Augmenting and Reasoning in Semantically Enriched Maps using Open Data

Christian Landsiedel
Institute for Automatic Control Engineering (LSR)
Technische Universität München
christian.landsiedel@tum.de

Dirk Wollherr
Institute for Automatic Control Engineering (LSR)
Technische Universität München
dw@tum.de

***Abstract*—Complex robotic tasks require the use of knowledge that cannot be aquired with the sensor repertoire of a mobile, autonomous robot alone. For robots navigating in urban environments, geospatial open data repositories such as OpenStreetMap provide a source for auch knowledge. We propose the integration of a 3D metric environment representation with the semantic knowledge from such a data base, and describe an application where road network information from OpenStreetMap is used to improve road geometry information determined from laser data. This approach is evaluated on a challenging data set of the Munich inner city.**

***Index Terms*—spatial reasoning; hybrid maps; scene understanding**

I. Introduction

As tasks devolved to robots become ever more complex and encompass more domains, also demands towards their understanding of relationships and autonomy are growing. Different sources of knowledge that can be tapped for a higher-level understanding of concepts and tasks, which is desirable for a more intuitive and user-friendly interaction with a robot, have been explored. Human interaction partners themselves have been used as a knowledge source for example in the IURO project (Wollherr et al. [1], Figure 1). Other approaches have considered the augmentation of robot knowledge used ontological models in databases that can be shared for learning and usage by different robots [2, 3]. In this work, we consider OpenStreetMap, a community-driven online mapping framework, as a source for semantic information for robots moving autonomously in an urban environment. We propose the extension of a hybrid map, which includes a 3D occupancy grid as well as information about objects in the environment, with semantic and topological information from this data base.

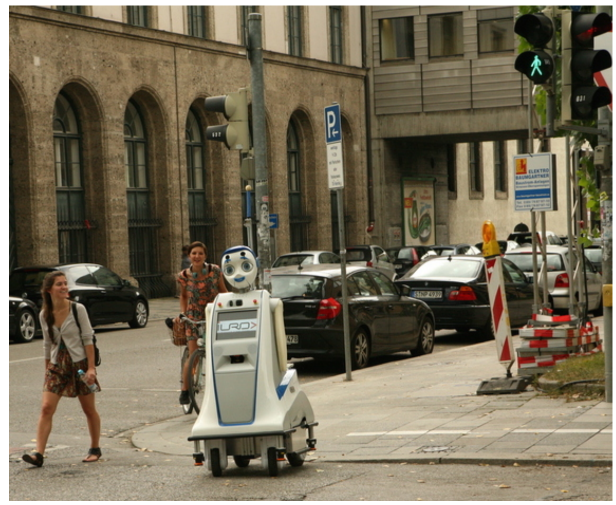


Fig. 1: The robot IURO [4] in an urban environment.

There are multiple reasons why a tighter integration between robot mapping frameworks with data repositories like OpenStreetMap is beneficial. For once, these repositories contain manually selected and curated information, which ensures that it is specified on a level that is understandable to humans and thus usable in interaction, for example for giving or receiving route instructions. Furthermore, even state-of the art scene understanding algorithms primarily rely on assigning labels on a per-pixel or per-region basis, and can have problems at determining distinctions between objects where this distinction happens primarily on a semantic level, i.e. two adjoining rooms with different functions in a space that is not clearly separated, or a building where different parts serve a different purpose. These will be hard to distinguish based on sensor data alone, but the information might be readily available as a bounding box in the OpenStreetMap annotation. Furthermore, the benefit of robots using open databases could be mutual. The sensor repertoire used in robot mapping approaches will provide up-to-date metric spatial information in the near future, which can be uploaded to Open Data repositories for sharing with humans and other robots.

This paper describes applications and possibilities offered by integrating 3D laser maps with rich semantic and geospatial Open Data repositories. As one such application scenario, it is described how road network information from OpenStreetMap can be used to improve understanding of street geometry based on 3D laser data. The approach is evaluated on a challenging data set covering an area in downtown Munich, showing a considerable increase in accuracy over the baseline.

II. Related Work

Data retrieved from OpenStreetMap, in particular the information about the topology and layout of the road network, has been used for multiple robotics and related applications. An important requirement for the use of geospatial data is knowledge about the location of the robot on a global map, i.e., a solution for the localization problem. Hentschel and Wagner [5] describe a localization method that uses building outlines from OpenStreetMap, which are matched to corresponding features in 3D laser scans. Additionally, the work covers route planning on the OSM route network, and robot behavior control for the robot car's lights based on semantic attributes from OSM. Brubaker et al. [6] use the OSM route network to localize based on visual odometry data. The localization problem is modeled as a dynamic network, where the state is a position related to the current route segment, and the visual slam trace is the input for filtering. Floros et al. [7] perform localisation on the OSM route network with visual slam and a GPS initial guess. The result of visual odometry is used as input to a particle filter, where the distribution is pruned based on comparison to the OSM road network.

Baatz et al. [8] use 3D building geometries from sources similar to OpenStreetMap and vanishing point detection to rectify and align training and query images for place recognition tasks. Li et al. [9] use 3D point clouds for global registration of images. The 3D point clouds are generated with Structure of Motion techniques from geotagged monocular images. Ruchti et al. [10] describe localization of a robot on the OpenStreetMap global map based on classification of 3D laser scan point clouds in road and non-road regions in a SLAM framework.

III. OpenStreetMap Data Model and Relevant Data

The data model of OpenStreetMap is a graph-like structure, where the basic building blocks are *nodes*, *ways* and *relations*. *Nodes* represent points on the map and are characterized by their latitude and longitude, as well as an optional elevation. *Ways* connect nodes to form open or closed paths and represent spatial entities like the path followed by railroad tracks, building outlines or the area covered by a football field. *Relations* describe higher-level characteristics of sets of nodes and ways, like all buildings belonging to an university campus, or the complete set of roads followed by a bus route. All instances of these three building blocks are identified by globally unique identifiers. Moreover, arbitrary tags can be applied to each instance of these data types, although there is an established set of tags and values that is largely adhered to, which can be used to automatically extract semantic information.

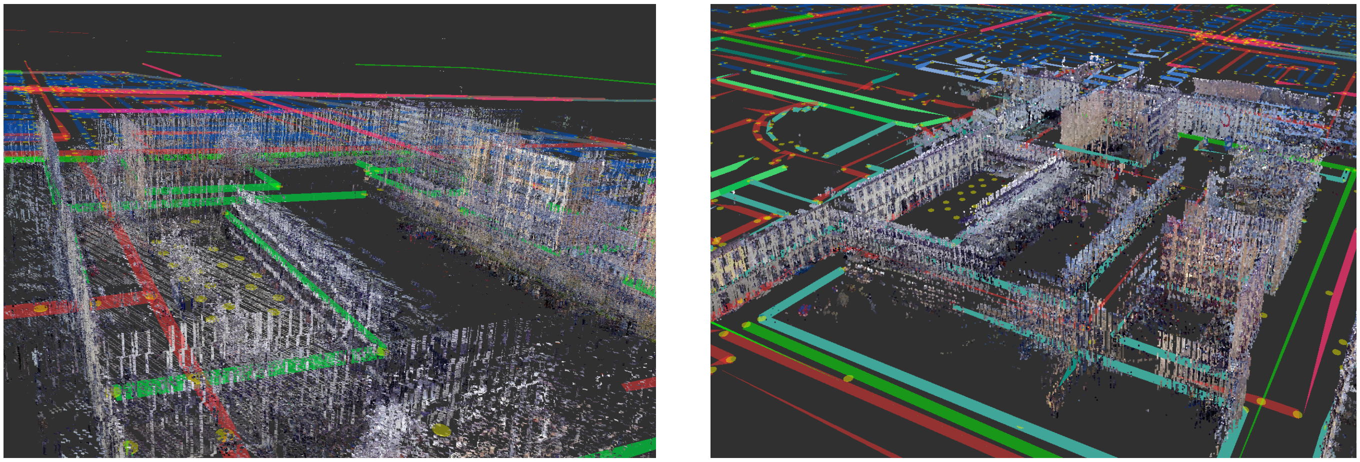
Many features from OpenStreetMap can be easily transferred to a metric map used for robot applications provided that the transformation between the different global coordinate systems are known. Different localization approaches to address this problem have been proposed as summarized in Section II, and this transformation is assumed as known for the purposes of the work presented here. In this case, the mapping of spatial locations allows the transfer of features between the two maps, for example for route planning based on street addresses in an occupancy grid derived from sensor data, or for identifying all buildings belonging to a particular ensemble in a 3D map, as exemplarily displayed in Figure 2.

IV. Street Width Estimation

The approach for street geometry information presented here is related to the work described by Ruchti et al. [10], where cells of a 3D laser-based map are classified point-by-point in order to enable localization of a robot in a road network like OpenStreetMap. In the presented work, additionally, the modelling imposes a strong geometric consistency constraint – road cells have to be adjacent and located in a strip around the street center. In keeping with most other urban scene segmentation approaches, the term 'street' is understood as the area of the road that is driven on, the sidewalk as well as possible parking space on the side of the road. Estimation of street width using information about the location of the street center from Open Data sources is also treated by Yuan and Cheriyadat [11], where aerial images are used as sensor data in combination with a road vector network, and by Chen et al. [12], where street geometry is inferred on the basis of high-resolution multispectral remote sensing satellite imagery. Geiger et al. [13] present an approach for urban scene understanding based on a generative model for street geometry and topographic information that is based on 3D data constructed from stereo vision recorded by a vehicle traveling on the road.

For this work, this topological information is provided largely from the road network from OpenStreetMap. The goal is to augment this graph with additional metric information in the form of street width, which is largely not existing in the OpenStreetMap database.

This relies on the road network data being available and sufficiently accurate. This is the case for the regions considered in the evaluation of this paper, and has also been found to suffice for the different purposes of the other works that use road network data, and perform evaluation on data from other parts of the world. However, street *width*, even though the infrastructure (an xml tag defined for the purpose of annotating it) exists, is not annotated



(a) 3D data set overlayed with OpenStreetMap road network

(b) Buildings on the TUM campus, extracted from OpenStreetMap building outlines

Fig. 2: Examples for combinations of 3D laser data with additional RGB information and information from OpenStreetMap. The visualizations of OpenStreetMap data in this paper are created with software based on the `open_street_map` ROS package [1].

often. In the data set used for evaluation in this paper, only one street segment is annotated with a width tag in the OpenStreetMap data base.

A. *Modelling Street Geometry Information*

The approach for estimating street width from 3D laser data followed here is largely a two-step process. Basis for the estimation is the road network from OpenStreetMap, which provides approximate road center lines subdivided into segments of varying length, within which the road is assumed to be straight. In order to reconcile this information with a metric 3D representation such as our laser map, two parameters need to be estimated for each road segment s_i: The vertical offset d_i of the actual road center from the vector connecting the waypoints $WP_{i,s}$ and $WP_{i,e}$ defining the road segment in the road network, and the width of the road around this actual center line. The directions of the road segments from OpenStreetMap are assumed to be in keeping with the actual topology of the environment. This model for the layout is displayed in Figure 4.

B. *Inferring Street Geometry from 3D Laser Data and Road Network Information*

In a first step, local geometric and appearance features are classified to a label set of road/non-road by a baseline classifier. For this task, a candidate environment of a predefined width around each segment center line from the road network is retrieved from the 3D map. This section of the map is then discretized in the ground plane, and for each resulting patch a set of local features is computed. The patch width and the maximum distance of a road side from the road network center line are chosen as 0.5 m and 15 m, respectively.

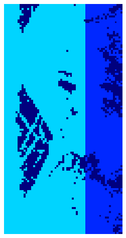
The feature set contains fairly standard geometric and appearance-based features, namely the mean, median, standard deviation and absolute range of the z-coordinates of all points projected to each patch, as well as the same statistics for the intensity values recorded for each point and the polar angle of a normal vector computed for a small neighborhood around each point. Using these features, a baseline classifier is trained to separate between road and non-road patches. We use a Support Vector Machine with a linear kernel and parameters estimated in a 5-fold cross validation scheme.

The classification results then provide candidate information for the second step, which introduces a strong global geometric constraint on the inferred route geometry, i.e. that is has straight parallel side lines. In addition, a prior distribution over road width is used. These constraints are formulated in a probabilistic fashion as a Bayesian network, where the probability of a pair of offset and width values, given the classifier estimates for each patch $Y_s = \{y_p\}$, can be written as

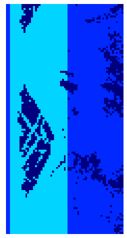
$$P(w, d|Y) \propto \prod_{p \in P} p(x_{w,d}(p)\ |y(p))p(w), \tag{1}$$

where the subscript s has been dropped. In this expression, $x_{w,d}(p)$ refers to a class assignment where road/nonroad labels have been assigned according to the road geometry determined by w and d for each patch p, and $y(p)$ is the

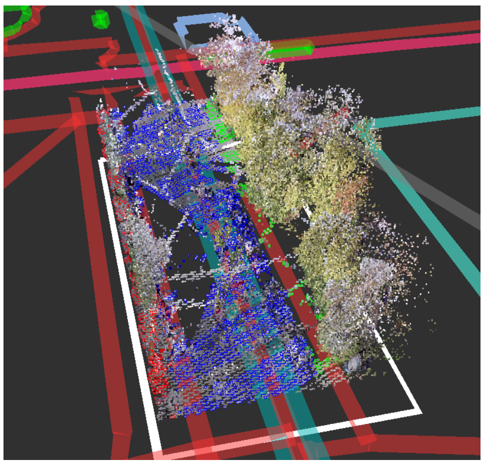
[1] authored by Jack O'Quinn, https://github.com/ros-geographic-info/open_street_map.



(a) Ground truth annotation



(b) Classification result



(c) Section of 3D map with overlayed ground truth annotation

Fig. 3: Example annotated road segment. For a) and b), light patches are road pathches, darker patches are non-road and for very dark patches, not features are available. For c), road and sidewalk points are overlaid in red, green and blue.

classifier output for the same patch. The distribution for the classification error $p(x_{w,d}(p)\,|y(p))$, is derived from the confusion matrix of the baseline classifier, such that each patch where the class assignment has to change between the classification result and the geometrically constrained solution incurs a penalty into the total probability of that constrained solution. The prior distribution of road widths is modeled as a Gaussian, the parameters of which are estimated from the same training data as the ones of the SVM classifier. For the offset between inferred street center and the road center from the road network, a non-informative prior is used.

Given this model, road offset and width are determined as the parameters which maximise the probability density function as

$$w^*, d^* = \underset{w,d}{\operatorname{argmax}}\, P(w, d|Y_s) \tag{2}$$

The complexity of this second inference step is quadratic in the number of bins allowed for discretization of the space around the route segment, and can be computed exhaustively for the parameters chosen for the evaluation, although a dynamic programming approach could also be followed if fast performance is required.

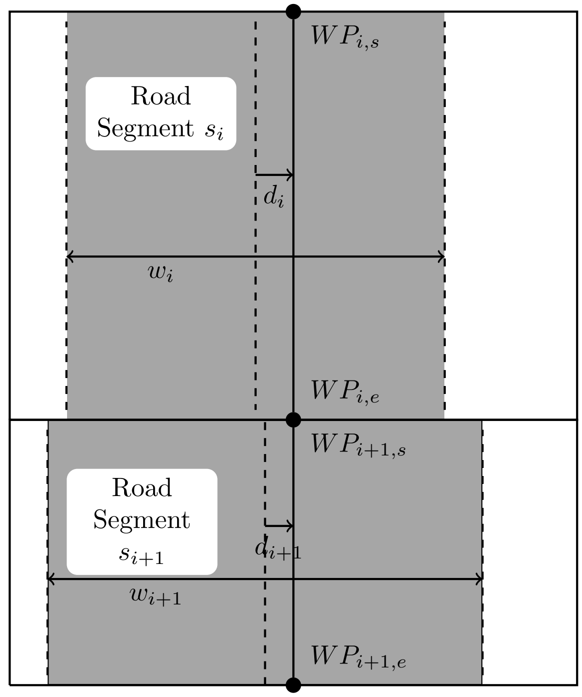


Fig. 4: Illustration of road position and width model

V. Experimental Evaluation

A. Munich Urban 3D Data Set

The data set that was used for experiments is in part overlapping with the one described by Wollherr et al. [1]. It consists of 80 high-resolution laser range finder scans in 3D, acquired with an Z+F 5010C laser range finder, of an area in downtown Munich around the university campus. Additionally, laser intensity and RGB channels are recorded. In this data set, object instances are manually segmented and annotated for object classes such as *building*, *street*, *sidewalk* or *car*, as well as for qualitative spatial relations, such as *left of* or *behind*, between objects. The data set provides a challenging environment for scene understanding tasks, since it incorporates a considerable range of different environments, such as residential streets with parked and artefacts of moving cars, tunnels, and cobbled or gravelled streets closed for motor vehicles. Additionally, the laser scans are taken from positions on the sidewalk, such that in many cases the ground plane is not observable because of occlusions or dynamic objects blocking visibility at the time of registering the laser scan.

	Precision	Recall	F_1-Score	width RMSE
Baseline (SVM)	0.82	0.82	0.82	14.33
SVM+Geometry	0.91	0.87	0.87	5.23

TABLE I: Database recall metrics and root mean square error of the estimated road widths for the baseline classifier and for the solution including geometric constraints

B. Registration of Point Cloud Data with OpenStreetMap

Since the data set is recorded sequentially with no ground truth information about the absolute robot position at the time of recording a scan, nor about the relative movement of the sensor between scans, a registration step is necessary to obtain a complete 3D representation of the area covered by the union of the different laser scans. To this end of estimating the transformations between the sensor positions for each recorded 3D scan, registration with multiple iterations of the 3D Iterative Closest Point algorithm as introduced by Besl and McKay [14], with the maximum allowed correspondence distance decreasing with each iteration was carried out beginning from a rough initial guess. Also for the lack of ground truth position data of the laser data, manual alignment of the 3D data with an export of OpenStreetMap data for the region covered by the laser data was carried out.

Since handling the complete data for the combination of all laser scans is intractable, the data was filtered and downsampled using the RMAP algorithm introduced by Khan et al. [15]. This procedure produces a denoised occupancy grid at a variable resolution, where the grid size was chosen to 0.1 m for the experiments in this paper.

For the evaluation of the street geometry estimation, a set of 34 route segments with a total length of about 1 km annotated in OpenStreetMap was chosen, and the regions surrounding them were manually segmented and annotated as road and non-road regions for the training of the classifiers and for evaluation of the results. An example route segment is illustrated in Figure 3.

C. Experiments

The method for road geometry estimation described above was evaluated on this augmented Munch 3D Urban Data Set. A summary of the results in terms of per-patch retrieval of the correct labels, expressed as precision, recall and F_1 score, as well as the root mean square error of the estimated road widths, is given in Table I.

It can be seen that introducing the geometric constraint improves retrieval metrics of labels for individual regions, as well as it also decreases the error in the street width estimation.

An analysis of the failure modes on particular segments where street geometry estimates were unsatisfactory revealed that many of them were located in tunnels, which were underrepresented in the training set as to learn proper classification for these environments, where the 3D layout of the environment is rather different from a generic urban road scene. Additionally, the data set also contains streets of different categories (i.e., residential urban roads as well as gravel roads closed for general traffic and without sidewalks), which again are quite different in nature from a generic scene. In order to further improve the geometry estimation results, more qualitative information from OpenStreetMap could be used, for example by building and employing different models for roads of different categories, or road segments that are annotated as tunnels.

VI. Conclusion

In this paper, we have argued the benefits of including information from open geospatial repositories in hybrid maps. The application of road classification and road width estimation, a parameter which is often missing in OpenStreetMap and could be added automatically from 3D maps, has shown that including a geometric constraint based on OpenStreetMap data provides a considerable improvement over a baseline solution based on classification alone. Experiments have been carried out on a challenging data set, where laser scans have been recorded from the sidewalk, so that the full width of the road is often occluded, and which contains a widely varying array of road types including tunnels. With the increase in mobile robot platforms navigating in urban scenarios that are equipped with a 3D laser scanners, it is to be expected that different avenues for use of additional information will be explored.

There are several directions in which the work presented here can be extended. Especially in the vein of improving urban scene interpretation by using mapping data from OpenStreetMap would be the use of information about additional properties of roads such as traversability and the existence of bike paths and sidewalks. Furthermore, it can be expected that knowledge about the type of street from the annotation as *residential*, *primary*, *secondary* etc. will be useful if separate models are built and conditioned on the different types of environment.

Acknowledgements

The authors would like to acknowledge Mustafa Sezer, who worked on registration of the individual laser scans to a joint point cloud, Sheraz Khan for providing assistance with the handling of point clouds and the RMAP library, and Christopher Bayer and Roderick de Nijs for collecting a data set of road information and providing road width prior probability distributions.

References

[1] D. Wollherr, S. Khan, C. Landsiedel, and M. Buss, “The interactive urban robot IURO: Towards robot action in human environments,” in *Int. Symposium on Experimental Robotics*, 2014.

[2] M. Tenorth and M. Beetz, “Knowrob: A knowledge processing infrastructure for cognition-enabled robots,” *Int. J. Rob. Res.*, vol. 32, no. 5, pp. 566–590, Apr. 2013.

[3] L. Riazuelo, M. Tenorth, D. Di Marco, M. Salas, D. Gálvez-López, L. Mosenlechner, L. Kunze, M. Beetz, J. D. Tardos, L. Montano *et al.*, “Roboearth semantic mapping: A cloud enabled knowledge-based approach,” *IEEE Trans. on Automation Science and Engineering*, vol. 12, no. 2, pp. 432–443, 2015.

[4] M. Buss, D. Carton, S. Khan, B. Kühnlenz, K. Kühnlenz, R. de Nijs, A. Turnwald, and D. Wollherr, “IURO – Soziale Mensch-Roboter-Interaktion in den Straßen von München,” *at – Automatisierungstechnik*, 2015.

[5] M. Hentschel and B. Wagner, “Autonomous robot navigation based on OpenStreetMap geodata,” in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*. IEEE, 2010, pp. 1645–1650.

[6] M. A. Brubaker, A. Geiger, and R. Urtasun, “Lost! Leveraging the crowd for probabilistic visual self-localization,” in *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 3057–3064.

[7] G. Floros, B. van der Zander, and B. Leibe, “OpenStreetSLAM: Global vehicle localization using OpenStreetMaps,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation*. IEEE, 2013, pp. 1054–1059.

[8] G. Baatz, K. Köser, D. Chen, R. Grzeszczuk, and M. Pollefeys, “Leveraging 3D city models for rotation invariant place-of-interest recognition,” *Int. Journal of Computer Vision*, vol. 96, no. 3, pp. 315–334, 2012.

[9] Y. Li, N. Snavely, D. Huttenlocher, and P. Fua, “Worldwide pose estimation using 3D point clouds,” in *European Conf. Computer Vision*. Springer, 2012, pp. 15–29.

[10] P. Ruchti, B. Steder, M. Ruhnke, and W. Burgard, “Localization on OpenStreetMap data using a 3D laser scanner,” in *Proc. of the IEEE Int. Conf. on Robotics & Automation*, Seattle, Washington, USA, May 2015.

[11] J. Yuan and A. M. Cheriyadat, “Road segmentation in aerial images by exploiting road vector data,” in *Proc. of the IEEE Int. Conf. on Computing for Geospatial Research and Application*. IEEE, 2013, pp. 16–23.

[12] B. Chen, W. Sun, and A. Vodacek, “Improving image-based characterization of road junctions, widths, and connectivity by leveraging OpenStreetMap vector map,” in *Proc. of the IEE Int. Geoscience and Remote Sensing Symposium*, July 2014, pp. 4958–4961.

[13] A. Geiger, M. Lauer, and R. Urtasun, “A generative model for 3D urban scene understanding from movable platforms,” in *IEEE Conf. Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 1945–1952.

[14] P. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[15] S. Khan, A. Dometios, C. Verginis, C. Tzafestas, D. Wollherr, and M. Buss, “RMAP: a rectangular cuboid approximation framework for 3D environment mapping,” *Autonomous Robots*, pp. 1–17, 2014.