# Proposal of Distributed Intelligent Robot Network Based on Mental Image Directed Semantic Theory

Masao Yokota

*Faculty of Information Engineering*
*Fukuoka Institute of Technology.*
E-mail: yokota@fit.ac.jp

**Abstract**

*An ideal ubiquitous computing environment can be a network system of such intelligent and human-friendly robots that never appear in front of humans except when needed. In this paper the distributed intelligent robot network (DIRN) is proposed as one kind of wireless sensor and actor networks (WSAN) consisting of one brain node and numerous sensor and actor nodes with human-friendly interfaces. In order to realize well-coordinated DIRNs, it is very important to develop a systematically computable knowledge representation language universal for any kind of device as well as efficient networking technologies. As a candidate for this purpose, the multimedia description language* $L_{md}$ *was evaluated by applying it to simulation of DIRN-world interaction.*

## 1. Introduction

At present, the realization of wireless sensor and actor networks (WSANs) is one of the challenging topics in the concerned research fields, and a considerable number of important issues have been proposed especially from the viewpoint of networking [11]-[13]. From the viewpoint of AI (Artificial Intelligence), a WSAN can be considered as an intelligent robot system with distributed sensors and actuators that can gather information of high density and perform appropriate actions upon its environment over wide areas.

The distributed intelligent robot network (DIRN) as shown in Fig.1, we propose here as one kind of WSAN, consists of one brain node and numerous sensor and actor nodes with human-friendly interfaces. It is assumed, for example, that sensors and actuators can collaborate autonomously to perform appropriate actions just like reflexive actions in humans and that the brain node works exclusively for complicated computation based on profound knowledge in order to control the other kinds of nodes, to communicate with people, etc.

In order to realize well-coordinated DIRNs, it is very important to develop a systematically computable knowledge representation language [14] as well as efficient networking technologies [13]. This type of language is indispensable to *knowledge-based* processing such as *understanding* sensory events, *planning* appropriate actions and *knowledgeable* communication even with humans, and therefore it needs to have at least a good capability of representing spatio-temporal events that correspond to humans'/robots' sensations and actions in the real world.

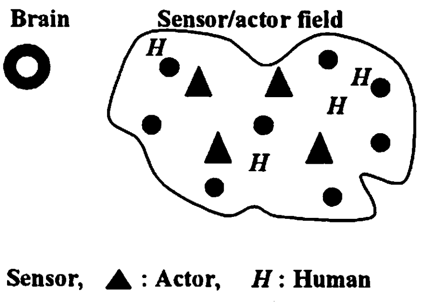


**Fig.1 Physical architecture of DIRNs.**

Traditionally macro-commands such as 'move(10meters)' were employed for deploying sensors/motors. However, these commands were very specific to the devices and apt to have miscellaneous variants such as 'move(10meters, quickly)' and 'move(quickly, 10meters, leftward)', which is very inconvenient for communications especially between devices unknown to each other. Therefore, it is very important to develop such a language as is universal among all kinds of equipments.

Yokota, M. et al have proposed a semantic theory for natural languages so called 'Mental Image Directed Semantic Theory (MIDST)' [1], [2]. In MIDST, word concepts are associated with omnisensual mental images of the external or physical world and are formalized in an intermediate language $L_{md}$ [14]. This language is employed for many-sorted first-order predicate logic with five types of terms. The most remarkable feature of $L_{md}$ is its capability of formalizing both temporal and spatial event concepts on the level of human sensations while the other similar knowledge representation languages are designed to describe the logical relations among conceptual primitives such as words [3], [4].

The $L_{md}$ 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 [1], [5], [14]-[16] and there is a feedback loop between them for their mutual refinement, unlike other similar ones [6], [7].

In this paper we focus on the semantic processing of sensory and action data represented in the formal language $L_{md}$, simulating the interactions between robots and their environments including humans.

## 2. A Brief description of $L_{md}$

MIDST treats word meanings in association with mental images, not limited to visual but omnisensual, 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" with the duration $[t_i, t_f]$ 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').

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

The formula is called 'Atomic Locus Formula' whose first two arguments are sometimes referred to as 'Event Causer (EC)' and 'Attribute Carrier (AC)', respectively.

The intuitive interpretation of the expression (1) 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 ≠ 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 and g=Gs, the locus indicates monotonous change or constancy of the attribute in time domain and that in space domain, respectively. 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.

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A12,Gt,k)\wedge bus(y) \qquad (2)$$

(S2) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,Tokyo,Osaka,A12,Gs,k)\wedge road(y) \qquad (3)$$

The duration of an atomic locus, suppressed in the atomic locus formula, corresponds to the time-interval over which the Focus of the Attention of the Observer (FAO) is put on the corresponding phenomenon outside.

MIDST has employed 'tempo-logical' connectives representing both logical and temporal relations between loci. A tempo-logical connective $K_i$ is defined by (4), where $\tau_i$, $\chi$ and $K$ refer to one of the temporal relations indexed by '$i$', locus, and an ordinary binary logical connective such as the conjunctive '$\wedge$', respectively. This is more natural and economical than explicit indication of time intervals, considering that people do not consult chronometers all the time in their daily lives.

Here are introduced two examples of tempo-logical connectives, namely, 'SAND' and 'CAND'. The expression (5) is the conceptual description of the English verb "fetch", implying such a temporal event that 'x' goes for 'y' and then comes back with it, where '$\Pi$' and '$\bullet$' 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'.

$$\chi_1 K_i \chi_2 \Leftrightarrow (\chi_1 K \chi_2) \wedge \tau_i(\chi_1, \chi_2) \qquad (4)$$

$$(\exists x,y,p1,p2,k)\ L(x,y,p1,p2,A12,Gt,k)$$
$$\bullet (L(x,y,p2,p1,A12,Gt,k)$$
$$\Pi L(x,y,p2,p1,A12,Gt,k))\wedge x\neq y\wedge p1\neq p2 \qquad (5)$$

Furthermore, a very important concept called 'Empty Event (EE)' and symbolized as 'ε' is introduced. An EE stands for nothing but for time elapsing and is explicitly defined as (6) with the attribute 'Time Point (A34)'. It is essentially significant for the MIDST that ***every temporal relation can be represented by a combination of Empty Events, SANDs and CANDs***. For example, (7) represents '$X_1$ during $X_2$'.

According to this scheme, the duration $[p, q]$ of an arbitrary locus X can be expressed as (8).

$$\varepsilon \Leftrightarrow (\exists x,y,p,q,g,k)\ L(x,y,p,q,A34,g,k) \qquad (6)$$

$(\varepsilon_1 \bullet X_1 \bullet \varepsilon_2)\ \Pi X_2$ (7)

$X\ \Pi\ \varepsilon(p,q)$ (8)

The sensory event description scheme presented here is also valid for word concepts of actions because any action must be measured with certain sensors including kinetic ones in order for its analytical or parameterized formalization. That is, ***grounding words on actions is equivalent to grounding words on sensations of actions.*** We have already validated this by experiments on cross-media translation between human motions and texts [15], [16], and on robot manipulations by texts described in 6.

# 3. Specification of the world for a DIRN

'The world for a DIRN' (***W***) refers to 'the set of matters observable for the DIRN' and is defined by (xx) as the union of the set of its nodes (***D***) and the set of the objects in its environment. The set ***D*** is the union of the sets of a brain node ({B}), sensor nodes (***Se***) and actor nodes (***Ac***) as represented by (xx) while the set ***O*** includes possibly humans and the other DIRNs.

$$W = D \cup O \quad (9)$$

$$D = \{B\} \cup Se \cup Ac \quad (10)$$

'A constituent $C_k$ of the world for a DIRN' (i.e., $C_k \in W$) can be specified by the loci in the attribute spaces distinguishable by the sets of Attributes and Standards unique to the DIRN.

### 3.1 Specification of objects

An object in the environment of a DIRN (i.e., $C_k \in O$) can be characterized by the loci of its structure and so on. For example, the characteristics of a tree 'C1' in the environment can be represented by such a locus formula as (11), reading its height (A03) is between 4m and 5m, its location (A12) is in the park 'C2',....For another example, a road 'C3' that runs from a town 'C4' to a town 'C5' via a town 'C6' can be defined by (12), where 'Me' is the standard 'Meter'.

$$tree(C1) \Leftrightarrow (\exists x,h,k,...)\ L(x,C1,h,h,A03,Gt,Me) \wedge (4m<h<5m) \wedge L(x,C1,C2,C2,A12,Gt,k) \wedge park(C2) \wedge ... \quad (11)$$

$$road(C3) \Leftrightarrow (\exists x,k,...)\ L(x,C3,C4,C6,A12,Gs,k) \bullet L(x,C3,C6,C5,A12,Gs,k) \wedge town(C4) \wedge town(C5) \wedge town(C6) \quad (12)$$

### 3.2 Specification of a sensor node

A sensor node (i.e., $C_k \in Se$) can be specified by the loci of its structure and its collectable sensory data. Firstly, a sensor can be distinguished by the definition (13) from another kind of constituent where 'data(y)' is replaceable by (14), reading that a sensor 'x' is what takes in some data 'y' from some constituent 'z'.

$$(\lambda x)sensor(x) \Leftrightarrow (\lambda x \exists y,z,g1,k1)\ L(x,y,z,x,p,q,A12,g1,k1) \wedge data(y) \quad (13)$$

$$(\exists z,z1,...,z_n,a,g,k,p1,...,p_{n+1})\ (y=L(z1,z,p1,p2,a,g,k) \bullet ... \bullet L(z_n,z,p_n,p_{n+1},a,g,k)) \quad (14)$$

The formula (14) implies that 'y' is a locus in the attribute space referred to by the attribute 'a' and the standard 'k' unique to the sensor. For example, a thermometer 'C7' with the measurable range [-10℃, +100℃] can be characterized by (15) with the attribute 'temperature (A28)' and the standard of 'Celsius (Ce)'.

$$(\exists y,z,g1,k1,z,z1,...,z_n,p1,...,p_{n+1})\ L(C7,y,z,C7,p,q,A12,g1,k1) \wedge (y=L(z1,z,p1,p2,A28,Gt,Ce) \bullet ... \bullet L(z_n,z,p_n,p_{n+1},A28,Gt,Ce) \wedge (-10℃ \le p1 \le +100℃) \wedge ... \wedge (-10℃ \le p_{n+1} \le +100℃)) \quad (15)$$

### 3.3 Specification of an actor node

An actor (i.e., $C_k \in Ac$) can be specified by the loci of its structure, performable actions and, if any sensors with it, collectable sensory data. For example, a tanker 'C8' with the coverage [0km,100km] can be characterized by (16) with the attribute 'mileage (A17)'.

$$(\exists x,r)L(C8,x,0,r,A17,Gt,Me) \wedge (0km \le r \le 100km) \wedge liquid(x) \quad (16)$$

### 3.4 Specification of the brain node

The brain node (i.e., *B*) can be specified by the loci of commonsense knowledge and the world knowledge including such specifications of the other constituents as mentioned above. For example, (17) is an example of commonsense knowledge, reading that ***<u>a matter has never different values of an attribute at a time</u>***.

$$L(x,y,p1,q1,a,g,k)\ \Pi L(z,y,p2,q2,a,g,k) .\supset. p1=p2 \wedge q1=q2 \quad (17)$$

The intelligence of the brain node must be conscious of all about the other constituents but can be unconscious of the structure (e.g., hardware configuration) and the computational performance specification (e.g., CPU speed) of itself because they are what only meta-systems such as OS and meta-brain node have to concern. In our case, the brain node is a high-performance multimedia computer with the OS WINDOWS/XP and our intelligent system IMAGES-M installed [14].

## 4. Systematic computation of $L_{md}$

The integrated multimedia understanding system IMAGES-M works as the main intelligence of the brain node of a DIRN. The intelligence of each sensor or actuator is a small-scaled IMAGES-M adapted for its specialized function. IMAGES-M has employed locus formulas as intermediate conceptual representations, through which it can integrally understand and generate sensor data, speech, visual image, text, and action data.

IMAGES-M 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 (18) represented in $L_{md}$, where 'A29' is the attribute 'taste'. This locus formula can be translated into the English sentence S3 by TPS, but it is semantically anomalous because a 'desk' has ordinarily no taste.

$$(\exists x)L(\_,x,Sweet,Sweet,A29,Gt,\_)\wedge desk(x) \qquad (18)$$

(S3) The desk is sweet.

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

Firstly, assume the commonsense knowledge of "desk" as (19) in KB, where 'A39' refers to the attribute 'vitality', and the special symbols '*' and '/' are defined as (20) and (21) representing 'always' and 'no value', respectively.

$$(\lambda x)\ desk(x) \Leftrightarrow (\lambda x)\ (\ldots L^*(\_,x,/,/,A29,Gt,\_)$$
$$\wedge \ldots \wedge L^*(\_,x,/,/,A39,Gt,\_) \wedge \ldots) \qquad (19)$$
$$X^* \Leftrightarrow (\forall p,q) X\ \Pi\ \varepsilon(p,q) \qquad (20)$$
$$L(\ldots,/,\ldots) \Leftrightarrow \sim(\exists p)\ L(\ldots,p,\ldots) \qquad (21)$$

Secondly, the postulates expressed by (22) and (23) in KB are utilized. The formula (22) means that ***if one of two loci exists every time interval, then they can coexist***. The formula (23) states that ***a matter has never different values of an attribute at a time***.

$$X \wedge Y^*\ .\supset.\ X\ \Pi\ Y \qquad (22)$$
$$L(x,y,p,q,a,g,k)\ \Pi L(z,y,r,s,a,g,k)\ .\supset.\ p=r \wedge q=s \qquad (23)$$

Lastly, IE detects the semantic anomaly of "sweet desk" by using (19)-(23). That is, the formula (24) below is finally deduced from (29)-(23), which violates the commonsense given by (23), that is, " $Sweet \neq /$ ".

$$L(\_,x,Sweet,Sweet,A29,Gt,\_)\ \Pi L(\_,x,/,/,A29,Gt,\_) \qquad (24)$$

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. [14].

## 5. Interaction between a DIRN and its world

A DIRN is to solve some kinds of problems in its world. Such problems can be classified roughly into two categories as follows.

(CP) Creation Problem:
e.g.) house building, food cooking,
and
(MP) Maintenance Problem:
e.g.) fire extinguishing, room cleaning.

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

### 5.1 Definition of a problem and a job for a DIRN

A DIRN must determine its job to solve a problem in the world. In general, the DIRN needs to interpolate some transit event $X_T$ between the two events, namely, 'Current Event ($X_C$)' and 'Goal Event ($X_G$)' as shown by (25).

$$X_C \bullet X_T \bullet X_G \quad (25)$$

According to this formalization, a problem $X_P$ is defined as $X_T \bullet X_G$ and a job for the DIRN 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 by the inference employing such a postulate as (26) implying 'Continuity in attribute values'. Therefore, the problem $X_P$ in (27) is to be inferred as (28).

$$L(x,y,p1,p2,a,g,k) \bullet L(z,y,p3,p4,a,g,k) .\supset. p3=p2 \quad (26)$$

$$L(x,y,p1,p2,a,g,k) \bullet X_P \bullet L(z,y,p3,p4,a,g,k) \quad (27)$$

$$L(z',y,p2,p3,a,g,k) \bullet L(z,y,p3,p4,a,g,k) \quad (28)$$

### 5.2 CP finding and solving

Consider a verbal command such as S5 uttered by a human. Its interpretation is given by (29) as the goal event $X_G$. If the current event $X_C$ is given by (30), then (31) with the transit event $X_T$ underlined can be inferred as the problem corresponding to S5.

(S5) Keep the temperature of 'room C9' at 20.

$$L(z,C9,20,20,A28,Gt,k) \wedge room(C9) \wedge (z \in O) \quad (29)$$

$$L(x,C9,p,p,A28,Gt,k) \wedge room(C9) \quad (30)$$

$$\underline{L(z1,C9,p,20,A28,Gt,k)} \bullet L(z,C9,20,20,A28,Gt,k) \wedge room(C9) \wedge (z1 \in O) \quad (31)$$

For this problem, the DIRN is to execute a job deploying a certain thermometer and actors 'z1' and 'z'. The selection of the actor 'z1' is performed as follows:

*If 20-p <0 then z1 is a cooler, otherwise*
*if 20-p >0 then z1 is a heater, otherwise*
*20-p =0 and no actor is deployed as z1.*

The selection of 'z' is a job in case of MP described in the next section.

### 5.3 MP finding and solving

In general, the goal event $X_G$ for an MP is that for another CP such as S5 given possibly by humans and solved by the DIRN in advance. That is, the job in this case is to autonomously restore the goal event $X_G$ created in advance to the current event $X_C$ as shown in (32), 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 (29) in advance, then $X_T$ is also represented as the underlined part of (31) while $X_{-T}$ as (33). Therefore the job here is quite the same that was described in the previous section.

$$X_G \bullet X_{-T} \bullet X_C \bullet X_T \bullet X_G \quad (32)$$

$$L(z1,C9,20,p,A28,Gt,k) \wedge room(C9) \wedge (z1 \in O) \quad (33)$$

## 6. Application to robot manipulation

The intelligent system IMAGES-M, still under development, is intended to facilitate integrated multimedia information understanding, including cross-media operations. 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 (34) 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 (35) and (36), respectively.

$$L(John,John,Walkf\text{-}1,Walkf\text{-}m,A11,Gt,AIBO) \wedge L(John,John,Wavelh\text{-}1,Wavelh\text{-}n,A11,Gt,AIBO) \quad (34)$$

$$Walkf=\{Walkf\text{-}1, Walkf\text{-}2,\ldots,Walkf\text{-}m\} \quad (35)$$

$$Wavelh=\{Wavelh\text{-}1, Wavelh\text{-}2,\ldots, Wavelh\text{-}n\} \quad (36)$$

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 (37) which is the goal event here.

$$L(John,John,Walkf\text{-}1,Walkf\text{-}m,A11,Gt,AIBO) \bullet \underline{L(John,John,Walkf\text{-}m,Wavelh\text{-}1,A11,Gt,AIBO)} \bullet L(John,John,Wavelh\text{-}1,Wavelh\text{-}n,A11,Gt,AIBO) \quad (37)$$

Thirdly, (38) is to be inferred, where the transit event, underlined, is interpolated between the current event and the goal event $X_G$ (=(37)).

$$L(John,John,p1,p2,A11,Gt,AIBO) \bullet \underline{L(John,John,p2,Walkf\text{-}1,A11,Gt,AIBO)} \bullet X_G \quad (38)$$

Finally, (38) is interpreted into a series of joint rotations in the AIBO as shown in Fig.2.

At the present stage, verbal commands to robots are put in through the keyboard, and translated into locus formulas and in turn into actions. In near future, robots will come to be activated by voice through SPU (Speech Processing Unit) of IMAGES-M.

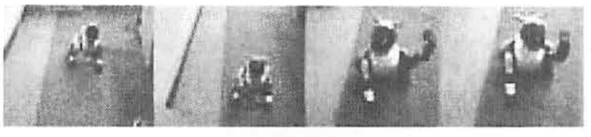

**Fig.2.** AIBO (Sony) behaving in accordance to the command 'Walk forward and wave your left hand'

## 7. Discussion and conclusion

The cross-references between texts in several languages (Japanese, Chinese, Albanian and English) and pictorial patterns such as maps were almost successfully implemented on IMAGES-M [14]. This leads to the conclusion that employment of atomic locus formulas has made the logical expressions of event concepts remarkably computable and has proved to be very adequate to systematize cross-media operations such as simulation of DIRN-world interactions. This is due to their medium-freeness and good correspondence with the performances of miscellaneous devices, which in turn implies that locus formula representation may make it easier for the devices to share a task than macro-command representation.

From the simulation results, we conclude that $L_{md}$ can be a universal language for WSANs including DIRNs. The future works of our research are as follows.

(P1) Augmentation of processing units

(P2) Conceptual analysis and description of such a group of words that can form commonsense knowledge

## Acknowledgements

This work was partially funded by the Grants from Computer Science Laboratory, Fukuoka Institute of Technology and Ministry of Education, Culture, Sports, Science and Technology, Japanese Government, numbered 14580436 and 17500132.

## References

[1] M.Yokota: "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.

[2] M. Yokota, et al: "Mental-image directed semantic theory and its application to natural language understanding systems", Proc. of NLPRS'91, pp.280-287, 1991.

[3] J.F. Sowa: Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

[4] G.P. Zarri: "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.

[5] S.Oda, M.Oda & M.Yokota : "Conceptual Analysis Description of Words for Color and Lightness for Grounding them on Sensory Data", Trans.of JSAI,16-5-E,pp.436-444, 2001.

[6] R.W. Langacker : Concept, Image and Symbol, Mouton de Gruyter, Berlin/New York, 1991.

[7] G.A. Miller & P.N.: Johnson-Laird : Language and Perception, Harvard University Press, 1976.

[8] A.Yamada, et.al.: Reconstucting spatial image from natural language texts, in Proc. of Coling 90, Nantes, 1992

[9] P.Olivier & J.Tsujii: A Computational View of the Cognitive Semantics of Spatial Expressions, Proc. of ACL 94, Las Cruces, 1994.

[10] G. Adorni, M. Di Manzo, & F. Giunchiglia: Natural Language Driven Image Generation, Proc. of COLING 84, pp. 495-500, 1984

[11] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, & E. Cayirci: "Wireless sensor networks: a survey", Computer Networks, Vol.38, No.4, pp. 393-422, March 2002.

[12] M. Haenggi: "Mobile Sensor-Actuator Networks: Opportunities and Challenges", Proc. of 7th IEEE Int. Workshop, Frankfurt,Germany, pp. 283-290, July 2002.

[13] I. F. Akyildiz & I.H. Kasimoglu: "Wireless Sensor and Actor_ Networks: Research Challenges", Ad Hoc Networks, 2, pp. 351-367, 2004.

[14] M.Yokota & G.Capi: "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.

[15] M.Amano, et al : "Linguistic interpretation of human motion based on Mental Image Directed Semantic Theory," Proc. of IEEE AINA2005, pp.139-144, Taipei , 2005.

[16]. M.Amano, et al : "Cross-media translation between human motions and texts based on Mental Image Directed Semantic Theory," Proc. of IEEE ICDCSworkshop2005, pp.707-713, Ohio, 2005