluation using real data is presented. This shows significant improvements compared to classical snapshot data association methods.

The next section reviews related work in localization with perception sensors and maps. Section III introduces the estimation scheme based on Kalman filtering and the observations considered in this paper. Section IV details the association process and section V describes the adjustment step. Finally, section VI presents experimental results obtained from data recorded on public roads using an experimental vehicle.

II. RELATED WORK

The matching problem comes up in many fields of robotics. Image based detectors are typically able to extract descriptors of the detected features. These descriptors are compared to those stored in the map, making the matching problem easier.

To alleviate the effect of bad associations, the localization problem is solved using robust estimation techniques such as RANSAC [11]. This assumes that the best association needs to maximize the number of matches which is not necessarily the case when the map is missing features.

Descriptors can also be found for point clouds. The authors in [3], [4] exploit both point cloud descriptors and the relative distances between mapped and observed features to provide a matching solution. They build a graph where each connection represents two coherent (in terms of relative distance) associations. The associations forming the biggest clique are then used for localization.

Matching is also a problem studied in SLAM applications. Although consecutive observations can be matched fairly easily since the map is built based on those same observations, SLAM approaches still need to find accurate matches for loop closure tasks. In [8], the authors build upon the iSAM2 framework and introduces multi-hypothesis factors to account for matching ambiguities. This method provides multiple estimates of the variable depending on the selected hypothesis. iSAM2 is not suitable for applications where real-time estimation is critical and MH-iSAM2 significantly increases the computational cost making such method unsuited for autonomous vehicles.

Variations of iSAM have been proposed to enable real-time estimation of the state. Kaess et al. [9] separate the problem in two: a filtering part operating in real-time and a smoothing part the periodically integrate the information from the smoother. This framework assumes, however, that the factors introduced in the graph are accurate, thus that any ambiguity has already been dealt with in the front-end.

Another way used to account for ambiguous association is the use in the factor graph of non-Gaussian noise models. The authors in [15], [2] proposes to approximate the multi-modal Gaussian by whichever mode is most likely at the estimation point. This approach is dangerous for real-time applications as, although the dominant mode might improve during the iterative estimation process, the first choice will influence the most recent estimate significantly. This method is not suited for applications where the real-time estimate is the most important and the rest of the trajectory is somewhat irrelevant. Others have also proposed to use non-Gaussian models [18].

Although approaches used in SLAM have shown great promise, they usually do not enable high frequency state estimation which is critical for autonomous driving applications. Moreover, approaches that attempt to solve this issue [9] expect input measurements to be unambiguous. Because in SLAM applications the map is built incrementally using the same detection system used for localization, the ambiguous association problem typically occurs rarely at loop closure. Using highly detailed third party maps results in ambiguities at almost every observation. For these reasons, this work uses Kalman filtering in order to guarantee real-time estimation and relies on a trajectory adjustment before selecting matches. This step is done without approximating the multi-modal noise model and is made tractable by only estimated a 2D rigid transformation of the trajectory instead of solving the complete state space as would be done using a SAM framework.

III. LOCALIZATION FILTER

A. Filtering Scheme

Localization is performed using extended Kalman filtering to estimate the vehicle state $\mathbf{x}_k = \begin{bmatrix} x_k, y_k, \theta_k, v_k, \omega_k \end{bmatrix}^T$ composed of the vehicle two dimensional pose, *i.e.*, position and heading, along with its longitudinal speed and yaw rate. The filter combines three kinds of measurements: Dead-Reckoning (DR) measurements which provide the relative motion of the vehicle from one time sample to the next, GNSS measurements giving coarse estimates of the position and vehicle heading, and perception measurements from a camera and a 3D lidar that are used to localize the vehicle within a map through Kalman's updates. The filtering is derived classically as follows:

$$
\begin{array}{rrcl}
\text{Prediction:} & \widehat{\mathbf{x}}_{k|k-1} & = & f_k\left(\widehat{\mathbf{x}}_{k-1|k-1}\right) \\
& \boldsymbol{P}_{k|k-1} & = & \boldsymbol{F}_k \boldsymbol{P}_{k-1|k-1} \boldsymbol{F}_k^T + \boldsymbol{Q}_k \\
\text{Update:} & \tilde{\mathbf{y}}_k & = & \mathbf{z}_k - h_k\left(\widehat{\mathbf{x}}_{k|k-1}\right) \\
& \boldsymbol{S}_k & = & \boldsymbol{H}_k \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^T + \boldsymbol{R}_k \\
& \boldsymbol{K}_k & = & \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^T \boldsymbol{S}_k^{-1} \\
& \widehat{\mathbf{x}}_{k|k} & = & \widehat{\mathbf{x}}_{k|k-1} + \boldsymbol{K}_k \tilde{\mathbf{y}}_k \\
& \boldsymbol{P}_{k|k} & = & \left(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k\right) \boldsymbol{P}_{k|k-1}
\end{array}
$$

where $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$ are the Jacobian matrices of respectively the evolution model f_k and the observation model h_k.

All the measurements are used as observations. Since GNSS data fusion with DR is well studied in the literature, it is not detailed in this paper. The measurements used for the DR estimation of the vehicle are a yaw rate gyro, 4 wheel encoders and the steering wheel angle. The observation models used for these sensors follows Ackerman geometry and are described in-depth in [21].

Perception measurements from cameras or lidars provide localization information relatively to the sensor reference frame. To be used for absolute localization, these measurements have to be matched to a map encoding absolute position. In this work, we suppose to have a HD map containing road markings, represented as polylines, *i.e.* sequences of line segments, and road signs, represented by points localized at their geometric center. An intelligent camera is used for road markings detection and a lidar for road sign detection. Through the rest of the paper, the HD map $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots\}$ is represented as a set of map features $\mathbf{m}_i$ which can either be a line segment (road marking) or a point (road sign).

B. Observation Models for the Camera and the Lidar

1) Road Markings: ADAS use off-the-shelf smart cameras that are able to return lateral distances to neighboring road markings. These observations are valuable for localization as they provide a lateral constraint for state estimation.

Let $\mathbf{m}_i = \left[\left(x_i^A, y_i^A\right), \left(x_i^B, y_i^B\right)\right]$ be a map feature representing a line segment $[AB]$, where $\left(x_i^A, y_i^A\right)$ and $\left(x_i^B, y_i^B\right)$ are the geo-referenced coordinates of points A and B. Let $\mathbf{x}_k$ be the vehicle state at time k. The road marking observation
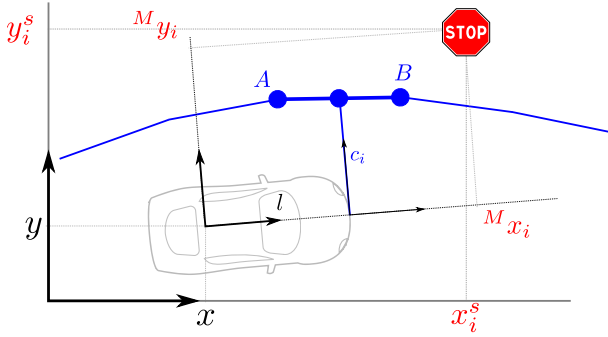
Fig. 1: Schema of the vehicle with a road sign observation (red) and a ground feature observation (blue). The lateral distance c_i from the front bumper to the ground feature is measured as well as the 2D position $\begin{bmatrix} ^M x_i & ^M y_i \end{bmatrix}^T$ of the road sign in the mobile frame.

model h_i^m of the camera measurement c_i is given by [19] :

$$c_i = h_i^m\left(\mathbf{x}_k\right) \tag{1}$$
$$= \frac{\left(l\sin\theta_k + y_k - y_i^A\right)x_i^{AB} - \left(l\cos\theta_k + x_k - x_i^A\right)y_i^{AB}}{x_i^{AB}\cos\theta_k + y_i^{AB}\sin\theta_k}$$

where $x_i^{AB} = x_i^B - x_i^A$, $y_i^{AB} = y_i^B - y_i^A$ and l is the longitudinal distance between the camera and vehicle frame as pictured in Figure 1.

2) Road Signs: Lidars are especially well suited for road sign detection as they are made to be reflective for visibility purpose. This results in a high returned intensity. A detector can therefore be built by thresholding the lidar point cloud to keep only intensities over a threshold $I_{\min}$. The resulting points can be grouped into clusters by Euclidean clustering. The centroid $\left(^M x_i, {}^M y_i\right)$ of each cluster can then be found by fitting a bounding box to the cluster using principal component analysis (PCA).

Now, let $\mathbf{m}_i = \left(x_i^S, y_i^S\right)$ be a map feature representing a geo-referenced road sign. The road signs observation model h_i^s of the lidar measurement is defined as

$$\begin{bmatrix} ^M x_i \\ ^M y_i \end{bmatrix} = h_i^s(\mathbf{x}_k) = \begin{bmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{bmatrix} \begin{bmatrix} x_i^S - x_k \\ y_i^S - y_k \end{bmatrix} \tag{2}$$

IV. DATA ASSOCIATION PROCESS

Associating the detected objects to their corresponding map features is a critical step. Map features can improve localization accuracy significantly but require to be correctly associated. The effects of a bad association are twofold. The estimation error increases while the uncertainty decreases because of the new measurement. Then, because the state has been corrected using bad observation, subsequent observations can be wrongly matched, worsening the problem.

At a given time k, let $\mathbf{Z}^{(k)} = \left\{\mathbf{z}_1^{(k)}, \mathbf{z}_2^{(k)}, \ldots\right\}$ be the set of observed features, which can be a mix of road markings and signs. The goal of the data association process is to find for each observation $\mathbf{z}_j^{(k)}$ its corresponding map feature $\mathbf{m}_i \in \mathbf{M}$.

A. Unique Nearest Neighbor

Nearest neighbor association is widely used to associate high level observations to map features. In this paper, the
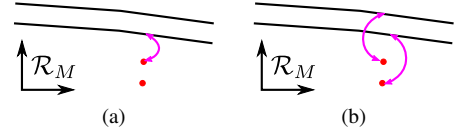

Fig. 2: Difference between a UNN (a) and Hungarian (b) association methods. The UNN method matches an observation to its closest feature and does not match if the feature is already taken, thus the top observation (red) is associated to the bottom lane marking (black). The Hungarian method, associates every observation in order to minimize the sum of association distances. The two observations are thus associated to their correct lane marking.

nearest neighbor association is performed using Mahalanobis distances with the different map features. This is done to account for the uncertainty propagation of the predicted measurements. It is also useful for rejecting unlikely associations.

Hence, an observation $\mathbf{z}_j^{(k)}$ is associated with the map feature $\mathbf{m}_i$ such that

$$\mathbf{m}_i = \operatorname*{arg\,min}_{\mathbf{m}_i \in \mathbf{M}} \left(\sqrt{\mathbf{y}_{i,j}^T \boldsymbol{S}_{i,j}^{-1} \mathbf{y}_{i,j}}\right) \tag{3}$$

where $\mathbf{y}_{i,j} = \mathbf{z}_j^{(k)} - h_i\left(\hat{\mathbf{x}}_k\right)$, $\hat{\mathbf{x}}_k$ is an estimate of the vehicle state and $\boldsymbol{S}_{i,j} = \boldsymbol{H}_i \boldsymbol{P}_k \boldsymbol{H}_i^T + \boldsymbol{R}_j$ with $\boldsymbol{H}_i$ the Jacobian of h_i, $\boldsymbol{P}_k$ and $\boldsymbol{R}_j$ the covariance matrices of respectively the state estimate and the measurement.

This method allows multiple observations to be associated with the same feature. To prevent this, when multiple observations are associated with the same feature, only the observation with the lowest Mahalanobis distance is kept. We refer to this method as Unique Nearest Neighbor (UNN).

B. Hungarian Association (Kuhn-Munkres)

UNN can either associate an observation with its closest feature, or not associate the observation if that feature is closer to another observation. Munkres algorithm [14] performs a global matching considering all observation at time k. Like the UNN method, it only allows features to be associated with a single observation. However, it does not necessarily associate an observation with its closest features (Figure 2). The method finds matches that minimize the summed Mahalanobis distances of all matches. Hence, at time k for an observation set $\mathbf{Z}^{(k)}$ of size n, it will find the subset $\mathbf{M}_k \subseteq \mathbf{M}$ containing n features so that

$$\mathbf{M}_k = \operatorname*{arg\,min}_{\mathbf{M}_k \subseteq \mathbf{M}} \left(\sum_{\mathbf{m}_i \in \mathbf{M}_k} \sqrt{\mathbf{y}_{i,j}^T \boldsymbol{S}_{i,j}^{-1} \mathbf{y}_{i,j}}\right) \tag{4}$$

where $\mathbf{m}_i$ is the feature associated to the observation $\mathbf{z}_j^{(k)}$.

C. Association Rejection

To avoid potential wrong matches, the maximum Mahalanobis distance allowed for an association to be valid is limited. Observations that are too far from their associated feature are discarded. This step is done through a rejection step that only allows association with a small enough distance.
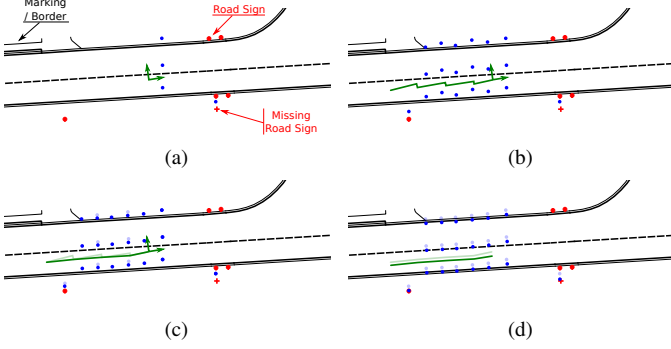
Fig. 3: Steps of the matching algorithm. Blue: measurements, Green: pose estimates, Red: road signs (dots are mapped, the cross is not), Black: markings. (a) The measurements available at the current time. (b) Buffered observations and estimates. (c) Smoothed trajectory. (d) Adjusted trajectory used for the final matching.

For an observation $\mathbf{z}_j^{(k)}$ associated with a map feature $\mathbf{m}_i$, the observation is considered valid if

$$\mathbf{y}_{i,j}^T \boldsymbol{S}_{i,j}^{-1} \mathbf{y}_{i,j} < F_{\chi^2}^{-1}(1-\alpha; d) \qquad (5)$$

where F_{χ^2} is the cumulative distribution function of a χ^2 law with a degree of freedom $d = 1$ for road markings and $d = 2$ for road signs. The value $\alpha \in [0, 1]$ quantifies the probability of rejecting the association (i, j) when it is actually valid.

V. BUFFERED POSE ADJUSTMENT FOR AN IMPROVED ASSOCIATION

When using in real-time high level features, such as road markings or signs, the observation set $\mathbf{Z}^{(k)}$ at time k is of limited size. Having few observations can lead to wrong data association as many ambiguities can arise due to a large pose uncertainty. Moreover, if the perception modules are asynchronous, the markings from the camera and the signs from the lidar are likely to arrive at different times.

The final objective is to use the exteroceptive measures with the best association in the update of the filter. Consider for instance the real situation of Figure 3 (a) where a missing road sign is detected but unfortunately there are other signs nearby. In order to limit association ambiguities, the observations are not matched in a snapshot manner. Buffers containing the observations are used with filtered state estimates (see Figure 3 (b)). Hence, the matching is performed using more observations thus limiting the risk of ambiguous matching. The observation buffer provides a larger number of measurements and allows a globally consistent matching, which is known to be more robust.

Thanks to this approach another problem can be addressed. Estimation errors which are time-correlated can cause incorrect matchings. A Kalman smoothing is first applied on the buffer to correct the jaggedness of the filtered estimates (see Figure 3 (c)). Then, an adjustment of the whole trajectory over the buffer is made to compensate as much as possible the filter errors to find the best data association. Figure 3 (d) shows the final result. It can be noticed that the faulty road sign measurement has been sufficiently shifted away such that a wrong matching will be rejected in the last filter update.

The observation buffer is processed at a low frequency (e.g. every $250ms$) since it cannot always be treated within the filter update period. Once the association of all the exteroceptive measurements is done, the new matches are provided to the Kalman filter estimator. The localization filter is run again starting from the first matched observation to provide an improved real-time localization. This is explained more in-depth in Section V-C.

Let K be the most recent time of the observation buffer, with which the matching is solved. Let $\mathbf{Z}_K = \left\{ \mathbf{Z}^{(k)} | k \in [\![K - S, K]\!] \right\}$ be the set of all the observations over a buffer of size S. The goal is to associate all the elements of $\mathbf{Z}_K$ simultaneously. The following sections detail the steps used to achieve this. The vehicle trajectory is first smoothed. It is then adjusted using the observations. The adjusted trajectory is finally used to match observations to map features.

A. Pose Smoothing

Instead of using directly the filtered state estimates $\widehat{\mathbf{x}}_{k|k}$, with $k \in [\![K - S, K]\!]$, we propose to first compute better state estimates over the buffer using a Kalman smoother [17]. To do so, a backward pass is performed as follows:

$$\widehat{\mathbf{x}}_{k|K} = \widehat{\mathbf{x}}_{k|k} + \boldsymbol{J}_k \left(\widehat{\mathbf{x}}_{k+1|K} - \widehat{\mathbf{x}}_{k+1|k} \right) \qquad (6)$$

$$\boldsymbol{P}_{k|K} = \boldsymbol{P}_{k|k} + \boldsymbol{J}_k \left(\boldsymbol{P}_{k+1|K} - \boldsymbol{P}_{k+1|k} \right) \boldsymbol{J}_k^T \qquad (7)$$

where

$$\boldsymbol{J}_k = \boldsymbol{P}_{k|k} \boldsymbol{F}_{k+1}^T \boldsymbol{P}_{k+1|k}^{-1} \qquad (8)$$

Thus, from the filtered estimates $\left\{ \widehat{\mathbf{x}}_{k|k} | k \in [\![K - S, K]\!] \right\}$, smoothed state estimates $\widehat{\mathbf{X}}_K = \left\{ \widehat{\mathbf{x}}_{k|K} | k \in [\![K - S, K]\!] \right\}$ are obtained. These are more accurate with less uncertainty which will be very useful for the data association process.

B. Pose Adjustment

An additional way to improve the state estimation is to maximize the likelihood of all the states in the buffer given all the observations:

$$L \left(\mathbf{x}_{K-S}, \dots, \mathbf{x}_K \left| \mathbf{Z}^{(K-S)}, \dots, \mathbf{Z}^{(K)} \right. \right) \qquad (9)$$

Maximizing this likelihood function in equation (9) for all the poses is the buffer is computationally. Moreover, the temporal coherence between the successive state estimates coming from the evolution model could be degraded. Instead, we propose to estimate a unique 2D local rigid transformation, i.e., a translation and a rotation, $\boldsymbol{\delta} = \left[\delta_x, \delta_y, \delta_\theta \right]$ to all the state estimates $\widehat{\mathbf{x}}_{k|K}$ so that the likelihood of the resulting states is maximized. Let $\widetilde{\mathbf{X}}_K(\boldsymbol{\delta}) = \{ \widetilde{\mathbf{x}}_k | k \in [\![K - S, K]\!] \}$ be the δ-adjusted state estimates defined as follows:

$$\widetilde{\mathbf{x}}_k = \begin{bmatrix} \widetilde{x}_k \\ \widetilde{y}_k \\ \widetilde{\theta}_k \end{bmatrix} = \begin{bmatrix} \cos\delta_\theta & -\sin\delta_\theta & 0 \\ \sin\delta_\theta & \cos\delta_\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{x}_{k|K} \\ \widehat{y}_{k|K} \\ \widehat{\theta}_{k|K} \end{bmatrix} + \begin{bmatrix} \delta_x \\ \delta_y \\ \delta_\theta \end{bmatrix} \quad (10)$$

A new likelihood function is defined as

$$L \left(\boldsymbol{\delta} | \mathbf{Z}_K \right) = L \left(\widetilde{\mathbf{x}}_{K-S}, \dots, \widetilde{\mathbf{x}}_K \left| \mathbf{Z}^{(K-S)}, \dots, \mathbf{Z}^{(K)} \right. \right) \qquad (11)$$

which can be rewritten as

$$
\begin{aligned}
L\left(\boldsymbol{\delta}\,|\mathbf{Z}_K\right) &= \prod_{k\in[\![K-S,K]\!]} L\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{Z}^{(k)}\right) \qquad (12)\\
&= \prod_{k\in[\![K-S,K]\!]}\prod_{\mathbf{z}_j^{(k)}\in\mathbf{Z}^{(k)}} L\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right)
\end{aligned}
$$

The observation $\mathbf{z}_j^{(k)}$ is related to the state estimate $\widetilde{\mathbf{x}}_k$ via the map features $\mathbf{M}=\{\mathbf{m}_1,\mathbf{m}_2,\ldots\}$. Suppose that $\mathbf{z}_j^{(k)}$ is one of the observations at time k associated to the map feature $\mathbf{m}_i$, then the corresponding likelihood is expressed as follows with a Gaussian assumption:

$$
L_{i,j}\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right) = \frac{1}{\sqrt{(2\pi)^d|\widetilde{\boldsymbol{S}}_{i,j}|}} \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{y}}_{i,j}^T\widetilde{\boldsymbol{S}}_{i,j}^{-1}\widetilde{\boldsymbol{y}}_{i,j}\right) \quad (13)
$$

with

$$
\widetilde{\boldsymbol{y}}_{i,j}=\mathbf{z}_j^{(k)}-h_i\left(\widetilde{\mathbf{x}}_k\right)\ ,\quad \widetilde{\boldsymbol{S}}_{i,j}=\boldsymbol{H}_i\boldsymbol{P}_{k|K}\boldsymbol{H}_i^T+\boldsymbol{R}_j \quad (14)
$$

and $d=1$ for road markings and $d=2$ for road signs.

Because the association between $\mathbf{z}_j^{(k)}$ and $\mathbf{m}_i$ is not known at this stage, all the map features $\mathbf{M}$ in the vicinity have to be considered. From the law of total probability, we have

$$
L\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right) = \sum_{\mathbf{m}_i\in\mathbf{M}} L_{i,j}\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right)\mathbb{P}_{i,j} \qquad (15)
$$

where $\mathbb{P}_{i,j}$ is the probability that $\mathbf{z}_j^{(k)}$ corresponds to $\mathbf{m}_i$. Without any additional information such as appearance or shape cues on the detections or on the features, there is no reason to favor one association over another. Therefore, a discrete uniform distribution over $\mathbb{P}_{i,j}$ is chosen.

The likelihood function in equation (15) does not take into consideration the fact that the sensors may detect features that either do not exist or have not been mapped. To account for this, we propose to add a non-association probability $\mathbb{P}_{\emptyset,j}$ and a likelihood $L_{\emptyset,j}$ which leads to

$$
L\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right) = \sum_{\mathbf{m}_i\in\mathbf{M}} L_{i,j}\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right)\mathbb{P}_{i,j}+L_{\emptyset,j}\mathbb{P}_{\emptyset,j} \quad (16)
$$

In practice, maximizing the likelihood in equation (12) may lead to a large adjustment $\boldsymbol{\delta}$, especially when there are some spatial invariance, *i.e.*, along a straight lane. In order to constrain the spacial adjustment, we propose to reformulate the maximum likelihood estimation into a maximum *a posteriori* (MAP) one. For that purpose, we set the *a priori* distribution over $\boldsymbol{\delta}$ as $\boldsymbol{\delta}\sim\mathcal{N}\left(0;\boldsymbol{P}_{K|K}\right)$. We chose arbitrarily the most recent covariance matrix estimate $\boldsymbol{P}_{K|K}$ from the buffer as a representative measure of the pose uncertainty. The resulting MAP estimation is then defined as

$$
\begin{aligned}
\widehat{\boldsymbol{\delta}} &= \underset{\boldsymbol{\delta}}{\arg\max}\left(L\left(\boldsymbol{\delta}\,|\mathbf{Z}_K\right)\mathbb{P}\left(\boldsymbol{\delta}\right)\right) \qquad (17)\\
&= \underset{\boldsymbol{\delta}}{\arg\max}\left(L\left(\boldsymbol{\delta}\,|\mathbf{Z}_K\right)\exp\left(-\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{P}_{K|K}^{-1}\boldsymbol{\delta}\right)\right)
\end{aligned}
$$

In practice, it is more convenient to minimize the negative log likelihood

$$
\begin{aligned}
\widehat{\boldsymbol{\delta}} &= \underset{\boldsymbol{\delta}}{\arg\min}\left(-\log L\left(\boldsymbol{\delta}\,|\mathbf{Z}_K\right)-\log\mathbb{P}\left(\boldsymbol{\delta}\right)\right) \qquad (18)\\
&= \underset{\boldsymbol{\delta}}{\arg\min}\left(\sum_{k\in[\![K-S,K]\!]}\sum_{\mathbf{z}_j^{(k)}\in\mathbf{Z}^{(k)}}-\log L\left(\widetilde{\mathbf{x}}_k\,\Big|\mathbf{z}_j^{(k)}\right)\right.\\
&\left.\qquad+\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{P}_{K|K}^{-1}\boldsymbol{\delta}\right)
\end{aligned}
$$

This minimization problem is solved using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which is an approximation of the Newton method. The algorithm typically converges in fewer than 10 iterations but a maximum number of iterations is also set such that the matching step always ends in time to provide matches.

C. Matching and Integration in the Filter

The adjusted trajectory is an improvement over the filtered trajectory. However, as quantifying the resulting uncertainty attached to it is not easy, it is only used to improve the matching of features. This matching is then used in the filter.

Once the δ-adjusted trajectory has been found, it can be used to associate observations by applying methods presented in Section IV. The newly associated observations can now be used in the filtering scheme.

Because the matching cannot always be performed within the filter update period, it is instead run at a higher period Δt_m. Therefore, when using the buffer pose adjustment method, road sign and road marking observation are not used directly to estimate the state. The state is estimated without these observations for some time before they are associated and can be used.

The matching starts at time sample K using observations and state estimates from time $K-S$ to K. In parallel, the filter keeps estimating states without map feature observations to always get a real-time estimate. When the matching is done, the filter might have estimated states beyond time K. The filter is then run again starting at time $K-S$ up to current time to account for the new matches that have been made.

VI. EXPERIMENTAL RESULTS

A. Experimental Setup

The method has been evaluated on real data. The sensor data was recorded using an experimental Renault ZOE, see Figure 4, equipped with a u-blox M8T GNSS, an intelligent camera by Mobileye providing up to four simultaneous road marking measurements at 3.7 Hz, and a Velodyne VLP-32C lidar used to detect road signs at 10 Hz. The vehicle was also equipped with a Novatel SPAN-CPT that combines Real Time Kinematic (RTK) GNSS and high accuracy IMU data to provide centimeter level accuracy. This system is only used as a ground truth to evaluate the localization quality. The sensor data were recorded using the ROS framework and all sensors except the camera were synchronized to the GNSS time. The non-accurate synchronization of the camera has
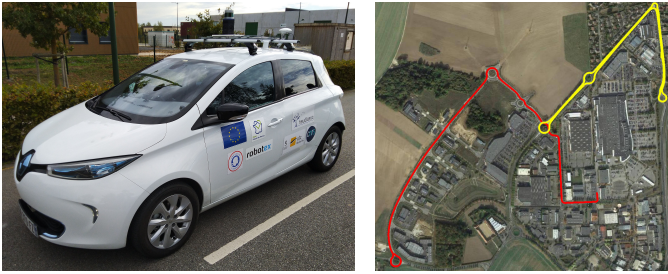
Fig. 4: Left: Experimental vehicle with GNSS, lidar and camera. Right: trajectory used to test the algorithm (red: Rambouillet 1, yellow: Rambouillet 2).

not been found to be detrimental to the localization as these lateral measurements do not change significantly as the vehicle moves.

The data has been recorded on open roads through the commercial district of the city of Rambouillet, France. The trajectory used to evaluate in detail the method consists in several stretches of straight roads separated by roundabouts, as shown in red on Figure 4. Further results will be presented in section VI-E on the yellow trajectory and on trajectories recorded in Compiègne. The Mobileye camera was able to detect road markings only during the straight portions. The road sign detector provided measurements along the entire trajectory, although there were more road signs close to intersections and roundabouts.

The map used in this experiment is a HD map that references road markings as line segments and road signs as points in two dimensions. The map was built by a third party map provider and is expected to have centimeter level of accuracy.

Several configurations have been tested in these experiments. Results are presented when the two matching strategies (UNN and HG) are applied in real-time. In this case, measurements are used as they are received and matched directly to be used in the next filter update. The same matching strategies are also applied after the method presented in this paper has been used (Buffered UNN and Buffered HG). The observations are not used directly but are saved to be matched every Δt_m. The entire matching step has been found to take less than 50 ms to compute in most cases. Δt_m was chosen high enough (250 ms) such that the matching always ends within this period. In both cases, the localization is performed at 50 Hz. The buffer length is chosen at 5s ($S = 250$), the reason for this choice is explained in section VI-D.

Thanks to a test trajectory with a ground truth, the model uncertainty of the camera has been found to increase linearly with the lateral distance to the observed markings. Therefore, the measurement covariance matrix is chosen as $\boldsymbol{R}^m = \left[(\lambda \cdot c_i)^2 \right]$, where λ is a scaling factor (chosen at 0.1). We have observed that the uncertainty of road sign measurements can be simply modeled as a diagonal matrix $\boldsymbol{R}^s = \sigma_s^2 \cdot \mathbf{I}_{2 \times 2}$. This has been estimated by a statistical analysis and found $\sigma_{RS} = 0.2$m.

As the Mobileye camera is a black-box from which the raw images are not available, evaluating the correctness of the association is impossible. Instead, the proportion of observation that can be matched is studied as well as its effect on
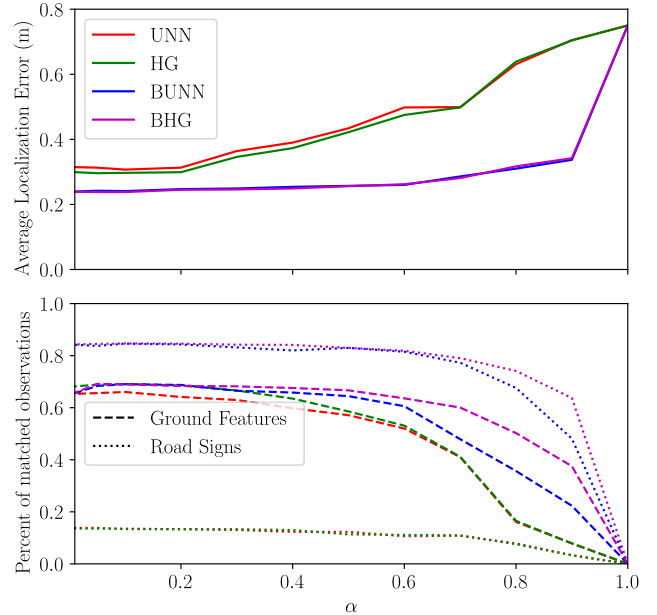

Fig. 5: Localization error and observation availability depending on the rejection rate α. Results are given for the UNN, Hungarian (HG), Buffered UNN (BUNN) and Buffered HG (BHG). The solid lines show the localization error. The dashed line (road markings) and dotted lines (road signs) describe the percentage of matched features.

localization accuracy. Finally, the influence of the buffer size is assessed.

B. Observation Availability

The number of observations that can be matched to features directly affect the localization accuracy. As expected, increasing the rejection factor α results in fewer observations for every configuration. However, our method performs better as the adjustment step reduces the distances to features and therefore allows matching more map features that will be used in the filter. Moreover, since the number of matches remains almost constant until the rejection rate reaches $\alpha = 0.7$ (see Figure 5), its value can be set high which limits the chance of bad associations.

The availability of road marking feature matches is similar for every method. The UNN approach performs slightly worse than the Hungarian association as it only matches if an observation is closest to the features. When two side-by-side ground features are detected, the estimation error can cause the UNN approach to only match a single observation while the Hungarian approach can match both, see Figure 2. Using the proposed method, the advantage of using the Hungarian method is reduced as the adjustment step compensates for the estimation error.

Please note that the road sign matching is greatly improved using an observation buffer and an adjustment step. The adjustment step enables to reduce the observation residuals used for the matching, limiting the effect of the non-linearity of the model. Also, because all observations contained in the buffer are being matched, observations that could not have
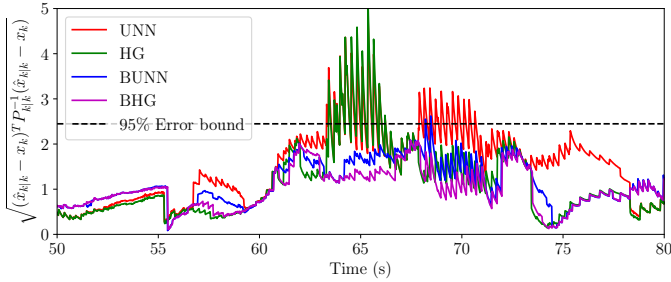
Fig. 6: Consistency plot for the four configurations. It shows the covariance weighted error (without unit) over a $30 \, s$ stretch of the test trajectory for the four tested configurations. Samples above the threshold line indicate a loss of consistency of the estimate. The results are presented for a rejection rate of $\alpha = 0.05$ for every configuration. Although decreasing the rejection rate α should increase the number of observations thus increase the accuracy, it also results in bad matches for methods UNN (red) and HG (green).
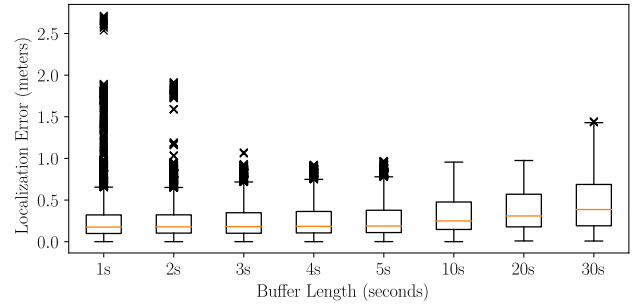

Fig. 7: Distribution of the localization error depending of the buffer size (the rejection rate is chosen at $\alpha = 0.5$). Buffers too short lead to ambiguous associations thus larger errors. Additionally, a rigid transformation cannot properly adjust the trajectory leading to fewer matches and decreased accuracy.

been matched earlier might be matched in the following iterations of the matching process. For road signs association, the Hungarian method does not provide significant improvement compared to UNN.

C. Localization accuracy and consistency

If the matching can be correctly done, a higher number of observations will result in a greater accuracy. This is true for all configurations, as the rejection rate α is chosen bigger, the number of matches decreases and the localization error increases. The observation error increases steadily when the UNN and Hungarian method are used alone. Because of the adjustment step, the number of matches for the buffered configurations only starts to decrease for high rejection rates. The localization error follows a similar pattern: it increases slowly until $\alpha = 0.6$ at which point the number of matches drops and the error increases.

When the matching is not buffered, the Hungarian method results in errors smaller by around 2 cm. Its ability to associate more road markings explains this improvement. However, when matching using an adjusted trajectory, the two methods are not noticeably different whichever rejection risk is chosen. Hence the UNN method is preferred as it is computationally less expensive to run.

Configurations without adjustment are less accurate because of the fewer measurements that can be matched. To increase the number of matches, a solution would be to lower the rejection rate. This, however, increases the change of erroneous matching. For rejection rate lower than 0.05%, a road sign is incorrectly matched greatly affecting the consistency of the state estimate, as shown in Figure 6. Since consistency is very important for the integrity of localization for autonomous vehicle navigation, we believe that snapshot matching approaches are not well suited in this context.

Given the localization for different rejection rate, α is chosen at 0.5 such that the maximum association distance is as small as possible while still keeping some margin to avoid risking the loss of accuracy occurring for $\alpha > 0.6$. With this value, the localization error averages at 0.28 m and has not been found to go beyond 1 m at any point of the trajectory.

D. Buffer size influence

Our method has two components that contribute to improving matching, the optimization step enables to reduce residuals, making matching easier thus increasing the number of observations. The second component is the use of past observations in the optimization step. Indeed, the adjustment step is dependent on the number of observations it has available. If the optimization is performed in a snapshot fashion using only to the current observations, the optimization converges toward the closest map feature to the observations. This would not help the matching as when a single road sign is observed, it would be matched to the closest feature. Hence, erroneous matchings would occur. Using a buffer of observations provides a more detailed picture of what is observed thus constraining the optimization and preventing the most recent measurements to be the sole influence on the matching.

The length of the buffer affects the number of observations considered. A longer buffer should therefore result in fewer ambiguous matchings. The buffer should, however, be small enough as the smoother state estimates are transformed in a rigid manner by the optimization, see equation (10).

Figure 7 shows the localization error of the vehicle for different buffer lengths. It can be seen that buffers shorter than 3 s are not sufficient as outliers start to appear. This is due to the lack of unambiguous measurements in small buffers to compensate for ambiguous ones. In this case, the error is due to an erroneous road sign matching in a situation similar to that described in Figure 3. With small buffers, there are not enough measurements to constrain the adjustment in orientation and laterally and the observed road sign is matched to the wrong feature (the observed sign was not referenced in the map).

Longer buffers enable to better adjust the trajectory but to a point. For buffers longer than 10 s, the localization error starts to increase. This is the result of the assumption made for the adjustment that the trajectory only needs to be moved rigidly. Hence, too long buffers result in higher residuals after adjustment. This causes more observations to be rejected, finally affecting the localization error. For a rejection rate of $\alpha = 0.5$, a buffer length of 5 s has been found to be sufficient to enable unambiguous matches while still providing enough measurements.

	BUNN	UNN		HG	
α	0.5	0.05	0.5	0.05	0.5
Rambouillet 1	**0.25**	0.31	0.42	0.3	0.43
Rambouillet 2	**0.68**	1.25	2.75	1.26	2.75
Compiègne 1	**0.75**	0.94	0.97	0.88	0.96
Compiègne 2	**0.57**	0.65	0.73	0.60	0.73
Compiègne 3	**0.77**	0.78	0.83	0.78	0.82

TABLE I: Average localization error (in meter) for each method tested on different trajectories.

E. Evaluation using different datasets

To further evaluate the performance of the proposed approach, it has been tested on another recording done in Rambouillet and on three other experiments carried out in another city (Compiègne) obtained with a similarly equipped vehicle. The five trajectories amount to 21 km of roads (50 minutes). Table I contains the average localization error for different datasets. It can be seen that in all cases the proposed method performs better (by 16% on average) although sometime marginally (Compiègne 3). The parameters used for the buffered method were kept the same as those identified with the first Rambouillet trajectory. Even with a low α, the UNN and HG methods rarely achieve accuracies similar to our approach. These results also show that optimized parameters on a particular trajectory lead to a good accuracy in different environments which shows a good robustness of the method.

VII. CONCLUSION

In this paper, an adjustment method aiming at improving features matching for map-aided localization has been presented. The quality of the localization is greatly improved by using signs and markings in buffers. Indeed, adjusting the trajectory before matching increases the number of road signs that can be matched sixfold. Through the minimization step described in this work, the adjustment can be made without making assumptions on the observations correspondence to map features. This also enables to set stricter rejection rate while keeping a good measurement availability. The quality of the adjustment step depends on the number of observations it uses. This work shows that by only keeping measurements from the last 5 seconds, the adjustment can converge satisfactorily and not result in wrong associations that would affect the localization error. Because of the increased number of observations that are matched, the accuracy of the localization system is also improved. In urban environment, the localization error averages 0.66 meters and rarely goes beyond a meter.

REFERENCES

[1] P. Biber and W. Strasser. The normal distributions transform: a new approach to laser scan matching. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[2] Kevin Doherty, David Baxter, Edward Schneeweiss, and John Leonard. Probabilistic Data Association via Mixture Models for Robust Semantic SLAM. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[3] Renaud Dube, Daniel Dugas, Elena Stumm, Juan Nieto, Roland Siegwart, and Cesar Cadena. SegMatch: Segment based place recognition in 3D point clouds. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5266–5272, Singapore, Singapore, May 2017.

[4] Renaud Dube, Mattia G. Gollub, Hannes Sommer, Igor Gilitschenski, Roland Siegwart, Cesar Cadena, and Juan Nieto. Incremental-Segment-Based Localization in 3-D Point Clouds. *IEEE Robotics and Automation Letters*, 3(3):1832–1839, July 2018.

[5] Farouk Ghallabi, Marie-Anne Mittet, Ghayath El-Haj-Shhade, and Fawzi Nashashibi. LIDAR-Based High Reflective Landmarks (HRL)s For Vehicle Localization in an HD Map. In *IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4412–4418, Auckland, New Zealand, October 2019.

[6] Dominique Gruyer, Rachid Belaroussi, and Marc Revilloud. Map-aided localization with lateral perception. In *IEEE Intelligent Vehicles Symposium Proceedings*, pages 674–680, MI, USA, June 2014.

[7] A. Hata and D. Wolf. Road marking detection using lidar reflective intensity data and its application to vehicle localization. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 584–589, Oct 2014.

[8] Ming Hsiao and Michael Kaess. MH-iSAM2: Multi-Hypothesis iSAM Using Bayes Tree and Hypo-Tree. In *IEEE International Conference on Robotics and Automation (ICRA)*, page 7, Montreal, QC, Canada, May 2019.

[9] Michael Kaess, Stephen Williams, Vadim Indelman, Richard Roberts, John J Leonard, and Frank Dellaert. Concurrent Filtering and Smoothing. In *15th International Conference on Information Fusion*, pages 1300–1307, July 2012.

[10] Dongwook Kim, Beomjun Kim, Taeyoung Chung, and Kyongsu Yi. Lane-Level Localization Using an AVM Camera for an Automated Driving Vehicle in Urban Environments. *IEEE/ASME Transactions on Mechatronics*, 22(1):280–290, February 2017.

[11] Henning Lategahn and Christoph Stiller. Vision-Only Localization. *IEEE Transactions on Intelligent Transportation Systems*, 15(3):1246–1257, June 2014.

[12] Jesse Levinson and Sebastian Thrun. Map-based precision vehicle localization in urban environments. In *Robotics: Science and Systems*, 2007.

[13] Ellon Mendes, Pierrick Koch, and Simon Lacroix. ICP-based pose-graph SLAM. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 195–200, Lausanne, Switzerland, October 2016.

[14] James. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[15] Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. *The International Journal of Robotics Research*, 32(7):826–840, 2013.

[16] Jan-Hendrik Pauls, Tobias Strauss, Carsten Hasberg, Martin Lauer, and Christoph Stiller. Can We Trust Our Maps? An Evaluation of Road Changes and a Dataset for Map Validation. In *21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2639–2644, Maui, HI, November 2018.

[17] H. E. Rauch, C. T. Striebel, and F. Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, August 1965.

[18] David M. Rosen, Michael Kaess, and John J. Leonard. Robust incremental online inference over sparse factor graphs: Beyond the Gaussian case. In *IEEE International Conference on Robotics and Automation*, pages 1025–1032, Karlsruhe, Germany, May 2013.

[19] Zui Tao, Philippe Bonnifait, Vincent Frémont, Javier Ibañez-Guzmán, and Stéphane Bonnet. Road-centred map-aided localization for driverless cars using single-frequency GNSS receivers. *Journal of Field Robotics*, 34(5):1010–1033, 2017.

[20] R. Toledo-Moreo, D. Betaille, F. Peyret, and J. Laneurit. Fusing GNSS, Dead-Reckoning, and Enhanced Maps for Road Vehicle Lane-Level Navigation. *IEEE Journal of Selected Topics in Signal Processing*, 3(5):798–809, October 2009.

[21] A. Welte, P. Xu, and P. Bonnifait. Four-wheeled dead-reckoning model calibration using RTS smoothing. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 312–318, May 2019.

[22] R. W. Wolcott and R. M. Eustice. Fast LIDAR localization using multiresolution gaussian mixture maps. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2814–2821, May 2015.