Semantics in Human Localization and Mapping

Aidos Sarsembayev, Antonio Sgorbissa

University of Genova, Dept. DIBRIS
Via Opera Pia 13, 16145 Genova, Italy
aidos.sarsembayev@edu.unige.it , antonio.sgorbissa@unige.it

Abstract. In this paper, we present an approach to human navigation and mapping using semantic techniques. The main idea is to enhance a map produced by a mapping algorithm with additional contextual information. The key advantage of our approach is the presence of a human in the loop, that can provide reliable semantic information. Several simulations of the proposed algorithm were conducted using a simulation framework developed particularly for this research.

1 Introduction

Human navigation attracts some attention over the last decade. A number of works addressed this problem by applying various techniques from different research fields such as Mobile Robotics and Artificial Intelligence. In this work we focus on the application of semantic techniques in human navigation and mapping.

A number of works were using semantic techniques in various forms and have achieved significant results. For instance, the authors of [1],[2],[3] and [4] propose to use indoor features, such as walls, stairs, doors and door-signs as landmarks of the environment for building the maps.

The above mentioned works focus on mobile robots mapping and navigation, whereas [5,6,7,8] and [9] present a semantic mapping approach for humans, which is more relevant to our case. In [5], the mapping system builds a map by classifying the indoor environment into places and transitions and transitions between places, after processing the images acquired from a catadioptric camera worn by a human. The authors of [6] and [7] propose a human-centered navigation system which uses an ontology repository and the human profile in order to perform navigation based on the user's physical and perceptual/cognitive characteristics. In [8] a semantic approach to SLAM leveraging the Android smartphones' sensors is presented. The system detects nearby landmarks in order to reset the dead reckoning errors. The PlaceSLAM, presented in [9], extends the well-known odometry based SLAM algorithm - FootSLAM. This approach increases the accuracy of FootSLAM by adding additional contextual information to a map. This information is acquired by prompting a user to describe the place that she sees.

In this paper, we propose an approach which tends to solve the correspondence (also known as loop-closing) problem in mapping process applying semantic techniques. Our approach uses graph structure as representation of places and spatial relations between them. By these means, it can be identified as topological mapping with the use of semantic techniques. The key advantage of our approach is the presence of a human in the loop, that can provide semantic information in reliable way. In order to test our approach, a naive algorithm for topological map-building has been adopted: this allows us to measure the impact of semantic information on loop closure even in absence of complex SLAM optimization techniques.

2 Semantic Mapping

Semantic Mapping In our approach we model a world as a graph G. This graph contains nodes which represent places of the environment that the agent explores. More formally $G = (N, E)$, where N denotes a set of nodes $N = \{ n_1, n_2, n_3 \dots n_i \}$, and E is a set of directed edges that signify a path traveled by an agent and a linkage between semantic objects on a map. Each node n_i carries supplementary information about its position X_i in a Cartesian coordinate system, and a set of labels L_i denoting objects or environmental features discovered in this place. For instance, the label may include the typology of relevant buildings that can be found in the corresponding place (e.g. *Church, Hospital, Shop*). As the reader can imagine, one purpose of such structure is that the information labeling nodes might be used to minimize the possible errors in localization of a person on a map, by recognition of already visited or known objects, buildings and other environmental features.

Knowledge base The presence of a human in the loop cannot guarantee an absolute precision in semantic information. In different time steps, the human may describe the same place at different levels of abstraction, she may use synonyms for the same environmental feature or even another feature. Some examples are "*shop*", "*bakery*", "*traffic lights*" and so on. At the first visit, the human may detect a "*shop*", at the second visit to the same place she may recognize a "*bakery*" (subclass of *shop*), at the third visit a "*traffic lights*" (a different feature in the same place). To handle ambiguities and hierarchical relationships between semantic objects, we suggest to use ontology techniques, modeling an environment as a hierarchy of environmental features connected to each other depending on the topology of the environment. We use the structure of the ontology to define a metrics that measures *"how far"* two labels are in a semantic sense comparing the classes that these labels belong to. The rationale is that the farther two classes are in the ontology tree, the less likely it is that the two labels are meant to represent the same environmental features in the real-world.

Localization algorithm The approach proposed in this article may be adapted to different starte-of-the-art SLAM approach: however, for sake of simplicity, here we suggest an algorithm that is based on Markov Localization for detecting the most probable node in which the person is located, and simply add a new node to the map whenever the probability of being in a pre-existing node is below a given threshold. As usual in Markov Localization we estimate the pose of the agent in two steps: action prediction and observation update.

In the first step, the probability distribution over the nodes at time t is updated by applying the control A_T starting from each node N_{t-1}. Notice that, as the agent moves in the real world, the path is updated using dead reckoning; however, the probability distribution is updated only when the agent thinks to have reached a new node, i.e., the agent recognizes environmental features that are worth being recorded and tells the system what she can see in its current location.

$$P(N_t = n_i | O_{0:T}, A_{1:T}) = \sum_{N_{t-1}} P(N_t = n_i | N_{t-1}, A_T) \times P(N_{t-1} | O_{0:T-1}, A_{1:T-1}) \tag{1}$$

The second step is an update phase, when we obtain posterior probability (observation update vector), incorporating measurement acquired by the agent according to the observation model. At this stage, the probability of the observation at time t does not depend on previously acquired observations. We use our ontology as the reference for observation model.

$$P(N_t = n_i | O_{0:T}, A_{1:T}) = \alpha' P(O_T | N_t = n_i) \times P(N_t = n_i | O_{0:T-1}, A_{1:T}) \tag{2}$$

From a practical point of view, if the pre-existing and observed labels belong to the same class in the ontology (e.g., *Shop-Shop*), the probability of the observation is equal to 1.0. If the two labels have a direct parent-child relationship (e.g., *Supermarket-Shop*), the probability of the observation is 0.5. If the two labels have a sibling relationship (e.g., *Supermarket-Bakery*) the probability of the observation is 0.25. In those cases when the two labels do not have any relationship the probability is relatively small value that leaves a minimum probability of observation, which we set to 0.05.

Notice also that whenever a new label is observed in a node n_i, the new label is added to the set L_i, thus increasing the probability of a matching during future observations.

3 Simulation

In order to assess the difference between using and not using semantic information during the mapping process, we have performed over 140 experimental simulations in our specialized simulation framework. We have created 18 artificial maps simulating a real world. These maps may be classified according to their difficulty level from simple to complex on the basis of their topology. We consider those maps containing many places situated close to each other as complex. Vice versa, simple map's nodes are rather distant from each other.

The agent explores the environment moving in a Manhattan-like way, horizontally and vertically. We assume that the agent has sensors for measuring odometry, which return its heading and path with a given error. The errors in measurements are generated randomly, basing on the angle deviation (e.g. from -3.0 to 3.0 degrees). We repeated the experiments with different error distributions, with a 0 mean and a standard deviation ranging from 1 to 7. The error is accumulated during the entire process, thus making localization more complicated. Once the agent reaches a place, it updates the system with a label describing it. We assume that the user should provide the system with verbal information through a microphone, however for the simulation process we have implemented an algorithm which randomly generates the labels from a prepared list that contains the labels for all classes that we have in our knowledge base. Figure 1 shows that our algorithm recognizes previously visited places (orange circle) also in those cases when a node has multiple environmental features belonging to the different classes. In the current example, the node 4 contained a set of labels *Pizzeria, Pizzeria, Clinic*. The new label *Pizzeria* is equal to at least one label in the set, thus the observation probability in this case is equal to 1.0.

In order to evaluate the performance of the map building algorithm in presence of semantic information we use the number of created nodes. We assume that, if all loops are correctly closed, the number of nodes in the map shall be equal to the nodes in the environment after the whole environment has been explored. In presence of errors, the map will likely contain a higher number of nodes, since the algorithm adds a new node to the map whenever a node is not recognized as an already visited node. The results of the simulations for each type of maps are presented in the Table 1. Each cell of the Table reports the ratio between the number of nodes created with or without Semantic information divided by the actual number of nodes in the map after a number S of "node-to-node" navigation steps: a lower ratio means better performance. Notice that this ratio may increase by increasing the number of navigation steps: then the number S is set to a value which is proportional to the number of nodes in the map, i.e., if a map has N nodes the number of steps will be $5 \times N$.

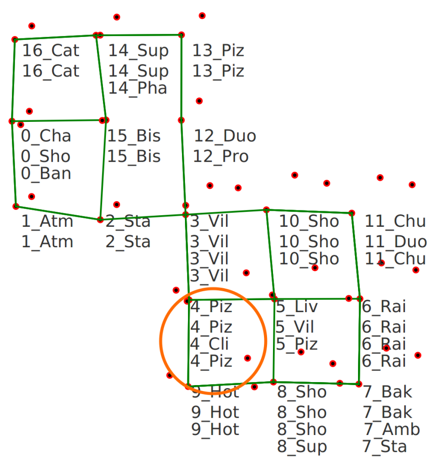

Fig. 1. The labels belonging to the multiple classes

4 Conclusion

In this extended abstract, we briefly presented a method for enhancing the process of topological mapping with semantic information. The results obtained

Table 1. The results of the simulations.

Errors range	Simple		Medium		Hard	
	Semantic	NonSemantic	Semantic	NonSemantic	Semantic	NonSemantic
-1 to 1	1.09	1.72	1.07	1.40	1.05	1.21
-3 to 3	1.29	2.33	1.36	2.40	1.55	3.85
-5 to 5	1.36	2.79	1.69	3.44	2.52	4.84
-7 to 7	1.57	3.26	2.87	4.65	3.20	4.89

from the simulation process show that the presence of semantic information allows for improving the number of loops that are correctly closed during navigation. Thus, we can state that the implementation of semantic information can significantly improve the quality of localization and mapping process.

References

1. Case, C., Bipin S., Adam C., Andrew Y. Ng: Autonomous sign reading for semantic mapping. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE (2011) 3297-3303.
2. Rogers, J.G., Trevor, A.J., Nieto-Granda, C., Christensen, H.I.: Simultaneous localization and mapping with learned object recognition and semantic data association. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, IEEE (2011) 1264-1270.
3. Nüchter, A., Joachim Hertzberg: Towards semantic maps for mobile robots. Robotics and Autonomous Systems 56, no. 11 (2008): 915-926.
4. Iocchi, L., and Pellegrini, S.: Building 3d maps with semantic elements integrating 2d laser, stereo vision and imu on a mobile robot. In: 2nd ISPRS International Workshop 3D-ARCH. 2007.
5. Rituerto, A., Murillo, A.C., Guerrero, J.J.: Semantic labeling for indoor topological mapping using a wearable catadioptric system. In: Robotics and Autonomous Systems 62, no. 5 (2014): 685-695.
6. Tsetsos, V., Anagnostopoulos, C., Kikiras, P., Hasiotis, P. Hadjiefthymiades, S.: A human-centered semantic navigation system for indoor environments. In: Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on, IEEE (2005) 146-155.
7. Tsetsos, V., Anagnostopoulos, C., Kikiras, P. Hadjiefthymiades, S.: Semantically enriched navigation for indoor environments. International Journal of Web and Grid Services 2, no. 4 (2006): 453-478.
8. Abdelnasser, H., Mohamed, R., Elgohary, A., Alzantot, M.F., Wang, H., Sen, S., Choudhury, R.R., Youssef, M.: SemanticSLAM: Using environment landmarks for unsupervised indoor localization. IEEE Transactions on Mobile Computing 15, no. 7 (2016): 1770-1782.
9. Robertson, P., Angermann, M. Khider, M.: Improving simultaneous localization and mapping for pedestrian navigation and automatic mapping of buildings by using online human-based feature labeling. In: Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION, 365-374.
10. Keshavdas, S., Zender, H., Kruijff, G.J.M., Liu, M. Colas, F.: Functional Mapping: Spatial Inferencing to Aid Human-Robot Rescue Efforts in Unstructured Disaster Environments. In: AAAI Spring Symposium: Designing Intelligent Robots. 2012.