

Leveraging User Activities and Mobile Robots for Semantic Mapping and User Localization

Stefano Rosa
University of Oxford
Oxford, UK
stefano.rosa@cs.ox.ac.uk

Xiaoxuan Lu
University of Oxford
Oxford, UK
xiaoxuan.lu@cs.ox.ac.uk

Hongkai Wen
University of Warwick
Warwick, UK
hongkai.wen@dcswarwick.ac.uk

Niki Trigoni
University of Oxford
Oxford, UK
niki.trigoni@cs.ox.ac.uk

ABSTRACT

This work proposes a probabilistic framework for combining high level information such as user activities, from a human user wearing a smart watch, and probabilistic information such as room connectivity from an assistive mobile robot for semantic mapping and user room level localization in domestic environments. The main idea is to leverage the semantic information provided by the user activities and the accurate metric map created by an assistive robot. The conceptual information is modeled as a probabilistic chain-graph. The user is equipped with only a smart watch, and we detect complex activities and a coarse trajectory using inertial data. We perform activity detection using a Long Short-Term Memory Recurrent Neural Network. The robot is equipped with an RGB-D camera, and creates a topological map of the environment. Both the user and the robot build a conceptual map composed by room categories on top of the low-level trajectory. When the robot and the user meet, the user's conceptual map is fused with the robot's conceptual map. The robot is able to match activities with types of rooms, learning a semantic representation of the environment over time (room types), while the user is able to be localized at room level by exploiting the precise map built by the robot. Preliminary ongoing tests show the feasibility of the approach.

Keywords

semantic maps; activity recognition; human-robot collaboration

1. INTRODUCTION

High-level understanding of the environment is still an open problem in robotics. Simultaneous Localization And Mapping (SLAM) systems that go beyond basic geometry

reconstruction to obtain a high level understanding of the environment (e.g., semantic, affordances, high-level geometry) will be required for robust robotics applications in the near future.

We consider applications where the robot is operating in a domestic environment (up until now, almost exclusively inhabited by humans). In such environments, concepts such as room types and activities are important, not only because of the interaction with humans but also for abstracting spatial knowledge.

In [1] the authors proposed a semantic SLAM algorithm for users wearing wearable sensors, by including detected activities in a particle filter SLAM approach; but in their system the user is wearing many inertial sensors, and the approach does not exploit interaction with robots.

In [2] a method is proposed for tagging maps with objects. The object's position is inferred by detecting user's activities and user's location, but the detected activities are not used in the map estimation and there is no information exchange between the robot and the user.

[3] presents a conceptual model for semantic map representation, with different levels of abstraction, from sensor data to concepts, such as rooms, with associated properties, such as shape, appearance, and detected objects. The layered structure of the spatial knowledge is used for reasoning at the semantic level. We start from the semantic mapping framework presented in [3] and extend it to a human-robot interaction scenario by adding the presence of a user and exploiting user activities as room properties. Our approach is opportunistic, as the robot and the user only exchange information about their respective map representations when they meet. We exploit meeting episodes since they are a strong constraint on the position of robot and user (when they meet it is most probable that they are in the same room).

The user is endowed with only a smart watch and it is updating a conceptual map composed by rooms and activities, detected using inertial data from the watch. The robot is able to recognize the user with the onboard camera using a simple detector based on *Histograms of Oriented Gradients* (HOG) features and a linear *Support Vector Machine*. When the robot and the user meet (we call it a *rendezvous* episode), the user's map is fused to the robot's map. Activity information is propagated to the robot to infer room categories over time. Moreover, room transition probabili-

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

HRI '17 Companion March 06-09, 2017, Vienna, Austria

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4885-0/17/03.

DOI: <http://dx.doi.org/10.1145/3029798.3038343>

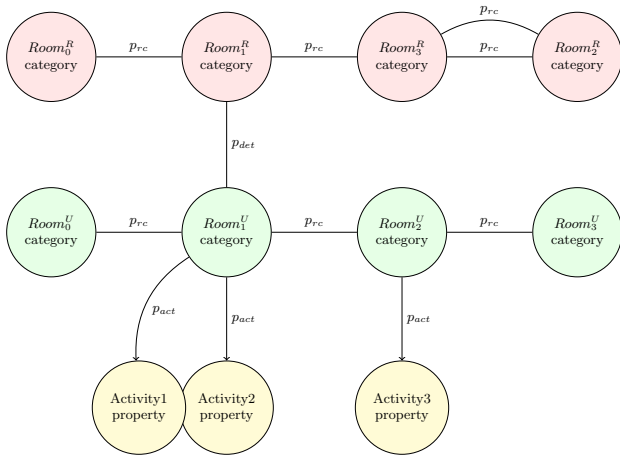


Figure 1: The structure of the combined chain graph model for the user and the robot after a rendezvous. Red nodes represent the robot’s conceptual map; green nodes represent the user’s conceptual map; yellow nodes are activity detections.

ties computed by the robot are used to do reasoning on the user’s map, in order to obtain room-level localization of the user over time.

2. SEMANTIC MAP REPRESENTATIONS

As in [3], at the lower level a SLAM algorithm creates a metric map of the environment using the robot sensors. A topological map is then built on top of the metric map and the continuous space is discretized into areas called places, that are added each meter along the trajectory. Places connect to other places via paths. We use a template-based door detector on range data. Based on detected doors, places are clustered into rooms (room is used in a broad sense, since it also represents corridors). For the user, we assume a coarse trajectory can be obtained using *Pedestrian Dead Reckoning*. At a higher abstraction level locations are grouped using the concept of room. Rooms tend to share similar semantics and are assigned semantic categorical labels (e.g. kitchen, bathroom, corridor, etc.).

The conceptual map is represented using a *chain graph*, which can model both directed and undirected relationships. As in [3], the chain graph is then converted to a *factor graph* and solved using the GTSAM library. An example of the combined chain graph for the robot and the user after a rendezvous event is shown in Figure 1. $p_{rc}(\cdot, \cdot)$ is the room transition probability; $p_{act}(\cdot|\cdot)$ is the activity conditional probability given room type; $p_{det}(\cdot, \cdot)$ is the probability of the user and the robot being in the same room if the robot is detecting the user. We set these probabilities to fixed values in the current work. Ongoing work will be devoted to learning the different joint and conditional probability values by analyzing user’s daily data over time.

3. ACTIVITY RECOGNITION

We use a *Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN)* for recognizing complex user activities from inertial data. We use raw accelerometer and gyroscope data as inputs. The sampling frequency is 100 Hz.

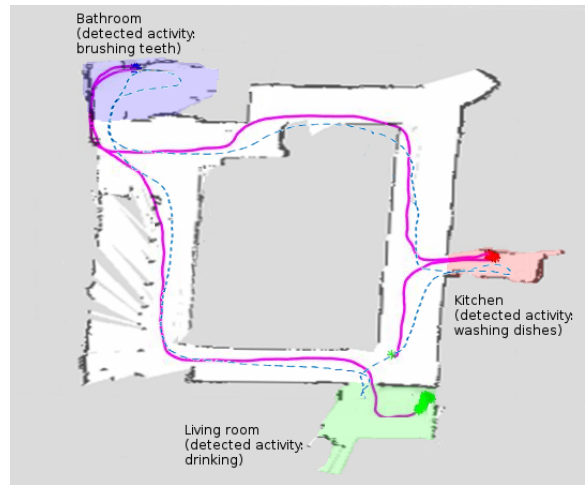


Figure 2: The experimental setup.

The network has 300 neurons in the hidden layer. We use a window of 3 seconds with 50% overlap; the network is trained on a dataset of four activities for a total time of 20 minutes, but we plan to increase the number of activities in order to 15-20 in future work.

4. EXPERIMENTS

Preliminary experiments have been carried out. The user is equipped with a Sony Smartwatch 3 smart watch, connected via Wi-Fi to the robot. The robot is a Turtlebot 2 equipped with a Microsoft Kinect camera. The robot is using the Kinect as a distance sensor to create a 2D metric map of the environment, and also to detect the user using the RGB video stream. The mapping system is implemented in ROS.

We detect four activities (walking, drinking, brushing teeth, washing dishes). Classification accuracy is 89.3%. Note that the last three activities are complex compared to usual activities such as walking, running, etc.

A preliminary experiment is shown in Figure 2. The user’s trajectory is shown in magenta; robot’s trajectory in dotted blue; detected activities are shown in red (washing dishes), blue (brushing teeth) and green (drinking). The user was visiting the rooms in anti-clockwise direction, while the robot was moving clockwise. The robot finally detects the user in the living room (bottom right).

5. REFERENCES

- [1] M. Hardegger, D. Roggen, A. Calatroni, and G. Tröster. S-smart: A unified bayesian framework for simultaneous semantic mapping, activity recognition, and tracking. *ACM Trans. Intell. Syst. Technol.*, 2016.
- [2] G. Li, C. Zhu, J. Du, Q. Cheng, W. Sheng, and H. Chen. Robot semantic mapping through wearable sensor-based human activity recognition. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 5228–5233. IEEE, 2012.
- [3] A. Pronobis and P. Jensfelt. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3515–3522. IEEE, 2012.