International Journal of Automation and Computing 15(5), October 2018, 559-569
DOI: 10.1007/s11633-018-1139-6

A Selective Attention Guided Initiative Semantic Cognition Algorithm for Service Robot

Huan-Zhao Chen Guo-Hui Tian Guo-Liang Liu

School of Control Science and Engineering, University of Shandong, Jinan 250061, China

Abstract: With the development of artificial intelligence and robotics, the study on service robot has made a significant progress in recent years. Service robot is required to perceive users and environment in unstructured domestic environment. Based on the perception, service robot should be capable of understanding the situation and discover service task. So robot can assist humans for home service or health care more accurately and with initiative. Human can focus on the salient things from the mass observation information. Humans are capable of utilizing semantic knowledge to make some plans based on their understanding of the environment. Through intelligent space platform, we are trying to apply this process to service robot. A selective attention guided initiatively semantic cognition algorithm in intelligent space is proposed in this paper. It is specifically designed to provide robots with the cognition needed for performing service tasks. At first, an attention selection model is built based on saliency computing and key area. The area which is highly relevant to service task could be located and referred as focus of attention (FOA). Second, a recognition algorithm for FOA is proposed based on a neural network. Some common objects and user behavior are recognized in this step. At last, a unified semantic knowledge base and corresponding reasoning engine is proposed using recognition result. Related experiments in a real life scenario demonstrated that our approach is able to mimic the recognition process in humans, make robots understand the environment and discover service task based on its own cognition. In this way, service robots can act smarter and achieve better service efficiency in their daily work.

Keywords: Service robot, cognition computing, selective attention, semantic knowledge base, artificial neural network.

1 Introduction

Service robot has a huge market potential as it is very useful for ordinary families. There still remain a few technology challenges constraining its development like cognition obstacle. At present, the comprehension of service robot is insufficient. Normally, only simple information (like finding obstacles or planning a route) is perceived. Due to the lacking of knowledge, the functions and applications of service robot are highly restricted. Performing high level service tasks (like serving a drink or cleaning floor) requires robots fully, precisely and appropriately particularizing their basic control programs. Robot intelligence could be improved by imitating human brain system, just like the application of semantic knowledge. It is usually considered an effective way[1]. This paper focuses on the algorithm of understanding and reusing deep semantic knowledge in service robot. In this way, service robot can deeply understand the environment and provide better services. The knowledge of computing provides a solid basis for building a cognition system for service robot. Here, we are trying to transfer the human cognition process to service robot.

Intelligent space is deployed to capture information in a mutual way. It is able to improve both perception skills and executive skills of service robots. The core idea of intelligent space is to distribute the sensors and actuators in the environment. It is a space where users can interact with computers and robots, and get useful services from them. Fig. 1 illustrates the intelligent space architecture, which is utilized in order to improve the perception skill and operational skill of service robots.

Fig. 1 Overview of intelligent space system. The core idea of intelligent space is to distribute the sensors and actuators in the environment.

Accurately understanding of the world is a major topic of robotics. In our research, we propose a cognition algorithm based on both visual way and key area in intelligent space. On one hand, applying artificial neural network enables to recognize common objects from visual information. On the other hand, using key area method in

Research Article
Special Issue on Intelligent Control and Computing in Advanced Robotics
Manuscript received January 20, 2018; accepted June 4, 2018; published online September 3, 2018
Recommended by Guest Editor Jun-Zhi Yu

intelligent space is very suitable for capturing human behavior of the domestic environment. Selective visual attention is capable of directing attention rapidly towards objects of interest in the environment[2]. As a result, we propose a novel selective attention model in order to choose the useful information relative to service task while ignoring the mass of useless information. We are trying to extract semantic knowledge from the information and build a hierarchical knowledge base. In our research, knowledge is described in description logic, while representing declarative expressions using some standard ontology languages. Knowledge processing is proposed to make robots be able to bridge the gap between vague task descriptions and the detailed information needed to actually perform those tasks.

2 Related work

We present some related work which includes robot semantic knowledge, intelligent space and selective attention. The detail of related work is described in this section.

2.1 Semantic robot knowledge

In recent years, the research of robot knowledge has made a great progress. Ontology plays a critical role in representing knowledge for robots and other domains[3]. Ontology is a traditional and popular way to represent knowledge and taxonomy. It is capable of defining the structure of knowledge from many different domains. Verbs represent relations between objects and nouns represent the classes of objects. Suh et al.[4] proposed an ontology based multi-layered robot knowledge framework called OMRKF. All the classes are built in a structure of 3 ontology layers and 3 knowledge level system. Furthermore, a novel OUR-K system for robot knowledge is proposed based on OMRKF. OUR-K system describes the service robot knowledge that is required to integrate the low-level knowledge about the common objects, metric maps, perceptual features and primitive behaviours with high-level knowledge. The high-level knowledge covers the objects, semantic maps, etc. Normally, it associates with the semantic knowledge classes on service task. Moreover, this OUR-K framework enables a robot to perform some work by using some supposed simple rules. The related rules are based on semantic knowledge classes and relations among different knowledge classes[5]. Wongpatikaseree et al.[6] introduced the novel context-aware infrastructure ontology for behaviour recognition in smart home (SH) domain. This model is built for behaviour recognition in SH. Wongpatikaseree et al.[6] also proposed a method for distinguishing activities based on object-based and location-based semantic concepts.

2.2 Intelligent space

The intelligent space understands the user by detecting user′s behavior through the sensors and provides information to the user. In the target space, a user is observed by some distributed sensors (like cameras, ultrasound, microphones, etc) connected to a network. Also, the status of intelligent devices can also be recorded. At the same time, some actuators can also be manipulated to provide service to users or robot. Robot is assumed as a physical agent of the space to provide physical services for the human. Lately researchers have transferred the intelligent space from concepts of ubiquitous computing to applications in the domestic environment. On the other way, a group of actuators (like robots, personalized wheelchairs) are also installed in the environment that allows executing service task and human-machine interaction. Finally, all the sensory and actuation machines make the supervisor system. The system is capable of helping agents to analyze the cognition information and make plans[7]. There are two major approaches of intelligent space as a key component between environment and related applications. Some approaches aim at assisting cooperation between human, co-workers and robots, as well as trying to personalize the environment to provide better service. The other approaches aim at improving user′s quality of life and reducing costs of energies. In recent years, the research on service robot and intelligent space integration has made a great progress world wide[8–10]. A hybrid cloud framework based on intelligent space was proposed[11]. Cloud computing is used to improve the computing and communication skills for intelligent space. This has created a great combination of service robot and intelligent space. Due to complete ubiquitous sensor network, intelligent space is introduced to perceive environment as well as user behavior in this paper.

2.3 Selective attention model

Selective attention model is proposed originally for rapid scene analysis. This kind of model is usually inspired by the mechanism and neuron architecture of the human visual system[12]. In the traditional approaches, a series of feature maps in different scales would be composed into a unique topographical map referred as saliency map. Then, an artificial neural network is built to assemble the saliency map and find out the focus of attention (FOA). The model has also been implemented considering the difference of Gaussian filter and Gabor filter. This operation has been biologically demonstrated[13]. The traditional computing process is as follows: First, the stimuli features were extracted out of the surrounding environment. Next, the saliency map in 3-dimensional (3D) formation is built based on the previous stimuli features. The saliency of the specific environment can be represented by building a saliency map. This has been demonstrated as a regular and effective way for bottom-up direction saliency computing. Third, the inhibition of return (IoR) mechanism is applied to dynamically switch FOA

in the computing process. The result of saliency computing is referred as FOA. It could be transferred from one present place to another potential place, while the original FOA is depressed. Fourth, referring to FOA attention, an eye movement is performed. This operation determines where to look. Normally, this step is known as the most challenging problem between attention placement with related coordination system. Finally, subsequent scene understanding and object recognition takes advantage of finding FOA. So the traditional selective attention computing model normally consists of these 5 key steps[14].

A series of saliency computing models are proposed based on classical Itti's model. For example, Hou et al.[15, 16] proposed an image signature of saliency computing, referred as image descriptor. Some novel selective attention models are proposed based on new methods like deep neural network in recent years. Zhao et al.[12] proposed a multi context deep learning framework where deep neural network is used to describe saliency for salient object detection in images. Wang et al.[17] presented a saliency detection algorithm based on integrating both global search and local estimation. The deep neural network also plays a key role in the computing process[17]. For local estimation operation, local saliency is detected by using a deep neural network (DNN-L). The method collects the local patch features to determine saliency score of each pixel in images[18, 19]. Zhang et al.[20] used the saliency detection algorithm to extract a representative set of patches of salient regions. This makes an unsupervised learning framework for classification task[20]. Many successful popular models of bottom-up direction attention computing are based on the saliency map approach. The differences in these approaches are the ways on how to process the sensory information and how to extract saliency from them. For domestic service robot, there are two main shortages in traditional attention mechanism. At first, blocking is a big trouble when considering the complex environment of the domestic environment. Due to the user's movement, it is not always easy to find the saliency objects directly. Second, human behavior is a key component of robot's attention for service task. Human behavior deserves more interests in saliency computing process.

Unlike previous studies focusing on low-level cognition this paper proposes a cognitive method for acquiring high-level semantic. Intelligent space is implemented to improve the robot's ability to comprehensively perceive the environment. Selective attention is used as a prior knowledge of the cognitive process. So salient and task-related objects are initiatively selected as the input of cognition. Cognitive results are semantically represented and transferred as robot knowledge. Methods on semantic knowledge processing are applied in our research for storing and reasoning knowledge in multi domains. Research progress in related fields laid the foundation for this research.

3 A selective attention guided initiative semantic cognition algorithm

In this chapter, a selective attention guided initiative semantic cognition algorithm is proposed. First, an attention selection model is presented which aims to find FOA. FOA contains potential information about service tasks. Then, objects in FOA are recognized based on deep neural network and multi-instance learning. Cognitive results exist in the form of simple semantics. At last, a knowledge base is built to represent and store semantic knowledge. Also a corresponding reasoning engine is built. Based on semantic knowledge and cognitive result, some corresponding solutions can be produced in order to support service.

3.1 Selective attention model

Human visual system has powerful capabilities to quickly locate significant items in a complex environment. In this chapter, we build a selective attention model based on visual saliency computing and event item as the general procedure is illustrated in Fig. 2. The input information consists of three types three forms as color image, depth image and discrete sensor data. Depth image is obtained by infrared sensors from Microsoft Kinect. Color images and depth images are jointly used to compute visual saliency from the images. According to intelligent space, discrete sensor data about users or facilities can be used to generate event items.

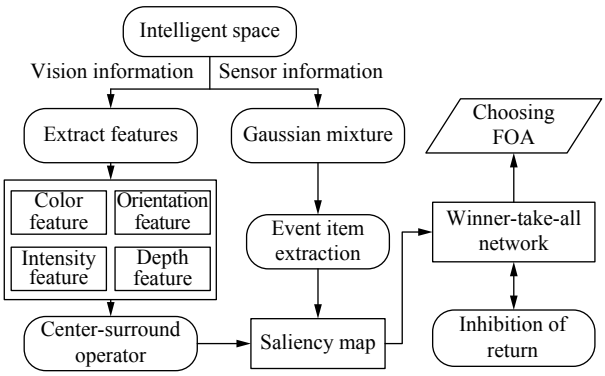

Fig. 2 Flow chart of selective attention process

3.1.1 Visual saliency computing

Color images are obtained from ceiling camera in the intelligent space platform. As r, g and b refer to the red, green and blue channels of one image. The intensity image I is produced as $\dfrac{I=(r+g+b)}{3}$. Then, a Gaussian pyramid $I(\sigma)$ is created with the scale $\sigma \in [0, 1, 2, \cdots, 8]$. We use center-surround mechanism of visual receptive field from human visual system in the research. $\ominus$ represents the cross-scale difference from different maps. This computing process is highly related the sensitive features

and showed in a series of maps $I(c,s)$, with scales c and s:

$$I(c,s) = |I(c) \ominus I(s)|. \tag{1}$$

By using Gabor filter, we represent local orientation from I as $O(\sigma,\theta)$, where $\sigma \in [0,1,2,\cdots,8]$ represents the scale and $\theta \in [0°, 45°, 90°, 135°]$ is the preferred orientation. Orientation feature maps $O(c,s,\theta)$ are encoded as a group local orientation contrast between the center and surround scales: $O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)|$.

Then, we obtain depth images based on infrared camera from Microsoft Kinect. At first, speckle pattern in different scales are captured by Kinnect. They are used as the primary reference pattern. Next, the infrared camera detects the objects and obtains its corresponding speckle pattern. The depth information is obtained by matching primary or reference pattern and using triangulation method of ground resistance test. Finally, it generates a depth image to describe the spatial information by using 3D reconstruction and local offset. The environment is demarcated in intelligent space. A group of area is preset with given weight. So when users or robots appear in the area, they will get a high score to reflect their saliency. After a normalization operation, we can get depth map m_d to describe the spatial relationship of the environment.

3.1.2 Event item saliency computing

We defined the Event item to resemble discrete sensor data on human behavior and facilities extracted from intelligent space. Fig. 3 illustrated the intelligent space platform where we can collect sensing information related to service task.

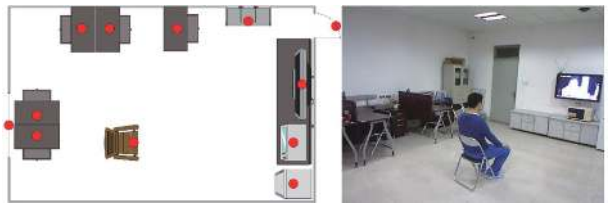

Fig. 3 Illustration of Key area in intelligent space. For a real target environment, discrete sensors (red knots) are used to capture status of users or facilities. Color versions of the figures in this paper are available online.

Discrete sensor data is transferred as event item and added to the selective attention model. The concept Key area has been proposed in our previous research on intelligent space. The area which is high related to some corresponding behavior is referred as Key area. With the help of intelligent space platform, objects in key area could be well observed. We labeled some key areas (like chairs and doors) in target environment. Event item is defined as a unique score assigned to every possible state triggered by sensor in key area. This kind of score is defined as an event item and assumed to be highly relative to human behavior. For example, the user is assumed to be cooking if he is detected near the stove. Human behavior is one of the key focuses for service robot. Robots should provide services based on their understanding of human behavior. This could be treated as prior knowledge and created as a top-down computing of attention selection for service robot. By using event item from sensor network of intelligent space, information about human behavior is obtained and mixed in saliency computing process.

When one user is active in the object environment, relative sensors could perceive his activity or relative facilities. Event items are used to describe the sensor information. The set of all event items is $\{e_1, e_2, e_3, \cdots, 8e_N\}$, suppose there are N observations. Here, we believe the output of the observation follows Gaussian mixture model M. The probability density of X is

$$p(x|M) = \sum_{n=1}^{N} \pi_n N(x|\mu_n, \sigma_n) \tag{2}$$

where π_n is mixture coefficient. $\sum_{n=1}^{N} \pi_n = 1$. μ_n and σ_n are average and variance of the n-th Gaussian component. Gaussian distribution $N(x|\mu_n, \sigma_n)$ is used to represent one specific event item. We built a feature map E to describe event items from intelligent space which are highly related to service task.

3.1.3 Saliency map generation

In this chapter, a saliency map is built by assembling feature maps to describe the saliency of one specific environment. For one specific saliency map, there is always a salient area with the maximum score. By common sense, this area deserves the focus of servers just like the definition of FOA in our research. The saliency map is built as a layer of artificial neurons.

Here we get the weight for every part of saliency amount based on the computing of entropy. The entropy can be obtained by the following equation:

$$e_s = \sum_{k=1}^{4} e_k. \tag{3}$$

The saliency map S is calculated by accumulation of static and dynamic feature map as

$$S = \frac{e_1}{e_s} m_c + \frac{e_2}{e_s} m_i + \frac{e_3}{e_s} m_o + \frac{e_4}{e_s} m_d + E. \tag{4}$$

In this way, features are accumulated by its entropy to ensure the result to be more precise and comprehensive. Event item E is also added to the computing process, so that discrete event information is also concerned in the saliency map.

3.1.4 Choosing focus of attention

Here we build a Winner-take-all (WTA) neuron network to choose the focus of attention (FOA). The neurons from WTA continuously receive excitatory inputs

from previous saliency map and are all independent from each other. If one neuron is firing and referred as a winner one, it is believed its corresponding area is salient. Next, the mechanism inhibition of return (IoR) is applied in our research to dynamically switch FOA. FOA could be shifted to the location of the winner neurons among WTA neurons while the saliency map is refreshing. This operation based on both WTA and IoR has been proved biologically plausible in human visual psychophysics.

In order to locate the FOA, the winning position:

$$FOA = \text{argmax}S. \tag{5}$$

In this way, the FOA which is highly relevant to service task can be picked out. This mechanism will be verified by experiments in the following chapter.

3.2 Recognition

In this chapter, we propose a mixed recognition method based on visual information computing and key area recognition of intelligent space. The FOA generated from the previous section is used in region proposal operation. Objects in FOA are emphasized during the recognition process. In visual recognition step, we build a convolutional auto-encoder to extract and encode features. Then a multi-instance learning classifier is proposed to recognize. By applying key area, user activities and status of facilities are used to recognize human behavior.

3.2.1 Visual recognition

We propose a visual recognition method based on a convolutional auto-encoder (CAE). Local features could be well preserved due to the weights sharing of the neural network covering all areas in CAE. The latent representation could be used for linear combination of simple image patches for reconstruction. So we build a convolutional neural network in order to capture latent representation h for describing images.

For an input image x, the encoder generates a latent representation h as k-th feature map:

$$h^k = \sigma(x * W^k + b^k). \tag{6}$$

Here b is the bias and σ is the scaled hyperbolic tangent as the activation function. The symbol $*$ denotes the 2D convolution operator. Then the reconstruction operation is

$$y = \sigma\left(\sum_{x \in H} h^k * W^k + c\right). \tag{7}$$

There is one bias c for each input channel. Here H identifies the latent feature maps and W identifies the weights in both dimensions. The cost function is the mean squared error (MSE):

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}(x_i - y_i)^2. \tag{8}$$

The training process based on BP algorithm is

$$\frac{\delta h}{\delta y} = x * \delta h^k + \bar{h}^k * \delta y \tag{9}$$

where δh and δy respectively are the deltas of the hidden states and the reconstruction.

After training process, we are trying to build multiple instance bags based on images captured by intelligent space. In this way, the recognition question is transferred into a multi instance learning (MIL) problem. For an image X, it will be input to feature extractor and turned into latent representation h. The latent representation is treated as a bag $\vec{h} = h_1, h_2, h_3, \cdots, h_n$. Features in the bag are treated as the instance.

Then, an image instance library is built, which is relevant to specific objects in domestic environment. Fig. 4 illustrates the image library and takes service robot as an example. A group of images is captured from different shooting angle or distance. The images are captured in local environment, so they contain local features like lighting or background. Then the images are input to feature extractor mentioned previously to achieve latent representation.

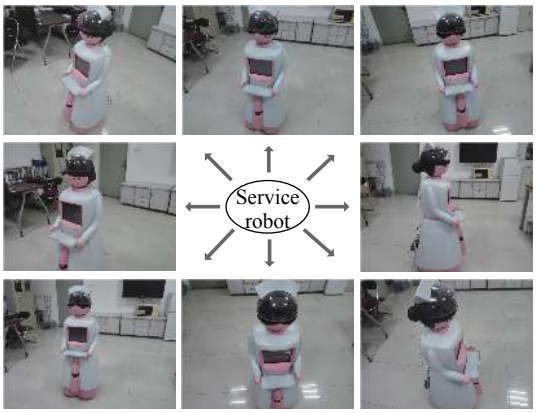

Fig. 4 Illustration of instances for service robot. For each specific object, the image is captured under different conditions.

Suppose there are N bags $\{B_1, B_2, \cdots, B_N\}$, the i-th bag is composed of $a(i)$ instances $\{B_{11}, B_{12}, B_{13}, \cdots, B_{1a}\}$. Each instance B_{ij} is a d dimensional feature vector $[B_{ij1}, B_{ij2}, B_{ij3}, \cdots, B_{ijd}]^{\mathrm{T}}$, where its mark set is $\Gamma = \{l_1, l_2, l_3, \cdots, l_N\}$. Here we mark space as $\psi = \{positive, negative\}$, Instance space Ω is $\{B_1, B_2, \cdots, B_N\}$. So, the training dataset is

$$D = \langle B, \Gamma \rangle = \{\langle B, l_i \rangle | i = 1, 2, 3, \cdots, N\}. \tag{10}$$

For an known instance space Ω, bags $B_i = \{B_1, B_2, \cdots, B_N\}$ are consist of instances, where

$i = 1, 2, 3, \cdots, N$. The training dataset $D = \langle B, \Gamma \rangle$ is

$$f : \{B_1, B_2, B_3, \cdots, B_N\} \to \psi. \tag{11}$$

The goal of the instance learning is to build a classifier $\hat{f}$ which aims at predicting the labels of new bags. A normal back propagation neural network is built to be the classifier. At first, the global error function at the level of the bags is defined as

$$E = \sum_{i=1}^{N} E_i \tag{12}$$

where E_i is the error in B_i. Assume that $B_i = +$ means B_i is a positive bag, while $B_i = -$ means a negative bag. The error is defined as follows:

$$E_i = \begin{cases} \min\limits_{i \le j \le M_j} E_{ij}, & \text{if } B_i = + \\ \max\limits_{i \le j \le M_j} E_{ij}, & \text{if } B_i = -. \end{cases} \tag{13}$$

The error E_{ij} for each instance is defined as

$$E_{ij} = \begin{cases} 0, & \text{if } (B_i = +) \text{ and } (0.5 \le o_{ij}) \\ 0, & \text{if } (B_i = -) \text{ and } (o_{ij} \le 0.5) \\ \dfrac{1}{2}(o_{ij} - 0.5)^2, & \text{otherwise} \end{cases} \tag{14}$$

where o_{ij} is the actual output of B_{ij}.

Next, we apply back-propagation algorithm to train the neural network. The proposed neural network already contained p input nodes, one output node and one hidden layer. Here we choose sigmoid function as the activation function. For each epoch, the training bags are processed and fed into the network one by one. E_{ij} could be produced according to (14) when the instance B_{ij} is fed. If E_{ij} is zero for a positive bag B_i, all the rest instances of B_i are not fed to the network in this epoch and the weights in the network will not be updated for B_i. Otherwise E_i is computed according to (13). The weights in the network would be updated according to its updating mechanism after all instances of B_i are fed. Then $B_{i,j+1}$ is fed to the network and the training process is repeated until the global error E decreases to some preset threshold or the number of epoches increases to some preset threshold.

3.2.2 Key area recognition

In this chapter, we build a human behavior classification method in Key area. Human behavior is believed highly related to his location, as well as the status of corresponding facilities. Take an office scene as an example. One user is assumed to be intended to leave the room with high probability if he is detected near a door. If the user is near a sofa, he might intend to take a break. In our research, we have built an intelligent space platform which is capable of capturing both human location and the status of facilities. Next a spatial grid has been installed in the same environment with intelligent space. The user trajectory in the space can be described using sequences of grid. The key area recognition process is illustrated in Fig. 5.

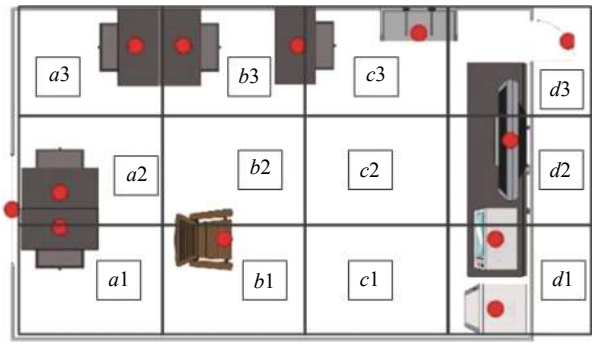

Fig. 5 Illustration of intelligent space grid installation. By region division using grids, human trajectory can be measured.

We built a set of $\Omega_p = p_1, p_2, p_3, \cdots, p_n$, which represents the sequences in the spatial grid. We also build $\Omega_s = s_1, s_2, s_3, \cdots, s_k$ to represent the status of facilities like TV, refrigerator, etc. Intelligent space could record user's movement if he is wandering in target environment. Then we build a rule base to classify the human behavior and to generate corresponding semantic representation. For a fixed environment, a group of proper rules can accurately reveal human behavior and improve recognition efficiency. The rules are defined as

$$\{\boldsymbol{p}, \pm s\} \to behavior. \tag{15}$$

The detailed examples can be seen in Table 1. In this way, the sensor information directly triggers corresponding behavior in specific scenario.

Table 1 Examples of rule base

Number	Status sequences	Behavior
1	$\{a_3, b_3, c_3, d_3, door\_on\}$	Exit
2	$\{d_3, c_3, b_3, a_3, table01\_on\}$	Work
3	$\{a_3, a_2, b_2, b_1, tv1\_on\}$	Entertain
4	$\{b_1, b_2, a_2, a_3, table01\_on\}$	Work

4 Knowledge storage and semantic reasoning

In previous chapter, objects are well recognized and transferred into semantic format. In this chapter, we describe the overall knowledge by building a knowledge base which is illustrated in Fig. 6. There are two main components as class and instance to represent knowledge. They are arranged in a taxonomic structure and in hundreds of classes. Users could extend the set of ontologies

by deriving new classes from old ones. The class level contains abstract terminological knowledge in a taxonomic structure like the types of objects, events and actions. The instances represent concrete physical objects or actually performed actions. Then the properties are used to link between classes and instances. All relations are in (Subject, Object, Property) triples formation. We define three ontology components as *environment*, *service* and *robot*.

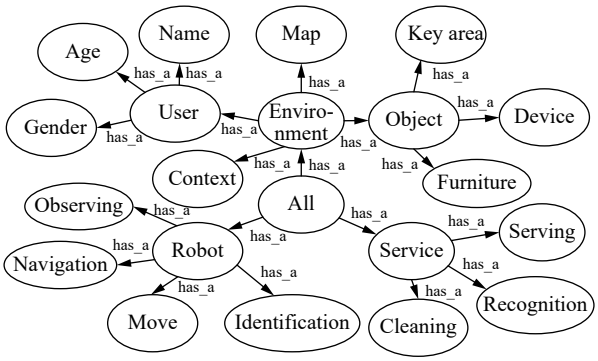

Fig. 6 Illustration of knowledge base hierarchy

In our approach, the class of *environment* includes some key components for service task like *users*, *objects*, *contexts*, *behaviour*, *intelligent space*, etc. *objects* describes the classes of objects with their semantic properties as well as the knowledge about how to recognize and manipulate them. An object is described by its features like color, material, shape, especially the function is also attached to the specific object. The *map* class contains a map of the environment. In our work, the map is mixed with topological map and semantic map. The topological map in our approach composes of a few nodes or links in order to describe the environment. Maps can be created by many different kinds of sensors like laser sensors and cameras, which could be stored in 2D or 3D format. Our approach provides several classes to describe different kinds of maps. The semantic map consisted of localized object instances. The robot is able to use the *map* to locate objects and could be updated. The links and areas can be arranged for robot path planning. The *context* class contains some context knowledge for service robot. The behaviour class contains behaviour knowledge for users in the environment. The *user* class contains some knowledge for common users and guests, including their identification knowledge, personal hobbies and medicine knowledge. The *intelligent space* class contains some knowledge about facilities in the intelligent space platform.

The class of *service* contains some key components of semantic description for service robot knowledge recognition and execution. These classes are similar to a dictionary for describing service tasks. By connecting with service class service robot is able to obtain the sequences of steps to perform a specific task. For example, an action navigation may have the properties like *fromLocation* and *toLocation*. This process is realized by class restriction. Moreover, this class is independent of robot, which makes the ontology more unified and generic between types of robots.

The class of *robot* describes the semantic properties of the robot including the identification components or actions for each specific robot. In our research, a robot is defined to be a physical agent which is equipped with sensors and actuators. The class *component* covers hardware (sensors and actuators) and software (related programs) of robots. The class *identification* describes identification of the robot (like functions and hardware configurations). The class of *actions* describes actions intuitively performed by robots to execute tasks or shift the scenes.

Then we build a rule based reasoning engine based on description logic (DL). Fig. 7 illustrates the reasoning mechanism. In this paper, DL is used to design rules containing two functions as computing class affiliation and computing the most specific classes an individual belongs to. By using rules, we can obtain semantic knowledge on high-level environment or users from cognitive result, and then find service need and make service plans. Cognitive results are used to design a specific semantic query in the beginning, as service robot treat the users as core of the service. The query is mainly directed at the user's status or intention. Then, it is sent to the knowledge engine and transferred into query statement in query analysis step. Against specific query, knowledge about user's demand and corresponding service can be assembled. Next, for a specific robot to provide the service, a detailed solution on how to perform the task will be generated. At last, the solution is sent back to service robots for guiding robot to provide service.

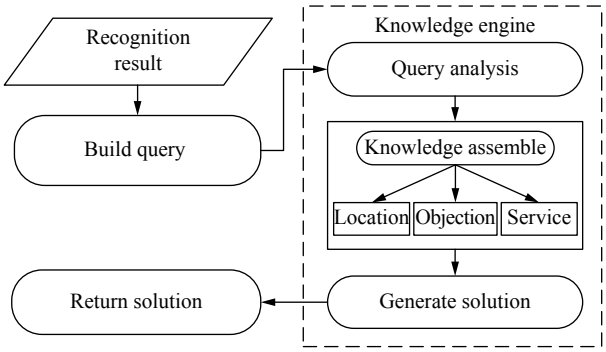

Fig. 7 Illustration of reasoning mechanism

5 Experiments

A set of experiments in real scenario are performed in order to verify the selective attention guided initiative semantic cognition algorithm. At beginning we built an intelligent space platform located in a real scenario. Sensors

(like switch sensors, motion sensors and pressure sensors) are installed around the key areas such as the television, refrigerator, window, door, computer desk and the chairs. We perform an experiment using the intelligent space platform to test the general performance of this algorithm.

Selective attention experiment. The selective attention experiment was first performed. Fig. 8 illustrates the performance of this environment. In first two rows, color images and depth images are captured in the scenario of family life. The third row are saliency maps generated by the proposed saliency computing method in Section 3. The saliency part from both image and event can be revealed in the saliency part. Hence, we draw a red rectangle around the saliency part of the saliency map in the fourth column.

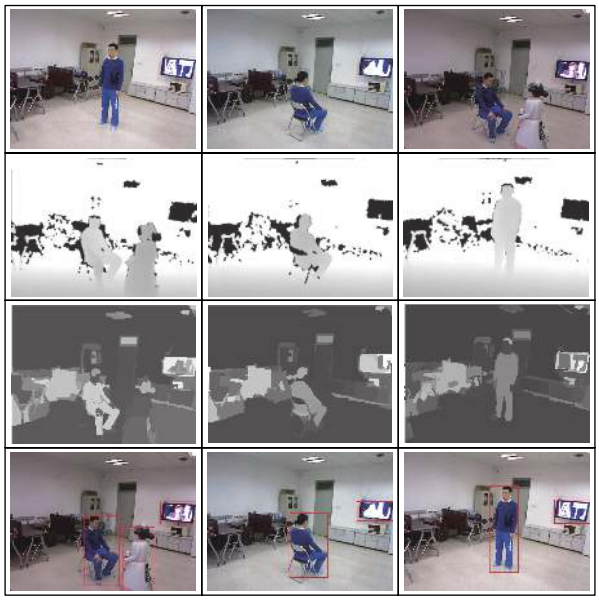

Fig. 8 Experiment on selective attention model

From Fig. 8, the most obvious part in the environments is chosen to be FOA. We can see that user, TV and service robot are chosen to be FOA. It is obvious that these objects are all highly relative to service task. As other information can be ignored, the cognition system can be more effective. According to Fig. 8, we can see that our approach is capable of computing the saliency of the environment and making a right choice by common sense.

Recognition experiment. The recognition method based on artificial neural network and key area is proposed to recognize objects. We built an instance image library for the personalized environment. A group of instances based on users or common objects in service scenario are produced and included in the library. For a specific object, we obtain a set of images at different angles and with different environments. A few instances of the common objects in this paper are illustrated in Fig. 9.

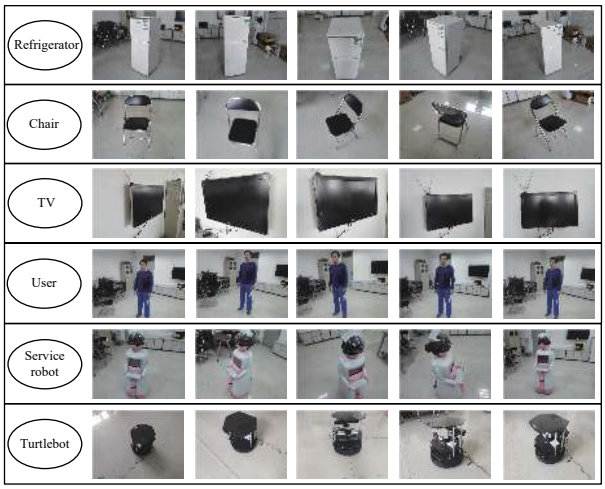

Fig. 9 Demos of instance library

Instance library contains common objects in a personalized environment. Saliency parts of the image are picked out and recognized through selective attention model. The result is illustrated in Fig. 10. It shows that user, service robot and facilities can be recognized correctly in the image. Due to the guidance of selective attention model, only part of the targets associated with the service task are selected and recognized. The results show that the method is capable of choosing the salient things relative to service task and recognize them.

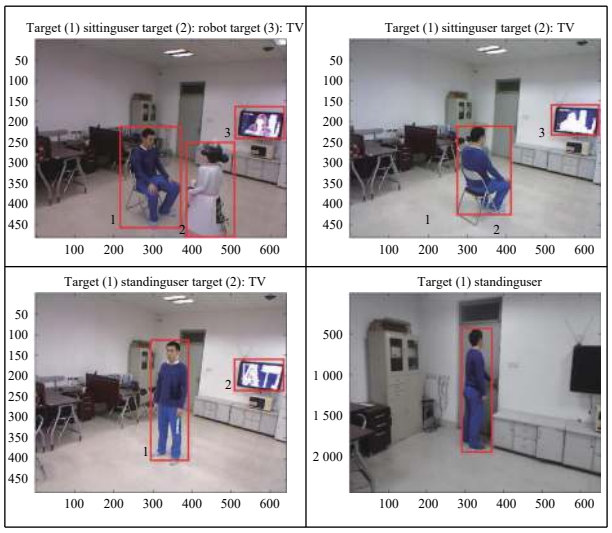

Fig. 10 Demonstration of recognition

Table 2 illustrates summary of the comparative experiment results. It contains the results from both our approach and traditional approaches (like R-CNN, fast R-CNN and faster R-CNN). Due to the attention selection operation, only useful objects related service work is transferred to recognition process. This operation not only increases the accuracy, but also decreases the processing time. The result shows that our approach is cap-

Table 2 Summary of experimental results for recognition

Methods	Target number	Precision	Time
This paper	5	99.2%	0.17s
R-CNN	20	46.2%	35s
Fast R-CNN	23	55.5%	0.49s
Faster R-CNN	22	59.8%	0.28s

able of recognizing a few useful objects in a short time.

Reasoning experiment. At last, we built a knowledge graph based on the taxonomy described previously. Fig. 11 illustrated the process of reasoning experiment. The knowledge graph mainly includes some aspects of users, service, robot and corresponding relations. Then, we applied a reasoning engine which aims at utilizing knowledge to produce solutions. Suppose the *user*1 is found to be in the room and sitting on a specific chair *chair*1, then service robot should understand user status and then go close to offer some drink. First, the user behavior is recognized, as well as the status of TV. We can see that the user is in the state *user_is_watching_Tv* and needs some drink. The service *serve_water* is required. Components of this specific task on *find_cup*, *delivery_cup* and *to_user* are assembled. The robot could get a service solution about locations, destination, objects, etc. Finally, the robots prepare to offer service. Based on the specific function of service robot, motions and actions (like *delivery*) could be produced. Then, detailed actions are produced and *robot*1 is able to execute specialized motions (like *locate*, *move_to*) to finish service task.

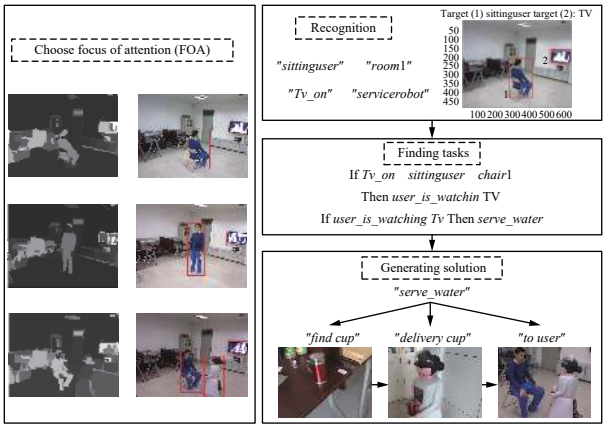

Fig. 11 Demonstration of finding a task

6 Conclusions

The service robot urgently needs to perceive the external environment and actively provide service operations based on the result of understanding. This greatly improves the robot′s intelligence and service capabilities. The cognitive ability of robot′s is still a bottleneck. Simulating human biological cognitive systems to improve the cognitive ability of robots is an important challenge. This helps to significantly improve the robot′s ability to process information in a complex home environment for cognition. We are trying to simulate human cognition on serving robots with the help of intelligent space. In order to enhance the robot′s service function and overcome cognitive bottlenecks, we propose a selective attention guided initiative semantic cognition algorithm.

Comparing with traditional research on robot cognition, the main contributions of this paper are as follows. At first, selective attention model is applied in our research to guide the recognition process. On one hand, it could handle the complexity of the perceptual information. A mechanism to locate the image regions for further processing is considered as potential work. On the other hand, it is capable of supporting action decisions and plans. Robots face similar domestic environment as humans do: Move to one specific location and carry some objects in a given time. So that a mechanism that finds the potential parts of the environment and decides what to do next is crucial. Since robots usually operate in fixed domestic environment as humans, it is reasonable to mimic the human attention system to execute these tasks. This kind of operation could greatly improves the efficiency of recognition skills in service robots. Secondly, we proposed a selective attention guided recognition algorithm based on mixing of MIL and key area. In this way robots are capable of simultaneously recognizing objects or users in one specific image. Due to the relationship between human behavior and his spatial position, methods on key area in intelligent space are able to recognize user behavior accurately and timely. At last, we developed a semantic knowledge representation for a service robot to assist in the development and certification of effective foundations for sensing, mobility, SLAM, planning and interaction with users in daily life. Furthermore, the semantic knowledge covers service, objects and environments. This algorithm makes it universal in different domains in daily life. A series of rules are also proposed describing relations between ontology. The class of service plays a key role in integration of different aspects in the environment and users. As is known, the major function of service robot is serving the users. Deep understanding skills based on semantic cognition could enable robots provide smarter and high-level service.

There still remain a few shortages in our research. As our approach relies on intelligent space, we need to verify our approach in a given environment with intelligent space platform. However, the experiment scenario is in real-life. But, the scenario and background are still comparatively simple. The ability of extracting semantic knowledge in our approach is inadequate. So, the robot is smart enough to initiatively find knowledge not completely by itself. In the future, more biological theories and new discoveries will be implemented in our research

to optimize the cognition process. In that way, we will build a model to capture high-level intention of the user and make corresponding plans. As we have built a knowledge base, we will propose a top-down knowledge driven attention selection process. In that way, we can take priori knowledge into service task. Referring new methods like LSTM and recurrent neural network (RNN), we will propose new methods to obtain semantic knowledge with more details and being in a higher level.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 61773239, 91748115 and 61603213), Natural Science Foundation of Shandong Province (No. ZR2015FM007), and Taishan Scholars Program of Shandong Province.

References

[1] T. J. Huang. Imitating the brain with neurocomputer a "New" way towards artificial general intelligence. *International Journal of Automation and Computing*, vol. 14, no. 5, pp. 520–531, 2017. DOI: 10.1007/s11633-017-1082-y.

[2] X. L. Fu, L. H. Cai, Y. Liu, J. Jia, W. F. Chen, Z. Yi, G. Z. Zhao, Y. J. Liu, C. X. Wu. A computational cognition model of perception, memory, and judgment. *Science China Information Sciences*, vol. 57, no. 3, pp. 1–15, 2014. DOI: 10.1007/s11432-013-4911-9.

[3] H. Guan, H. J. Yang, J. Wang. An ontology-based approach to security pattern selection. *International Journal of Automation and Computing*, vol. 13, no. 2, pp. 168–182, 2016. DOI: 10.1007/s11633-016-0950-1.

[4] I. H. Suh, G. H. Lim, W. Hwang, H. Suh, J. H. Choi, Y. T. Park. Ontology-based multi-layered robot knowledge framework (OMRKF) for robot intelligence. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, USA, pp. 429–436, 2007. DOI: 10.1109/IROS.2007.4399082.

[5] G. H. Lim, I. H. Suh, H. Suh. Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 3, pp. 492–509, 2011. DOI: 10.1109/TSMCA.2010.2076404.

[6] K. Wongpatikaseree, M. Ikeda, M. Buranarach, T. Supnithi, A. O. Lim, Y. S. Tan. Activity recognition using context-aware infrastructure ontology in smart home domain. In *Proceedings of the 7th International Conference on Knowledge, Information and Creativity Support Systems*, IEEE, Melbourne, Australia, pp. 50–57, 2012. DOI: 10.1109/KICSS.2012.26.

[7] J. H. Lee, N. Ando, H. Hashimoto. Intelligent space for human and mobile robot. In *Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Atlanta, USA, pp. 784, 1999. DOI: 10.1109/AIM.1999.803269.

[8] K. Morioka, H. Hashimoto. Appearance based object identification for distributed vision sensors in intelligent space. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, pp. 199–204, 2004. DOI: 10.1109/IROS.2004.1389352.

[9] P. Steinhaus, M. Strand, R. Dillmann. Autonomous robot navigation in human-centered environments based on 3D data fusion. *Eurasip Journal on Advances in Signal Processing*, vol. 2007, Article number 86831, 2007. DOI: 10.1155/2007/86831.

[10] C. Losada, M. Mazo, S. Palazuelos, D. Pizarro, M. Marron. Multi-camera sensor system for 3D segmentation and localization of multiple mobile robots. *Sensors*, vol. 10, no. 4, pp. 3261–3279, 2010. DOI: 10.3390/s100403261.

[11] H. Z. Chen, G. H. Tian, F. Lu, G. L. Liu. A hybrid cloud robot framework based on intelligent space. In *Proceedings of the 12th World Congress on Intelligent Control and Automation*, IEEE, Guilin, China, pp. 2996–3001, 2016. DOI: 10.1109/WCICA.2016.7578487.

[12] R. Zhao, W. L. Ouyang, H. S. Li, X. G. Wang. Saliency detection by multi-context deep learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Boston, USA, pp. 1265–1274, 2015. DOI: 10.1109/CVPR.2015.7298731.

[13] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1165, 1985. DOI: 10.1364/JOSAA.2.001160.

[14] L. Itti, C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001. DOI: 10.1038/35058500.

[15] X. D. Hou, L. Q. Zhang. Saliency detection: A spectral residual approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007. DOI: 10.1109/CVPR.2007.383267.

[16] X. D. Hou, J. Harel, C Koch. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012. DOI: 10.1109/TPAMI.2011.146.

[17] L. J. Wang, H. C. Lu, X. Ruan, M. H. Yang. Deep networks for saliency detection via local estimation and global search. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 3183–3192, 2015. DOI: 10.1109/CVPR.2015.7298938.

[18] T. S. Chen, L. Lin, L. B. Liu, X. N. Luo, X. L. Li. DISC: Deep image saliency computing via progressive representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1135–1149, 2016. DOI: 10.1109/TNNLS.2015.2506664.

[19] J. T. Pan, E. Sayrol, X. Giro-I-Nieto, K. McGuinness, N. E. O'Connor. Shallow and deep convolutional networks for saliency prediction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 598–606, 2016. DOI: 10.1109/CVPR.2016.71.

[20] F. Zhang, B. Du, L. P. Zhang. Saliency-guided unsupervised feature learning for scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015. DOI: 10.1109/TGRS.2014.2357078.

Huan-Zhao Chen received the B. Sc. degree from Binzhou University, China in 2011, and the M. Sc. degree from Beijing University of Technology, China in 2014. He is currently a Ph. D. degree candidate in School of Control Science and Engineering, Shandong University, China.

His research interests include service robot, robot recognition, semantic knowledge processing and reasoning.

E-mail: drwonkaa@gmail.com

ORCID iD: 0000-0003-4667-1291

Guo-Hui Tian received the B. Sc. degree from Department of Mathematics, Shandong University, China in 1990, the M. Sc. degree from Department of Automation, Shandong University of Technology, China in 1993, and the Ph. D. degree from School of Automation, Northeastern University, China in 1997. He studied as a post doctoral researcher in School of Mechanical Engineering of Shandong University from 1999 to 2001, and studied as a visiting professor in Graduate School of Engineering of Tokyo University, Japan from 2003 to 2005. He was a lecturer from 1997 to 1998 and an associate professor from 1998 to 2002 in Shandong University, China. At present, he is a professor in School of Control Science and Engineering, Shandong University, China. And also he is the vice director of the Intelligence Robot Specialized Committee of Chinese Association for Artificial Intelligence, the vice director of the Intelligent Manufacturing System Specialized Committee of Chinese Association for Automation, and the member of the IEEE Robotics and Automation Society.

His research interests include service robot, intelligent space, cloud robotics, and brain-inspired intelligent robotics.

E-mail: g.h.tian@sdu.edu.cn (Corresponding author)
ORCID iD: 0000-0001-8332-3064

Guo-Liang Liu received the B. Sc. degree from Shandong Normal University, China in 2005, the M. Sc. degree from National University of Defense Technology, China in 2007, and the Ph. D. degree from University of Goettingen, Germany in 2012. He studied as a post-doctoral researcher in School of Control Science and Engineering, Shandong University from 2014 to 2016. At present, he is an associate professor in School of Control Science and Engineering, Shandong University, China.

His research interests include service robot, intelligent space, and SLAM.

E-mail: liuguoliang@sdu.edu.cn