Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 184103, 10 pages
doi:10.1155/2012/184103

Research Article

Aware Computing in Spatial Language Understanding Guided by Cognitively Inspired Knowledge Representation

Masao Yokota

Department of System Management, Fukuoka Institute of Technology, Fukuoka 811-0295, Japan

Correspondence should be addressed to Masao Yokota, yokota@fit.ac.jp

Received 23 January 2012; Accepted 29 March 2012

Academic Editor: Keitaro Naruse

Mental image directed semantic theory (MIDST) has proposed an omnisensory mental image model and its description language L_{md}. This language is designed to represent and compute human intuitive knowledge of space and can provide multimedia expressions with intermediate semantic descriptions in predicate logic. It is hypothesized that such knowledge and semantic descriptions are controlled by human attention toward the world and therefore subjective to each human individual. This paper describes L_{md} expression of human subjective knowledge of space and its application to aware computing in cross-media operation between linguistic and pictorial expressions as spatial language understanding.

1. Introduction

The serious need for more human-friendly intelligent systems has been brought by rapid increase of aged societies, floods of multimedia information over the WWW, development of robots for practical use, and so on. For example, it is very difficult for people to exploit necessary information from the immense multimedia contents over the WWW. It is still more difficult to search for desirable contents by queries in different media, for example, text queries for pictorial contents. In this case, intelligent systems facilitating cross-media references are helpful and worth developing. In this research area so far, it has been most conventional that conceptual contents conveyed by information media such as languages and pictures are represented in computable forms independent of each other and translated via so-called "transfer" processes which are often ad hoc and very specific to task domains [1–3].

In order to systematize cross-media operation, however, it is needed to develop such a computable knowledge representation language for multimedia contents that should have at least a good capability of representing spatiotemporal events perceived by people in the real world. For this purpose, mental image directed semantic theory (MIDST) has proposed a model of human mental image and its description language L_{md} (Language for mental-image description) [4]. This language is capable of formalizing human omnisensory mental images (equal to multimedia contents, here) in predicate logic, while other knowledge description schema [5, 6] are too coarse or linguistic (or English-like) to formalize them in an integrative way as intended here. L_{md} is employed for many-sorted predicate logic and has been implemented on several versions of the intelligent system IMAGES [4, 7] and there is a feedback loop between them for their mutual refinement unlike other similar theories [8, 9].

As detailed in the following sections, MIDST was rigidly formalized as a deductive system [10] in the formal language L_{md}, which is remarkably distinguished from other work (e.g., [5, 8]). However, its application to computerized systems is another thing because computational cost of logical formulas is very high in general. In fact, however, the deductive system contains a considerable number of theses or postulates much easier to realize in imperative programming (e.g., in C) than in declarative programming (e.g., in Prolog) because L_{md} expressions normalized by atomic locus formulas are very suitable to structure and operate in table so-called Hitree [11]. Conventionally, it is as well convinced that hybrid computation based on both the programming paradigms is more flexible and efficient than that based on only one of them. This is also the case

for each version of IMAGES so far and therefore the author has been promoting to replace declarative programs with imperative ones considering the benefit of L_{md} expression. This paper focuses as well on the hybrid computation guided by L_{md} expression and 3D map data, here so-called partially symbolized direct knowledge of space (PSDKS), in cross-media operation between linguistic and pictorial expressions as spatial language understanding. That is, static spatial relations among objects as 3D map data for imperative programming are utilized as well as those in L_{md} for declarative programming.

The remainder of this paper is organized as follows. Section 2 presents the omnisensory mental image model and its relation to the formal language L_{md}. Section 3 describes representation of subjective spatial knowledge in L_{md}. In Sections 4 and 5 are sketched several cognitive hypotheses on mental images for their systematic computation. Section 6 describes the systematic cross-media operation based on L_{md} expression. Section 7 gives the details of direct knowledge of space. In Section 8, is described an example of cross-media operation by IMAGES. Some discussion and conclusion are given in the final section.

2. Mental Image Model and L_{md}

An attribute space corresponds with a sensory system and can be compared to a certain measuring instrument just like a barometer, thermometer or so, and the loci represent the movements of its indicator. A general locus is to be articulated by "Atomic Locus" over a certain absolute time interval $[t_i, t_f]$ as depicted in Figure 1 and formulated as (1) in L_{md}, where the interval is suppressed because people are not aware of absolute time (nor always consult a chronograph).

$$L(x, y, p, q, a, g, k). \tag{1}$$

This is a formula in many-sorted predicate logic, where "L" is a predicate constant with five types of terms: "Matter" (at "x" and "y"), "Value (of Attribute)" (at "p" and "q"), "Attribute" (at "a"), "Pattern (of Event)" (at "g"), and "Standard" (at "k"). Conventionally, Matter variables are headed by "x", "y," and "z."

This formula is called "Atomic Locus Formula" whose first two arguments are sometimes referred to as "Event Causer (EC)" and "Attribute Carrier (AC)," respectively, while ECs are often optional in natural concepts such as intransitive verbs. By the way, hereafter, the terms at AC and Standard are often replaced by "_" when they are of little significance to discern one another. The parameters "g" and "k" cannot be denoted explicitly in Figure 1 because their roles vary drastically depending on its interpretation.

The intuitive interpretation of (1) is given as follows.

> *"Matter " x" causes Attribute " a" of Matter " y" to keep ($p = q$) or change ($p \neq q$) its values temporally ($g = G_t$) or spatially ($g = G_s$) over an absolute time-interval, where the values " p" and " q" are relative to the standard " k". "*

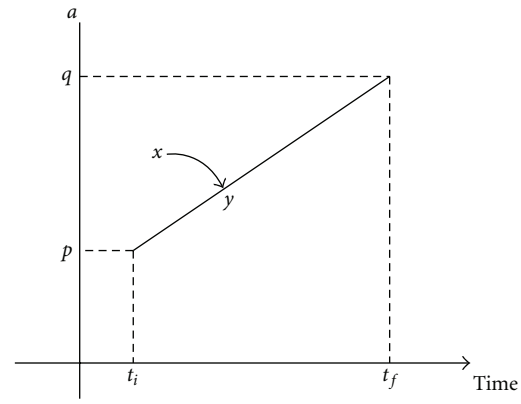

Figure 1: Graphical interpretation of Atomic Locus—the curved arrow indicates the abstract effect from "x" to "y."

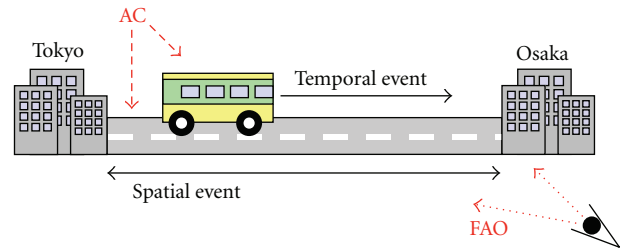

Figure 2: FAO movements and Event types.

When $g = G_t$ and $g = G_s$, the locus indicates monotonic change or constancy of the attribute in time domain and that in space domain, respectively. The former is called "temporal change event" and the latter, "spatial change event," which are assumed to correspond with temporal and spatial gestalt in psychology, respectively. For example, the motion of the "bus" represented by (S1) is a temporal change event and the ranging or extension of the "road" by (S2) is a spatial change event whose meanings or concepts are formulated as (2) and (3), respectively, where "A_{12}" denotes the attribute "Physical Location". These two formulas are different only at the term "Pattern."

(S1) The bus runs from Tokyo to Osaka.

$$(\exists x, y, k)L(x, y, \mathrm{Tokyo}, \mathrm{Osaka}, A_{12}, G_t, k) \wedge \mathrm{bus}(y). \tag{2}$$

(S2) The road runs from Tokyo to Osaka.

$$(\exists x, y, k)L(x, y, \mathrm{Tokyo}, \mathrm{Osaka}, A_{12}, G_s, k) \wedge \mathrm{road}(y). \tag{3}$$

The difference between temporal and spatial change event concepts can be attributed to the relationship between the Attribute Carrier (AC) and the Focus of the Attention of the Observer (FAO). To be brief, FAO is fixed on the whole AC in a temporal change event but runs about on the AC in a spatial change event. Consequently, as shown in Figure 2, the bus and the FAO move together in the case of (S1) while FAO

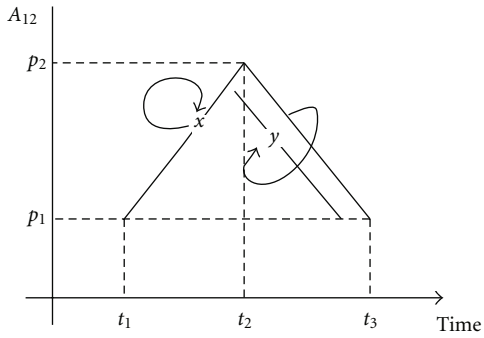

Figure 3: Conceptual image of "fetch."

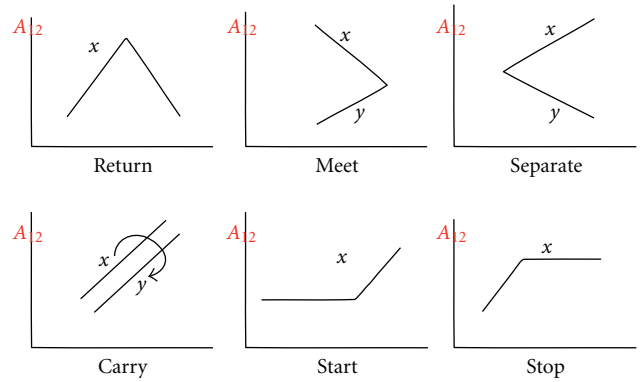

Figure 4: Event patterns of physical location (A_{12}).

solely moves along the road in the case of (S2). That is, *all loci in attribute spaces correspond one to one with movements or, more generally, temporal change events of FAO.*

Articulated loci are combined with tempological conjunctions, where "SAND ($\wedge_0$)" and "CAND ($\wedge_1$)" are most frequently utilized, standing for "Simultaneous AND" and "Consecutive AND", conventionally symbolized as "$\sqcap$" and "$\cdot$," respectively. The formula (4) refers to a temporal change event depicted as Figure 3, implying that "x" goes to some location and then comes back with "y" and corresponding to such a verbal expression as "x fetches y from some location":

$$(\exists x, y, p_1, p_2, k)L(x, x, p_1, p_2, A_{12}, G_t, k) \cdot (L(x, x, p_2, p_1, A_{12}, G_t, k) \sqcap L(x, y, p_2, p_1, A_{12}, G_t, k)) \wedge x \neq y \wedge p_1 \neq p_2. \quad (4)$$

As easily imagined, an event expressed in L_{md} is compared to a movie film taken through a floating camera where both temporal and spatial extensions of the event are recorded as a time sequence of snapshots because it is necessarily grounded in FAO's movement over the event. This is one of the most remarkable features of L_{md}, clearly distinguished from other knowledge representation languages (KRLs).

The attribute spaces for humans correspond to the sensory receptive fields in their brains. At present, about 50 attributes and 6 categories of standards concerning the physical world have been extracted from thesauri. Event patterns are the most important for our approach and have been already reported concerning several kinds of attributes [4, 7]. Figure 4 shows several examples of event patterns in the attribute space of "physical location (A_{12})."

3. Representation of Subjective Spatial Knowledge

MIDST can provide human knowledge pieces with flat L_{md} expressions as human mental images, not concerning whether they are concepts meant by certain symbols (i.e., semantic) or not. Therefore, such a distinction is not denoted explicitly hereafter. There are assumed two major hypotheses

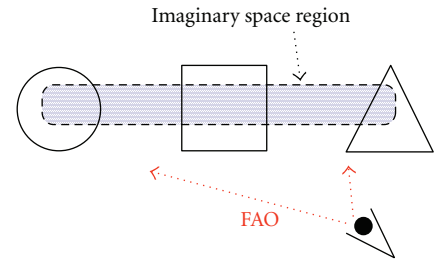

Figure 5: Row as spatial change event.

on mental image. One is that mental image is in one-to-one correspondence with FAO movement as mentioned above. And, the other is that it is not one-to-one reflection of the real world. It is well known that people perceive more than reality, for example, so-called "Gestalt" in psychology. A psychological matter here is not a real matter but a product of human mental functions, including Gestalt and abstract matters such as "society" and "information" in a broad sense. For example, Figure 5 concerns the perception of the formation of multiple objects, where FAO runs along an imaginary object so called "*Imaginary Space Region (ISR).*" This spatial change event can be verbalized as (S3) using the preposition "between" and formulated as (5) or (6), corresponding also to such concepts as "row," and "line-up," where A_{13} denotes the attribute "Direction".

Employing ISRs and the 9-intersection model [12], all the topological relations between two objects can be formulated in such expressions as (7) or (8) for (S4), and (9) for (S5), where "In," "Cont," and "Dis" are the values "inside", "contains" and "disjoint" of the attribute "Topology (A_{44})" with the standard "9-intersection model ($K_{9\mathrm{IM}}$)," respectively. Practically, these topological values are given as 3×3 matrices with each element equal to 0 or 1 and therefore, for example, "In" and "Cont" are transposes each other. That is, Cont = In^T.

(S3) The square is between the triangle and the circle.

(S4) Tom is in the room.

(S5) Tom exits the room.

$$
\begin{aligned}
&(\exists x_1, x_2, x_3, y, p, q)\\
&\quad (L(\_, y, x_1, x_2, A_{12}, G_s, \_) \sqcap L(\_, y, p, p, A_{13}, G_s, \_))\\
&\quad \cdot (L(\_, y, x_2, x_3, A_{12}, G_s, \_) \sqcap L(\_, y, q, q, A_{13}, G_s, \_))\\
&\quad \wedge\ \mathrm{ISR}(y) \wedge p = q \wedge \mathrm{triangle}(x_1)\\
&\quad \wedge \mathrm{square}(x_2) \wedge \mathrm{circle}(x_3),
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
&(\exists x_1, x_2, x_3, y, p)\\
&\quad (L(\_, y, x_1, x_2, A_{12}, G_s, \_) \cdot L(\_, y, x_2, x_3, A_{12}, G_s, \_))\\
&\quad \sqcap L(\_, y, p, p, A_{13}, G_s, \_) \wedge \mathrm{ISR}(y)\\
&\quad \wedge \mathrm{triangle}(x_1) \wedge \mathrm{square}(x_2) \wedge \mathrm{circle}(x_3),
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
&(\exists x, y) L(\mathrm{Tom}, x, y, \mathrm{Tom}, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{In}, \mathrm{In}, A_{44}, G_t, K_{9\mathrm{IM}}) \wedge \mathrm{ISR}(x) \wedge \mathrm{room}(y),
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
&(\exists x, y) L(\mathrm{Tom}, x, \mathrm{Tom}, y, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{Cont}, \mathrm{Cont}, A_{44}, G_t, K_{9\mathrm{IM}})\\
&\quad \wedge \mathrm{ISR}(x) \wedge \mathrm{room}(y),
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
&(\exists x, y, p, q) L(\mathrm{Tom}, \mathrm{Tom}, p, q, A_{12}, G_t, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, y, \mathrm{Tom}, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{In}, \mathrm{Dis}, A_{44}, G_t, K_{9\mathrm{IM}}) \wedge \mathrm{ISR}(x)\\
&\quad \wedge \mathrm{room}(y) \wedge p \neq q.
\end{aligned}
\tag{9}
$$

With a special attention, the author has analyzed a considerable number of spatial terms over various kinds of English words such as prepositions, verbs, adverbs, and so forth, categorized as "Dimensions," "Form," and "Motion" in the class "SPACE" of the Roget's thesaurus [13], and found that almost all the concepts of spatial change events can be defined in exclusive use of five kinds of attributes for FAOs, namely, "Physical location (A_{12})," "Direction (A_{13})," "Trajectory (A_{15})," "Mileage (A_{17})," and "Topology (A_{44})."

4. Hypothetical Operations upon Mental Images

People can transform their mental images in several ways such as mental rotation [14]. Here are introduced and defined 3 kinds of mental operations, namely, "reversing," "duplicating," and "converting."

4.1. Image Reversing. It is easy for people to imagine the reversal of an event just like "rise" versus "sink." This mental operation is here denoted as "R" and recursively defined as O_R, where χ_i stands for a image. The reversed values p^R and q^R depend on the properties of the attribute values p and q. For example, $p^R = p$, $q^R = q$ for A_{12}; $p^R = -p$, $q^R = -q$ for A_{13}; $p^R = p^T$, $q^R = q^T$ for A_{44}.

O_R:

$$
\begin{aligned}
\left(\chi_1 \cdot \chi_2\right)^R &\Longleftrightarrow \chi_2^R \cdot \chi_1^R,\\
\left(\chi_1 \sqcap \chi_2\right)^R &\Longleftrightarrow \chi_1^R \sqcap \chi_2^R,\\
L^R(x, y, p, q, a, g, k) &\Longleftrightarrow L\left(x, y, q^R, p^R, a, g, k\right).
\end{aligned}
\tag{10}
$$

4.2. Image Duplicating. Humans can easily imagine the repetition of an event just like "visit twice" versus "visit once." This operation is also recursively defined as O_n, where "n" is an integer representing the frequency of an image χ.

O_n:

$$
\begin{aligned}
\chi^n &\Longleftrightarrow \chi \quad (n = 1),\\
\chi^n &\Longleftrightarrow \chi \cdot \chi^{n-1} \quad (n > 1).
\end{aligned}
\tag{11}
$$

4.3. Image Converting. We can convert temporal and spatial change event images each other and this is the reason why it is easy for us to understand instantly such an expression as (S2). This mental operation is here denoted as "C" and recursively defined as O_C, which will help a robot to cope with such a somewhat queer expression as "The road jumps up at the point. Be careful!".

O_C:

$$
\begin{aligned}
\left(\chi_1 \cdot \chi_2\right)^C &\Longleftrightarrow \chi_1^C \cdot \chi_2^C,\\
\left(\chi_1 \sqcap \chi_2\right)^C &\Longleftrightarrow \chi_1^C \sqcap \chi_2^C,\\
L^C(x, y, p, q, a, g, k) &\Longleftrightarrow L\left(x, y, p, q, a, g^C, k\right),
\end{aligned}
\tag{12}
$$

where $g^C = G_s$ for $g = G_t$ and $g^C = G_t$ for $g = G_s$.

5. Hypothetical Properties of Mental Images

Properties or laws of mental images as spatial knowledge pieces are formalized in L_{md} and introduced as postulates and their derivatives in a deductive system [10] to be employed in theorem proving there. Here are described two examples of such postulates, namely, "Postulate of Reversibility of Spatial Change Event" and "Postulate of Partiality of Matter."

5.1. Postulate of Reversibility of Spatial Change Event. As already mentioned in Section 2, all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal change events of the FAO. Therefore, the L_{md} expression of an event is compared to a movie film recorded through a floating camera over the event. And this is why (S6) and (S7) can refer to the same scene in spite of their appearances, where what "sinks" or "rises" is the FAO as illustrated in Figure 6 and whose conceptual descriptions are given as (13) and (14), respectively, where "A_{13}," "↑," and "↓" refer to the attribute "Direction" and its values "upward" and "downward" (practically as 3D unit vectors), respectively.

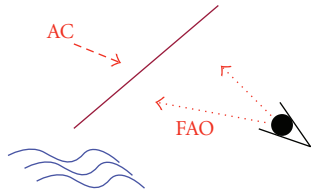

Figure 6: Slope as spatial change event.

(S6) The path *sinks* to the brook.

(S7) The path *rises* from the brook.

$$(\exists y,z,p)L(\_,y,p,z,A_{12},G_s,\_)\sqcap L(\_,y,\downarrow,\downarrow,A_{13},G_s,\_)$$
$$\wedge\ \text{path}(y)\wedge\text{brook}(z)\wedge z\neq p, \tag{13}$$

$$(\exists y,z,p)L(\_,y,z,p,A_{12},G_s,\_)\sqcap L(\_,y,\uparrow,\uparrow,A_{13},G_s,\_)$$
$$\wedge\ \text{path}(y)\wedge\text{brook}(z)\wedge z\neq p. \tag{14}$$

Such a fact is generalized as P_{RS} (postulate of reversibility of spatial change event), where χ_s and χ_s^R are an image and its "reversal" for a certain spatial change event, respectively, and they are substitutable with each other because of the property of "$\equiv_0$." This postulate can be one of the principal inference rules belonging to people's common-sense knowledge about geography.

P_{RS}:

$$\chi_s^R\cdot\equiv_0\chi_s. \tag{15}$$

This postulation is also valid for such a pair of (S8) and (S9) as interpreted approximately into (16) and (17), respectively. These pairs of conceptual descriptions are called equivalent in the P_{RS}, and the paired sentences are treated as paraphrases each other.

(S8) Route A and Route B separate at the city.

(S9) Route A and Route B meet at the city.

$$(\exists p,y,q)L(\_,\text{Route\_}A,p,y,A_{12},G_s,\_)$$
$$\sqcap L(\_,\text{Route\_}B,q,y,A_{12},G_s,\_)\wedge\text{city}(y)\wedge p\neq q, \tag{16}$$

$$(\exists p,y,q)L(\_,\text{Route\_}A,y,p,A_{12},G_s,\_)$$
$$\sqcap L(\_,\text{Route\_}B,y,q,A_{12},G_s,\_)\wedge\text{city}(y)\wedge p\neq q. \tag{17}$$

Of course, P_{RS} is as well applicable to such an inference that "if x is to the right of y, then y is to the left of x," which is conventionally based on a considerably large set of such *linguistic* axioms as (18) regardless of *time*. Furthermore, it is notable that there are an infinite number of directions without good correspondence with single words such as "right."

$$(\forall x,y)\text{right}(x,y)\supset\text{left}(y,x),$$
$$(\forall x,y)\text{under}(x,y)\supset\text{above}(y,x). \tag{18}$$

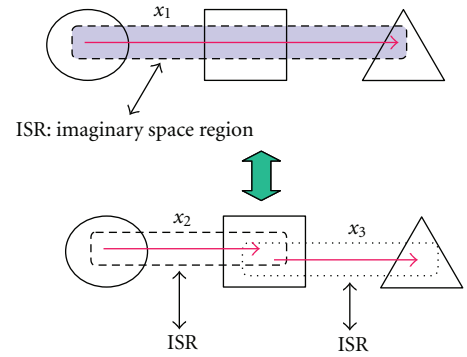

Figure 7: Partiality of ISR—the arrows represent the directions of FAO.

5.2. Postulate of Partiality of Matter. Any matter is assumed to consist of its parts in a structure (i.e., spatial change event) and generalized as P_{PM} (postulate of partiality of matter) here. For example, Figure 7 shows that an ISR x_1 can be deemed as a complex of ISRs x_2 and x_3.

P_{PM}:

$$(\forall y,x_1,p,q,a,k)L(y,x_1,p,q,a,G_s,k)$$
$$\cdot L(y,x_1,q,r,a,G_s,k).\supset_0(\exists x_2,x_3)L(y,x_2,p,q,a,G_s,k)$$
$$\sqcap L(y,x_3,q,r,a,G_s,k),$$

$$(\forall y,x_2,x_3,p,q,a,k)$$
$$L(y,x_2,p,q,a,G_s,k)\sqcap L(y,x_3,q,r,a,G_s,k)$$
$$.\supset_0(\exists x_1)L(y,x_1,p,q,a,G_s,k)\cdot L(y,x_1,q,r,a,G_s,k). \tag{19}$$

We often refer to parts of an image especially for deductive inference upon it. For example, we can easily deduce from Figure 7 (Top) the two facts "the square is to the left of the triangle" and "the circle is to the left of the square." As its reversal, we can merge these two partial images into one meaningful image such as Figure 7 (Bottom). That is, P_{PM} is very useful to compute static spatial relations that are expressed by English spatial terms and conventionally formalized by a large set of such *linguistic* axioms as (20) regardless of *time* just like the case of P_{RS}. Furthermore, it is notable that the reversals of these axioms (i.e., $(\forall x,y,z)$ between $(y,z,x)\supset w(y,x)\wedge w(z,y)$) do not always exist in good correspondence with words (e.g., "left" for the predicate w).

$$(\forall x,y,z)\text{left}(y,x)\wedge\text{left}(z,y)\supset\text{between}(y,z,x),$$
$$(\forall x,y,z)\text{under}(y,x)\wedge\text{under}(z,y)\supset\text{between}(y,z,x). \tag{20}$$

Besides its orthodox usage above, P_{PM}, in cooperation with P_{RS}, can be utilized for translating such a paradoxical sentence as "The Andes Mountains run north and south." into such a plausible interpretation as "Some part of the Andes Mountains run north (from somewhere) and the other part run south."

6. Cross-Media Translation

As easily understood by its definition, an atomic formula corresponds with a pair of snapshots at the beginning and the ending of a monotonic change in an attribute. Viewed from pictorial representation, temporal and spatial change events correspond to animated and still pictures, respectively. Furthermore, the L_{md} expression of a spatial change event as the locus of FAO can be related to the sequence of pen-down and pen-up in line drawing. This section describes cross-media translation in general, focusing on that between text and map, one kind of still picture, as the core of spatial language understanding.

6.1. Functional Requirements. Systematic cross-media translation here is defined by the functions (F1)–(F4) as follows.

(F1) To translate source representations into target ones as for contents describable by both source and target media. For example, positional relations between/among physical objects such as "in", "around." are describable by both linguistic and pictorial media.

(F2) To filter out such contents that are describable by source medium but not by target one. For example, linguistic representations of "taste" and "smell" such as "sweet candy" and "pungent gas" are not describable by usual pictorial media although they would be seemingly describable by cartoons, and so forth.

(F3) To supplement default contents, that is, such contents that need to be described in target representations but not explicitly described in source representations. For example, the shape of a physical object is necessarily described in pictorial representations but not in linguistic ones.

(F4) To replace default contents by definite ones given in the following contexts. For example, in such a context as "There is a box to the left of the pot. The box is red. ...," the color of the box in a pictorial representation must be changed from default one to red.

For example, the text consisting of such two sentences as "There is a hard cubic object" and "The object is large and gray" can be translated into a still picture in such a way as shown in Figure 8.

6.2. Formalization. According to the MIDST, any content conveyed by an information medium is assumed to be associated with the loci in certain attribute spaces and in turn the world describable by each medium can be characterized by the maximal set of such attributes. This relation is conceptually formalized by (21), where Wm, Am_i, and F mean "the world describable by the information medium m," "an attribute of the world," and "a certain function for determining the maximal set of attributes of Wm," respectively,

$$F(\mathrm{Wm}) = \{\mathrm{Am}_1, \mathrm{Am}_2, \ldots, \mathrm{Am}_n\}. \tag{21}$$

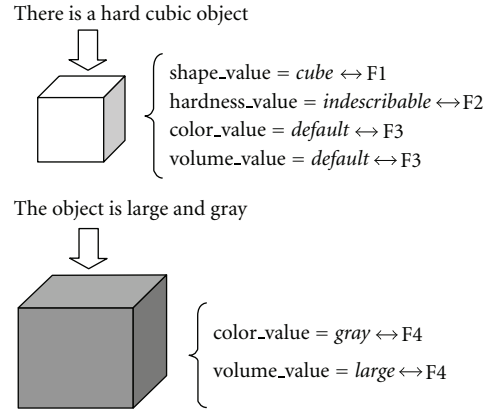

Figure 8: Systematic cross-media translation.

Considering this relation, cross-media translation is one kind of mapping from the world describable by the source medium (ms) to that by the target medium (mt) and can be defined by the following equation:

$$Y(S_{\mathrm{mt}}) = \psi(X(S_{\mathrm{ms}})), \tag{22}$$

where S_{ms}: maximal set of attributes of the world describable by the source medium ms, S_{mt}: maximal set of attributes of the world describable by the target medium mt, $X(S_{\mathrm{ms}})$: L_{md} expression about the attributes belonging to S_{ms}, $Y(S_{\mathrm{mt}})$: L_{md} expression about the attributes belonging to S_{mt}, and ψ: function for transforming X into Y, so called, "L_{md} expression paraphrasing function."

The function ψ is designed to clear all the requirements (F1)–(F4) by inference processing at the level of L_{md} expression.

6.3. L_{md} Expression Paraphrasing Function ψ. In order to realize the function (F1), a certain set of "*Attribute paraphrasing rules* (*APRs*)," so called, are defined *at every pair of source and target media*. The function (F2) is realized by detecting L_{md} expressions about *the attributes without any corresponding APRs* from the content of each input representation and replacing them by *empty events* [10].

For (F3), *default reasoning* is employed. That is, such an inference rule as defined by (23) is introduced, which states if *X is deducible and it is consistent to assume Y then conclude Z*. This rule is applied typically to such instantiations of X, Y, and Z as specified by (24) which means that the indefinite attribute value " p" with the indefinite standard " k" of the indefinite matter " y" is substitutable by the constant attribute value " P" with the constant standard "K"of the definite matter "$O\#$" of the same kind "M":

$$X \circ Y \longrightarrow Z, \tag{23}$$

$$\begin{aligned}&\{X/(L(x, y, p, p, A, G, k) \wedge M(y))\\ &\quad \wedge (L(z, O\#, P, P, A, G, K) \wedge M(O\#)),\\ &\quad Y/p = P \wedge k = K, Z/L(x, y, P, P, A, G, K) \wedge M(y)\}.\end{aligned} \tag{24}$$

Table 1: APRs for text-picture translation (A_{12}: physical location, A_{13}: direction, A_{17}: mileage, A_{10}: volume, A_{11}: shape, A_{32}: color, A_{44}: topology).

APRs	Correspondences of attributes (text : picture)	Value conversion schema (text ↔ picture)
APR-01	$A_{12} : A_{12}$	$p \leftrightarrow p'$
APR-02	$\{A_{12}, A_{13}, A_{17}\} : A_{12}$	$\{p, d, l\} \leftrightarrow p' + l'd'$
APR-03	$\{A_{11}, A_{10}\} : A_{11}$	$\{s, v\} \leftrightarrow v's'$
APR-04	$A_{32} : A_{32}$	$c \leftrightarrow c'$
APR-05	$\{A_{12}, A_{44}\} : A_{12}$	$\{p_a, m\} \leftrightarrow \{p'_a, p'_b\}$

The function (F4) is realized quite easily by memorizing the history of applications of default reasoning.

6.4. Attribute Paraphrasing Rules for Text and Picture. Five kinds of APRs for this case are shown in Table 1 where $p, s, c, \ldots$ and $p', s', c', \ldots$ are linguistic expressions and their corresponding pictorial expressions of attribute values, respectively. Further details are as follows.

(i) APR-02 is used especially for a sentence such as "The box is 3 meters to the left of the chair." The symbols p, d and l correspond to "the location of the chair," "left," and "3 meters," respectively, yielding the pictorial expression of "the location of the box," namely, "$p' + l'd'$."

(ii) APR-03 is used especially for a sentence such as "The pot is big." The symbols $s$and v correspond to "the shape of the pot (default value)" and "the volume of the pot ("big")," respectively. In pictorial expression, the shape and the volume of an object is inseparable and therefore they are represented only by the value of the attribute "shape", namely, $v's'$.

(iii) APR-05 is used especially for a sentence such as "The cat is in the box." The symbols p_a, p_b and m correspond to "the location of the desk," "the location of the cat," and "in," respectively, yielding a pair of pictorial expressions of the locations of the two objects.

7. Direct Knowledge of Space

Partially symbolized direct knowledge of space (PSDKS in short) introduced here is one of the data structures for imperative programming in IMAGES as well as Hitree [11]. PSDKS is a map for directional and metric relations among objects while Hitree is intended to be a complete substitute of L_{md} expression. That is, the relation between L_{md} expression and PSDSK is what is formalized by APR-02 in Table 1. For example, consider the scene of a room shown in Figure 9, where the FAO is posed on the formation of the flower-pot, box lamp, chair, and cat. PSDKS here does not mean any kind of live image perceived by a human (or snapshot by a system) at a time point but somewhat abstract 3D map resulted from its recognition as depicted in Figure 10.

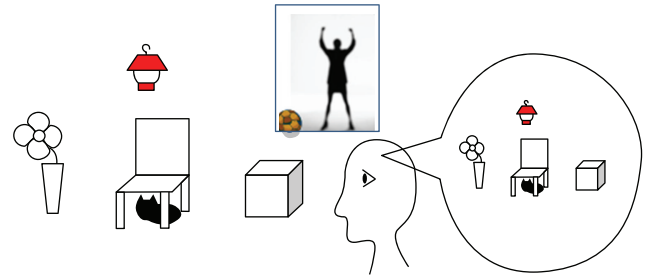

Figure 9: Scene of a room and its live image in human.

That is, PSDKS is defined as a set of points representing the 3D locations (i.e., A_{12}) of the involved objects linked to the corresponding L_{md} expression and therefore directly reusable for computation without recognizing them unlike the memory of their live image or snapshot.

In turn, consider verbalization of the PSDKS. In this case, any system must be forced to articulate it in accordance with existing word concepts and may utter such a set of sentences (S10)–(S13). These are to be generated from such L_{md} expressions as (25)–(28), respectively, where I_n, Fp, Ch, Bx, Lp and Ct stand for ISR, flower-pot, chair, box lamp, and cat, respectively.

(S10) The chair is 3 meters to the right of the flower-pot.

(S11) The flower-pot is 6 meters to the left of the box.

(S12) The lamp hangs above the chair.

(S13) The cat lies under the chair.

$$L(\_, I_1, \mathrm{Fp}, \mathrm{Ch}, A_{12}, G_s, \_) \sqcap L(\_, I_1, \rightarrow, \rightarrow, A_{13}, G_s, \_) \sqcap L(\_, I_1, 3m, 3m, A_{17}, G_s, \_), \tag{25}$$

$$L(\_, I_2, \mathrm{Bx}, \mathrm{Fp}, A_{12}, G_s, \_) \sqcap L(\_, I_2, \leftarrow, \leftarrow, A_{13}, G_s, \_) \sqcap L(\_, I_2, 6m, 6m, A_{17}, G_s, \_), \tag{26}$$

$$L(\_, I_3, \mathrm{Ch}, \mathrm{Lp}, A_{12}, G_s, \_) \sqcap L(\_, I_3, \uparrow, \uparrow, A_{13}, G_s, \_), \tag{27}$$

$$L(\_, I_4, \mathrm{Ch}, \mathrm{Ct}, A_{12}, G_s, \_) \sqcap L(\_, I_4, \downarrow, \downarrow, A_{13}, G_s, \_). \tag{28}$$

Even only for directional and metric relationships between two objects out of the five objects in Figure 10, there can be at least 20 ($=_5P_2$) expressions in English including (S10)–(S13) that correspond with such formulas in conventional logic as (29)–(32), respectively.

$$\mathrm{right}(\mathrm{Ch}, \mathrm{Fp}, 3\_\mathrm{meters}), \tag{29}$$

$$\mathrm{left}(\mathrm{Fp}, \mathrm{Bx}, 6\_\mathrm{meters}), \tag{30}$$

$$\mathrm{above}(\mathrm{Lp}, \mathrm{Ch}), \tag{31}$$

$$\mathrm{under}(\mathrm{Ct}, \mathrm{Ch}). \tag{32}$$

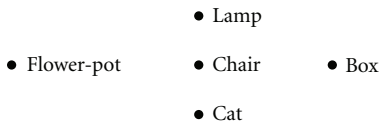

Figure 10: PSDKS resulted from the live image in Figure 9.

This fact implies that conventional declarative programs must employ numerous theses including the axioms (18) and (20) even for solving rather simple problems associated with this scene such as "What is between the box and the flower-pot?". The meaning of this question is conventionally notated as (33). However, it must be noted that the axioms like (18) and (20) cannot be applied to the assertions (29)–(32) for the answer to this question (i.e., $?x$).

On the contrary, it is much easier to search in the PSDKS for the event pattern specified by the L_{md} expression (34) for the question. This formula, a locus of FAO, can be procedurally interpreted as the command "Find "$?x$" by scanning *straight* from the *box* to the *flower-pot*." In case of understanding (S10)–(S13), the system is to apply APR-02 to (25)–(28) and synthesize the partial scenes into one whole scene similar to (not always the same as) the PSDKS shown in Figure 10, that is to say, *reconstructed* direct knowledge of space:

$$\mathrm{between}(?x, \mathrm{Bx}, \mathrm{Fp}), \tag{33}$$

$$\begin{aligned}&(L(\_, y, \mathrm{Bx}, ?x, A_{12}, G_s, \_) \cdot L(\_, y, ?x, \mathrm{Fp}, A_{12}, G_s, \_))\\ &\quad \sqcap L(\_, y, p, p, A_{13}, G_s, \_).\end{aligned} \tag{34}$$

At summarization of this section, PSDKS is very much compact in memory size compared with conventional declaration about space and L_{md} expression can systematically indicate how to search PSDKS for an event pattern.

8. Implementation

IMAGES-M, the last version of intelligent system IMAGES, has recently adopted the multiparadigm language Python in place of PROLOG to facilitate both declarative and imperative programming. IMAGES-M is one kind of expert system with five kinds of user interfaces besides the inference engine (IE) and the knowledge base (KB) as follows.

(i) Text Processing Unit (TPU).

(ii) Speech Processing Unit (SPU).

(iii) Picture Processing Unit (PPU).

(iv) Action Data Processing Unit (ADPU).

(v) Sensory Data Processing Unit (SDPU).

These user interfaces can mutually convert information media and L_{md} expressions in the collaboration with IE and KB, and miscellaneous combinations among them bring forth various types of cross-media operations. The further details about mutual conversion between language and picture can be found in other papers (e.g., [15, 16]).

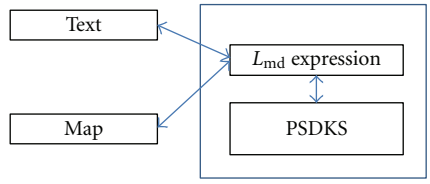

Figure 11: Text-map operation via L_{md} expression and PSDKS.

Figure 12: Transactions between human user and IMAGES-M while text understanding, map composition and question-answering on the map (At headers: "u…" = human user, "s…" = IMAGES-M).

The methodology mentioned above has been implemented on IMAGES-M for spatial language understanding. Here, distinguished from others, spatial language understanding is defined as cross-media operation between spatial language and map such as mutual translation and question-answering between them. The author has confirmed that the hybrid program in Python employing L_{md} expression mainly and PSDKS auxiliarly as shown in Figure 11 is much more flexible and efficient than the previous one [4] in PROLOG for solving problems expressed in spatial language.

Here is presented an example of cross-operation between text and picture performed by IMAGES-M.

IMAGES-M understood the human user's assertions or questions and answered them in picture or word. Figure 12 shows the transactions exchanged between the human user and the system, where the headers "u…." and "s…." stand for the human user's inputs and the system's responses,

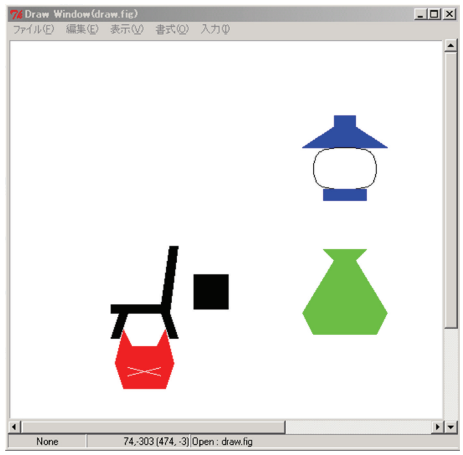

Figure 13: Map finally composed by IMAGES-M for u0001–u0006.

respectively. IMAGES-M can accept 3 kinds of natural language besides English, namely, Japanese (e.g., u0002, u0008 and s0029), Chinese (e.g., u0007 and s0026 in Pinyin) and Albanian (e.g., u0003, u0010 and s0035) as shown in Figure 12, where

u0002 = "The cat is 1 m under the chair,"

u0003 = "The cat is red,"

u0008 = "What is between the chair and the pot?,"

s0029 = "Box,"

u0007 = "Is the cat red?,"

s0026= "yes,"

u0010 = "Is the box between the cat and the lamp?,"

s0035 = "yes."

The map shown in Figure 13 was the final version of those which IMAGES-M composed at each of the user's assertions. IMAGES-M interpreted the assertions u0001–u0006 into L_{md}, and in turn into map and PSDKS (exactly, reconstructed PSDKS), where the system updated them assertion by assertion, responding so by s0002–s0022. In the process of text to map, default reasoning about color, and so forth. was performed in such a way as shown in Figure 8, where only the default locations of the objects within the map are significant for PSDKS.

On the other hand, during the question-answering (i.e., u0007-s0035), IMAGES-M translated each of the user's questions (i.e., u0007–u0010) into L_{md} and consulted the reconstructed PSDKS about Location (A_{12}) within the map or the corresponding L_{md} expression about the other attributes such as Color (A_{32}). In this process, the postulates P_{RS} and P_{PM} were utilized as procedures in Python, which could reduce remarkably the number of axioms such as (18) and (20) that are necessarily employed in conventional systems.

9. Discussion and Conclusion

MIDST is still under development and intended to provide a formal system, represented in L_{md}, for natural semantics of space and time. This formal system is one kind of applied predicate logic consisting of axioms and postulates subject to human perceptive processes of space and time, while the other similar systems in Artificial Intelligence [17–19] are objective, namely, independent of human perception and do not necessarily keep tight correspondences with natural language. This paper showed that L_{md} expressions can contribute to aware computing of spatial relations leading to representational and computational cost reduction in aid of Partially Symbolized Direct Knowledge of Space (PSDKS) while some further quantitative elaboration is needed on this point.

The author has already reported that cross-media operation between texts in several languages (Japanese, Chinese, Albanian, and English) and pictorial patterns like maps were successfully implemented on IMAGES-M [4]. As detailed in this paper, IMAGES-M has recently adopted the multiparadigm language Python in place of PROLOG to facilitate both declarative and imperative programming, and the author has confirmed that the hybrid program in Python employing L_{md} expression mainly and PSDKS auxiliarly is much more flexible and efficient than the previous one in PROLOG for solving problems expressed in spatial language. To our best knowledge, there is no other system (e.g., [20, 21]) that can perform cross-media operations in such a seamless way as described here. This leads to the conclusion that L_{md} has made the logical expressions of event concepts remarkably computable and has proved to be very adequate to systematize cross-media operations. This adequacy is due to its medium-freeness and its good correspondence with the performances of human sensory systems in both spatial and temporal extents while almost all other knowledge representation schemes are ontology-dependent, computing- unconscious or spatial-change-event unconscious (e.g., [8, 9]).

The author deems that aware science or technology is still on the way to maturation and therefore that now it should foster various kinds of approaches. The model of human cognition employed in MIDST is formalized based on declarative knowledge representation in symbolic logic which has almost been discarded in this research area so far and instead certain approaches based on procedural knowledge representation has been prevalent. The author's very intention here is to present some prospective possibility of his original theory MIDST in aware science. The example presented in Section 8 is rather simple but one of the most complicated spatial relations displayable in this version of the intelligent system IMAGES-M because it was programmed exclusively to check the efficacy of PSDKS. Another extended version of the system is now under construction and some examples of further complicated human-system interaction in natural language have already been presented in another paper [15].

Our future work will include establishment of learning facilities for automatic acquisition of word concepts from

sensory data [7] and human-robot communication by natural language under real environments [22].

Acknowledgment

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, nos. 14580436, 17500132, and 23500195.

References

[1] A. Yamada, A. Yamada, H. Ikrda et al., "Reconstructing spatial image from natural language texts," in *Proceedings of the 15th International Conference on Computational Linguistics (COLING '90)*, Nantes, France, 1992.

[2] P. Olivier and J. Tsujii, "A computational view of the cognitive semantics of spatial expressions," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics (ACL '94)*, Las Cruces, New Mexico, 1994.

[3] G. Adorni, M. Di Manzo, and F. Giunchiglia, "Natural language driven image generation," in *Proceedings of the 10th International Conference on Computational Linguistics (COLING '84)*, pp. 495–500, 1984.

[4] M. Yokota and G. Capi, "Cross-media operations between text and picture based on mental image directed semantic theory," *WSEAS Transactions on Information Science and Applications*, vol. 2, no. 10, pp. 1541–1550, 2005.

[5] J. F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole, Pacific Grove, Calif, USA, 2000..

[6] 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*, vol. 3, pp. 231–253, 1997.

[7] S. Oda, M. Oda, and M. Yokota, "Conceptual analysis and description of words for color and lightness for grounding them on sensory data," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 16, no. 5, pp. 436–444, 2001.

[8] R. W. Langacker, *Concept, Image and Symbol, Mouton de Gruyter*, Berlin, Germany, 1991.

[9] G. A. Miller and P. N. Johnson-Laird, *Language and Perception*, Harvard University Press, 1976.

[10] M. Yokota, "Systematic formulation and computation of subjective spatiotemporal knowledge based on mental image directed semantic theory: toward a formal system for natural intelligence," in *Proceedings of the 6th International Workshop on Natural Language Processing and Cognitive Science (NLPCS '09)*, pp. 133–143, Milan, Italy, May 2009.

[11] M. Yokota, "Towards awareness computing under control by world knowledge grounded in sensory data," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '10)*, pp. 769–775, October 2010.

[12] B. M. Shariff, M. J. Egenhofer, and D. M. Mark, "Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms," *International Journal of Geographical Information Science*, vol. 12, no. 3, pp. 215–245, 1998.

[13] P. Roget, *Thesaurus of English Words and Phrases*, J.M. Dent & Sons Ltd, London, UK, 1975.

[14] R. Shepard and J. Metzler, "Mental rotation of three-dimensional objects," *Science*, vol. 171, no. 3972, pp. 701–703, 1971.

[15] M. Yokota, "Systematic analysis and synthesis of human subjective knowledge of space and time for intuitive human-robot interaction," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '11)*, pp. 208–215, 2011.

[16] M. Yokota, "Towards artificial communication partners with a multiagent mind model based on mental image directed semantic theory," in *Humanoid Robots*, B. Choi, Ed., pp. 333–364, I-Tech Press, 2009.

[17] J. F. Allen, "Towards a general theory of action and time," *Artificial Intelligence*, vol. 23, no. 2, pp. 123–154, 1984.

[18] D. V. McDermott, "A temporal logic for reasoning about processes and plans," *Cognitive Science*, vol. 6, no. 2, pp. 101–155, 1982.

[19] Y. Shoham, "Time for actions: on the relationship between time, knowledge, and action," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 954–959, Detroit, Mich, USA, 1989.

[20] J. P. Eakins and M. E. Graham, "Content-based Image Retrieval: A report to the JISC Technology Applications Programme," Institute for Image Data Research, University of Northumbria at Newcastle, 1999.

[21] M. L. Kherfi, D. Ziou, and A. Bernardi, "Image retrieval from the World Wide Web: issues, techniques, and systems," *ACM Computing Surveys*, vol. 36, no. 1, pp. 35–67, 2004.

[22] M. Yokota, M. Shiraishi, and G. Capi, "Human-robot communication through a mind model based on the mental image directed semantic theory," in *Proceedings of the 10th International Symposium on Artificial Life and Robotics (AROB '05)*, pp. 695–698, Oita, Japan, 2005.