Large Scale Graph-based SLAM
using Aerial Images as Prior Information

Rainer Kümmerle   Bastian Steder   Christian Dornhege
Alexander Kleiner   Giorgio Grisetti   Wolfram Burgard

Department of Computer Science, University of Freiburg, Germany

Abstract— To effectively navigate in their environments and accurately reach their target locations, mobile robots require a globally consistent map of the environment. The problem of learning a map with a mobile robot has been intensively studied in the past and is usually referred to as the simultaneous localization and mapping (SLAM) problem. However, existing solutions to the SLAM problem typically rely on loop-closures to obtain global consistency and do not exploit prior information even if it is available. In this paper, we present a novel SLAM approach that achieves global consistency by utilizing publicly accessible aerial photographs as prior information. Our approach inserts correspondences found between three-dimensional laser range scans and the aerial image as constraints into a graph-based formulation of the SLAM problem. We evaluate our algorithm based on large real-world datasets acquired in a mixed in- and outdoor environment by comparing the global accuracy with state-of-the-art SLAM approaches and GPS. The experimental results demonstrate that the maps acquired with our method show increased global consistency.

I. INTRODUCTION

The ability to acquire accurate models of the environment is widely regarded as one of the fundamental preconditions for truly autonomous robots. In the context of mobile robots, these models typically are maps of the environment that support different tasks including localization and path planning. The problem of estimating a map with a mobile robot navigating through and perceiving its environment has been studied intensively and is usually referred to as the simultaneous localization and mapping (SLAM) problem.

In its original formulation, SLAM does not require any prior information about the environment and most SLAM approaches seek to determine the most likely map and robot trajectory given a sequence of observations without taking into account any prior information about the environment. How-ever, there are certain scenarios, in which one wants a robot to autonomously arrive at a specific location described in global terms, for example, given by a GPS coordinate. Consider, for example, rescue or surveillance scenarios in which one requires specific areas to be covered with high probability to minimize the risk of potential casualties. Unfortunately, GPS typically suffers from outages or occlusions so that a robot only relying on GPS information might encounter substantial positioning errors. Even sophisticated SLAM algorithms can-not fully compensate for these errors as there still might be lacking constraints between some observations combined with large odometry errors that introduce a high uncertainty in the

current position of the vehicle. However, even in situations with substantial overlaps between consecutive observations, the matching processes might result in errors that linearly propagate over time and lead to substantial absolute errors. Consider, for example, a mobile robot mapping a linear structure (such as a corridor of a building or the passage between to parallel buildings). Typically, this corridor will be slightly curved in the resulting map. Whereas this is not critical in general as the computed maps are generally locally consistent [13], they often still contain errors with respect to the global coordinate system. Thus, when the robot has to arrive at a position defined in global coordinates, the maps built using a standard SLAM algorithm might be sub-optimal.

In this paper, we present an approach that overcomes these problems by utilizing aerial photographs for calculating global constraints within a graph-representation of the SLAM problem. In our approach, these constraints are obtained by matching features from 3D point clouds to aerial images.

Compared to traditional SLAM approaches, the use of a global prior enables our technique to provide more accurate solutions by limiting the error when visiting unknown regions. In contrast to approaches that seek to directly localize a robot in an outdoor environment, our approach is able to operate reliably even when the prior is not available, for example, because of the lack of appropriate matches. Therefore, it is suitable for mixed indoor/outdoor operation. Figure 1 shows a motivating example and compares the outcome of our approach with the ones obtained by applying a state-of-the-art SLAM algorithm and a pure localization method using aerial images.

The approach proposed in this paper relies on the so called graph formulation of the SLAM problem [18, 22]. Every node of the graph represents a robot pose and an observation taken at this pose. Edges in the graph represent relative transforma-tions between nodes computed from overlapping observations. Additionally, our system computes its global position for every node employing a variant of Monte-Carlo localization (MCL) which uses 3D laser scans as observations and aerial images as reference maps. The use of 3D laser information allows our system to determine the portions of the image and of the 3D scan that can be reliably matched by detecting structures in the 3D scan which potentially correspond to intensity variations in the aerial image.

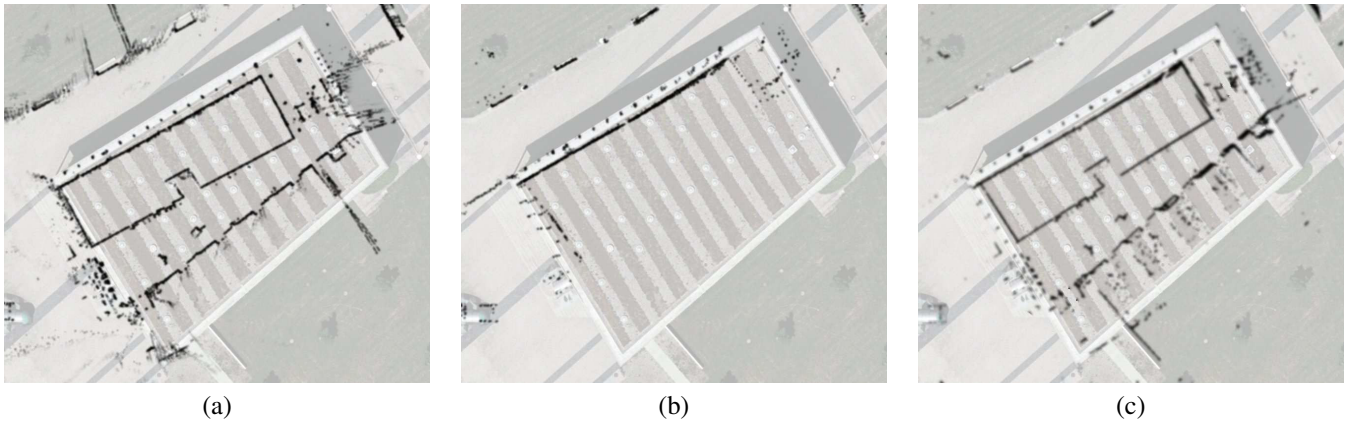GPS is a popular device for obtaining accurate position

Fig. 1.   Motivating example comparing standard SLAM (a), localization using aerial imagery as prior information (b), and our combined approach (c). Note the misalignment relative to the outer wall of the building in (a). Whereas the localization applied in (b), which relies on aerial images, yields proper alignments, it cannot provide accurate estimates inside the building. Combining the information of both algorithms yields the best result (c).

estimates. Whereas it has also been used to localize mobile vehicles operating outdoors, the accuracy of this estimate is usually not accurate enough to obtain a precise map. Generally, the position estimate provided by GPS substantially decreases when the robot moves close to buildings or in narrow streets. Our approach to deal with this problem is to use aerial images and to match measurements obtained by the robot to obtain an accurate estimate of the global position.

The approach proposed in this paper works as follows: we apply a variant of Monte Carlo localization [3] to localize a robot by matching 3D range scans to aerial images of the environment. To achieve this, our approach selects the portions of the scan and of the image which can be reliably matched. These correspondences are added as constraints in a graph-based formulation of the SLAM problem. Note that our system preserves the flexibility of traditional SLAM approaches and can also be used in absence of the prior information. However, when the prior is available our system provides highly accurate solutions also in pathological datasets (i.e., when no loop closures take place). We validate the results with a large-scale dataset acquired in a mixed in- and outdoor environment. We furthermore compare our method with state-of-the-art SLAM approaches and with GPS.

This paper is organized as follows. After discussing related work, we will give an overview over our system followed by a detailed description of the individual components in Section III. We then will present experiments designed to evaluate the quality of the resulting maps obtained with our algorithm in Section IV. In this section, we furthermore compare our approach with a state-of-the-art SLAM system that does not use any prior information.

II. RELATED WORK

SLAM techniques for mobile robots can be classified according to the underlying estimation technique. The most popular approaches are extended Kalman filters (EKFs) [16, 23], sparse extended information filters [7, 26], particle filters [19], and least square error minimization approaches [18, 9, 12].

The effectiveness of the EKF approaches comes from the fact that they estimate a fully correlated posterior about landmark maps and robot poses [16, 23]. Their weakness lies in the strong assumptions that have to be made upon both, the robot motion model and the sensor noise. If these assumptions are violated, the filter is likely to diverge [14, 27].

An alternative approach designed to find maximum likelihood maps is the application of least square error minimization. The idea is to compute a network of relations given the sequence of sensor readings. These relations represent the spatial constraints between the poses of the robot. In this paper, we also follow this approach. Lu and Milios [18] first applied this technique in robotics to address the SLAM problem by optimizing the whole network at once. Gutmann and Konolige [12] proposed an effective way for constructing such a network and for detecting loop closures while running an incremental estimation algorithm.

All the SLAM methods discussed above do not take into account any prior knowledge about the environment. On the other hand, several authors addressed the problem of utilizing prior knowledge to localize a robot outdoors. For example, Korah and Rasmussen [15] used image processing techniques to extract roads on aerial images. This information is then applied to improve the quality of GPS paths using a particle filter by calculating the particle weight according to its position relative to the streets. Leung *et al.* [17] presented a particle filter system performing localization on aerial photographs by matching images taken from the ground by a monocular vision system. Correspondences between aerial images and ground images have been detected by matching line features. These have been generated from aerial images by a Canny edge detector and Progressive Probabilistic Hough Transform (PPHT). A vanishing point analysis for estimating building wall orientations was used on the monocular vision. In contrast to laser-based approaches, their method maximally achieved an average positioning accuracy within several meters. Ding *et al.* [4] use vanishing point analysis to extract 2D corners from aerial images and inertial tracking data, and they also extract

2D corners from LiDAR generated depth maps. The extracted corners from LiDAR are matched with those from the aerial image in a multi-stage process. Corresponding matches are used to gain a fine estimation of the camera pose that is used to texture the LiDAR models with the aerial images. Chen and Wang [2] use an energy minimization technique to merge prior information from aerial images and mapping. Mapping is performed by constructing sub-maps consisting of 3D point clouds, that are constrained by relations. Using a Canny edge detector, they compute a vector field from the image that models force towards the detected edges. The sum of the forces applied to each point is used as an energy measure in the minimization process, when placing a sub-map into vector field of the image. Dogruer *et al.* [5] utilized soft computing techniques for segmenting aerial images into different regions, such as buildings, roads, and forests. They applied MCL on the segmented maps. However, compared to the approach presented in this paper, their technique strongly depends on the color distribution of the aerial images since different objects on these images might share similar color characteristics.

Früh and Zakhor [10] introduced the idea of generating edge images out of aerial photographs for 2D laser-based localization. As they stated in their paper, localization errors might occur if rooftops seen on the aerial image significantly differ from the building footprint observed by the 2D scanner. The method proposed in this paper computes a 2D structure from a 3D scan, which is more likely to match with the features extracted from the aerial image. This leads to an improved robustness in finding location correspondences. Additionally, our system is not limited to operate in areas where the prior is available. In this cases, our algorithm operates without relevant performance loss compared to SLAM approaches which do not utilize any prior. This allows our system to operate in mixed indoor/outdoor scenarios.

Sofman *et al.* [24] introduced an online learning system predicting terrain travel costs for unmanned ground vehicles (UGVs) on a large scale. They extracted features from locally observed 3D point clouds and generalized them on overhead data such as aerial photographs, allowing the UGVs to navigate on less obstructed paths. Montemerlo and Thrun [20] presented an approach similar to the one presented in this paper. The major difference to our technique is that they used GPS to obtain the prior. Due to the increased noise which affects the GPS measurements this prior can result in larger estimation errors.

III. Graph-SLAM with Prior Information from Aerial Images

Our system relies on a graph-based formulation of the SLAM problem. It operates on a sequence of 3D scans and odometry measurements. Every node of the graph represents a position of the robot at which a sensor measurement was acquired. Every edge stands for a constraint between the two poses of the robot. In addition to direct links between consecutive poses, it can integrate prior information (when

Fig. 2. The graph representation used by our approach. In contrast to the standard approach, we additionally integrate global constraints (shown in yellow / light gray) given by the prior information.

available) which in our case is given in form of an aerial image.

This prior information is introduced to the graph-SLAM framework as global constraints on the nodes of the graph, as shown in Figure 2. These global constraints are absolute locations obtained by Monte Carlo localization [3] on a map computed from the aerial images. These images are captured from a viewpoint significantly different from the one of the robot. However, by using 3D-scans we can extract the 2D information which is more likely to be consistent with the one visible in the reference map. In this way, we can prevent the system from introducing inconsistent prior information. To initialize the particle filter, we draw the particle positions from a Gaussian distribution, where the mean was determined by GPS. We use 1,000 particles to approximate the posterior.

In the following we explain how we adapted MCL to operate on aerial images and how to select the points in the 3D scans to be considered in the observation model. Subsequently we describe our graph-SLAM framework.

A. Monte Carlo Localization

To estimate the pose $\mathbf{x}$ of the robot in its environment, we consider probabilistic localization, which follows the recursive Bayesian filtering scheme. The key idea of this approach is to maintain a probability density $p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{0:t-1})$ of the location $\mathbf{x}_t$ of the robot at time t given all observations $\mathbf{z}_{1:t}$ and all control inputs $\mathbf{u}_{0:t-1}$. This posterior is updated as follows:

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{0:t-1}) = \qquad\qquad (1)$$
$$\alpha \cdot p(\mathbf{z}_t \mid \mathbf{x}_t) \cdot \int p(\mathbf{x}_t \mid \mathbf{u}_{t-1}, \mathbf{x}_{t-1}) \cdot p(\mathbf{x}_{t-1}) \, d\mathbf{x}_{t-1}.$$

Here, α is a normalization constant which ensures that $p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{0:t-1})$ sums up to one over all $\mathbf{x}_t$. The terms to be described in Eqn. (1) are the *prediction model* $p(\mathbf{x}_t \mid \mathbf{u}_{t-1}, \mathbf{x}_{t-1})$ and the *sensor model* $p(\mathbf{z}_t \mid \mathbf{x}_t)$. One contribution of this paper is an appropriate computation of the sensor model in the case that a robot equipped with a 3D range sensor operates in a map generated from a birds-eye view.

For the implementation of the described filtering scheme, we use a sample-based approach which is commonly known

as *Monte Carlo localization* (MCL) [3]. MCL is a variant of particle filtering [6] where each particle corresponds to a possible robot pose and has an assigned weight $w^{[i]}$. The *belief update* from Eqn. (1) is performed according to the following two alternating steps:

1) In the **prediction step**, we draw for each particle with weight $w^{[i]}$ a new particle according to $w^{[i]}$ and to the prediction model $p(\mathbf{x}_t \mid \mathbf{u}_{t-1}, \mathbf{x}_{t-1})$.

2) In the **correction step**, a new observation $\mathbf{z}_t$ is integrated. This is done by assigning a new weight $w^{[i]}$ to each particle according to the sensor model $p(\mathbf{z}_t \mid \mathbf{x}_t)$.

Furthermore, the particle set needs to be re-sampled according to the assigned weights to obtain a good approximation of the pose distribution with a finite number of particles. However, the re-sampling step can remove good samples from the filter which can lead to particle impoverishment. To decide when to perform the re-sampling step, we calculate the number N_{eff} of effective particles according to the formula proposed in [6]

$$N_{\text{eff}} \;=\; \frac{1}{\sum_{i=1}^{N} \left(\widetilde{w^{[i]}}^2 \right)}, \qquad (2)$$

where $\widetilde{w^{[i]}}$ refers to the normalized weight of sample i and we only re-sample if N_{eff} drops below the threshold of $\frac{N}{2}$ where N is the number of samples. In the past, this approach has already successfully been applied in the context of the simultaneous localization and mapping (SLAM) problem [11].

So far we described the general framework of MCL. In the next section, we will describe our sensor model for determining the likelihood $p(\mathbf{z} \mid \mathbf{x})$ of perceiving the 3D scan $\mathbf{z}$ from a given robot position $\mathbf{x}$ within an aerial image.

B. Sensor Model for 3D Range Scans in Aerial Images

The task of the sensor model is to determine the likelihood $p(\mathbf{z} \mid \mathbf{x})$ of a reading $\mathbf{z}$ given the robot is at pose $\mathbf{x}$. In our current system, we apply the so called endpoint model or likelihood fields [25]. Let z_k be the k-th measurement of a 3D scan $\mathbf{z}$. The endpoint model computes the likelihood of z_k based on the distance between the scan point z_k' corresponding to z_k re-projected onto the map according to the pose $\mathbf{x}$ of the robot and the point in the map d_k' which is closest to z_k' as:

$$p(\mathbf{z} \mid \mathbf{x}) \;=\; f(\|z_1' - d_1'\|, \ldots, \|z_k' - d_k'\|). \qquad (3)$$

By assuming that the beams are independent and the sensor noise is normally distributed we can rewrite (3) as

$$f(\|z_1' - d_1'\|, \ldots, \|z_k' - d_k'\|) \;\propto\; \prod_j e^{\frac{(z_j' - d_j')^2}{\sigma^2}}. \qquad (4)$$

Since the aerial image only contains 2D information about the scene, we need to select a set of beams from the 3D scan, which are likely to result in structures, that can be identified and matched in the image. In other words, we need to transform both, the scan and the image in a set of 2D points which can be compared via the function $f(\cdot)$.

To extract these points from the image we employ the standard Canny edge extraction procedure [1]. The idea behind this is, that if there is a height gap in the aerial image, there will often also be a visible change in intensity in the aerial image. This intensity change will be detected by the edge extraction procedure. In an urban environment, such edges typically correspond to borders of roofs, trees, fences, or other structures. Of course, the edge extraction procedure returns a lot of false positives that do not represent any actual 3D structure, like street markings, grass borders, shadows, and other flat markings. All these aspects have to be considered by the sensor model. Figure 3 shows an aerial image and the extracted canny image along with the likelihood-field.

To transform the 3D scan into a set of 2D points which can be compared to the canny image, we select a subset of points from the 3D scan and consider their 2D projection in the ground plane. This subset should contain all the points which may be visible in the reference map. To perform this operation we compute the z-buffer [8] of a scan from a bird's eye perspective. In this way we discard those points which are occluded in the bird's eye view from the 3D scan. By simulating this view, we handle situations like overhanging roofs, where the house wall is occluded and therefore is not visible in the aerial image in a more sophisticated way.

The regions of the z-buffer which are likely to be visible in the canny image are the ones which correspond to relevant depth changes. We construct a 2D scan by considering the 2D projection of the points in these regions. This procedure is illustrated by the sequence of images in Figure 4.

An implementation purely based on a 2D scanner (like the approach proposed by Früh and Zakhor [10]) would not account for occlusions due to overhanging objects. An additional situation where our approach is more robust is in the presence of trees. In this case a 2D view would only sense the trunk, whereas the whole crown is visible in the aerial image.

In our experiments, we considered variations in height of 0.5 m and above as possible positions of edges that could also be visible in the aerial image. The positions of these variations relative to the robot can then be matched against the Canny edges of the aerial image in a point-by-point fashion and in a similar way like matching of 2D-laser scans against an occupancy grid map.

This sensor model has some limitations. It is susceptible to visually cluttered areas, since it then can find random correspondences in the Canny edges. There is also the possibility of systematic errors, when a wrong line is used for the localization, e.g., in the case of shadows. Since we use position tracking, this is not critical, unless the robot moves through such areas for a long period. The main advantages of the end point model in this context are that it ignores possible correspondences outside of a certain range and implicitly deals with edge points that do not correspond to any 3D structure.

The method, of course, also depends on the quality of the aerial images. Perspective distortions in the images can easily introduce errors. However, in our experiments we did not find
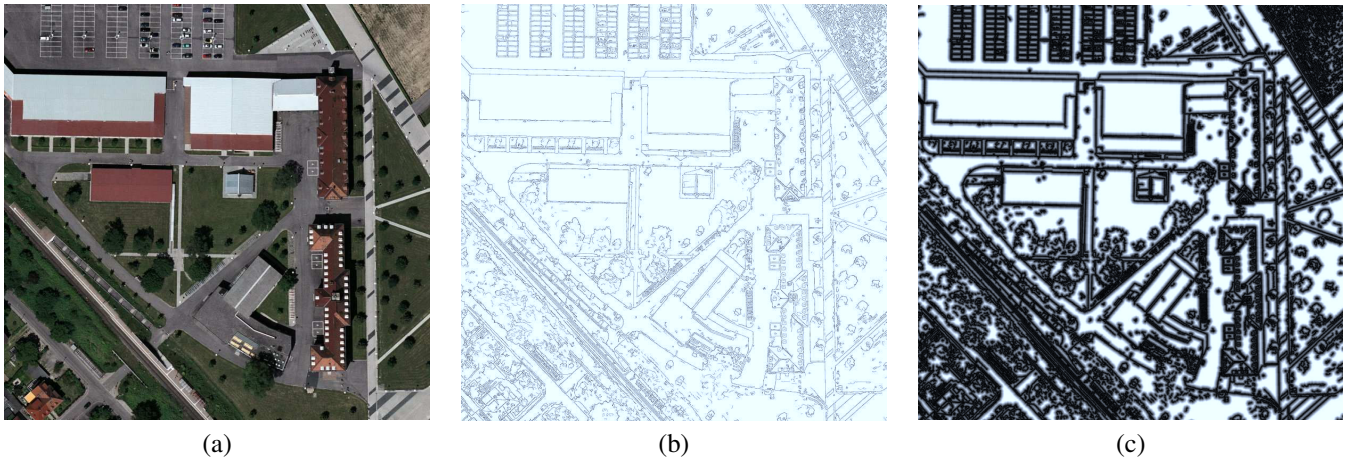
Fig. 3. Google Earth image of the Freiburg campus (a), the corresponding Canny image (b), and the corresponding likelihood field computed from the Canny image (c). Note that the structure of the buildings and the vertical elements is clearly visible despite of the considerable clutter.
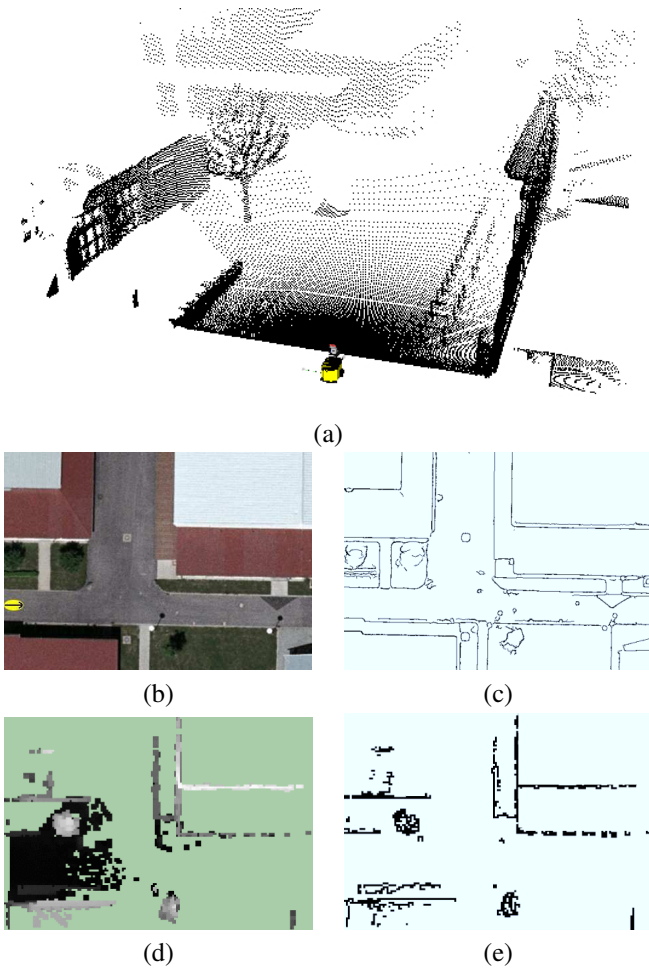

Fig. 4. A 3D scan represented as a point cloud (**a**), the aerial image of the corresponding area (**b**), the Canny edges extracted from the aerial image (**c**), the 3D scene from (a) seen from the top (**d**) (gray values represent the maximal height per cell, the darker a pixel, the lower the height, and the green/bright gray area was not visible in the 3D scan), and positions extracted from (d), where a high variation in height occurred (**e**).

evidence that this is a major complicating factor.

Note that we employ a heuristic to detect when the prior is not available, i.e., when the robot is inside of a building or under overhanging structures. This heuristic is based on the 3D perception. If there are range measurements whose endpoints are above the robot, no global constraints from the localization are integrated, since we assume that the area the robot is sensing is not visible in the aerial image. While a more profound solution regarding place recognition is clearly possible, this conservative heuristic turned out to yield sufficiently accurate results.

C. Graph-based Maximum Likelihood SLAM

This section describes the basic algorithm for obtaining the maximum likelihood trajectory of the robot. We use a graph-based SLAM technique to estimate the most-likely trajectory, i.e., we seek for the maximum-likelihood (ML) configuration like the majority of approaches to graph-based SLAM.

In our approach, each node $\mathbf{x}_i$ models a robot pose. The spatial constraints between two poses are computed by matching laser scans. An edge between two nodes i and j is represented by the tuple $\langle \delta_{ji}, \Omega_{ji} \rangle$, where δ_{ji} and Ω_{ji} are the mean and the information matrix of the measurement respectively. Let $e_{ji}(\mathbf{x}_i, \mathbf{x}_j)$ be the error introduced by the constraint $\langle j, i \rangle$. Assuming the independence of the constraints, a solution to the SLAM problem is given by

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n)^* = \underset{(\mathbf{x}_1, \ldots, \mathbf{x}_n)}{\operatorname{argmin}} \sum_{\langle j, i \rangle} e_{ji}(\mathbf{x}_i, \mathbf{x}_j)^T \Omega_{ji} e_{ji}(\mathbf{x}_i, \mathbf{x}_j). \quad (5)$$

To account for the residual error in each constraint, we can additionally consider the prior information by incorporating the position estimates of our localization approach. To this end, we extend Eqn. (5) as follows:

$$(\mathbf{x}_1, \ldots, \mathbf{x}_n)^* = \underset{(\mathbf{x}_1, \ldots, \mathbf{x}_n)}{\operatorname{argmin}} \sum_{\langle j, i \rangle} e_{ji}(\mathbf{x}_i, \mathbf{x}_j)^T \Omega_{ji} e_{ji}(\mathbf{x}_i, \mathbf{x}_j)$$
$$+ \sum_{i \in \mathcal{G}} (\mathbf{x}_i - \hat{\mathbf{x}}_i)^T R_i (\mathbf{x}_i - \hat{\mathbf{x}}_i), \quad (6)$$

where $\hat{\mathbf{x}}_i$ denotes the position as it is estimated by the localization using the bird's eye image and R_i is the information matrix of this constraint. In our approach, we compute R_i based on the distribution of the samples in MCL. We use non-linear conjugate gradient to efficiently optimize Eqn. (6). The result of the optimization is a set of poses which maximizes the likelihood of all the individual observations. Furthermore, the optimization also accommodates the prior information about the environment to be mapped whenever such information is available. In particular, the objective function encodes the available pose estimates as given by our MCL algorithm described in the previous section. Intuitively the optimization deforms the solution obtained by the relative constraints path to maximize the overall likelihood of all the observations, including the priors. Note that including the prior information about the environment yields a globally consistent estimate of the trajectory.

If the likelihood function used for MCL is a good approximation of the global likelihood of the process, one can refine the solution of graph-based SLAM by an iterative procedure. During one iteration, new global constraints are computed based on the gradient of the likelihood function estimated at the current robot locations. The likelihood function computed by our approach is based on a satellite image from which we extract "potential" 3D structures which may be seen from the laser. Whereas the heuristics described in Section III-B reject most of the outliers, there may be situations where our likelihood function approximates the true likelihood only locally. We observed that applying the iterative procedure discussed above may worsen the initial estimate generated by our approach.

IV. Experiments

The approach described above has been implemented and evaluated on real data acquired with a *MobileRobots Powerbot* with a *SICK LMS* laser range finder mounted on an *Amtec* wrist unit. The 3D data used for the localization algorithm has been acquired by tilting the laser up and down while the robot moves. The maximum translational velocity of the robot during data acquisition was 0.35 m/s. This relatively low speed allows our robot to obtain 3D data which is sufficiently dense to perform scan matching without the need to acquire the scans in a stop-and-go fashion. During each 3D scan the robot moved up to 2 m. We used the odometry to account for the distortion caused by the movement of the platform. Figure 5 depicts the setup of our robot. Although the robot is equipped with an array of sensors, in the experiments we only used the devices mentioned above.

A. Experiment 1 - Comparison to GPS

This experiment aims to show the effectiveness of the localization on aerial images compared with the one achievable with GPS. We manually steered our robot along a 890 m long trajectory through our campus, entering and leaving buildings. The robot captured 445 3D scans that were used for

Fig. 5. The robot used for carrying out the experiments is equipped with a laser range finder mounted on a pan/tilt unit. We obtain 3D data by continuously tilting the laser while the robot moves.
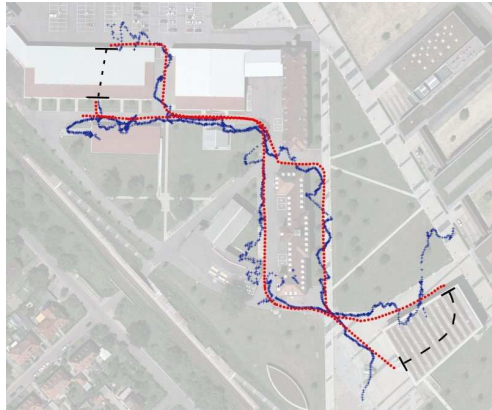

Fig. 7. Comparison between GPS measurements (blue crosses) and global poses from the localization in the aerial image (red circles). Dashed lines indicate transitions through buildings, where GPS and aerial images are unavailable.

localization. We also recorded the GPS data for comparison purposes. The data acquisition took approximately one hour.

Figure 7 compares the GPS estimate with the one obtained by MCL on the aerial view. The higher error of the GPS-based approach is clearly visible. Note that GPS, in contrast to our approach, does not explicitly provide the orientation of the robot.

B. Experiment 2 - Global Map Consistency

The goal of this experiment is to evaluate the ability of our system to create a consistent map of a large mixed in- and outdoor environment and to compare it against a state-of-the-art SLAM approach similar to the one proposed by Olson [21].

We evaluate the global consistency of the generated maps obtained with both approaches. To this end, we recorded five data sets by steering the robot through our campus area. In each run the robot follows approximately the same trajectory. The trajectory of one of these data sets as it is estimated by our approach and a standard graph-SLAM method is shown in Figure 6.

For each of the two approaches (our method using the aerial image and the graph-based SLAM technique that uses no prior
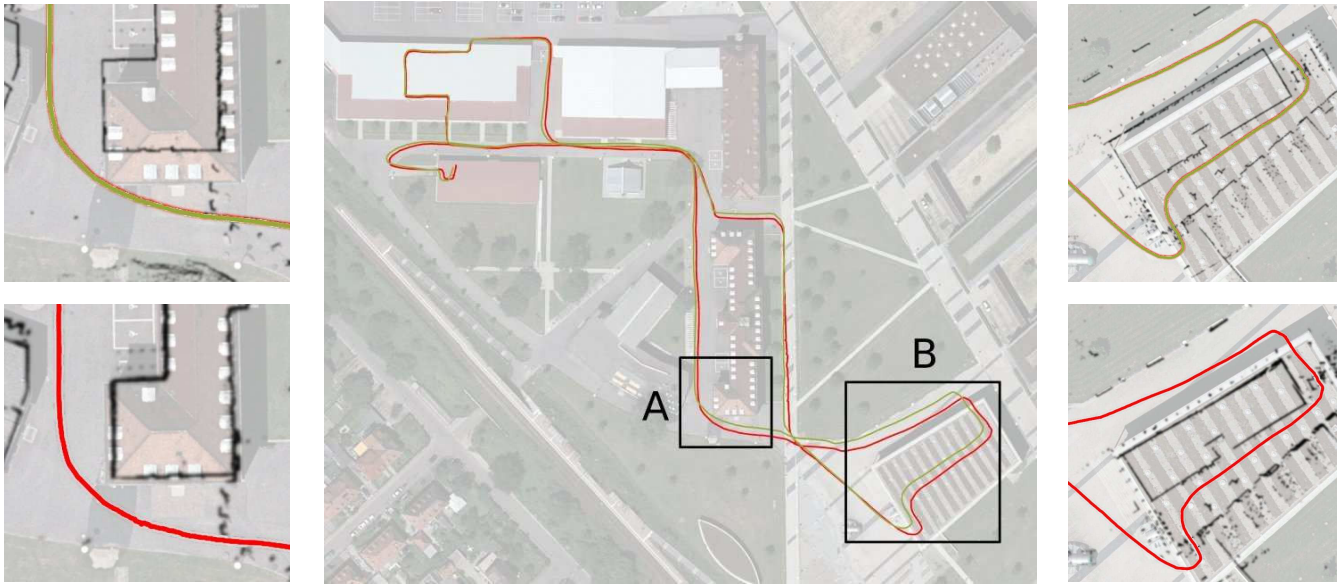
Fig. 6. Comparison of our system to a standard SLAM approach in a complex indoor/outdoor scenario. The center image shows the trajectory estimated by the SLAM approach (bright/yellow) and the trajectory generated by our approach (dark/red) overlaid on the Google Earth image used as prior. On the left and right side, detailed views of the areas marked in the center image are shown, each including the trajectory and map. The upper images show the results of the standard SLAM approach; detail A on the left and B on the right. The lower images show the results of our system (A on the left side and B on the right). It is clearly visible, that, in contrast to the SLAM algorithm without prior information, the map generated by our approach is accurately aligned with the aerial image.
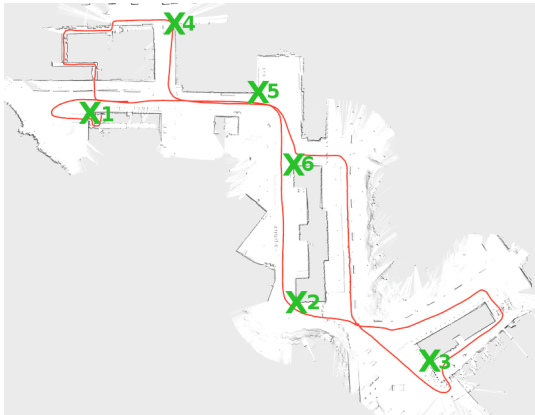
Fig. 8. The six points (corners on the buildings) we used for evaluation are marked as crosses on the map.

Fig. 9. Error bars ($\alpha = 0.05$) for the estimated distances between the six points used for evaluation of the map consistency.

information) we calculated the maximum likelihood map by processing the acquired data of each run.

For each of the five data sets we evaluated the global consistency of the maps by manually measuring the distances between six easily distinguishable points on the campus. We compared these distances to the corresponding distances in the maps (see Figure 8). We computed the average error in the distance between these points. The result of this comparison is summarized in Figure 9. As ground-truth data we used the so-called *Automatisierte Liegenschaftskarte* which is provided by the German land registry office. It contains the outer walls of all buildings where the coordinates are stored in a Gauss-Krüger reference frame.
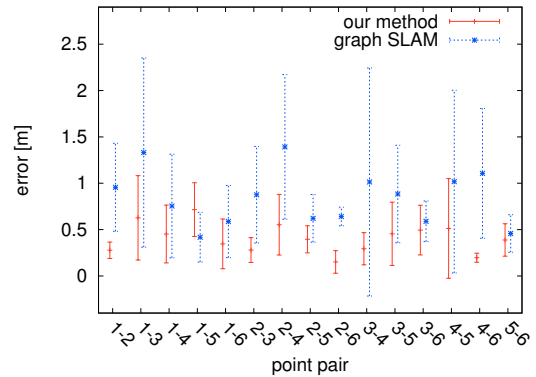
Compared to SLAM without prior information, our approach has a smaller error and it does not require frequent loop closures to limit the error of the estimate. Note that using our approach loop closures are not required to obtain a globally consistent map. Additionally, the standard deviation of the estimated distances is substantially smaller than the standard deviation obtained with a graph-SLAM approach that does not utilize prior information. Our approach is able to robustly estimate a globally consistent map on each data set.

C. Experiment 3 - Local Alignment Errors

Ideally, the result of a SLAM algorithm should perfectly correspond to the ground truth. For example, the straight wall of a building should lead to a straight structure in the resulting
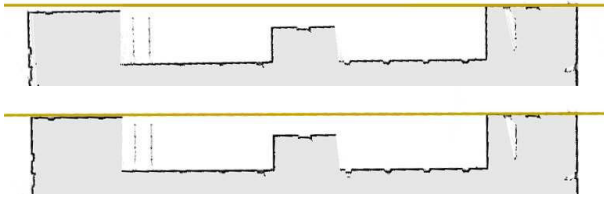
Fig. 10. Close up view of an outer wall of a building as it is estimated by graph-SLAM (top) and our method with prior information (bottom). In both images a horizontal line visualizes the true orientation of the wall. As can be seen from the image, graph-SLAM bends the building.

map. However, the residual errors in the scan matching process typically lead to a slightly bended wall. We investigated this in our data sets for both SLAM algorithms by analyzing an approximately 70 m long building on our campus. To measure the accuracy, we approximated the first part of the wall by a line and extended this line to the other end of the building. In a perfectly estimated map, both corners of the building are located on this line. Figure 10 depicts a typical result. On average the distance between the horizontal line and the corner of the building for graph-SLAM is 0.5 m whereas it is 0.2 m for our approach.

V. Conclusion

In this paper, we presented an approach to solve the SLAM problem in mixed in- and outdoor environments based on 3D range information and using aerial images as prior information. To incorporate the prior given by the aerial images into the graph-SLAM procedure, we utilize a variant of Monte-Carlo localization with a novel sensor model for matching 3D laser scans to aerial images. In this way, our approach can achieve global consistency without the need to close loops.

Our approach has been implemented and tested in a complex indoor/outdoor setting. Practical experiments carried out on data recorded with a real robot demonstrate that our algorithm outperforms state-of-the-art approaches for solving the SLAM problem that have no access to prior information. In situations, in which no global constraints are available, our approach is equivalent to standard graphical SLAM techniques. Thus, our method can be regarded as an extension to existing solutions of the SLAM problem.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[2] C. Chen and H. Wang. Large-scale loop-closing by fusing range data and aerial image. *Int. Journal of Robotics and Automation*, 22(2):160–169, 2007.

[3] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1998.

[4] M. Ding, K. Lyngbaek, and A. Zakhor. Automatic registration of aerial imagery with untextured 3d lidar models. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[5] C.U. Dogruer, K.A. Bugra, and M. Dolen. Global urban localization of outdoor mobile robots using satellite images. In *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, 2007.

[6] A. Doucet, N. de Freitas, and N. Gordan, editors. *Sequential Monte-Carlo Methods in Practice*. Springer Verlag, 2001.

[7] R. Eustice, H. Singh, and J.J. Leonard. Exactly sparse delayed-state filters. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2428–2435, 2005.

[8] J. D. Foley, A. Van Dam, K Feiner, J.F. Hughes, and Phillips R.L. *Introduction to Computer Graphics*. Addison-Wesley, 1993.

[9] U. Frese, P. Larsson, and T. Duckett. A multilevel relaxation algorithm for simultaneous localisation and mapping. *IEEE Transactions on Robotics*, 21(2):1–12, 2005.

[10] C. Früh and A. Zakhor. An automated method for large-scale, ground-based city model acquisition. *Int. Journal of Computer Vision*, 60:5–24, 2004.

[11] G. Grisetti, C. Stachniss, and W. Burgard. Improving grid-based SLAM with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2443–2448, 2005.

[12] J.-S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *Proc. of the IEEE Int. Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 1999.

[13] A. Howard. Multi-robot mapping using manifold representations. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 4198–4203, 2004.

[14] S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proc. of the American Control Conference*, pages 1628–1632, 1995.

[15] T. Korah and C. Rasmussen. Probabilistic contour extraction with model-switching for vehicle localization. In *IEEE Intelligent Vehicles Symposium*, pages 710–715, 2004.

[16] J.J. Leonard and H.F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(4):376–382, 1991.

[17] K.Y.K. Leung, C.M. Clark, and J.P. Huissoon. Localization in urban environments by matching ground level video images with an aerial image. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2008.

[18] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Journal of Autonomous Robots*, 4:333–349, 1997.

[19] M. Montemerlo, S. Thrun D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, pages 1151–1156, 2003.

[20] M. Montemerlo and S. Thrun. Large-scale robotic 3-d mapping of urban structures. In *Proc. of the Int. Symposium on Experimental Robotic(ISER)*, pages 141–150, 2004.

[21] E. Olson. *Robust and Efficient Robotic Mapping*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, June 2008.

[22] E. Olson, J. Leonard, and S. Teller. Fast iterative optimization of pose graphs with poor initial estimates. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2262–2269, 2006.

[23] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial realtionships in robotics. In I. Cox and G. Wilfong, editors, *Autonomous Robot Vehicles*, pages 167–193. Springer Verlag, 1990.

[24] B. Sofman, E. L. Ratliff, J.A. Bagnell, N Vandapel, and T. Stentz. Improving robot navigation through self-supervised online learning. In *Proc. of Robotics: Science and Systems (RSS)*, August 2006.

[25] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[26] S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous localization and mapping with sparse extended information filters. *Int. Journal of Robotics Research*, 23(7/8):693–716, 2004.

[27] J. Uhlmann. *Dynamic Map Building and Localization: New Theoretical Foundations*. PhD thesis, University of Oxford, 1995.