Robotics and Autonomous Systems 68 (2015) 47–58

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Robot semantic mapping through human activity recognition: A wearable sensing and computing approach

Weihua Sheng [a,*], Jianhao Du [a], Qi Cheng [a], Gang Li [a], Chun Zhu [a], Meiqin Liu [b], Guoqing Xu [c]

[a] School of Electrical and Computer Engineering, Oklahoma State University, Stillwater OK, 74078, USA
[b] College of Electrical Engineering, Zhejiang University, Hangzhou, 310027, China
[c] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China

HIGHLIGHTS

- A framework for robot semantic mapping through human activity recognition.
- Human activity recognition is realized through wearable motion sensors.
- Validated through both simulation and experiments.

ARTICLE INFO

Article history:
Received 22 September 2013
Received in revised form
9 November 2014
Accepted 4 February 2015
Available online 12 February 2015

Keywords:
Semantic map
Human activity recognition
Wearable sensor
Simultaneous localization and mapping (SLAM)
Information fusion

ABSTRACT

Semantic information can help robots understand unknown environments better. In order to obtain semantic information efficiently and link it to a metric map, we present a new robot semantic mapping approach through human activity recognition in a human–robot coexisting environment. An intelligent mobile robot platform called ASCCbot creates a metric map while wearable motion sensors attached to the human body are used to recognize human activities. Combining pre-learned models of activity–furniture correlation and location–furniture correlation, the robot determines the probability distribution of the furniture types through a Bayesian framework and labels them on the metric map. Computer simulations and real experiments demonstrate that the proposed approach is able to create a semantic map of an indoor environment effectively.

1. Introduction

1.1. Motivation

With the advancement of robotics research, it is predicted that in the near future robots will be capable of adapting to complex, unknown environments and interacting with humans to assist with various tasks in human's daily life, which include house cleaning, security, nursing, life-support, entertainment, etc. [1]. The coexistence of humans in the same environment could provide robots with valuable information regarding human behaviors, which can help 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 [2]. Existing formats such as metric map [3] and topological map [4] may be sufficient for basic tasks such as navigation. However, these maps do not contain any high level semantic understanding 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 room, but it does not indicate whether this room is an office, a kitchen, or a bedroom. It does not indicate the type of furniture or object either. As a matter of fact, this kind of semantic information is very important for robots to accomplish high-level tasks, especially in a smart home environment. Some of these high-level tasks can be *bring me a can of coke from the refrigerator in the kitchen*, *put my breakfast on the dining table in the kitchen*, etc. These tasks can be efficiently accomplished if the robot is able to perceive the semantic meaning of its surroundings.

* Corresponding author.
E-mail address: weihua.sheng@okstate.edu (W. Sheng).

http://dx.doi.org/10.1016/j.robot.2015.02.002
0921-8890/

This research is motivated by the goal of enabling such future smart home applications involving robots.

A semantic map can be manually created by labeling special objects or landmarks on a metric map. However, an automated semantic mapping method is highly desirable for an intelligent robot. In this paper, we aim to develop an approach for the robot to recognize the objects or landmarks in an indoor environment. Particularly, we are interested in furniture such as table, chair, bed, and shelf. Vision-based recognition techniques have been adopted in the literature for extracting these objects and landmarks [5]. The main drawback of these methods is their high computational complexity. Extracting features for pattern recognition from vision data is very demanding in terms of memory usage and data processing capacity. Furthermore, vision-based methods usually fail in an environment with low visibility. In this paper, we propose that *semantic information can be inferred from how human subjects interact with the objects and landmarks in the environments*. To recognize human activities we use wearable motion sensors. In this way, we can avoid the challenges associated with vision-based recognition approaches.

1.2. 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 that can acquire and use semantic information [8–11]. Traditionally, semantic mapping [12–15] is treated as a *robot-centric* task where the robot sees the environment via its sensors and fuses the collected sensor data into a multi-layer representation of spatial knowledge. The robot in [16] was able to build a map representation from both spatial and semantic perspectives, where a spatial hierarchy and a conceptual hierarchy are interrelated through the concept of anchoring. Sharing a similar spirit, the semantic mapping system proposed in [17] constructs a representation composed of layers representing maps at different levels of abstraction: metric, navigation, topological and conceptual. In [18,19] specific places of indoor environments are labeled based on the presence of key objects in them. Nuchter et al. developed approaches that extract semantic information from 3D models built from a laser scanner mounted on the robot [20]. In [21,22], semantic mapping was realized through autonomous detection and perception of objects and augmenting spatial metric maps with object information. Most of these existing works focused on object recognition using different attributes, such as color, shape, and texture. The relationship between objects and humans is largely ignored.

In recent years, in contrast to the above robot-centric way, researchers have investigated how to obtain semantic information through human–environment interactions. The concept of affordance [23] was first introduced into the computer vision community by Gibson to model the relationship between the human behavior and the environment. Grabner et al. [24] have proposed an affordance detector where functionality is handled as a cue complementary to appearance, rather than being considered after appearance-based detection. Delaitre et al. [25] have proposed a statistical descriptor of person–object interactions and demonstrated its benefits for recognizing objects and predicting human body poses in new scenes. Gupta et al. [26,27] presented a Bayesian approach to object and scene understanding by observing the human movement on objects. Their approach applies spatial and functional constraints on each of the perceptual elements for coherent semantic interpretation. Such constraints make it possible to recognize objects and actions when the appearances are not discriminative enough. Kjellström et al. [28] presented a method for categorizing manipulated objects and human manipulation actions in the context of each other. The robot learns the object affordance from how human interacts with objects.

Most of these existing semantic mapping techniques use vision data as the input. Although people may prefer passive vision sensors, they have several serious limitations [29]. First, they usually incur high computational cost which makes them hard for realtime implementation. Second, these techniques are usually subject to constant occlusions, lighting changes and other environmental factors which make them unreliable and not suitable for use in residential environments. Third, they are not easy to scale up due to the large amount of video or image data to process. In contrast, the approach proposed in this paper can provide a more efficient way to achieve object recognition through human–environment interaction. Our approach will use motion data collected by the wearable motion sensors instead of vision data from cameras.

1.3. Contributions

The main contributions of this paper are as follows. First, this paper proposes a probabilistic framework to learn the semantic information through human–environment interaction, which can be used to derive semantic maps for robotic applications. Second, this paper uses motion data from wearable sensors for human activity recognition, which overcomes some limitations inherited in traditional vision based human activity recognition. Third, the probabilistic approach enables incremental learning of the furniture types as more and more human–environment interactions are observed, which is conducted in a Bayesian way. Fourth, both computer simulations and experimental evaluations are conducted to evaluate the proposed framework and methodologies.

The rest of the paper is organized as follows. Section 2 gives the problem formulation. Section 3 presents the theoretical framework of the proposed wearable sensor-based semantic mapping approach. Section 4 describes the human activity recognition algorithm. In Section 5, computer simulations are conducted to evaluate the proposed theoretical framework. Section 6 presents the real experiments and results. Conclusions are given in Section 7.

2. Problem statement

The basic concept of robot semantic mapping through wearable sensor-based activity recognition is illustrated in Fig. 1. Initially, through the simultaneous localization and mapping (SLAM) [30] algorithm, the robot creates a metric map. It can also determine its own location and orientation in the map, which are denoted by $L_r(t) = [x_r(t), y_r(t), \theta_r(t)]$. We assume the map of the indoor area is divided into a total of K grids. Any location $[x, y]$ within the area is mapped to a grid index through the following function: $G = g([x, y])$, where $G \in \{1, \ldots, K\}$. To build a semantic map of the environment, the robot needs to not only detect the surrounding objects and their locations, but also label them. In this paper, we focus on the furniture (and some appliances as well). Therefore we define a furniture type probability distribution function $P_{G,t}(F)$ at grid G and time t. $F \in \{f_1, \ldots, f_M\}$ denotes the furniture type, where M is the total number of furniture types. Since we have no prior knowledge about the furniture distribution, we assume a uniform distribution $P_{G,0}(F) = \frac{1}{M}$, implying the least prior knowledge or the maximum entropy. As the semantic mapping process goes on over time, we expect to have a more informative probabilistic distribution $P_{G,t}(F)$ at time t, which results in a reduced entropy at that location.

Without complicated vision data processing, it is usually difficult for the robot to distinguish one object from another. The problem can be more severe when the environment is crowded with objects. However, there are other sources of information that

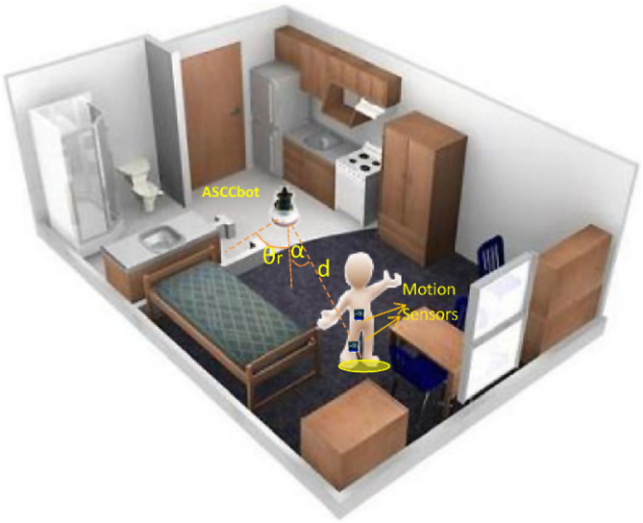


Fig. 1. The concept of robot semantic mapping through wearable sensor-based activity recognition. The robot identifies the furniture by recognizing human activities based on wearable motion sensors.

can be used to improve furniture recognition. First, in an indoor environment, the human activities are usually highly correlated with the furniture types. For example, we have the following activities associated with furniture: "*sitting*" on a "*chair*", "*lying*" on a "*bed*", etc. By recognizing the human activities around the furniture over time, the knowledge about the furniture type can be gradually learned. Second, the location of the furniture can also impose a constraint on possible furniture types. For example, a shelf or a bed is more likely to be placed against a wall. Such location information can be incorporated to improve the accuracy of semantic mapping. In this paper, we will demonstrate the feasibility of using such activity and location information for robot semantic mapping. While we currently deal with a limited number of furniture types, our framework can be extended to other object types as long as the human activities associated with these objects can be recognized.

3. Activity-based semantic mapping

In this section, we first give an overview of the proposed semantic mapping approach. Then we present the probabilistic theoretical framework used in our approach.

3.1. Overview

The proposed semantic mapping approach works as follows. A robot implements the SLAM algorithm to generate a metric map of the unknown indoor environment and localize itself. In the mean time, the robot finds the human subject's location through vision and recognizes his activity through motion sensors attached to the human subject. By fusing the location and activity information of the human subject using pre-learned models, the furniture type can be inferred. Contextual interpretation, e.g., "*This is an office room*", based on the furniture identified in the semantic map, will therefore be possible as well. It is worth noting that in the current experiment, we adopt color blob tracking through vision to simplify the human localization problem, which does not imply that we use vision for human activity recognition or object recognition. The vision based localization can be easily replaced with other human localization methods, such as using onboard lasers, wearable motion sensors, or distributed IR sensors.

The overall block diagram of the proposed semantic mapping approach is illustrated in Fig. 2. We adopt a data fusion strategy to integrate the two channels of information for semantic mapping. The first channel is human activity which can be recognized through a hierarchical recognition algorithm using data from wearable motion sensors. The second channel is the furniture location type which can be obtained through robot self localization and human detection. By fusing these two channels of information, a furniture probability distribution can be updated. This updated semantic map can also be used as the prior knowledge for the next iteration of furniture recognition, which is called *incremental map learning*. The furniture probability distribution can be used to derive the labels of the furniture.

Below we describe the detailed theoretical framework that fuses the two channels of information (activity and location) to update the furniture probability distribution.

3.2. Activity and furniture correlation

We model the relationship between the human activity and the furniture type. Let A_t denote the true human activity at time t and O_t denote the corresponding estimated activity based on the data from the motion sensors. The observation model is represented as $P(O_t|A_t)$, which gives the likelihood of getting O_t when the true activity is A_t. This model basically characterizes the accuracy of the activity recognition algorithm, which can be obtained experimentally. The activity recognition algorithm will be discussed in Section 4.

On the other hand, the human activity A is generally associated with the furniture type and this is represented by the probabilistic model $P(A|F)$. For example, when the furniture "*bed*" is given, the probability of "*lying*" and "*sitting*" is much higher than that of "*standing*". This probability model can be learned experimentally based on long term observations in human daily life. Based on the rule of total probability, we have:

$$P(O_t|F) = \sum_A P(O_t|A, F)P(A|F) = \sum_A P(O_t|A)P(A|F) \quad (1)$$

which gives the activity observation model for a given furniture type.

3.3. Location and furniture correlation

We assume that the robot can detect and localize the human subject. As shown in Fig. 1, $\theta_r(t)$ denotes the heading of the robot while $d(t)$ and $\alpha(t)$ denote the distance and bearing of the human subject with respect to the robot at time t. Then the location of the human subject $L_h(t) = [x_h(t), y_h(t)]$ can be calculated through trigonometry. The corresponding grid index is $G_t = g(L_h(t))$.

As mentioned before, there exists correlation between the location and the type of a furniture, which can be used to improve the accuracy of furniture recognition. As shown in Fig. 3, any 2D location in the map can be classified into one of five categories based on its relative positions to the walls: Type 1 (l_1): furniture is not adjacent to any wall; Type 2 (l_2): furniture is adjacent to a wall; Type 3 (l_3): furniture is around a corner; Type 4 (l_4): furniture is between two walls; Type 5 (l_5): furniture is surrounded by three walls.

Let $LT(t)$ denote the location type at time t. We have $LT(t) \in \{l_1, l_2, l_3, l_4, l_5\}$. The prior knowledge of location type to furniture correlation can be represented by the conditional probability $P(LT(t)|F)$. The distance threshold between an object and a wall can be set according to the environment, for example, 20 cm. It means that if an object is within 20 cm of a wall, it is considered to be adjacent to the wall.

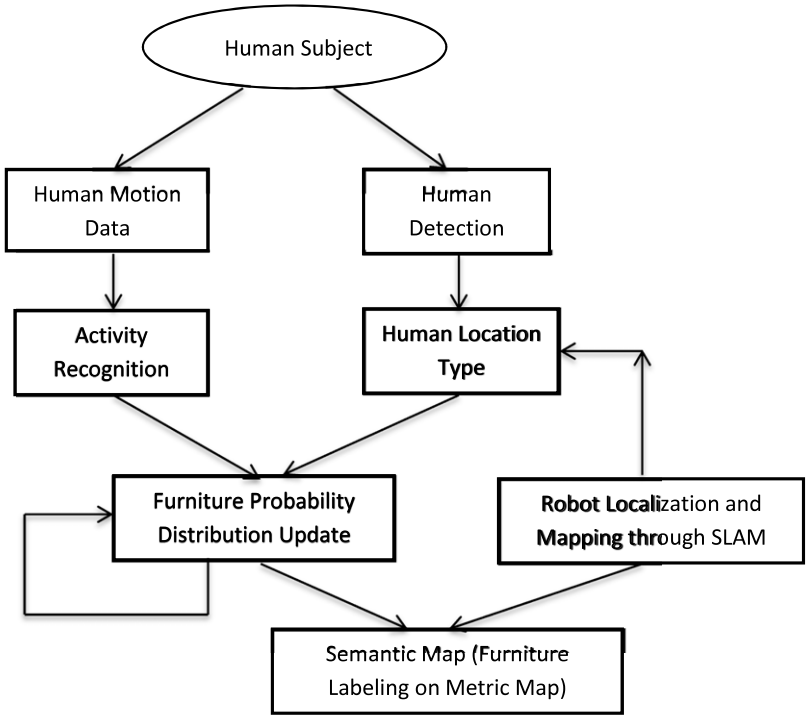


Fig. 2. The block diagram of the semantic mapping method.

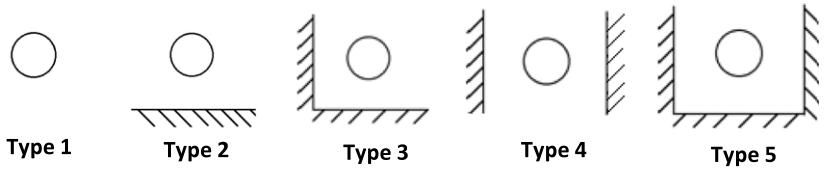


Fig. 3. Five location types. (The circle denotes the object.)

3.4. Furniture type inference

With the two channels of information, i.e., human activity and location, the knowledge regarding the furniture type can be updated. Specifically, using the Bayes rule, the posterior probability of the furniture type at time t and grid G_t can be updated as follows:

$$P_{G,t}(F|O_t, LT(t)) \propto P(O_t, LT(t)|F)P_{G,t-1}(F) = P(O_t|F)P(LT(t)|F)P_{G,t-1}(F).$$

The second equation is derived based on the conditional independence between the two channels of information. $P_{G,t-1}(F)$ is the prior knowledge of the furniture type at grid G based on all past information up to time $t-1$. $P_{G,t}(F|O_t, LT(t))$ can be written in a shorter form $P_{G,t}(F)$ for the next iteration of update.

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

$$F_{G,t} = \arg\max_F P_{G,t}(F). \quad (2)$$

4. Wearable sensor-based human activity recognition

Wearable sensor-based human activity recognition is a critical component in the proposed semantic mapping algorithm. As shown in Fig. 4, motion sensors are attached to the human subject to collect motion data and the activity recognition algorithm processes the data to identify activities. In this section, we first give an introduction to the wireless motion sensor node we developed and then we describe the activity recognition algorithm.

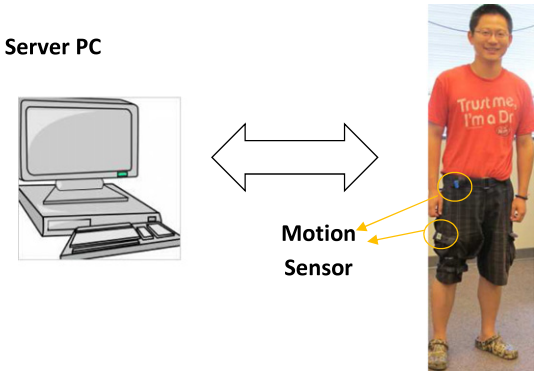


Fig. 4. The overview of the setup for daily activity recognition.

4.1. Wireless motion sensor

A compact, power-aware motion sensor is developed in our lab [31], which can sense human motion in terms of acceleration and angular rate at a sampling rate of 20 Hz. Fig. 5 shows the prototype of the sensor node, which consists of a VN-100 orientation sensor module from VectorNav, Inc. [32], an XBee RF module [33], a microcontroller, a 3-axis accelerometer and a small 3.3 V battery. The block diagram of the sensor node is shown in Fig. 6. The motion data consist of 3D orientation, acceleration, angular rate and magnetic field, which are sent to the PC through the XBee RF module. The dimension of the whole sensor node is 36 mm × 35 mm × 18 mm and the weight is about 40 g. This motion

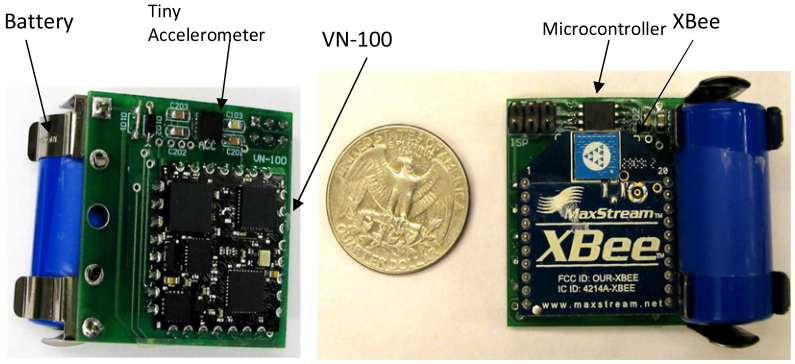


Fig. 5. The wireless motion sensor based on the VN-100 module. (Left: bottom view. Right: top view.)

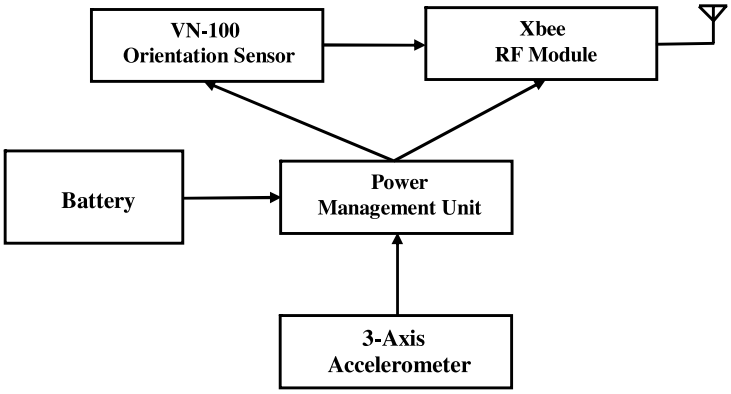


Fig. 6. The block diagram of the wearable motion sensor node.

sensor node can be used to collect motion data from various body parts of a human subject.

To extend the life time of the sensor node, a power management scheme is implemented. The basic idea behind the power management is that the VN-100 sensor node can be switched into the sleep mode when no significant motion is detected and switched back to the normal mode when there is significant motion. To facilitate this, a small low-power 3-axis accelerometer is used to detect motion or no motion, based on which the VN-100 node can be switched between the two modes. In this way, we are able to prolong the battery life of the motion sensor node from 5 to 14 h [31], which is sufficient to support daytime monitoring before recharging the battery at night.

4.2. Human activity recognition

Human activity can be recognized by machine learning algorithms that collect motion data, extract features and infer activity types. Depending on the types of activities to be recognized, a different number of motion sensors can be attached to the human body. For basic body activities such as "*sitting*", "*standing*", and "*walking*", two sensors attached to the waist and thigh are sufficient [34]. For more complex daily activities such as "*eating*" and "*cooking*", we need to attach additional sensors on the hands [35], which enable us to recognize hand gestures associated with the complex daily activities. Below we briefly describe the algorithms for recognizing body activities and hand gestures.

4.2.1. Body activity recognition

For body activity recognition, we adopt a hierarchical recognition algorithm which combines neural networks and hidden Markov models (HMM) [36]. The block diagram of the algorithm is shown in Fig. 7. There are basically two steps in the recognition algorithm: (1) coarse-grained classification and (2) fine-grained classification. In the coarse-grained classification step, raw motion data from two motion sensors on the waist and thigh respectively are processed to obtain the features (mean and variance), which are fed into the corresponding neural network NN_w and NN_t to get three types of coarse activities: *zero displacement activity*, *transitional activity* and *strong displacement activity*. A fusion module then combines the waist and thigh coarse activities to categorize the body activities [34]. In the fine-grained classification step, the sequential constraints of body activities are modeled using an HMM. A modified short-time Viterbi algorithm [37] is adopted to obtain the detailed body activity types. More details of the fine-grained classification can be found in [34].

To verify the effectiveness and evaluate the accuracy of the proposed algorithm, we conducted an experiment to recognize five body activities: "*sitting*", "*sitting-to-standing*", "*standing-to-sitting*", "*standing*", and "*walking*". For performance evaluation purpose, the human body activities are recorded by a camera as the ground truth, which are compared with the recognition results. Fig. 8 shows the activity recognition results of a 10-min experiment, which clearly demonstrates that the proposed hierarchical algorithm works effectively.

4.2.2. Hand gesture spotting and recognition

Hand gesture recognition is important to recognizing complex daily activities. Hand gestures are first spotted from other non-gesture movements. Since hand gestures exhibit different intensity levels in different complex activities, the parameters for gesture spotting have to adapt to the change of environments and body activities. For example, when a person is typing on a keyboard, the hand movement intensity is much less than that during cooking. Therefore, the classifiers need to be trained under different locations and body activities.

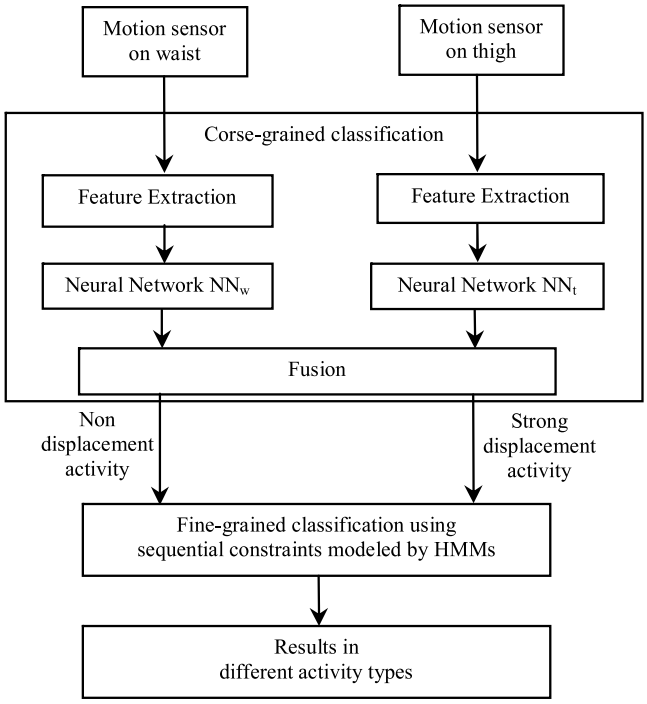


Fig. 7. The block diagram of the human body activity recognition algorithm [34].

To recognize the hand gestures, each gesture is represented by one HMM model, which is trained by a series of data recorded when the human subject repeatedly performs the same gesture. We label the data to train the HMMs. The EM (Expectation–Maximization) method is used to train the parameters of HMMs. In the recognition phase, a sliding window of 1 s length moves along the symbol sequence of the segmented gesture and the likelihood under each set of HMM parameters is estimated. We choose the model which maximizes the likelihood over other HMMs to be the recognized type as the output decision of the sliding window. Next, a decision based on majority voting is produced as the output of the HMMs for the segmented gesture. More detailed description of the hand gesture recognition algorithm can be found in our previous work [38].

5. Computer simulations

In order to evaluate the performance of the proposed semantic mapping approach, computer simulations are conducted first. In the simulation, we assume the human subject lives in an apartment as shown in Fig. 9. We consider five types of activity:

$a_1 =$ *lying on a bed*; $a_2 =$ *using a bath sink*;
$a_3 =$ *opening a refrigerator*; $a_4 =$ *eating at a table*;
$a_5 =$ *sitting on a sofa*.

Thus, $A \in \{a_1, a_2, a_3, a_4, a_5\}$ and $O \in \{a_1, a_2, a_3, a_4, a_5\}$. The activity observation model $P(O|A)$ is shown in Table 1. The probabilities in this table are obtained by observing the recognition results from the activity recognition algorithm. For example, the probability of observed activity $O = a_1$ given that the true activity $A = a_1$ (*lying on a bed*) is 0.89. Correspondingly, five types of furniture are considered here:

$f_1 =$ *bed*; $f_2 =$ *bath sink*; $f_3 =$ *refrigerator*;
$f_4 =$ *table*; $f_5 =$ *sofa*.

Thus, we have $F \in \{f_1, f_2, f_3, f_4, f_5\}$. Table 2 illustrates the relation between furniture type and activity. Here the probabilistic model $P(A|F)$ represents the likelihood of a human displaying certain hand gesture/body activities related to complex activity A when he is around furniture F. This model can be obtained through long term observation of human activities on various furniture types in real life. While for the simulation purpose, we conducted a simple experiment on five subjects in our mock apartment for a duration of 4 h and approximately determined these probability values. For example, the probability of activity $A = a_1$ (*lying on a bed*) given that furniture type $F = f_1$ (bed) is 0.60. The relation between the furniture type and the location type $P(LT|F)$ is shown

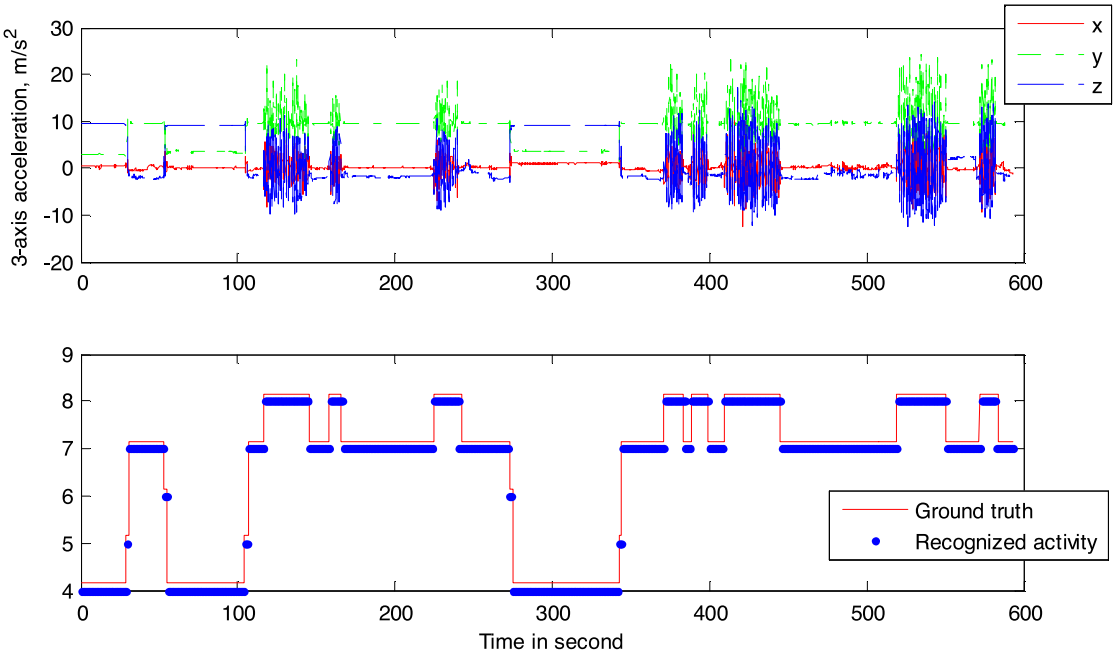


Fig. 8. The results of body activity recognition. Top: the sampled raw acceleration data from the waist sensor. Bottom: the recognized activity results. The labels are as follows: 4 sitting; 5 sitting-to-standing; 6 standing-to-sitting; 7 standing; 8 walking.

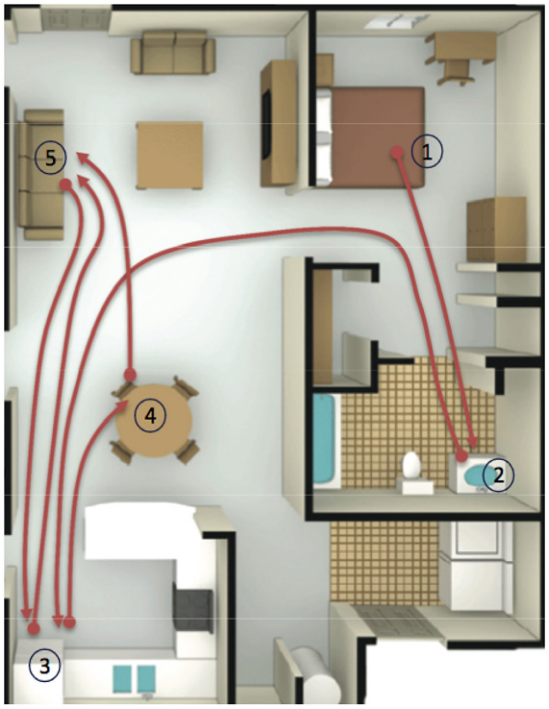


Fig. 9. The simulated activities and the corresponding furniture in an apartment.

Table 1
The activity observation $P(O|A)$.

		True activity A				
		a_1	a_2	a_3	a_4	a_5
Observed activity O	a_1	0.89	0.01	0.01	0.01	0.01
	a_2	0.01	0.80	0.14	0.01	0.02
	a_3	0.01	0.15	0.82	0.04	0.01
	a_4	0.04	0.02	0.02	0.84	0.11
	a_5	0.05	0.02	0.01	0.10	0.85

Table 2
Model $P(A|F)$.

		Furniture type F				
		f_1	f_2	f_3	f_4	f_5
Activity A	a_1	0.60	0.01	0.01	0.01	0.16
	a_2	0.01	0.65	0.30	0.03	0.02
	a_3	0.01	0.30	0.66	0.04	0.01
	a_4	0.05	0.02	0.02	0.70	0.10
	a_5	0.33	0.02	0.01	0.22	0.71

in Table 3. This correlation between furniture types and locations can be learned from observing the furniture arrangement in a sufficient number of real home environments.

The human subject conducts a sequence of the above five activities according to the order shown in Fig. 9, that is,

$$a_1(f_1) \rightarrow a_2(f_2) \rightarrow a_3(f_3) \rightarrow a_4(f_4) \rightarrow a_5(f_5) \rightarrow a_3(f_3) \rightarrow a_5(f_5).$$

Two sets of simulations are conducted. First, simulation is conducted to evaluate the incremental map learning process. Second, the performance of semantic mapping is evaluated in terms of the accuracy of furniture recognition after a certain number of visits to each furniture.

5.1. Incremental map learning

In this simulation, we evaluate how the incremental map learning process evolves as more activities are observed at the same furniture. We use entropy and mutual information as the measures for this purpose, which are defined as follows,

$H_{G,t}(F)$—the entropy of the furniture type at grid G and time t, which can be calculated as

$$H_{G,t}(F) = -\sum_F P_{G,t}(F) \log P_{G,t}(F).$$

$I_{G,t}(F)$—the mutual information gain by observing human activity at grid G and between time $t-1$ and time t, which can be calculated as

$$I_{G,t}(F) = H_{G,t-1}(F) - H_{G,t}(F).$$

The incremental map learning results are summarized in Table 4. As mentioned in Section 2, we initialize $P_{G,0}(F)$ with a uniform distribution indicating the least prior knowledge or the maximum entropy. In this table, F is the true furniture type. O is the activity observation. LT is the location type observation. $\hat{F}$ is the recognized furniture type. E is the entropy and MI is the mutual information. From this table, we can see that at time t_6 and t_7, the human subject visits the same places (refrigerator and sofa, respectively) again. After observing the same activity a second time, the entropy of the furniture type is significantly reduced. This confirms that the uncertainty regarding the type of the furniture is gradually reduced as more activities are observed at the same place.

Table 3
The location sensing model.

		Furniture type F				
		f_1	f_2	f_3	f_4	f_5
Location type LT	l_1	0.03	0.01	0.01	0.60	0.15
	l_2	0.50	0.40	0.47	0.30	0.50
	l_3	0.45	0.52	0.50	0.08	0.25
	l_4	0.01	0.05	0.01	0.01	0.06
	l_5	0.01	0.02	0.01	0.01	0.04

Table 4
Incremental map learning results.

	Time steps						
	t_1	t_2	t_3	t_4	t_5	t_6	t_7
F	f_1	f_2	f_3	f_4	f_5	f_3	f_5
O	a_1	a_2	a_3	a_4	a_5	a_3	a_5
LT	l_2	l_3	l_3	l_1	l_2	l_3	l_2
$\hat{F}$	f_1	f_2	f_3	f_4	f_5	f_3	f_5
E	1.134	1.236	1.234	0.428	1.624	**0.859**	**0.997**
MI	1.188	1.086	1.088	1.893	0.698	0.376	0.628

5.2. Accuracy of furniture recognition

To evaluate the accuracy of furniture recognition, we conducted simulations using the model parameters presented above. We assume that each furniture is visited five times. The recognition results for the five furniture types are shown in Table 5. From this table, we can clearly see that *table* has the highest accuracy due to its unique location type and the *eating* activity. On the other hand, *bath sink* and *refrigerator* have lower accuracy due to the similarity of the location type and the associated activities between them.

6. Experiments

In this section, first, we introduce the overall experimental setup and describe the robot platform in terms of hardware and software. Second, we discuss the experimental procedure. Finally, we present the obtained results.

Table 5
Confusion matrix for furniture recognition.

Decision $\hat{F}$	Furniture type F				
	f_1	f_2	f_3	f_4	f_5
f_1	0.8851	0.0018	0.0016	0.0004	0.1079
f_2	0.0009	0.7399	0.2571	0.0003	0.0020
f_3	0.0006	0.2558	0.7402	0.0003	0.0003
f_4	0.0007	0.0003	0.0004	0.9554	0.0411
f_5	0.1128	0.0023	0.0069	0.0435	0.8486
Accuracy	0.8851	0.7399	0.7402	0.9554	0.8486

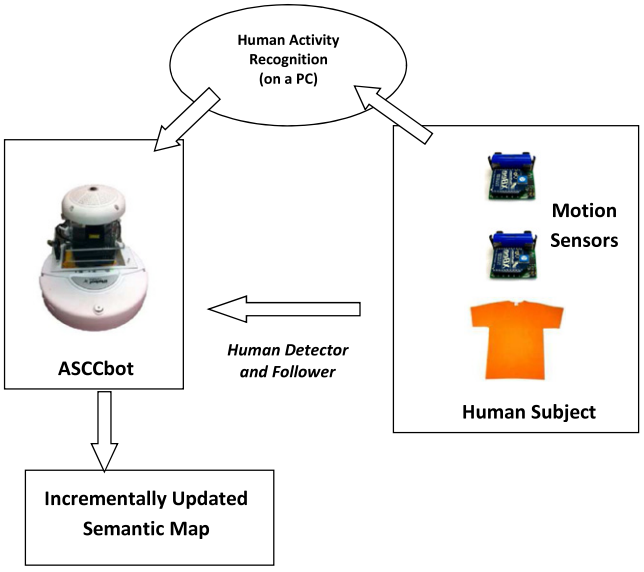


Fig. 10. The overall system setup for robot semantic mapping.

6.1. Experimental setup

Fig. 10 illustrates the experimental setup of the semantic mapping system, which mainly consists of a robot and a human subject wearing a small set of wireless motion sensors. We also use a desktop PC to run the activity recognition algorithms. The robot is the ASCCbot [39] developed in our lab, which is shown in Fig. 11. The ASCCbot is a compact, intelligent mobile platform with the following features: open-source, extendable, duplicable and equipped with various functionalities such as SLAM, human detection and human following. The ASCCbot is built on an iRobot Create [40] equipped with a mini-computer called *FitPC2* [41], a Hokuyo laser range finder (LRF) URG-04LX [42] and a Q24 panoramic camera [43]. The Q24 camera is capable of providing different views, including a panoramic view of the surrounding area. It provides a resolution up to 3 M pixels (2048 × 1536).

The software on the ASCCbot is built upon the Robot Operating System (ROS) [44], an open-source, meta-operating system for robots. It provides services such as hardware abstraction, low-level device control, message-passing between processes, and package management. Its distributed computing feature can also facilitate multi-agent applications. Several functionalities are implemented in the ASCCbot. First, the SLAM node runs in the background which creates a 2D metric map and localizes the robot. Second, a communication node is used to receive activity recognition results from the PC wirelessly. Additionally, there are two nodes that control the robot to automatically follow the human subject: the human detector and the human follower. The specific task of the human detector is to find the human subject and then calculate his location with respect to the ASCCbot.

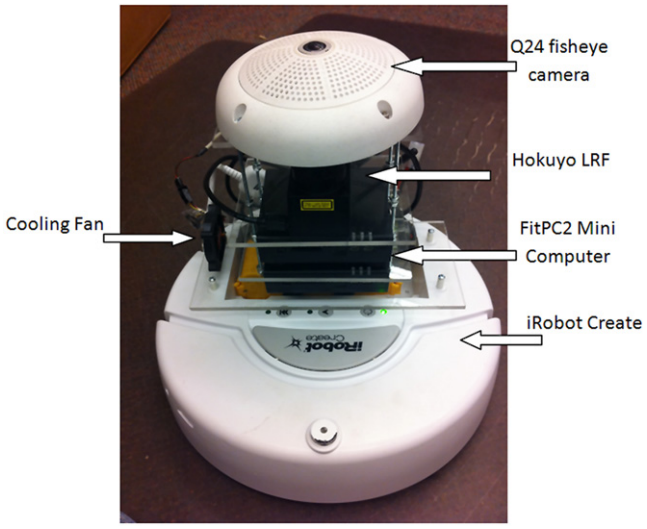


Fig. 11. The ASCCbot is used for semantic mapping.

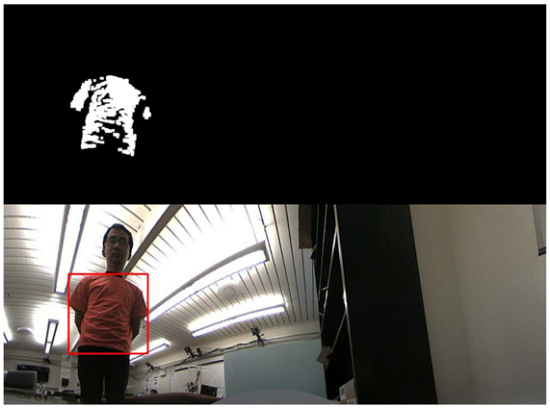


Fig. 12. Detection of a human subject using color segmentation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 13. The mock apartment setup for the first experiment.

Without losing the generality of our semantic mapping framework, we simplify the human detection problem by assuming the subject wears a colored T-shirt. In this way, the human detection problem is reduced to a color detection problem. With color segmentation, the orange T-shirt is detected as shown in Fig. 12. In the panoramic view of Q24, the angle and size of the detected region

Fig. 14. Some snapshots of the experiment. Left: sitting on a chair. Right: standing by a bookshelf.

will be output to the human follower for tracking purpose. When the human subject is performing certain activity at the location of certain furniture, the associated semantic information at that location will be updated on the metric map.

6.2. Experimental procedure

In the real experiments, we set up a mock apartment in the center of our lab. Inside the mock apartment, we deploy different types of furniture which will be recognized and labeled in the semantic map. The ASCCbot is put into the mock apartment. It creates a 2D metric map of the mock apartment using the 2D SLAM algorithm [30], follows the human subject wearing a color T-shirt, and localizes him on the 2D map. The wireless motion sensors are attached to the human subject. The raw data of the motion sensors are sent to a PC server where the activity recognition program runs. The human subject performs different activities corresponding to the furniture in the mock apartment.

The human detector and the human follower start working when the human subject appears in the view of the Q24 camera. The ASCCbot follows the human subject to the furniture location and receives the activity recognition results. When the ASCCbot obtains the pose estimate and the human activity recognition results, the semantic labels are generated according to Eq. (2), which reveals the most possible furniture type at that location. In order to better visualize the created semantic map, a 3D map of the mock apartment is created using a 6D SLAM algorithm [45] with a Kinect sensor [46]. All furniture are shown in the 3D map. We align the 2D map with the 3D map, which allows the labels of the furniture to appear at the right place on the semantic map.

6.3. Experimental results

We conducted two experiments to verify the proposed semantic mapping approach. In the first experiment, as shown in Fig. 13, four pieces of furniture were deployed in the mock apartment: a bench to mimic a bed, two chairs and a bookshelf. Two motion sensors were attached to the waist and thigh of the human subject, respectively.

The human subject conducted daily activities as follows. She first lay on the bench, until the robot arrived at the bench and updated the semantic information. Next she went to sit on the first chair (on the right side) and then stood by the bookshelf. After that, she went to sit on the second chair (on the left side). Then she returned to the first chair and its semantic label is updated to "Chair → returned". The last stop is the bookshelf, and the semantic label is updated to "Shelf → returned". Some of the snapshots of the experiment are shown in Fig. 14. The final semantic map is shown in Fig. 15 which verifies that the robot successfully derives the semantic information.

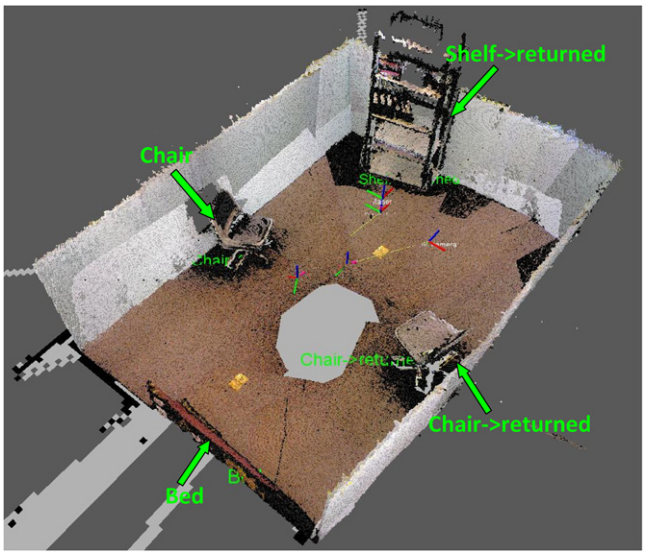


Fig. 15. The created semantic map of the mock apartment.

Fig. 16. The mock apartment setup for the second experiment.

In the second experiment, we put more furniture in the environment, which include a chair, a bench, a bookshelf, a computer desk, and a table to mimic both the kitchen and dining area. The whole setup is shown in Fig. 16. Six daily activities associated with these types of furniture are conducted: *eating at the dinning table*,

Fig. 17. Activities in the second experiment. (Top: eating, lying and cooking; Bottom: reading, sitting and using computer.)

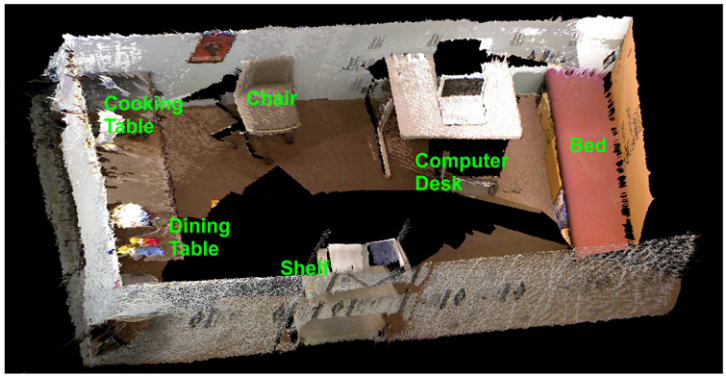


Fig. 18. The result of semantic mapping in the second experiment.

lying on the bench, *cooking in the kitchen*, *reading by the book shelf*, *sitting on the chair* and *using computer*. In order to recognize these daily activities, we used three motion sensors: one on the waist, one on the thigh and one on the right hand of the human subject. The sensor on the hand was used to distinguish several activities that involve the use of hands, such as *eating*, *cooking* or *using computer*. A detailed description of the activity recognition algorithm can be found in our previous work [35]. Some snapshots of the activities are shown in Fig. 17. The final result of the semantic mapping is shown in Fig. 18, where all the furniture pieces are correctly labeled.

6.4. Discussions

In the experiments, we have explored the use of a limited number of motion sensors to recognize several human daily activities and used them to help recognize the involved furniture. As mentioned before, wearable motion sensors have advantages over vision sensors in recognizing human activities. However, the limitation of motion sensors is that they can be obtrusive to the subject if there are many sensors attached to the human body. With a limited set of motion sensors, it may not be able to capture the full human motion, therefore limiting the number of activities and furniture that can be recognized. Also it is worth noting that the overall system performance is sensitive to the location of the sensors on the human body. It is desirable to have a more quantitative analysis on the relation between sensor locations and recognition accuracy.

In this paper, we focus on how to use the knowledge of how humans interact with the environment to help recognize the objects. It is true that many researchers have used RGB-D sensors to recognize the objects and label the scene. However, RGB-D sensors, like many other cameras, still have problems with background noise, lighting variations in home environments. In this paper, we try to provide a new perspective on inferring objects based on other important attributes, the affordance, or how humans interact with them. We believe that vision-based object recognition can be combined with our work to improve the accuracy, which is part of our future work.

7. Conclusions

In this paper, based on wearable sensor-based human activity recognition, an automated semantic mapping system is proposed. In this system, motion sensors are attached to a human subject for activity recognition. Activity observations and the location of the human subject are fused in a Bayesian framework to iteratively

update the semantic information on the metric map. The most likely furniture type is tagged to the metric map. Both computer simulations and real experiments demonstrate the effectiveness and accuracy of the proposed approach. This approach offers a new perspective for robot semantic mapping and can significantly reduce the difficulties involved in traditional vision-based object classification algorithms. In our future work, we will study how to combine both activity information and visual appearance for furniture recognition, which will make this framework work on 3D maps directly. We will also conduct experiments in a larger environment involving more complicated activities. Additionally we will investigate how to use the correlation between activity and furniture to improve the accuracy of activity recognition and the furniture recognition at the same time. This work is also part of our ongoing research that explores wearable computing technologies in human–robot interaction (HRI). We aim to develop more natural, human-centered, proactive HRI mechanisms.

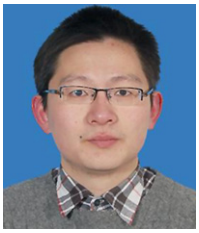
Acknowledgments

This project is supported by the National Science Foundation (NSF) Grants CISE/IIS 1231671, CISE/IIS/1427345, CISE/CNS 0916864, CISE/CNS MRI 0923238, National Natural Science Foundation of China (NSFC) Grants 61328302, 61222310 and the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT1408).

References

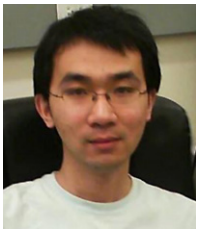
[1] C.C.Y. Weng, C. Sun, Toward the human–robot co-existence society: on safety intelligence for next generation robots, Int. J. Soc. Robot. 1 (4) (2009) 267–282.
[2] S. Thrun, Robotic mapping: a survey, in: G. Lakemeyer, B. Nebel (Eds.), Exploring Artificial Intelligence in the New Millennium, 2002.
[3] B. Kuipers, T. Levitt, Navigation and mapping in large-scale space, AI Mag. 9 (2) (1988) 28–43.
[4] D. Baker, Some topological problems in robotics, Math. Intelligencer 12 (1) (1990) 66–77.
[5] E. Menegatti, M. Wright, E. Pagello, A new omnidirectional vision sensor for the spatial semantic hierarchy, in: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2001.
[6] B. Kuipers, Modeling spatial knowledge, Cogn. Sci. 2 (1978) 129–153.
[7] R. Chatila, J. Laumond, Position referencing and consistent world modeling for mobile robots, in: IEEE International Conference on Robotics and Automation, ICRA, IEEE Computer Society Press, 1985.
[8] C. Theobalt, J. Bos, T. Chapman, A. Espinosa-romero, M. Fraser, G. Hayes, E. Klein, T. Oka, R. Reeve, Talking to godot: dialogue with a mobile robot, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2002, 2002, pp. 1338–1343.
[9] G.M. Kruijff, H. Zender, P. Jensfelt, H.I. Christensen, Situated dialogue and spatial organization: what, where…and why? Int. J. Adv. Robot. Syst. 4 (1) (2007) 125–138. Special Issue on Human–Robot Interaction.
[10] D. Gonzalez-Aguirre, J. Hoch, S. Rohl, T. Asfour, E. Bayro-Corrochano, R. Dillmann, Towards shaped-based visual object categorization for humanoid robots, 2011.
[11] H. Koppula, A. Anand, T. Joachims, A. Saxena, Labeling 3D scenes for personal assistant robots, in: Proc. RSS Workshop on RGB-D cameras, 2011.
[12] A. Pronobis, Semantic mapping with mobile robots (Ph.D. thesis), KTH Royal Institute of Technology, 2011.
[13] D.F. Wolf, G.S. Sukhatme, Semantic mapping using mobile robots, IEEE Trans. Robot. 24 (2) (2008) 245–258.
[14] S. Albrecht, T. Wiemann, M. Gunther, J. Hertzberg, Matching CAD object models in semantic mapping, in: Proc. ICRA 2011 Workshop Semantic Perception, Mapping and Exploration, 2011.
[15] S. Vasudevan, R. Siegwart, Bayesian space conceptualization and place classification for semantic maps in mobile robotics, Robot. Auton. Syst. (RAS) 56 (6) (2008).
[16] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernandez-Madrigal, J. Gonzalez, Multi-hierarchical semantic maps for mobile robotics, in: Proc. the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS05, 2005.
[17] H. Zender, O.M. Mozos, P. Jensfelt, G.-J.M. Kruijff, W. Burgard, Conceptual spatial representations for indoor mobile robots, Robot. Auton. Syst. (RAS) 56 (6) (2008).
[18] A. Rottmann, O. Martínez Mozos, C. Stachniss, W. Burgard, Semantic place classification of indoor environments with mobile robots using boosting, in: Proc. of the National Conference on Artificial Intelligence, AAAI, 2005, pp. 1306–1311.
[19] S. Vasudevan, V. Nguyen, Towards a cognitive probabilistic representation of space for mobile robots, in: IEEE International on Information Acquisition, ICIA, 2006.
[20] A. Nüchter, J. Hertzberg, Towards semantic maps for mobile robots, Robot. Auton. Syst. 56 (2008) 915–926.
[21] D. Meger, P.-E. Forssen, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J.J. Little, D.G. Lowe, C. George, An attentive semantic robot, Robot. Auton. Syst. (RAS) 56 (6) (2008).
[22] A. Nuchter, J. Hertzberg, Towards semantic maps for mobile robots, Robot. Auton. Syst. (RAS), 56 (11) (2008).
[23] J.J. Gibson, The Ecological Approach to Visual Perception, Houghton Mifflin, Boston, 1979.
[24] H. Grabner, J. Gall, L.J.V. Gool, What makes a chair a chair?, in: CVPR, 2011, pp. 1529–1536.
[25] V. Delaitre, D. Fouhey, I. Laptev, J. Sivic, A. Gupta, A. Efros, Scene semantics from long-term observation of people, in: Proc. 12th European Conference on Computer Vision, 2012.
[26] A. Gupta, S. Satkin, A.A. Efros, M. Hebert, From 3D scene geometry to human workspace, in: Computer Vision and Pattern Recognition, CVPR, 2011.
[27] A. Gupta, A. Kembhavi, L.S. Davis, Observing human–object interactions: Using spatial and functional compatibility for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (10) (2009) 1775–1789.
[28] H. Kjellström, J. Romero, D. Kragić, Visual object-action recognition: inferring object affordances from human demonstration, Comput. Vis. Image Underst. 115 (1) (2011) 81–90.
[29] S. Prince, Computer Vision: Models, Learning and Inference, Cambridge University Press, 2012.
[30] G. Dissanayake, P. Newman, S. Clark, H. Durrant-whyte, M. Csorba, A solution to the simultaneous localization and map building (SLAM) problem, IEEE Trans. Robot. Automat. 17 (2001) 229–241.
[31] S. Zhang, G. Li, W. Sheng, Development and evaluation of a compact motion sensor node for wearable computing, in: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2010.
[32] VectorNav Technologies, 2012. http://www.vectornav.com/.
[33] Digi International Inc., 2012. http://www.digi.com/.
[34] C. Zhu, W. Sheng, Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: IEEE International Conference on Robotics and Automation, 2009, pp. 2154–2159.
[35] C. Zhu, W. Sheng, Realtime recognition of complex human daily activities using human motion and location data, IEEE Trans. Biomed. Eng. 59 (9) (2012) 2422–2430.
[36] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[37] J. Bloit, X. Rodet, Short-time viterbi for online hmm decoding: evaluation on a real-time phone recognition task, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. 2008, pp. 2121–2124.
[38] C. Zhu, W. Sheng, Online hand gesture recognition using neural network based segmentation, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009, pp. 2415–2420.
[39] G. Li, J. Du, C. Zhu, W. Sheng, A cost-effective and open mobile sensor platform for networked surveillance, in: Proc. SPIE: Signal Data Process. Small Targets, 2011.
[40] IRobot inc., February 2013. http://www.irobot.com/.
[41] Fit-PC2, 2011. http://www.fit-pc.com/web/.
[42] Hokuyo laser, 2011. http://www.hokuyo-aut.jp/.
[43] Q24 camera, 2011. http://www.mobotix.com/.
[44] Robot Operating System (ROS), 2012. http://www.ros.org/wiki/.
[45] A. Nuchter, K. Lingemann, J. Hertzberg, H. Surmann, 6D slam-3D mapping outdoor environments, J. Field Robot. 24 (8–9) (2007).
[46] Kinect sensor, 2012. http://microsoft.com.

Weihua Sheng is currently an associate professor at the School of Electrical and Computer Engineering, Oklahoma State University (OSU), USA. He is the Director of the Laboratory for Advanced Sensing, Computation and Control (ASCC Lab, http://ascc.okstate.edu) at OSU. Dr. Sheng received his Ph.D degree in Electrical and Computer Engineering from Michigan State University in May 2002. He obtained his M.S and B.S. degrees in Electrical Engineering from Zhejiang University, China in 1997 and 1994, respectively. During 2002–2006, he taught in the Electrical and Computer Engineering Department at Kettering University (formerly General Motor Institute). He was promoted to associate professor there before he joined Oklahoma State University. He is the author of more than 140 papers in major journals and international conferences. Seven of them have won best paper or best student paper awards in major international conferences. His current research interests include mobile robotics, wearable computing, human–robot interaction and intelligent transportation systems. His research has been supported by US National Science Foundation (NSF), Department of Defense (DoD), Oklahoma Transportation Center (OTC)/Department of Transportation (DoT), etc. Dr. Sheng is a senior member of IEEE and served as an Associate Editor for IEEE Transactions on Automation Science and Engineering during 2010 to 2014.



Jianhao Du is a Ph.D. student at the School of Electrical and Computer Engineering, Oklahoma State University. He obtained his M.S. and B.S. degrees in Electrical Engineering from Zhejiang University, China in 2007 and 2010, respectively. His major interests lie in computer vision, camera network and mobile robotics.

Qi Cheng received the B.E. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China in July 1999. From 1999 to 2000, she worked as a System Engineer in Guoxin Lucent Technologies Network Technologies Co. Ltd., Shanghai, China. She received the M.S. and Ph.D. degrees in Electrical Engineering from Syracuse University, Syracuse, NY, in 2003 and 2006, respectively. Dr. Cheng joined Oklahoma State University, Stillwater, OK USA in 2006 and currently is an associate professor with the School of Electrical and Computer Engineering. Dr. Cheng has done extensive research in the areas of distributed detection and estimation, distributed change/fault/anomaly detection, statistical learning theory, communications and information theory and their applications to distributed sensor networks. Dr. Cheng's current area of interest mainly focuses on statistical signal processing and data fusion with applications in distributed sensor networks and wireless communications. Dr. Cheng is a Senior Member of IEEE and has served as Associate Editor for the IEEE Communications Letters since 2011. The work presented is partly supported by the Oklahoma Transportation Center and the RITA University Transportation Center Program.



Gang Li is a Ph.D. student at the School of Electrical and Computer Engineering, Oklahoma State University. His major interests lie in mobile robotics, sensor networks and human–robot interaction.

Chun Zhu received her Master's and Bachelor's degree from Tsinghua University, China. She currently is a Ph.D. candidate in the Laboratory for Advanced Sensing, Computation & Control, at Oklahoma State University, where she has worked to develop a robotic assisted living platform for the elderly care. Zhu's research interests include machine learning for recognizing and understanding human daily activity patterns, real-time measurement and analysis, embedded system application design and human–robot interaction using wearable computing.

Meiqin Liu received the B.E. and Ph.D. degrees in control theory and control engineering from Central South University, Changsha, China, in 1994 and 1999, respectively. She was a post-doctoral research fellow with the Huazhong University of Science and Technology, Wuhan, China, from 1999 to 2001. She was a visiting scholar with the University of New Orleans, New Orleans, LA, USA, from 2008 to 2009. She is currently a professor with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. She has authored more than 60 peer reviewed papers, including 33 journal papers. Her current research interests include neural network, robust control, multi-sensor network, and information fusion.

Guoqing Xu, Prof. Xu received his Ph.D. degree from Zhejiang University in 1994. He has worked as Associate Professor, Doctoral Supervisor, Director of the Department of Electrical Engineering, Director of Professional Committee of disciplines in Tongji University and Visiting Professor in The Chinese University of Hong Kong since 1997. Now he is a deputy director of CAS/CUHK Shenzhen Institute of Advanced Integration Technology and Professor of The Chinese University of Hong Kong. He has presided more than 20 important projects, which were awarded several prizes for Provincial Progress Awards. Meanwhile, Prof. Xu has published a monograph, more than 70 papers in international journals and conferences, and applied for 16 patents. His research interests include electric traction technology in vehicles, energy conversion and control technology of HEV, automobile electronic technology, vehicle intelligent technology and information processing automation and fault diagnosis technology. Now his projects are as follows: design and optimization of traction power system of high-speed EMU, motor and its control system of EV, electrical AC inverter system without position sensor, driving behavior recognition based intelligent anti-theft system and power train and integrated control system of intelligence omnibearing HEV(IOHEV).