Semantic Mapping for Mobile Outdoor Robots

Dagmar Lang, Susanne Friedmann, Jens Hedrich and Dietrich Paulus
Active Vision Group, University of Koblenz-Landau
56070 Koblenz, Germany
{dagmarlang, scfriedmann, jenshedrich, paulus}@uni-koblenz.de

Abstract

In this paper we present the concept and realization of a semantic mapping system for a mobile outdoor robot. Semantic maps aim to give robots the ability to gather semantic information about their environment, to store it, represent it for the user, and to perform high-level tasks based on the semantic information. The map is build by a system integrating the combination of object classification and common-sense knowledge. We validate the proposed semantic map representation on a real-world 3D point cloud dataset. The presented classification approach achieves an overall precision about 96 %. The semantic maps result into a data structure which offers the opportunity to solve complex task settings and can be integrated onto real robotic systems.

1 Motivation

In order to allow robots to solve high-level autonomous tasks, semantic mapping has been established by combining mapping, object classification, and common-sense knowledge into one map representation. This map representation can be used for example for autonomous driving or service robotics to optimize the search for objects such as traffic signs. With such a task, the robot normally has to search in a brute force manner everywhere in the map for the traffic sign. In contrast, if the robot has common-sense knowledge about the construction of road networks including the probability of objects in a certain place, it will search for traffic signs at road intersections first and can guide the search to places with lower probabilities afterwards.

Considering the creation of an outdoor map, cars are dynamic objects and sensor data, corresponding to cars, will be inserted into the map. This means that one obstacle will be inserted into the map, which is maybe not present during a second visit of this position. When using this map for path planning, paths can be blocked by the obstacle but at another time the path is passable and the planned path is much shorter.

Maps created from laser range data are represented as a mixture of metrical and topological data structures and are often modeled as point clouds. This representation does not take into account information about the objects in the map and their properties. A lot of methods have been established for object classification or place recognition in camera images, point clouds, or fused data for indoor and outdoor scenes. Often the results are represented as point clouds or images colored according to the classification result and doesn't take into account the properties of objects. In order to achieve this, the relationship between the perceived environment in the sensor data and the common-sense knowledge has to be established. A lot of methods are available for indoor environments, but for outdoor environments, there are only approaches available which establish the connection between object classification and common-sense knowledge.

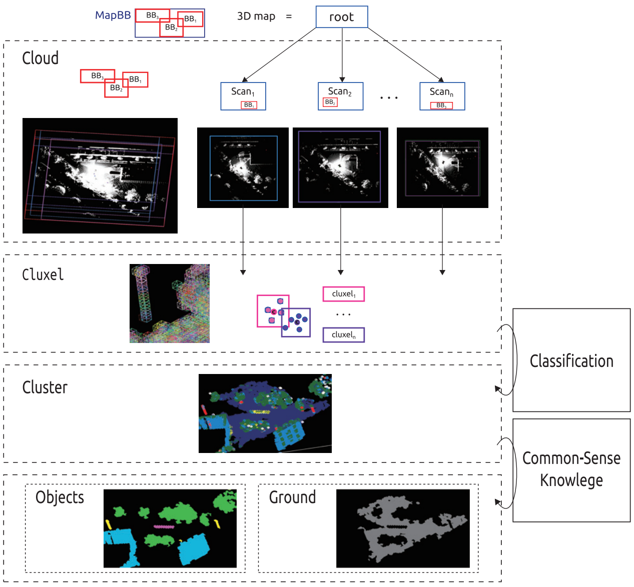

Figure 1: Overview of the data structures and methods to construct the semantic outdoor map. The data structures are highlighted with dotted lines and the main applied algorithms with solid lines.

In this paper we present a laser range data based realization of a semantic mapping approach for mobile outdoor robots based on the introduced map representation of Lang et al. [6] which satisfies the formal definition of semantic maps of Lang and Paulus [5].

A short overview on semantic indoor maps and object and place classification algorithms for outdoor environments is given in Section 2. We present the major contribution, the design and practical implementation of the semantic mapping system in Section 3 and the corresponding evaluation in Section 4. Finally, we draw our conclusion about this work in Section 5.

2 Related Work

For outdoor environments object classification and place recognition algorithms have been developed in order to create maps containing semantic information of the environment. Behley et al. [1] present a performance comparison of three different classification methods, ranging from a simple linear model to a more complex one based on probabilistic graphical models, and different features for point based classification of 3D point clouds. A multi-scale inference procedure

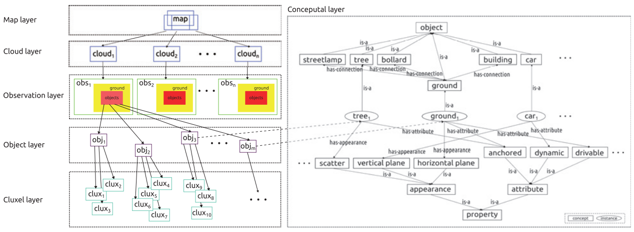

Figure 2: The overview of the semantic map representation is shown in this figure. In the left column the spatial hierarchy is described from coarse to fine (top to bottom) and on the right the conceptual hierarchy with the conceptual layer (note: the commonsense knowledge represented in the conceptual layer is only an example).

with a graphical model to capture the contextual relationship among 3D points by training point cloud statistics and to learn relational information over fine and coarse scales for different outdoor scenes was described by Xiong et al. [8]. Hu et al. [3] present an efficient and simple representation of the scene in conjunction with recent developments in structured prediction in order to obtain an efficient classification based on state-of-the-art methods. A 3D semantic outdoor mapping system with multi-label and resolution octree maps was presented by Lang et al. [4]. In order to classify large and small objects in the scene the classification is based on conditional random fields with a varying size of feature support regions and a scale invariant feature descriptor.

Lang and Paulus [5] present a formal definition of semantic maps. For indoor environments, different map representations have been established in order to create indoor semantic maps which satisfy the definition. Zender et al. [9] present a multi-layered spatial representation consisting of a metrical, navigational, topological and conceptual map. The first three layers are used for mapping and the last layer for reasoning based on an OWL-DL ontology and a description-logic reasoner. A large-scale semantic mapping approach for autonomous indoor service robots was presented by Pronobis and Jensfelt [7]. The semantic mapping system is divided into four layers on three hierarchy levels. The first hierarchy consists of the sensory layer, the second of the place and categorical layer, and the third one of the conceptual layer. The inference between the conceptual and the other layers is realized by a chain graph model. Lang et al. [6] present a semantic mapping system for outdoor mobile robots based on a spatial and conceptual hierarchy. The spatial hierarchy consists of five layers from coarse to fine representations: map, cloud, observation, object, and cluxel layer. The conceptual hierarchy consists of a conceptual layer. Based on the representation the authors presented a concept in order to create the semantic maps.

3 Outdoor Semantic Mapping System

The formal definition of semantic maps was presented by Lang and Paulus [5]. Based on that definition Lang et al. [6] propose a semantic map design for outdoor robotic tasks. This section presents the realization and evaluation of the system.

The system described in [6] is based on two parts: the construction and the representation of the semantic map. The pipeline for the construction is presented in Figure 1 and the representation in Figure 2. Since the design decisions and the relationship between the construction of the map and the map representation are already presented in [6], we will only describe the construction of the map and refer to the map representation for clarification.

Lang et al. [6] assume that mapping algorithms provide large-scale globally consistent 3D maps, which consist of point clouds and transformations in order to convert all clouds into one reference coordinate system. The calculation of the map is out of the scope of this paper, since there are a lot of different solutions available to create such 3D maps. The point clouds are treated as a set of point clouds, without the assumption that the clouds are consecutive clouds. This assumption allows to simulate the mapping process, loop closings, merging of two parts of semantic maps after loop closing, and the creation of a generic semantic mapping approach. Which satisfies that after testing and optimizing, the system can be used on a real operating robot.

The map is represented as graph and the root node (see Figure 1 and map layer in Figure 2), connects all point clouds available for semantic mapping and satisfies the condition that all clouds are in the same coordinate system. Each cloud is represented by a bounding volume in order to determine overlapping cloud parts (see cloud layer in Figure 2).

3.1 Construction of a Semantic Map

When the semantic mapping process starts the system gets the first point cloud. This cloud is then divided into smaller parts since classification methods applied to each 3D point have the drawback of long run times and a high memory consumption.

There are different methods available to represent 3D maps and subsample 3D data. Commonly used tools are 3D grids such as voxelgrids or 3D octrees. 3D grids and octrees are rigid and do not take into account the boundaries of objects.

The individual point cloud will be successively subdivided into so called *cluxels*, which is a hierarchy of bounding volumes. Cluxels are a composition of properties of clusters and voxels, since they are representing object boundaries more reliable, store 3D points, support fast neighbor, and cluster calculation.

The main goal of cluxels is to support a reliable point cloud representation and offer new opportunities to create object clusters based on the choice of the classification method and feature selection.

In order to create cluxels one point of the 3D point cloud is taken into account. The neighbors of that point are calculated based on a *kd* tree search within a predefined search radius. Then a cluxel is defined as a bounding volume around that point and its neighboring points. These points are marked as used similar to the calculation of the DBSCAN clustering [2]. This procedure will be repeated since there are no unused points in the cloud. Based on the calculation of cluxels, the bounding volume of neighboring cluxels can overlap each other, as shown in Figure 1. In the map representation the cluxels are stored in the cluxel layer (see Figure 2).

3.1.1 Semantic Classification

After the subdivision of the point cloud into cluxels, we want to classify them into *horizontal plane*, *vertical plane*, *column*, *small structure* and *scatter* labels as primitives. For classification we apply a supervised pairwise conditional random field (CRF) based on the description of Lang et al. [4]. The graph of the CRF is created by defining the Markov blanket based on intersecting cluxels.

As feature we calculate the histogram of oriented residuals (HOR) operator as suggested by [4], since the operator is scale invariant an can deal with different spatial scales of feature support regions. The feature support region for one cluxel $\mathbf{c}_i$ is obtained by calculating the intersecting cluxels $\mathbf{c}_{\mathcal{N}_j}$, $j = 1, \ldots, n$, to $\mathbf{c}_i$. Then, the next intersecting cluxels $\mathbf{c}_{\mathcal{N}_k}$, $k = 1, \ldots, m$ to each $\mathbf{c}_{\mathcal{N}_j}$ is calculated. In case of dense covered cluxels 3D points belonging to the neighbors and the neighbors of the neighbors are used to calculate the HOR operator. In order to calculate the HOR operator a center point $\mathbf{p}_s$ has to be defined. In our case, we use the center of $\mathbf{c}_i$ to define $\mathbf{p}_s$, which leads to the effect that $\mathbf{p}_s$ is not necessarily centered in relation to the 3D points taken into account for the HOR calculation. In addition, the coordinate system spanned for the feature calculation is also not centred in relation to $\mathbf{p}_s$. Based on that fact, the feature support region for each cluxel has different scales, which can be overcome since the HOR operator is calculated in a scale invariant manner. In this case, the calculation of the HOR operator and the calculation of the pairwise term of the CRF suffers from the effect of overlapping feature support regions.

As result of the classification we obtain cluxels with the labels *horizontal plane*, *vertical plane*, *column*, *small structure* and *scatter*. The main idea behind the clustering, and later when we apply the common-sense knowledge about outdoor scenes, is to create objects based on this primitive classes. In order to cluster cluxels with the same label we take one cluxel and calculate the intersecting cluxels and merge them, if they have the same label. This procedure is done since there are no cluxels left which can be merged. As results we obtain a hierarchy of cluxels called clusters.

3.1.2 Integration of Common-Sense Knowledge

The conceptual layer of the map representation (see Figure 2) models the common-sense knowledge. We now use the common-sense knowledge to create decision trees for each object instance by modeling the concepts of *attributes* and *appearance* of that objects in contrast to other objects and primitives.

In order to obtain the objects we first take into account object clusters with label *horizontal plane* in order to determine the *ground* with the decision tree. Based on the object cluster *ground* we calculate a plane and assume that all objects are arising from that plane. This assumption allows to separate objects such as trees which are close to each other.

Based on the example of a tree we want to obtain the *tree trunk* and *tree top* in order to merge the results into a label *tree*. The classification leads to reliable results for the primitives *column* and *scatter*, but there are also missclassifications at the border of the *tree top* as *building*. For a tree we first examine if the plane intersects with a *column* cluster and than we take into account other clusters above, intersecting with the *column* cluster. In case of a tree the *column* cluster should intersect with a *scatter* cluster. If the *scatter* cluster satisfies the knowledge represented in the decision tree in combination with the *column* cluster we combine them into one object with labels *tree top* and *tree trunk*. If there are misclassified clusters in the tree and the knowledge for the tree between *column* and *scatter* can be satisfied the misclassified cluster will be assigned to the most suitable object.

For other objects we use the same approach searching from the bottom to the top for intersecting clusters by incorporating common-sense knowledge defined in the decision trees. As results we obtain a set of bounding volumes called objects with the labels *ground*, *building*, *tree*, *small object*, *bush*, *street lamp* and *street sign*. This objects are depicted in the observation and object layer of the map representation (see Figure 2).

3.2 Extension of the Semantic Map

Since Section 3.1 describes only the construction of the semantic map based on one scan, in this subsection we will present the extension of the semantic map, when new scans arrive or after loop closing. Both cases include the merging of two semantic maps into one follow the approach described in the following. The results of the semantic mapping including the extension of the map are shown in Figure 3.

First, overlapping regions of the semantic map and the new scan will be determined. If the bounding volume of an object encloses parts of the cloud this points will be assigned to the object. If there are unclassified points after testing, this points will be subdivided into cluxels and this cluxels will be classified into clusters as described in Section 3.1.1. In order to take all information of the point cloud into account the graph and features will be calculated based on the cluxels belonging to the neighboring objects. In the next step, clusters will be calculated for the cluxels of the neighboring objects and the new cluxels. The cluster will be merged into objects as described in Section 3.1.2. The objects are represented in the object layer and their locations are stored in the observation layer (see Figure 2).

4 Experimental Results

In this section we describe the evaluation of the presented system. We test our semantic mapping approach on the Freiburg dataset.[1] The dataset was captured using a wheeled robot equipped with a SICK LMS laser range finder mounted on a pan-tilt unit and consists of 77 3D scans capturing an area of $292\,\mathrm{m} \times 167\,\mathrm{m} \times 28\,\mathrm{m}$. Each $360°$ scan was acquired in

[1] http://ais.informatik.uni-freiburg.de/projects/datasets/fr360/

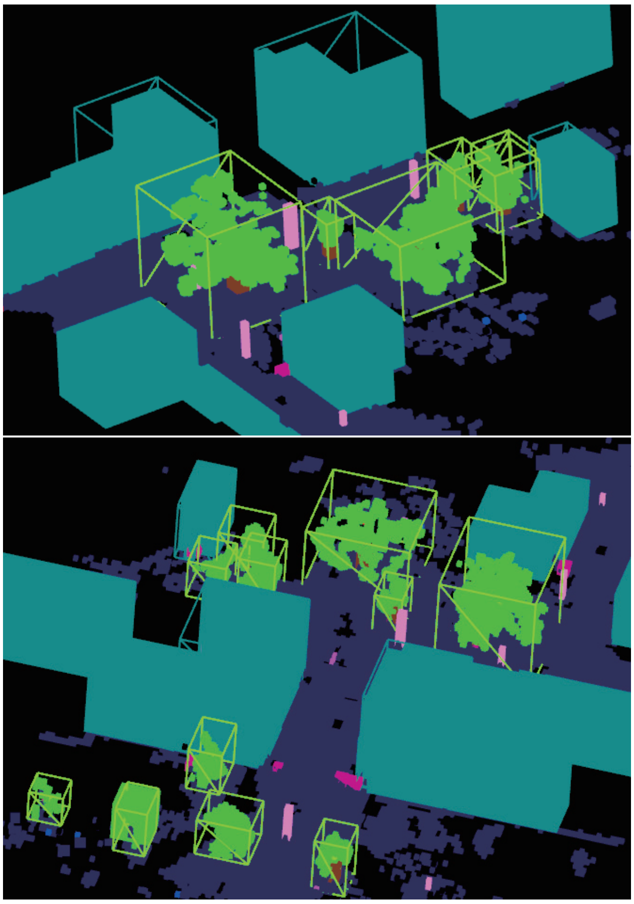

Figure 3: Semantic mapping results for the Freiburg dataset. In the left image the result after inserting two point clouds into the map is shown and in the right image after inserting the next one (note: the perspective change for better viewing). In the images the object bounding volumes are highlighted and the cluxels of the objects are colored according to the object label. The color-coding (best viewed in color: light blue = *building*, dark blue = *ground*, light green = *tree* respectively *tree top*, brown = *tree trunk*, light pink = *street lamp*, dark pink = *small object*).

a stop-and-go fashion and consists of 150,000-200,000 points.

The pairwise CRF will be trained with a set of hand labeled ground truth data. In order to calculate cluxels the search radius for the nearest neighbor search is set to 10 cm. The results of the CRF based classification is out of the scope of this evaluation since equivalent results for this dataset are presented in Lang et al. [4].

In this paper we want to evaluate how the classification results after the integration of the common-sense knowledge perform based on the whole dataset. The creation of the ground truth is based on hand labeled 3D point clouds with the labels *ground*, *building*, *tree*, *small object*, *bush*, *street lamp* and *street sign*. For evaluation we perform the cluster calculation described in Section 3.1.1 in order to create object bounding volumes for the whole map. The confusion matrix is calculated by counting if the bounding volumes of the classified map and the ground truth overlap by more than 80 %.

We calculated a confusion matrix and will present the results in the following text. The objects with the label *ground*, *building* and *bush* were all classified correct. For objects with the label *street lamp* 3 % of the objects were misclassified as *small object* since the merging from the bottom to the top failed after a few iterations. 75 % of the label *street sign* were misclassified as *tree*, since a *scatter* cluster occurred at the top of all clusters belonging to the object. Furthermore, only 8 % objects belonging to the label *small object* were misclassified as *bush*. For the label *tree* two trees were misclassified as one. But for example in the left image of Figure 3 the trees on the right are very close to each other and can be separated with the presented algorithm. In the evaluation we reach an overall precision of 96 %. It must be mentioned, that the results for the objects depend mainly on the results of the cluxel classification.

5 Conclusion and Future Work

In this paper we presented the realization of the semantic mapping system of Lang et al. [6]. This system allows to create semantic maps based on the integration of classification and common-sense knowledge for outdoor environments and offers the opportunity to simulate the mapping process, loop closings, and the merging of two semantic maps. In order to extend this proof of the design of the semantic map, for common-sense knowledge representation statistical methods such as Markov logical networks, which can be learned, can be integrated. Furthermore, the system can be adapted to an online mapping system which can deal with changing transformations between the point clouds in the cloud layer.

Acknowledgements

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG) under research contract PA 599/11-1.

References

[1] Behley et al.: *Learning to Hash Logistic Regression for Fast 3D Scan Point Classification*, Proc. of IROS, 2010

[2] Ester et al.: *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proc. of KDD, 1996

[3] Hu et al.: *Efficient 3-D Scene Analysis from Streaming Data*, Proc. of ICRA, 2013

[4] Lang et al.: *Semantic 3D Octree Maps based on Conditional Random Fields*, Proc. of MVA, 2013

[5] Lang and Paulus: *Semantic Maps for Robotics*, Proc. of ICRA, "Workshop on AI Robotics", 2014

[6] Lang et al.: *Definition of Semantic Maps for Outdoor Robotic Tasks*, Proc. of ROBIO, 2014

[7] Pronobis and Jensfelt: *Large-scale Semantic Mapping and Reasoning with Heterogeneous Modalities*, Proc. of ICRA, 2012

[8] Xiong et al.: *3-D Scene Analysis via Sequenced Predictions over Points and Regions*, Proc. of ICRA, 2011

[9] Zender et al.: *Conceptual Spatial Representations for Indoor Mobile Robots*, RAS, 56(6), pp. 493-502, 2008