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

Problem finding and solving based on Mental-image Description Language, L_{md}

Masao Yokota, Taiki Tomezuka, Yusuke Takubo, Tetsushi Oka and Kaoru Sugita

Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka-shi, 811-0295, Japan
(Tel : 81-92-606-5897; Fax : 81-92-606-8923)
(yokota@fit.ac.jp)

Abstract: Mental Image Directed Semantic Theory has proposed a formal language named “Mental-image Description Language, L_{md}”. This language can provide intermediate knowledge representation and has already been applied to the integrated multimedia understanding system IMAGES-M that can perform cross-media translation, question-answering, etc. This paper describes a systematic method based on L_{md} for intelligent robots to find and solve problems in their environments.

Keywords: Knowledge representation language, Problem finding, Problem solving.

I. INTRODUCTION

In recent years, there have been developed various types of real or virtual robots as artificial partners. However, they are to play their roles according to programmed actions to stimuli and have not yet come to perform as natural as ordinary people. In order to realize such artificial partners, it is very important to develop a systematically computable knowledge representation language [1] as well as connectionism-based technologies for unstructured data processing [2]. This type of language is indispensable to *knowledge-based* processing such as *understanding* sensory events, *planning* appropriate actions and *knowledgeable* communication even with ordinary people, and therefore it needs to have at least a good capability of representing spatiotemporal events that correspond to humans’ or robots’ sensations and actions in the real world. Most of conventional methods have provided robotic systems with such quasi-natural language expressions as ‘move(*Velocity*, *Distance*, *Direction*)’, ‘find(*Object*, *Shape*, *Color*)’, etc. for human instruction or suggestion, uniquely related to computer programs for deploying sensors/motors as their semantics [e.g., 3,4]. These expression schemas, however, are too linguistic or coarse to represent and compute sensory/motory events. This is also the case for AI planning (“action planning”) which deals with the development of representation languages for planning problems and with the development of algorithms for plan construction [5].

In order to solve this problem, Yokota, M. has employed the formal language so called ‘Language for Mental-image Description (L_{md})’ proposed in his original semantic theory ‘Mental Image Directed Semantic Theory (MIDST)’ [e.g., 6]. L_{md} was originally proposed for formalizing the natural semantics, that is, the semantics specific to humans, but it is general enough for the artificial semantics, that is, the semantics specific to each artificial device such as robot. This language has already been implemented on several types of computerized intelligent systems [e.g., 7] and there is a feedback loop between them for their mutual refinement, unlike other similar ones [e.g., 8].

This paper describes a systematic method based on L_{md} for intelligent robots to find and solve problems in their environments.

II. BRIEF DESCRIPTION OF L_{md}

MIDST treats word meanings in association with mental images, not limited to visual but omnisensory, modeled as “Loci in Attribute Spaces”. An attribute space corresponds with a certain measuring instrument just like a barometer, a map measurer or so and the loci represent the movements of its indicator. A general locus is to be articulated by “Atomic Locus” and formalized as the expression (1). This is a formula in many-sorted first-order predicate logic, where “L” is a predicate constant with five types of terms: “Matter” (at ‘x’ and ‘y’), “Attribute Value” (at ‘p’ and ‘q’), “Attribute” (at ‘a’), “Event Type” (at ‘g’) and “Standard” (at ‘k’).

$$L(x,y,p,q,a,g,k) \qquad (1)$$

This formula is called ‘Atomic Locus Formula’ and its intuitive interpretation is given as follows, where ‘matter’ refers to ‘object’ or ‘event’.

“Matter ‘x’ causes Attribute ‘a’ of Matter ‘y’ to keep (p=q) or change ($p \neq q$) its values temporally (g=Gt) or spatially (g =Gs) over a time-interval, where the values ‘p’ and ‘q’ are relative to the standard ‘k’.”

When $g=Gt$, the locus indicates monotonic change or constancy of the attribute in time domain and when $g=Gs$, that in space domain. The former is called a temporal event and the latter, a spatial event.

For example, the motion of the ‘bus’ represented by S1 is a temporal event and the ranging or extension of the ‘road’ by S2 is a spatial event whose meanings or concepts are formalized as (2) and (3), respectively, where the attribute is "physical location " denoted by 'A12'.

(S1) The bus runs from Tokyo to Osaka.

(S2) The road runs from Tokyo to Osaka.

(∃x,y,k)L(x,y,Tokyo,Osaka,A12,Gt,k)∧bus(y) (2)

(∃x,y,k)L(x,y,Tokyo,Osaka,A12,Gs,k)∧road(y) (3)

In order to represent both logical and temporal relations between loci, MIDST has employed 'tempo-logical' connectives [9] such that are defined by (4), where t_i, c and K refer to one of the temporal relations indexed by 'i', locus, and an ordinary binary logical connective such as the conjunction '∧', respectively.

$$c_1 K_i c_2 \leftrightarrow (c_1 K c_2) \grave{U} t_i(c_1, c_2) \quad (4)$$

Table 1 shows the definition of τ_i, where the conventional 13 types of temporal relations between two intervals are discriminated by the suffix 'i'(-6≤ i ≤6). For example, the expression (5) is the conceptual description of the English verb "fetch", depicted in Fig.4, implying such a temporal event that 'x' goes for 'y' and then comes back with it, where 'Π(=∧$_0$)' and '•(=∧$_1$)' are instances of the tempo-logical connectives, 'SAND' and 'CAND', standing for "Simultaneous AND" and "Consecutive AND", respectively. In general, a series of atomic locus formulas with such connectives is called 'Locus formula'.

(λx,y)fetch(x,y)↔(λx,y)(∃p1,p2,k)L(x,x,p1,p2,A12,Gt,k)•((L(x,x,p2,p1,A12,Gt,k)ΠL(x,y,p2,p1,A12,Gt,k))
∧x≠y∧p1≠p2 (5)

As indicated by the underline at (5), an event 'fetch(x,y)' is necessarily *finished by* an event 'carry(x,y)'. This fact can be formulated as (6), where '$\supset_{-4}$' is the '*implication* (⊃)' furnished with the temporal relation '*finished-by* (τ_{-4})' (See Table 1).

$$\text{fetch}(x,y) \supset_{-4} \text{carry}(x,y) \quad (6)$$

Table 1. List of temporal relations†

Definition of τ_i		Allen's notation [10]
$t_{11}=t_{21} \wedge t_{12}=t_{22}$	$\tau_0(\chi_1, \chi_2)$	equals(χ_1,χ_2)
	$\tau_0(\chi_2, \chi_1)$	equals(χ_2, χ_1)
$t_{12}=t_{21}$	$\tau_1(\chi_1, \chi_2)$	meets(χ_1, χ_2)
	$\tau_{-1}(\chi_2, \chi_1)$	met-by(χ_2, χ_1)
$t_{11}=t_{21} \wedge t_{12}<t_{22}$	$\tau_2(\chi_1, \chi_2)$	starts(χ_1, χ_2)
	$\tau_{-2}(\chi_2, \chi_1)$	started-by(χ_2, χ_1)
$t_{11}>t_{21} \wedge t_{12}<t_{22}$	$\tau_3(\chi_1, \chi_2)$	during(χ_1, χ_2)
	$\tau_{-3}(\chi_2, \chi_1)$	contains(χ_2, χ_1)
$t_{11}>t_{21} \wedge t_{12}=t_{22}$	$\tau_4(\chi_1, \chi_2)$	finishes(χ_1, χ_2)
	$\tau_{-4}(\chi_2, \chi_1)$	finished-by(χ_2, χ_1)
$t_{12}<t_{21}$	$\tau_5(\chi_1, \chi_2)$	before(χ_1, χ_2)
	$\tau_{-5}(\chi_2, \chi_1)$	after(χ_2, χ_1)
$t_{11}<t_{21} \wedge t_{21}<t_{12} \wedge t_{12}<t_{22}$	$\tau_6(\chi_1, \chi_2)$	overlaps(χ_1, χ_2)
	$\tau_{-6}(\chi_2, \chi_1)$	overlapped-by(χ_2, χ_1)

†The durations of χ_1 and χ_2 are [t_{11}, t_{12}] and [t_{21}, t_{22}], respectively.

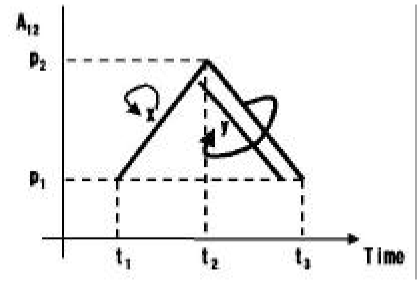

Fig.4. Locus of 'fetch'.

III. COMPUTATION ON L_{md}

Our intelligent system IMAGES-M [7] working on L_{md} is one kind of expert system equipped with five kinds of user interfaces for multimedia communication, that is, Sensor Data Processing Unit (SDPU), Speech Processing Unit (SPU), Image Processing Unit (IPU), Text Processing Unit (TPU), and Action Data Processing Unit (ADPU) besides Inference Engine (IE) and Knowledge Base (KB). Each processing unit in collaboration with IE performs mutual conversion between each type of information medium and locus formulas.

The fundamental computations on L_{md} by IMAGES-M are to detect semantic anomalies, ambiguities and paraphrase relations. These are performed as inferential operations on locus formulas at IE. Detection of semantic anomalies is very important to avoid succession of meaningless computations or actions. For an extreme example, consider such a report from certain sensors as (7) represented in L_{md}, where '…' and 'A29' stand for descriptive omission and the attribute 'Taste'. This locus formula can be translated into the English sentence S3 by TPU, but it is semantically anomalous because a 'desk' has ordinarily no taste.

(∃x,y,k)L(y,x,Sweet,Sweet,A29,Gt,k)∧desk(x) (7)

(S3) The desk is sweet.

These kinds of semantic anomalies can be detected in the following processes.

Firstly, assume the postulate (8) as the commonsense or default knowledge of "desk", stored in KB, where 'A39' is the attribute 'vitality'. The special symbol '*' represents 'always' as defined by (9), where 'ε(t1,t2)' is a simplified atomic locus formula standing for time elapsing with an interval [t1,t2]. Furthermore, '_' and '/' are anonymous variables employed for descriptive simplicity and defined by (10) and (10'), respectively.

(λx) desk(x) ↔ (λx) (…L*(_,x,/,/,A29,Gt,_)∧… ∧ L*(_,x,/,/,A39,Gt,_) ∧ …) (8)

X*↔(∀t1,t2)XΠε(t1,t2) (9)

X(_)↔ (∃u) X(u) (10)

X(/)↔ ~(∃u) X(u) (10')

Secondly, the postulates expressed by (11) and (12) in KB are utilized. The formula (11) means that ***if one of two loci exists every time interval, then they can coexist***. The formula (12) states that ***a matter has never different values with a standard of an attribute at a time***.

$$X \wedge Y^* . \supset . X \Pi Y \quad (11)$$

$$L(x,y,p,q,a,g,k) \Pi L(z,y,r,s,a,g,k) . \supset . \ p=r \wedge q=s \quad (12)$$

Lastly, IE detects the semantic anomaly of "sweet desk" by using (8)-(12). That is, the formula (13) below is finally deduced from (8)-(12), which violates the postulate (8), that is, " *Sweet* ≠ / ".

$$L(\_,x,Sweet,Sweet,A29,Gt,\_) \Pi L(z,x,/,/,A29,Gt,\_) \quad (13)$$

These processes above are also employed for dissolving syntactic ambiguities in people's utterances such as S4. IE rejects 'sweet desk' and eventually adopts 'sweet coffee' as a plausible interpretation.

(S4) Bring me the coffee on the desk, which is very sweet.

If multiple plausible interpretations of a text or another type of information are represented in different locus formulas, it is semantically ambiguous. In such a case, IMAGES-M will ask for further information in order for disambiguation.

Furthermore, if two different representations are interpreted into the same locus formula, they are paraphrases of each other. The detection of paraphrase relations is very useful for deleting redundant information, for cross-media translation, etc. [11].

IV. PROBLEM FINDING & SOLVING

Problems can be classified roughly into two categories as follows [11].

(CP) Creation Problem:
House building, food cooking, etc.

(MP) Maintenance Problem:
Fire extinguishing, room cleaning, etc.

An MP is relatively simple one that a robot can find and solve autonomously while a CP is relatively difficult one that is given to the robot, possibly, by humans and to be solved in cooperation with them.

A robot must determine its task to solve a problem in the world. The robot needs to interpolate some transit event X_T between the two events, namely, 'Current Event (X_C)' and 'Goal Event (X_G)' as shown in (14).

$$X_C \bullet X_T \bullet X_G \quad (14)$$

According to this formalization, a problem X_P is defined as $X_T \cdot X_G$ and a task for the robot is defined as its realization.

The events in the world are described as loci in certain attribute spaces and a problem is to be detected by the unit of atomic locus. For example, employing such a postulate as (15) implying 'Continuity in attribute values', the event X in (16) is inferred as (17).

$$L(x,y,p1,p2,a,g,k) \bullet L(z,y,p3,p4,a,g,k) . \supset . p3=p2 \quad (15)$$

$$L(x,y,p1,p2,a,g,k) \bullet X \bullet L(z,y,p3,p4,a,g,k) \quad (16)$$

$$L(z',y,p2,p3,a,g,k) \quad (17)$$

Consider a verbal command such as S5 uttered by a human. Its interpretation is given by (18) as the goal event X_G concerning the attribute of 'Height (A03)'. If the current event X_C is given by (19), then (20) with the transit event X_T underlined can be inferred as the problem corresponding to S5.

(S5) Keep 'balloon C7' flying 5-6 meters high.

$$L(z,C7,q,q,A03,Gt,k) \wedge balloon(C7) \wedge 5m \le q \le 6m \quad (18)$$

$$L(x,C7,p,p,A03,Gt,k) \wedge balloon(C7) \quad (19)$$

$$\underline{L(z1,C7,p,q,A03,Gt,k)} \bullet L(z,C7,q,q,A03,Gt,k) \wedge balloon(C7) \wedge 5m \le q \le 6m \quad (20)$$

For this problem, the DIRN is to execute a task deploying a certain height sensor and actors 'z1' and 'z'. The selection of 'z' is a task in case of MP described below while the actor 'z1' is selected as follows:

If 6m-p <0 then z1 is a sinker, otherwise
if 5m-p >0 then z1 is a raiser, otherwise
5m£p£6m and no actor is deployed as z1.

The goal event X_G for an MP is that for another CP such as S5 given possibly by humans and solved by the robot in advance. That is, the task in this case is to autonomously restore the goal event X_G created in advance to the current event X_C as shown in (21), where the transit event X_T is the reversal of such X_{-T} that has been already detected as 'abnormal' by the DIRN.

For example, if X_G is given by (18) in advance, X_T is also represented as the underlined part of (20) while X_{-T} as (22). Therefore the task here is quite the same that was described in the previous section.

$$X_G \bullet X_{-T} \bullet X_C \bullet X_T \bullet X_G \quad (21)$$

$$L(z1,C7,q,p,A03,Gt,k) \wedge balloon(C7) \quad (22)$$

At present, IMAGES-M, installed on a personal computer, can deploy SONY AIBOs, dog-shaped robots, as actors and gather information about the physical world through their microphones, cameras and tactile sensors. Communications between IMAGES-M and humans are performed though the keyboard, mouse, microphone and multicolor TV monitor of the personal computer.

Consider such a verbal command as S6 uttered to the robot, SONY AIBO, named 'John'.

(S6) John, walk forward and wave your left hand.

Firstly, late in the process of cross-media translation from text to AIBO's action, this command is to be interpreted into (23) with the attribute 'shape (*A11*)' and the values '*Walkf-1*' and so on at the standard of '*AIBO*', reading that John makes himself walk forward and wave his left hand. Each action in AIBOs is defined as an ordered set of shapes (i.e., time-sequenced snapshots of the action) corresponding uniquely with the positions of their actuators determined by the rotations of the joints. For example, the actions 'walking forward (*Walkf*)' and 'waving left hand (*Wavelh*)' are defined as (24) and (25), respectively.

L(John,John,Walkf-1,Walkf-m,A11,Gt,AIBO) ∧
L(John,John,Wavelh-1,Wavelh-n,A11,Gt,AIBO) (23)

Walkf={Walkf-1,Walkf-2,…,Walkf-m} (24)

Wavelh={Wavelh-1,Wavelh-2,…,Wavelh-n} (25)

Secondly, an AIBO cannot perform the two events (i.e., actions) simultaneously and therefore the transit event between them is to be inferred as the underlined part of (26) which is the goal event here.

L(John,John,Walkf-1,Walkf-m,A11,Gt,AIBO)•
<u>L(John,John,Walkf-m,Wavelh-1,A11,Gt,AIBO)</u>•
L(John,John,Wavelh-1,Wavelh-n,A11,Gt,AIBO) (26)

Thirdly, (27) is to be inferred, where the transit event, underlined, is interpolated between the current event and the goal event X_G (=(26)).

L(John,John,p1,p2,A11,Gt,AIBO)
<u>• L(John,John,p2,Walkf-1,A11,Gt,AIBO)•</u> X_G (27)

Finally, (26) is interpreted into a series of the joint angles in the AIBO. Figure 9 shows AIBO's standing up and turning right that cannot be done simultaneously.

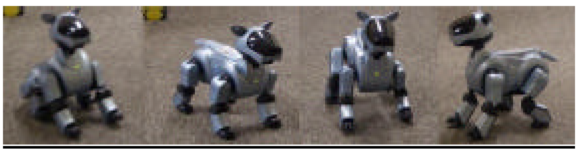

Fig.9. AIBO behaving in accordance to the command 'Stand up and turn right.'

V. CONCLUSION

This paper described about a novel method of problem finding and solving based on the formal language L_{md} where a problem is defined as the combination of a goal event X_G and a transit event X_T between the current event X_C and the goal event. The task sharing and assignment among sensors and actors are executed based on the information of a problem described as locus formulas in L_{md}. The most useful keys to task assignment are the attributes involved. About 50 kinds of attributes have been found in association with natural languages [6]. Furthermore, most of computations on L_{md} are simply for unifying (or identifying) atomic locus formulas and for evaluating arithmetic expressions such as '$p=q$', and therefore we believe that our formalismcan reduce the computational complexities of the others [e.g., 12,13] when applied to the same kinds of problems described here.

Our future work will include establishment of learning facilities for automatic acquisition of word concepts from sensory data and human-robot interaction by natural language under real environments.

REFERENCES

[1] Minsky,M.: "A Framework for Representing Knowledge," Report AIM--306, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1974.
[2] Brooks, R. A.: "A robust layered control system for a mobile robot", IEEE Journal of Robotics and Automation, Vol.RA-2, pp.14-23, 1986.
[3] Coradeschi, S. & Saffiotti, A.: "An introduction to the anchoring problem", Robotics and Autonomous Systems, 43, pp.85–96, 2003.
[4] Drumwright,E. Ng-Thow-Hing,V. & Mataric,M.J. "Toward a vocabulary of primitive task programs for humanoid robots", *Proceedings of International Conference on Development and Learning (ICDL06),* Bloomington IN, May 2006.
[5] Wilkins, D.E. & Myers K.L.: "A common knowledge representation for plan generation and reactive execution", *Journal of Logic and Computation*, Vol.5, No.6, pp.731-761, 1995.
[6] Yokota,M.: "An approach to natural language understanding based on a mental image model," Proc. of the 2nd International Workshop on Natural Language Understanding and Cognitive Science, pp.22-31, 2005.
[7] Yokota,M. & Capi,G.: "Cross-media Operations between Text and Picture Based on Mental Image Directed Semantic Theory," WSEAS Transactions on Information Science and Applications, 10-2, pp.1541-1550, 2005.
[8] Zarri,G.P.: "NKRL, a Knowledge Representation Tool for Encoding the 'Meaning' of Complex Narrative Texts", Natural Language Engineering - Special Issue on Knowledge Representation for Natural Language Processing in Implemented Systems, 3,pp.231-253, 1997.
[9] Yokota,M.: "Towards Intuitive Spatiotemporal Communication between Human and Ubiquitous Intelligence Based on Mental Image Directed Semantic Theory - A General Theory of Tempo-logical Connectives -", Proc. of the 4th International Conference on Ubiquitous Intelligence and Computing (UIC-07)(LNCS), pp.1066-1078, Hong Kong, China, July, 2007.
[10] Allen,J.F.: "Towards a general theory of action and time," Artificial Intelligence, *23-2*, pp.123-154, 1984.
[11] Yokota,M.: "Towards a Universal Language for Distributed Intelligent Robot Networking," Proc. of IEEE International Conference on Systems, Man and Cybernetics (SMC06)(Invited paper), pp. 2784-2789, Taipei, Oct. 2006.
[12] Shoham,Y.: "Time for actions: on the relationship between time, knowledge, and action," Proc. of IJCAI¥89, pp.954-959, 1989.
[13] P.Haddawy:"A logic of time, chance, and action for representing plans," Artificial Intelligence, 80-2,pp.243-308, 1996.