The Fourteenth International Symposium on Artificial Life and Robotics 2009 (AROB 14th '09),
B-Con Plaza, Beppu, Oita, Japan, February 5 - 7, 2009

Understanding User Commands by Evaluating Fuzzy Linguistic Information Based on Visual Attention

Buddhika Jayasekara Keigo Watanabe Kiyotaka Izumi
Department of Advanced Systems Control Engineering
Saga University
Saga 840-8502, Japan.

Abstract

This paper proposes a method for understanding the user commands based on visual attention. Visual attention system is implemented to evaluate the fuzzy linguistic information based on the environmental conditions. It is assumed that the corresponding distance value for a particular fuzzy linguistic command depends on the spatial arrangement of the surrounding objects. A fuzzy logic based voice command evaluation system (VCES) is proposed to assess the uncertain information in user commands. A situation of object manipulation for rearranging the users working space is simulated to illustrate the system. It is demonstrated with PA-10 robot manipulator.

1 Introduction

Natural human-human like interaction plays a major role in human friendly robotic system. Voice communication is significant in human-robot interaction. Human user may use subjective, uncertain information to convey their idea. Therefore, the ability of the robot companion to understand the uncertain information is crucial in effective human-robot interaction. An intelligent service robot would increase the functional capacity of the aged and possibly even improve their state of health. In addition, a successful human-friendly robot equipped with human-like voice communication capabilities will be able to help disabled people, to help in complex tasks such as surgery, etc. [1], [2].

The ability to understand the fuzzy linguistic information plays a key role in human-robot interaction. In Pulasinghe et al. [3], robot controlling by using rich voice commands such as "move little right" has been studied. Jayawardena et al. [4] proposed a natural language command based robot learning method using fuzzy coach-player systems. In evaluating the fuzzy linguistic information, they have assumed that the actual amount traversed as the response to a distance command depends on the distance traversed immediately before that. In addition, the system of understanding and quantifying the uncertain information in voice commands is predetermined. However, the ability to adapt the system for understanding the fuzzy linguistic information based on the environment conditions is vital. In addition, human learner may adjust their output for instructions including uncertain information by acquiring the environmental conditions through visual attention.

The capability to acquire the environmental conditions through vision is important in human-robot interaction. In addition, object identification plays a major role in an interaction with the environment. The object features and corresponding lexical symbols are taught to achieve the object identification. The natural language object references, which consist of combination of lexical symbols such as "small green cube," have been studied [5]. A vision system is studied to identify and locate objects for object manipulation tasks in natural, domestic environments [6]. But, these methods still fall sort of a system, which can evaluate the fuzzy linguistic information by acquiring environmental conditions through visual attention.

Therefore, this paper proposes a method to understand the fuzzy linguistic information by acquiring the environmental conditions through visual attention. The system overview is discussed in section 2. Next section 3 discusses the visual attention system for acquiring the environmental conditions. The evaluation process of voice commands is presented in section 4. Finally, summary is presented.

2 System Overview

The functional overview of the system is shown in Fig. 1. The system consists of fuzzy logic based voice command evaluation system (VCES), visual attention system (VAS), and interaction manager (IM). The voice interaction between human user and the robot is managed by the IM. The voice recognition and understanding are implemented by the IBM ViaVoice software development kit. Conversational grammar patterns and basic dialogue phrases are

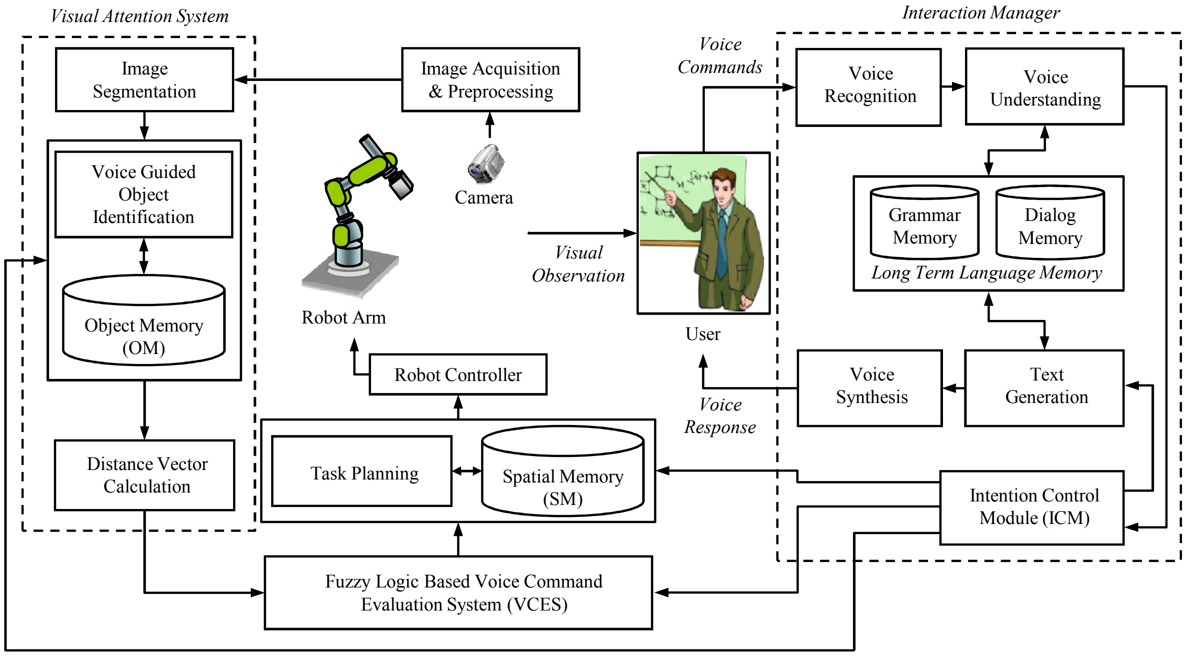

Figure 1: System overview.

stored in the long term language memory. The speech synthesis is a text-to-speech (TTS) conversion, which is implemented by using the Microsoft speech SDK.

The VAS is used to percept the spatial data of surrounding objects. VCES is introduced to evaluate the fuzzy linguistic information in the user commands. A task planner is deployed to identify the primitive behavior sequence, which is required to fulfill the task. A spatial memory (SM) is proposed to remember the corresponding position control vectors for visual-motor mapping. It is implemented by a neural map with a competitive layer [7]. Finally, the task planner guides the robot controller with the support of SM.

The presented system is capable of moving an existing object based on the user voice commands. In addition, new objects can be taught to the system by interactive dialogue.

3 Visual Attention System

VAS is introduced to capture the spatial arrangement of the objects in the surrounding area. First, the work space images are captured using a camera and they are preprocessed to remove the irregularities. Then, images are segmented to extract the objects. The corresponding object is identified based on the lexical symbol, which is included in the user command [5]. The object memory (OM) is used to store the direct mapping between the lexical symbol of the object and the feature vector. Here, average RGB color values and the Hu descriptors are considered as the feature set. The space around the corresponding object is divided into four main regions to compute the distance vector as in Fig. 2. The neighborhood is identified based on a ratio ζ, which determines the neighborhood region:

$$\zeta = \frac{r}{r} \quad (1)$$

Here, r is the radial distance to the nearest object in the region and r is the corresponding neighborhood. The average distance to the surrounding objects d_{avg} is calculated based on the average distances of each region as in (2):

$$d_{avg} = \psi^T d \quad (2)$$

where

$$\begin{aligned} \psi &= (\psi_1, \psi_2, \psi_3, \psi_4)^T, \ \sum_{i=1}^{4} \psi_i = 1 \\ d &= (d_1, d_2, d_3, d_4)^T \end{aligned}$$

Here, d_i, where $i = 1, 2, 3, 4$, is the average distance to the surrounding objects in the neighborhood of region i. ψ consists of the corresponding weighting factors for the regions. ψ_j, where $j = 1, 2, 3, 4$, is the corresponding weight for region j. In addition, the distance to the nearest object in the target direction X and the distance to the farthest object from all regions D are also obtained.

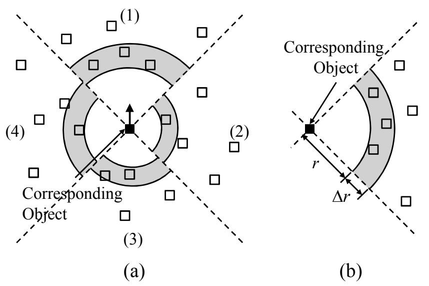

Figure 2: (a) and (b) illustrate the spatial arrangement of the corresponding object and the neighbor objects. Here, □s represent the neighbor objects of the corresponding object ■.

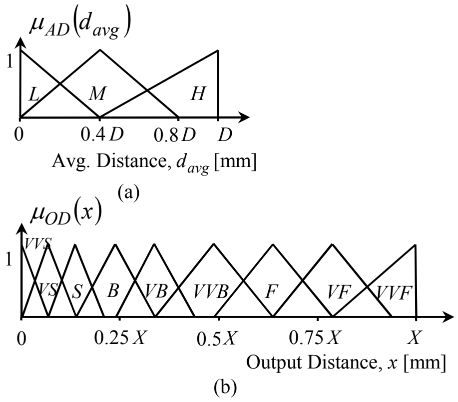

Figure 3: (a) and (b) represent the membership functions for average distance d_{avg} and output distance x. Fuzzy labels are defined by, L: Low, M: Medium, H: High for d_{avg} and VVS: Very Very Small, VS: Very Small, S: Small, B: Big, VB: Very Big, VVB: Very Very Big, F: Far, VF: Very Far, VVF: Very Very Far for x.

4 Voice Command Evaluation

VCES is implemented by using a fuzzy inference system to evaluate the fuzzy linguistic information in user commands based on the spatial arrangement of the surrounding objects. Here, it is assumed that the corresponding distance value for a particular fuzzy linguistic command depends on the spatial arrangement of the surrounding objects in the environment. The user command and the average distance to the surrounding objects are the inputs of VCES. The output is the corresponding output distance x_{out} for the user command. The membership functions for the average distance to the surrounding objects $\mu_{AD}(d_{avg})$ and the output distance $\mu_{OD}(x)$ are shown in Fig. 3. Here, the input spaces of $\mu_{AD}(d_{avg})$ and $\mu_{OD}(x)$

Input Memberships		Average Distance		
		Low	Medium	High
User Command	Very Little	*VVS*	*VS*	*S*
	Little	*S*	*B*	*VB*
	(Medium)	*VB*	*VVB*	*F*
	Far	*F*	*VF*	*VVF*

Figure 4: User commands are interpreted based on these fuzzy rules. Fuzzy labels are similar to that in Fig. 3.

Table 1: Basic actions and fuzzy predicates.

Action	Fuzzy predicate
move right	very little
move left	little
move forward	(medium)
move backward	far

are adjusted based on D and X respectively. This yields the adaptation of the system towards the environmental conditions. The considered rule base is given in Fig. 4. The possible set of actions and fuzzy predicates for action modification are shown in Table 1. The user command is structured by including the action and action modification as "Move" + <Lexical symbol of object> + <Fuzzy predicate> + <Direction component of action> (e.g. "Move blue Box very little backward").

5 Summary

The proposed system was implemented based on PA-10 robot manipulator. A working space of 0.5×0.8 m was used to manipulate the objects. A camera image of 1280×960 resolution was captured for each situation. The parameters of VAS are chosen as $\zeta = 0.5$ and $\psi = (0.3, 0.2, 0.3, 0.2)^T$. Here, ψ is selected by considering the natural human tendency. Normally, humans pay more attention to the objects in the moving direction. It is simulated by selecting higher values for ψ_1 and ψ_3. Different visual situations are achieved by changing the position of the idle objects in the environment and they are illustrated in Fig. 5. A set of user commands and the corresponding output movements are shown in Table 2. (a)–(d) visual situations are used to highlight the effectiveness of capturing the spatial arrangement by VAS. A situation with few actual objects is considered in (f).

A method has been proposed to understand the fuzzy

Table 2: User commands and corresponding output distances.

	User command	X (mm)	D (mm)	Average distances to objects (mm)				d_{avg} (mm)	x_{out} (mm)
				d_1	d_2	d_3	d_4		
(a)	Move blue Box little right	108.4	265.5	227.9	166.7	197.2	143.0	189.4	32.2
(b)		104.5	250.4	197.8	148.7	126.0	142.8	155.4	30.1
(c)		56.2	230.3	128.5	169.1	195.9	131.4	157.4	16.5
(d)		158.1	305.4	235.4	99.2	253.1	110.9	188.6	45.5
(e)	Move blue Box very little right	140.4	245.2	225.5	137.8	186.8	160.6	183.3	15.8
	Move blue Box right	140.4	245.2	225.5	137.8	186.8	160.6	183.3	81.5
	Move blue Box left	135.3	245.2	186.8	160.6	225.5	137.8	183.3	78.6
	Move blue Box little forward	70.9	245.2	160.6	225.5	137.8	186.8	172.0	20.9
(f)	Move Biscuit packet far left	81.5	402.7	236.4	204.1	328.4	261.6	262.6	72.9
	Move Biscuit packet little forward	112.3	402.7	204.1	328.4	261.6	236.4	252.7	32.4

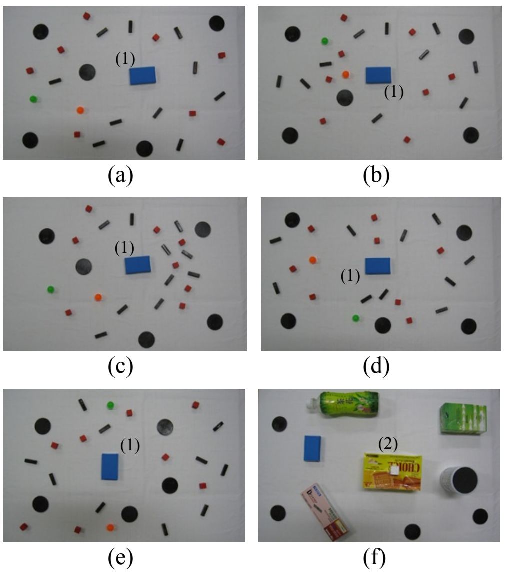

Figure 5: (a)–(f) represent up camera view of visual situations. The movement of blue Box (1) is considered in (a)–(e) and the movement of Biscuit packet (2) is considered in situation (f). Other idle objects are used to change the surrounding environment by changing the positions.

linguistic information by visual attention. The proposed system is an effective method to manipulate the objects based on user commands, which includes fuzzy linguistic information.

References

[1] C. Crangle and P. Suppes, *Language and Learning for Robots*, CSLI Publications, Stanford, CA, 1994.

[2] P. Menzel and F. D'Aluisio, *Robosapiens*, The MIT Press, England, 2000.

[3] K. Pulasinghe, K. Watanabe, K. Izumi, and K. Kiguchi, "A modular fuzzy-neuro controller driven by spoken language commands," *IEEE Trans. on Systems, Man, and Cybernetics-Part B:Cybernetics*, vol. 34, pp. 293–302, 2004.

[4] C. Jayawardena, K. Watanabe, and K. Izumi, "Knowledge acquisition by a sub-coach in a coach player system for controlling a robot," in *Proc. of the 4th Intl. Conf. on Advanced Mechatronics*, Hokkaido, Japan, Oct. 2004, pp. 601–606.

[5] C. Jayawardena, K. Watanabe, and K. Izumi, "Learning of object identification by robots commanded by natural language," in *Proc. of the 9th Intl. Conf. on Intelligent Autonomous Systems*, Tokyo, Japan, Mar. 2006, pp. 913–920.

[6] D. Kragic, M. Bjorkman, H.I. Christensen, and J. Eklundh, "Vision for robotic object manipulation in domestic settings," *Robotics and Autonomous Systems*, vol. 52, pp. 85–100, 2005.

[7] B. Jayasekara, K. Watanabe, and K. Izumi, "Controlling a robot manipulator with fuzzy voice commands guided by visual motor coordination learning," in *Proc. of the Intl. Conf. on Instrumentation, Control and Information Technology*, Tokyo, Japan, Aug. 2008, pp. 2540–2544.