2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Robot Semantic Mapping through Wearable Sensor-based Human Activity Recognition

Gang Li, Chun Zhu, Jianhao Du, Qi Cheng, Weihua Sheng and Heping Chen

Abstract— **Semantic information can help both humans and robots to understand their environments better. In order to obtain semantic information efficiently and link it to a metric map, we present a semantic mapping approach through human activity recognition in an indoor human-robot coexisting environment. An intelligent mobile robot platform can create a 2D metric map, while human activity can be recognized using motion data from wearable motion sensors mounted on a human subject. Combined with pre-learned models of activity-to-furniture type association and robot pose estimates, the robot can determine the distribution of the furniture types on the 2D metric map. Simulations and real world experiments demonstrate that the proposed method is able to create a reliable metric map with accurate semantic information.**

Index Terms— **semantic map, human activity recognition, wearable sensor, simultaneous localization and mapping (SLAM).**

I. INTRODUCTION

A. Motivation

With the advances of robotics research, it is predicted that within two decades robots will be capable of adapting to complex, unknown environments and interacting with humans to assist with various tasks in daily life, which include house cleaning, security, nursing, life-support, entertainment, etc [1]. The co-existence of human beings in the same environment could provide the robots valuable information through, for example, human-environment interaction. This helps robots understand the environment better and provide better service to humans.

Knowledge about the environment is usually encoded in the form of a map. The problem of how to represent, build, and maintain maps has been one of the most active robotics research areas in the last decade. Existing formats such as metric map [2], topological map [3] may be sufficient for basic tasks such as navigation. However, they do not contain any high level semantic description of the environment, which is critical for robots to perform higher level tasks in a human-robot coexisting environment. For instance, a metric map may represent the geometry of a house, but it does not carry the meaning of the geometric shapes, such as chairs, beds, tables, etc. It does not contain the information of the room type, such as a kitchen, or a bedroom. This kind of semantic information is very important for tasks such as "dimming the light in the bedroom if the human subject is sleeping".

Gang Li, Chun Zhu, Jianhao Du, Qi Cheng, Weihua Sheng are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA, email: weihua.sheng@okstate.edu

Heping Chen is with the Ingram School of Engineering, Texas State University, San Marcos, Texas 78666, USA. email: hc15@txstate.edu

A semantic map can be manually created by marking special landmarks on a metric map. However, it is highly desirable that the robot can conduct automated semantic mapping. In this paper, we will consider robotic semantic mapping of a residential environment. The robot need create a 2D map and tag the furniture in it, such as tables, chairs, beds, cabinets, etc. Vision-based recognition techniques have been adopted in literature for extracting these landmarks [4]. The drawback of these methods is their high computational complexity. Extracting features for pattern recognition from vision data is very demanding in terms of memory and data processing capacity. Furthermore, vision-based methods can be heavily influenced by the ambient lighting conditions and fail in environments with low visibility. To overcome the shortcomings of vision-based semantic mapping, in this paper, we propose a new approach for the robot to conduct semantic mapping. We argue that *semantic information can be inferred from how humans interact with the objects in the environment*. To recognize human activities we use wearable motion sensors. In this way, we can avoid the many difficulties inherited in vision-based object recognition for semantic mapping.

The proposed automated semantic mapping through wearable sensor-based human activity recognition is illustrated in Figure 1. A robot implements the simultaneous localization and mapping (SLAM) algorithm [5] to generate a 2D metric map of an unknown indoor environment. In the mean time, the robot localizes the human subject through onboard vision and identifies the activity the human subject is performing through motion sensors attached to the human subject. A pre-learned activity type-to-furniture type model is utilized to determine the furniture type. High level interpretation, e.g., "the room is an office", based on the furniture identified in the semantic map will therefore be possible. Given the learned furniture type, the prior knowledge of activity type-to-furniture type model can be furhter improved, which in turn can improve the accuracy in activity recognition.

The rest of the paper is organized as follows. The remaining of this section will discuss the related work. Section II presents the setup of the robot used in our research. Section III describes the wearable sensor-based human activity recognition. Section IV explains the proposed approach to semantic mapping. The simulations and real world experiments are presented in section V. Conclusions are given in Section VI.

Fig. 1. Illustration of human activity based semantic mapping.

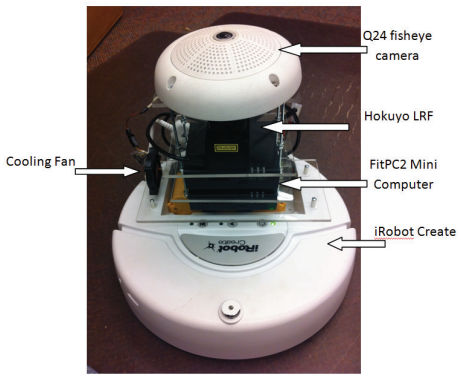

Fig. 2. The robot ASCCbot.

B. Related Work

The importance of including semantic information in robot maps has been recognized for a long time [6], [7]. In recent years, researchers have been developing robotic systems which can acquire and use semantic information [8]. Some of them obtained semantic information via a linguistic interaction with a human subject [9]. Some of them were limited to the classification of surface elements [10] like ceilings, floors, etc. In [11] and [12] specific places of indoor environments are labeled based on the presence of key objects in them through computer vision technology that extracts the necessary information from images. Jebari *et al.* utilized panoramic camera to extract high-level information through object recognition [13]. Other researchers focused on how to create a multi-hierarchical semantic map [14], [15], [16]. Their idea is useful if there are multiple rooms in the environment. In other works, researchers acquired semantic information through the recognition of 3D models. Nuchter *et al.* developed approaches to extracting semantic information from 3D models built from a laser scanner on the robot [17]. Nielsen *et al.* used snapshot technology to mix 2D images onto 3D objects [18]. All of these vision-based methods suffer from high computational cost, significant background noise and difficulty in segmentation. The approach proposed in this paper can provide a more efficient way to acquire semantic information through the interaction between human and his environment.

II. Robot Setup

The mobile robot we use is called ASCCbot, which is developed in the Advanced Sensing, Computation & Control Lab (ASCC Lab) at Oklahoma State University. It is a compact, intelligent mobile platform which is open, extendable, duplicable and equipped with various functionalities including object detection, SLAM and object following.

As shown in Figure 2, the ASCCbot is built on an iRobot Create with an Atom processor-based computer called FitPC2 [19], a Hokuyo laser range finder (LRF) [20] and a Q24 panoramic camera [21]. The iRobot Create is a platform designed for robot development and possesses a serial port through which sensor data can be read and motor commands can be issued using the iRobot Open Interface protocol. The FitPC2 is a small, light computer. The Hokuyo laser range finder URG-04LX is a USB-powered device which uses a laser beam to determine the distance to objects. It has a measuring range between 20 mm and 4094 mm, a scanning range of 240°, a scanning rate of 100 ms/scan, a distance accuracy of $\pm 3\%$ and an angle resolution of 0.36°. The fisheye camera (Q24) is capable of providing different views including a panoramic view so that it can cover the surrounding area of the mobile platform. The camera provides a highest resolution of 3 Megapixels and color images scalable from 160×120 to 2048×1536. The camera itself is a web server so that the stream of live images can be obtained by setting up a socket connection. The features of the camera (including resolutions, frame rates, etc.) can be adjusted by sending a web request. Moreover, the zooming and panning of the camera lenses can be done through the virtual PTZ function. External batteries are used to power the FitPC2 and Q24 while a USB-powered mini fan is used to cool the FitPC2.

The software system of the ASCCbot is built upon the Robot Operating System (ROS) [22], which is an open-source, meta-operating system for robots. It provides services similar to real operation systems, including hardware abstraction, low-level device control, implementation of commonly-used functionality, message-passing between processes, and package management. Its distributed computing feature can also facilitate multi-robot applications.

In the proposed approach, there are several functionalities that are necessary on the robot. First, the SLAM algorithm is running in the background which creates a 2D metric map and provides the robot pose estimates. The communication program is used to receive activity recognition results through Zigbee. The semantic mapping program updates the semantic information on the 2D metric map. All the above programs are separate ROS nodes running in a ROS network. Additionally, there are two basic nodes that control the robot to automatically follow the human subject: the

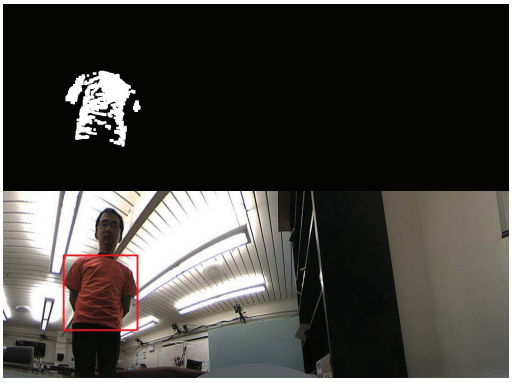

Fig. 3. Detection of human subject using color segmentation.

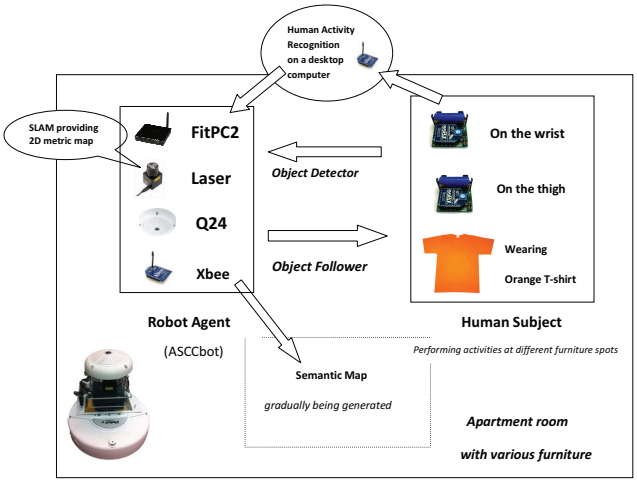

Fig. 4. Overall system setup for robot semantic mapping through human activity recognition.

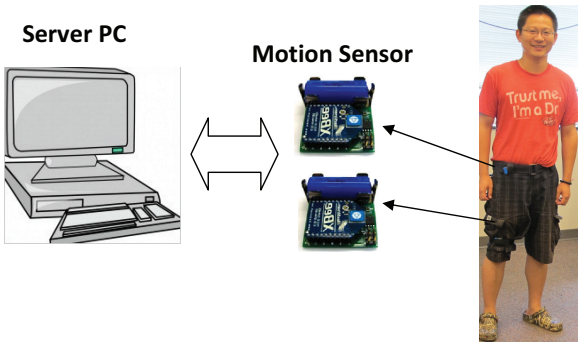

Fig. 5. The overview of the hardware platform for daily activity recognition.

human detector and the human follower.

The specific task of the human detector is to find the angle of the human subject with respect to the ASCCbot and roughly estimate the distance between them. To simplify the problem, on our platform, we designed a color tracking algorithm which can track the orange T-shirt the subject is wearing and send out motion control command to the ASSCbot. Color segmentation is used to find contiguous regions in which individual pixels share common characteristic. After applying the Gaussian filter and the morphology method (dilate and erode) to reduce the noise, the orange T-shirt region is detected as shown in Figure 3. In the panoramic view of Q24, the angle and size of the detected region will be output to the human follower node.

The human follower controls the robot to follow the human subject smoothly and stop when it is close enough to the human subject. With the angle of the subject and the relative size obtained from the human detector node, the human follower node basically tries to keep the subject in the middle of the view with respect to robot. When the human subject is performing certain activity near some furniture, the associated semantic information at that location will be updated on the 2D metric map through a Bayesian framework which will be described in Section IV.

The whole system setup for semantic mapping through human activity recognition is shown in in Figure 4;

III. Wearable Sensor-Based Human Activity Recognition

The hardware system for human activity recognition is shown in Figure 5. Two wearable wireless motion sensors [23] developed in our lab are used to collect motion data and transfer them to a server PC. The motion sensor is compact, power-aware and can provide fast sampling of human motion including acceleration, angular rate and magnetic field. The PC processes the data to recognize activities and sends the results to the robot.

The motion sensor node is developed based on a commercial orientation sensor called VN-100 [24]. The sensor nodes send data (3D acceleration, 3D angular velocity, orientation, and magnetic data) through Zigbee to a receiver on the PC for processing. Each motion sensor node has an ID to distinguish itself from others. Therefore, multi-person activity can also be tracked in this system. Since the position to attach the sensor is very important to activity recognition [25], we collected data using the sensors on different parts of the human body and found that the thigh and the waist are the best positions for body activity recognition.

The wearable motion sensor samples the 3D acceleration and 3D angular velocity at a rate of 20Hz. In the experiments we only collect the 3D acceleration as the raw data, which is sufficient for activity recognition. Features such as means and variances are extracted and further clustered into discrete observation symbols.

On the PC, a real-time program is used to collect data, extract features and recognize the human activity. We combine the neural networks and the hidden Markov models (HMMs) in the recognition algorithm. Its block diagram is shown in Figure 6. There are two steps in the recognition algorithm: (1) coarse-grained classification and (2) fine-grained classification. The coarse-grained classification step uses the outputs of two neural networks and generates a basic classification result. The fine-grained classification step generates the detailed activity types using a hidden Markov

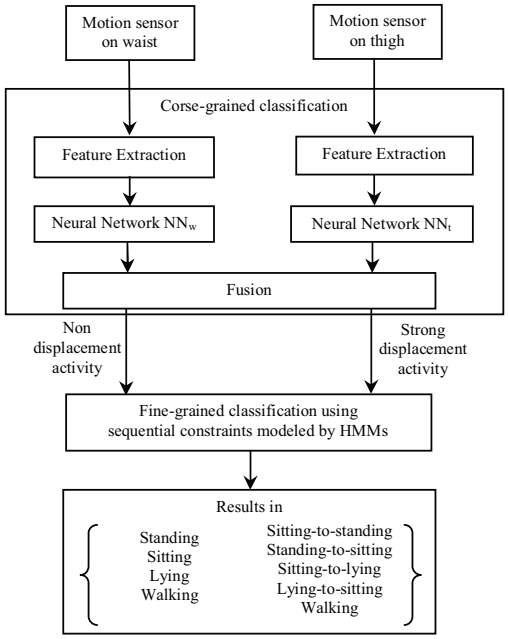

Fig. 6. The overview of the human activity recognition algorithm.

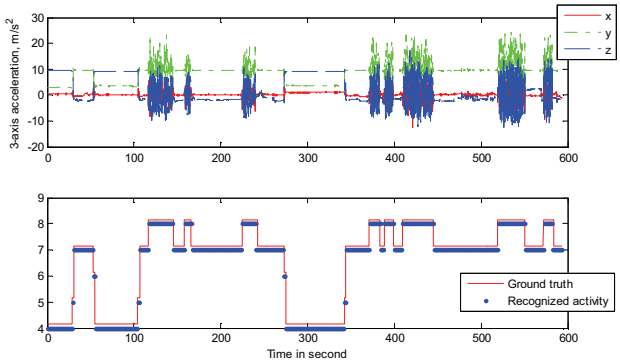

Fig. 7. The results of activity recognition. Activity result labels: 4: sitting; 5: sit-to-stand; 6: stand-to-sit; 7: standing; 8: walking.

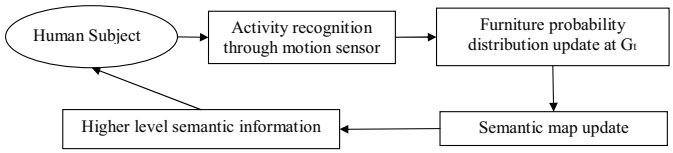

Fig. 8. The block diagram for robot semantic mapping.

model (HMM) approach.

In the coarse-grained classification module, raw data (3D acceleration) from two motion sensors are processed to obtain the features (mean and variance), which are fed into the corresponding neural network NN_w and NN_t for the sensors on waist and thigh, respectively. A fusion module integrates the individual types of waist and thigh activities and categorizes the human activities according to certain rules. For the details on the fusion of the two neural networks, please see our previous work [26].

In the fine-grained classification module, the sequential constraints of human daily activity are modeled using an HMM and a modified short-time Viterbi algorithm [27] is applied to realize real-time activity recognition in order to generate the detailed activity types. More details on the modeling using HMM can be found in [28].

In the training phase, the human subject carries a PDA and pushes a button to label the data. The label is then sent to the PC as the training target of the neural network. The neural network is trained through the back-propagation method. The HMM in the second step models the sequential constraints of the activities. The parameters of the HMM can be trained by observing the activity sequence of the human subject for a period of time.

In the testing phase, the human subject performs daily activities and the PC runs the proposed activity recognition algorithm. For the accuracy comparison purpose, the human activities are recorded through the Vicon motion capture system as the ground truth, which are compared with the recognition results. Figure 7 shows the accumulated results of a 10-minute experiment. Five types of daily activities are recognized using two motion sensor nodes attached to the waist and the thigh of the human subject. The effectiveness and accuracy of our approach are verified by the obtained results.

IV. Semantic Mapping

In this section, we describe the Bayesian framework for semantic mapping. At the beginning, the robot enters the unknown environment of interest. It first uses the SLAM algorithm to create a 2D metric map. To build the semantic map, it uses another source of information, i.e., the activity of human subject when he interacts with the environment. Generally, the human subject performs certain activities around certain furniture. For example, "sitting" on a "chair", "lying" on a "bed". By recognizing the human activities, the knowledge about the furniture type at the human's location can be learned. The semantic information can be acquired through an iterative fashion which increases the accuracy of the semantic map over the time. The block diagram of the semantic mapping process is shown in Figure 8.

At each time t, the human subject performs certain activity, e.g., sitting, lying. Let A_t denote the true activity type at t and O_t denote the corresponding estimated activity through the proposed activity recognition algorithm. The observation model is $P(O_t|A_t)$, which gives the likelihood of getting O_t when the true activity is A_t. This model is basically the accuracy of the activity recognition algorithm and can be learned beforehand.

On the other hand, the activity is associated with the furniture type through a probabilistic model $P(A_t|F)$. For example, when the furniture "bed" is given, the probability of "lying" and "sitting" are much higher than that of "standing". This knowledge can be obtained empirically or learned through observation.

TABLE I

THE ACTIVITY OBSERVATION $P(O|A)$.

		True activity A				
		1	2	3	4	5
Observed activity O	1	0.68	0.12	0.04	0.04	0.04
	2	0.20	0.76	0.04	0.04	0.04
	3	0.04	0.04	0.84	0.04	0.04
	4	0.04	0.04	0.04	0.68	0.12
	5	0.04	0.04	0.04	0.20	0.76

TABLE II

THE ACTIVITY TYPE-TO-FURNITURE TYPE MODEL $P(A|F)$.

		Furniture type F				
		1	2	3	4	5
Activity A	1	0.60	0.01	0.01	0.04	0.05
	2	0.02	0.75	0.01	0.04	0.05
	3	0.01	0.01	0.74	0.04	0.10
	4	0.35	0.21	0.23	0.87	0.10
	5	0.02	0.02	0.01	0.01	0.70

Based on the rule of total probability, we have:

$$P(O_t|F) = \sum_A P(O_t|A)P(A|F) \tag{1}$$

In this way, we obtain the activity observation model when the furniture type is given. We also have

$$P_t(F|O_t) \propto P(O_t|F)P_{t-1}(F) \tag{2}$$

where $P_{t-1}(F)$ is the prior knowledge of the furniture type based on all past information up to time $t-1$. The initial value of this probability can be set to be uniform distribution.

To determine the furniture type at time t, the maximum a *posteriori* probability criterion can be adopted:

$$\hat{F}_t = \arg\max_F P_t(F|O_t) \tag{3}$$

V. EXPERIMENT RESULTS

In order to evaluate the performance of the proposed approach, both simulation and real world experiments are carried out.

A. Simulation

In the simulations, we consider five types of activity: 1: lying, 2: typing, 3: eating, 4: sitting and 5: opening a door. Thus, $A_t \in \{1,2,3,4,5\}$ and $O_t \in \{1,2,3,4,5\}$. The observation model $P(O|A)$ is essentially the accuracy of the proposed activity recognition algorithm and is shown in Table I. For example, the probability of observing activity $O = 2$ given that the true activity $A = 2$ (typing) is 0.76. Five types of furniture are considered here: 1: bed, 2: computer desk, 3: dining table; 4: chair and 5: door. Thus, $F \in \{1,2,3,4,5\}$. Table II gives the activity type-to-furniture type model. For example, the probability of activity $A = 2$ (typing) given that furniture type $F = 2$ (computer desk) is 0.75.

In the simulations, we basically evaluate the Bayesian framework through simulated human activities. For each furniture type, one thousand activities observations O_t were generated. The recognition results for the five furniture types are shown in Table III. The integer numbers in the table are the numbers of decision $\hat{F}_t \in \{1,2,3,4,5\}$ corresponding to each furniture type, which is calculated according to Equation (3).From this table we can see that the Bayesian framework is effective in recognizing the furniture type.

TABLE III

SIMULATION RESULTS FOR FURNITURE TYPE RECOGNITION.

Decision	Furniture Type F				
	1	2	3	4	5
1	803	96	14	220	179
2	52	760	11	42	56
3	17	19	909	199	43
4	40	102	32	523	13
5	88	23	34	16	709
Accuracy	0.8030	0.760	0.909	0.5230	0.709

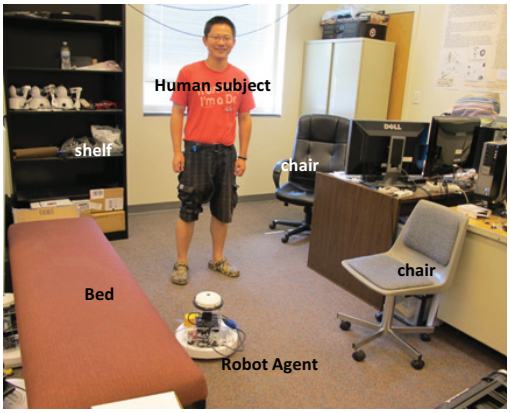

Fig. 9. The experiment in the mock apartment.

B. Real World Experiment

In the real world experiment, we set up a mock apartment room in our lab which is approximately 3 meters by 2 meters. There are a bench (as the bed), two chairs and a shelf as the furniture which need to be determined as can be seen in Figure 9.

The human subject wears an orange T-shirt and the ASC-Cbot is put into the mock apartment. All the ROS nodes are launched on the FitPC2 and two motion sensors are attached to the wrist and thigh of the human subject. The human detector node and the human follower node run when the orange T-shirt showed in the view of Q24 camera. The robot is controlled to follow the human subject around the mock apartment smoothly. When the human subject stops at one of the furniture and performs certain associated activities, the ASCCbot follows him to the furniture location, stops, receives the activity recognition results based on the wearable motion sensors and updates the semantic information on the map. During the real world experiment, the human subject walks around the mock apartment, sits on the chairs, stands beside the shelf and lies on bed. When the robot obtains the pose estimate and the human activity recognition results, the semantic labels are generated according to Equation (3),

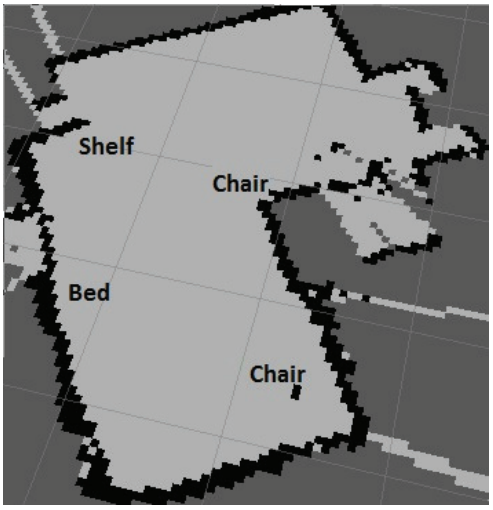

Fig. 10. The created semantic map of the mock apartment.

which reveals the most possible furniture type at that location. It is also possible to maintain a probabilistic distribution of the possible furniture at that location. After some activities are performed, it is clear that the generated semantic map in Figure 10 matches the real world situation shown in Figure 9.

VI. Conclusions

This paper develops an automated semantic mapping system through wearable sensor-based human activity recognition. In this system, motion sensors are attached to a human subject for activity recognition and the SLAM algorithm running on the robot generates a 2D metric map. Activity observations and the location of the human subject are combined to obtain the semantic information on the 2D map. The most possible furniture type is tagged to the 2D metric map. Both simulation results and real world experiment results demonstrate the effectiveness and accuracy of the proposed method. This method provides a new perspective for robotic semantic mapping and can significantly reduces the difficulties involved in traditional vision-based object classification algorithms.

ACKNOWLEDGMENTS

This project is partially supported by the NSF grant CISE/CNS 0916864, CISE/CNS MRI 0923238 and the DEPSCoR grant W911NF-10-1-0015.

References

[1] C. Chen Y. Weng and C. Sun. Toward the human-robot co-existence society: On safety intelligence for next generation robots. *International Journal of Social Robotics*, 1(4):267–282, 2009.

[2] B. Kuipers and T. Levitt. Navigation and mapping in large-scale space. *AI Magazine*, 9(2):28–43, 1988.

[3] D. Baker. Some topological problems in robotics. *The Mathematical Intelligence*, 12(1):66–77, 1990.

[4] E. Menegatti, M. Wright, and E. Pagello. A new omnidirectional vision sensor for the spatial semantic hierarchy, 2001.

[5] J. Aulinas, Y. Petillot, J. Salvi, and X. Llad. The slam problem: a survey, 2008.

[6] Benjamin Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2:129–153, 1978.

[7] R. Chatila and J. Laumond. Position Referencing and Consistent World Modeling for Mobile Robots. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Computer Society Press, March 1985.

[8] C. Theobalt et al. Talking to godot: Dialogue with a mobile robot. In *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002*, pages 1338–1343, 2002.

[9] G. M. Kruijff, H. Zender, P. Jensfelt, and H. I. Christensen. Situated dialogue and spatial organization: What, where... and why? *International Journal of Advanced Robotic Systems, Special Issue on Human-Robot Interaction*, 4(1):125–138, March 2007.

[10] Denis F. Wolf and Gaurav S. Sukhatme. Semantic Mapping Using Mobile Robots. *IEEE Transactions on Robotics*, 24(2):245–258, April 2008.

[11] Axel Rottmann, Oscar Martinez Mozos, Cyrill Stachniss, Wolfram Burgard, Íőscar MartÍłnez, Mozos Cyrill, and Stachniss Wolfram Burgard. Semantic place classification of indoor environments with mobile robots using boosting. In *in Proc. of the National Conference on Artificial Intelligence (AAAI*, pages 1306–1311, 2005.

[12] Shrihari Vasudevan and Viet Nguyen. Towards a cognitive probabilistic representation of space for mobile robots. In *in: IEEE International on Information Acquisition, ICIA*, 2006.

[13] I. Jebari and E. Battesti. Multi-sensor semantic mapping and exploration of indoor environments. In *Technologies for Practical Robot Applications (TePRA)*, 2011.

[14] C. Galindo, A. Saffiotti, S. Coradeschi, and P. Buschka. Multi-hierarchical semantic maps for mobile robotics. In *in Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, 2005*, pages 3492–3497, 2005.

[15] B. Kuipers, J. Modayil, P. Beeson, M. Macmahon, and F. Savelli. Local metrical and global topological maps in the hybrid spatial semantic hierarchy. In *in IEEE Int. Conf. on Robotics & Automation (ICRA-04*, pages 4845–4851, 2004.

[16] B. Kuipers, R. Froom, W. Lee, and D. Pierce. The semantic hierarchy in robot learning. In *Robot Learning*, pages 141–170. Kluwer Academic Publishers, 1993.

[17] Andreas Nüchter and Joachim Hertzberg. Towards semantic maps for mobile robots. *Robot. Auton. Syst.*, 56:915–926, November 2008.

[18] C. Nielsen, B. Ricks, D. Bruemmer, D. Few, and M. Walton. Snapshots for semantic maps. In *IEEE International Conference on Systems, Man and Cybernetics*, 2004.

[19] fit-PC2. http://www.fit-pc.com/web/, 2011.

[20] Hokuyo laser. http://www.hokuyo-aut.jp/, 2011.

[21] Q24. http://www.mobotix.com/, 2011.

[22] ROS wiki. http://www.ros.org/wiki/, 2010.

[23] S. Zhang, G. Li, and W. Sheng. Development and evaluation of a compact motion sensor node for wearable computing. In *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2010.

[24] VectorNav Technologies. http://www.vectornav.com/, 2011.

[25] U. Maurer, A. Smailagic, D.P.Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, pages 113–116, 2006.

[26] C. Zhu and W. Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *IEEE International Conference on Robotics and Automation*, pages 2154–2159, 2009.

[27] J. Bloit and X. Rodet. Short-time viterbi for online hmm decoding: Evaluation on a real-time phone recognition task. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 2121–2124, 2008.

[28] C. Zhu and W. Sheng. Recognizing human daily activity using a single inertial sensor. In *The 8th World Congress on Intelligent Control and Automation*, pages 282–287, 2010.