

SUPP: UNDERSTANDING MOVING PICTURE PATTERNS  
BASED ON LINGUISTIC KNOWLEDGE

Naoyuki Okada  
Department of Information Science  
and Systems Engineering  
Oita University  
Oita 870-11, Japan

This paper presents a system for understanding moving picture patterns based on linguistic knowledge. In order to deal with the difference of knowledge level between picture pattern and natural language, a hierarchical structure is adopted for the representation of knowledge. At a certain level knowledge data is organized as frame centred around the concept of verb. The concepts of approximately five thousand Japanese verbs were systematically classified to construct such a knowledge system.

The program reads a picture pattern sequence. Pairing a picture pattern with the successive one in the sequence, it recognizes primitive pictures in the pair. Then, checking dynamic relations among primitive pictures, it organizes the relations into verb-centred events. It makes inferences about similarity between two concepts of verbs. Finally it describes each event in Japanese and English.

## 1. INTRODUCTION

It is a matter of course that image and natural language understanding are very important aspects of artificial intelligence researches. The research works in both areas have remarkably increased in number in recent years, but the results obtained in one area have not been sufficiently used for the research works in the other area. There is much difference of the present knowledge level between image, especially real world image, and natural language understanding: the knowledge level for the former is low while one for the latter is high.

In the last decade or more we have worked at constructing a system named SUPP (System for Understanding Picture Patterns) in order to raise the level of image understanding up to the linguistic knowledge level and to fill the gap between the works in the two areas. This paper is an overview of SUPP[1]-[7]. The theory developed there agrees with much of Minsky's frames idea[8]. SUPP can be viewed as a specialization of frames idea in the domain of linguistic understanding of image sequence. It has the following two characteristics:

- a. A systematic and exhaustive investigation of the meaning of image is made from the linguistic viewpoint.
- b. A theoretical basis is given for dealing

-----  
The research reported herein was supported in part by the Japanese Ministry of Education under Grant in Aid for Scientific Research, No. 320709.

with the structure and the meaning of change, be it motion, deformation or so forth, occurring in image sequences.

## 2. CONCEPTUAL TAXONOMY OF JAPANESE VERBS

When an image or image sequence is given, infinite number of static or dynamic events can generally be observed in it. Suppose that the meaning of each event is described in natural language, the description sentences will amount to infinite number. An ordinary sentence is decomposed into simple sentences, each of which is governed syntactically and semantically by a verb. Since there are only finite number of verbs in each language, meanings of infinite number of events are roughly divided into meanings of those verbs and their interrelations. An object in the real world which is identified by a verb is called "matter." It arises accompanying things, events and attributes, which are called "constituents," so its concept can be regarded as the concept of dynamic or static relation among constituents.

After due consideration about correspondence with the real world, the concepts of approximately 5,000 Japanese verbs were classified according to an algorithm. The concepts are divided into two large classes: "simple matter concept" and "non-simple matter concept." The former can not be decomposed into elementary matter concepts while the latter can be decomposed. They can further be classified according to their structural patterns and semantic contents. The results are described in [2] and [3].

### 3. THE KNOWLEDGE SYSTEM

The knowledge system contains a visual, a conceptual, a linguistic and a thesaurus component.

#### 3.1 The Visual and the Conceptual Component

The visual component contains models of primitive pictures and syntactic rules among them. An example of model is shown in Fig. 4-(b). The rules are applied to picture pattern pairs, which Minsky called "before-after" frame-pairs. The conceptual component contains conceptual features, concepts and networks of concepts. A matter concept is expressed by  $[V : c_1c_2 \dots c_l d_1(E_1)d_2(E_2) \dots d_m(E_m)]$  where each  $c_i$  denotes a feature of matter itself and corresponds to a syntactic rule mentioned above. Each  $d_j()$  denotes the case or role of a constituent and must be filled by specific instance or concept of constituent. Features  $E_j (= e_{j1}e_{j2} \dots e_{jn})$  specify the conditions its assignment must meet. Networks are constructed among similar matter concepts.

#### 3.2 The Linguistic and the Thesaurus Component

The linguistic component consists of dictionaries for the production of Japanese and English sentences. The thesaurus component contains all the classified concepts in Chap. 2 and supports to develop other components.

The knowledge system is discussed in detail in [1], [5], [6] and [7].

### 4. THE UNDERSTANDING PROCESS

The process of understanding a sequence of two-dimensional line drawings (handwriting is allowed) is described. The overall system is indicated in Fig. 1.

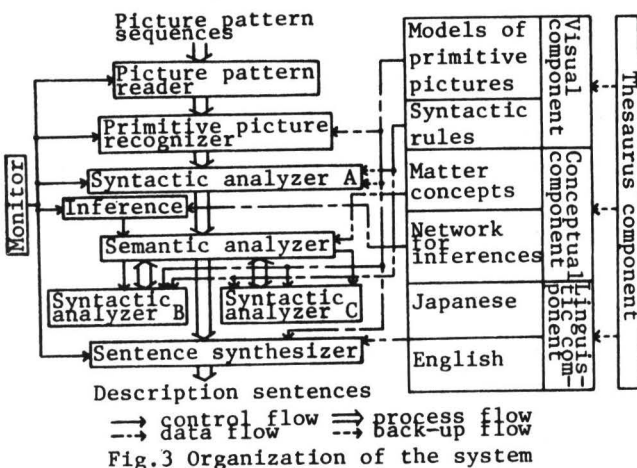


Fig.3 Organization of the system

#### 4.1 Primitive Picture Recognition and Syntactic Analysis

The picture pattern reader is a curve follower that traces line segments by octagonal scanning. The recognizer is based on Evans's matching program[9] for graph-like line drawings but is improved to handle noisy ones.

The syntactic analyzer A decomposes the complex picture, in which two or more primitive pictures may intersect or touch each other, and recognizes them according to Gestalt criteria. The syntactic analyzer B performs Boolean operations on quantized primitive pictures to check such a relation as "MAN INSIDE HOUSE." The syntactic analyzer C performs numerical operations on the data such as coordinates and transformational coefficients of primitive pictures. The programs mentioned above are described in detail in [4] and [5].

#### 4.2 Semantic Analyzer and Inference

The semantic analyzer detects the meaning of matter-centred change in picture pattern pairs by top-down analysis. Suppose that matter  $[V(s, o_\alpha) : c_1c_2 \dots c_l d_s(e_s) d_{o_\alpha}(e_{o_\alpha})]$  is directed by the Inference. The analyzer assigns the role of  $s$  to one of the primitive pictures, say  $P_s$ , after checking whether  $P_s$  meets  $e_s$ . It assigns the role of  $o_\alpha$  to another primitive picture  $P_{o_\alpha}$  in the same way. Then it analyzes each  $c_i$  by calling a correspondent sub-program in the syntactic analyzer B or C. If all the analyses end in success, the meaning of  $V(s, o_\alpha)$  is detected. The present Inference makes inferences about all the similar concepts in the network in depth-first order, directing each matter concept at a node to the semantic analyzer. Finally, the synthesizer produces Japanese and English simple sentences.

### 5. EXPERIMENTS

Almost all the programs were written in FORTRAN to run under the OSII/VS of the FACOM 230-38S medium scale computer at the Department of Information Science and Systems Engineering at Oita University. Experimental results [4]-[7] are shown in Fig. 2-5.

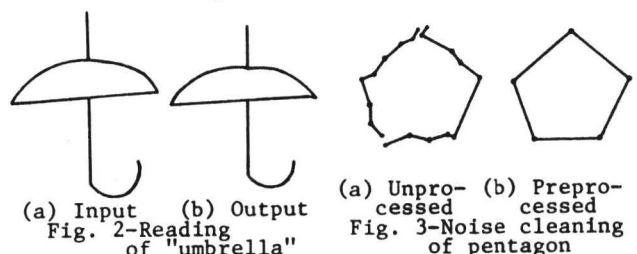
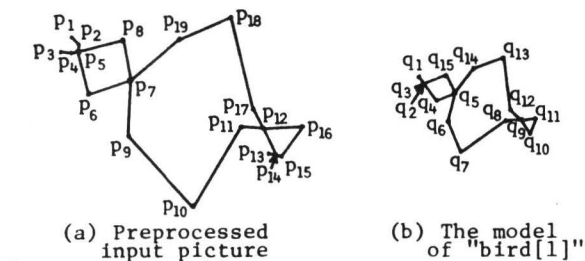


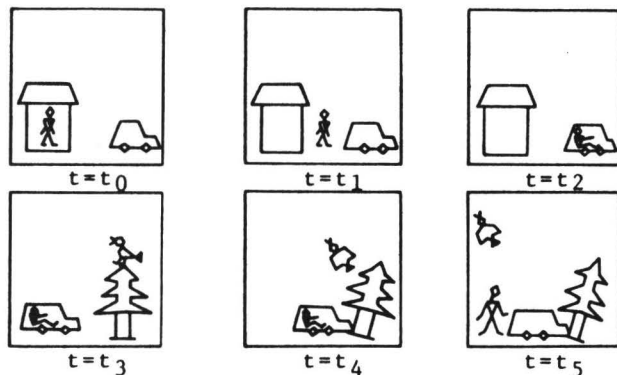
Fig. 2-Reading of "umbrella"

Fig. 3-Noise cleaning of pentagon



1. Pattern matching  
 $[p_7:q_5], [p_9:q_{14}], [p_{19}:q_6], [p_8:q_4], [p_6:q_{15}],$   
 $[p_{10}:q_{13}], [p_{18}:q_7], [p_5:q_2], [p_{11}:q_{12}],$   
 $[p_{17}:q_8], [p_{12}:q_9]$   
 $[p_{16}:q_{10}], [p_{12}:q_9], [p_{15}:q_{11}]$
2. Transformation and similarity  
Translation : (524.4, 483.7), scaling : 2.173  
times, rotation : 2.290 radian, reflection :  
in Y-axis. Similarity : 0.718( $\leq 1$ )

Fig. 4-Recognition of "bird[1]"



- (a) Input sequence
- $t=t_0$  \* OTOKO(4) GA IE NO UCHI NI ITA.  
\* THE MAN[4] WAS IN THE HOUSE.
- $t=t_1$  \* OTOKO(4) GA IE KARA DETA.  
\* THE MAN[4] WENT OUT OF THE HOUSE.
- $t=t_2$  \* OTOKO(4) GA KURUMA NI NOTTA.  
\* THE MAN[4] RODE INTO THE CAR.
- $t=t_3$  \* KURUMA GA HASHITTA.  
\* THE CAR RAN.
- $t=t_4$  \* KURUMA GA KI NI BUTSUKAITTA.  
\* THE CAR COLLIDED AGAINST THE TREE.
- $t=t_5$  TORI(1) GA KI KARA SATTA.  
THE BIRD[1] LEFT THE TREE.  
\* OTOKO(3) GA KURUMA KARA ORITA.  
\* THE MAN[3] ALIGHTED FROM THE CAR.

- (b) Output sentences
1. Description of  $t=t_0$  is for the picture pattern at  $t=t_0$ .
  2. Description of  $t=t_i$  ( $i \geq 1$ ) is for the picture pattern pair at  $t=t_i$  and  $t_{i-1}$ .
  3. The \* marks were assigned by man.

Fig. 5-Understanding of a picture pattern sequence

## 6. CONCLUSIONS

A system has been surveyed that understands moving picture patterns linguistically. Badler's approach[10] basically agrees with ours in handling verb-centred events but differs in that SUPP is based on a systematic analysis of almost all the verbs used for everyday Japanese while he notes only motion verbs in English. By SUPP image understanding research has come closer to natural language understanding research.

## ACKNOWLEDGEMENT

The author started to develop SUPP some ten years ago when he was at Kyushu University. The author wishes to express his gratitude to Prof. T.Tamati of Kyushu University for kind guidance and material support.

## REFERENCES

- [1] Okada, N. and Tamati, T. "Semantic Information of Natural Language and its Extraction and Classification." Trans. IECE Japan 52-C:10 (1969) 363-370.
- [2] Okada, N. and Tamati, T. "An analysis and Classification of "Simple Matter concepts" for Natural Language and Picture Interpretation." Trans. IECE Japan 56-D:9 (1973) 523-530.
- [3] Okada, N. and Tamati, T. "An Analysis and Classification of "Non-Simple Matter Concepts" for Natural Language and Picture Interpretation." Trans. IECE Japan 56-D:10 (1973) 591-598.
- [4] Okada, N. and Tamati, T. "Automated Editing of Fuzzy Line Drawings for Picture Description." Trans. IECE Japan 57-A:3 (1974) 216-223.
- [5] Okada, N. and Tamati, T. "Interpretation of the Meaning of Picture Patterns and its Description in Natural Language — Primitive Picture Recognition and Syntactic Analysis." Trans. IECE Japan J59-D:5 (1976) 323-330.
- [6] Okada, N. and Tamati, T. "Interpretation of the Meaning of Moving Picture Patterns and its Description in Natural Language—Semantic Analysis." Trans. IECE Japan J59-D : 5 (1976) 331-338.
- [7] Okada, N. and Miura, A. "A System for Linguistic Interpretation of Picture Patterns." Rep. Fac. Engng. Oita University No. 5(1979)61-70.
- [8] Minsky, M. "A Framework for Representing Knowledge." In The Psychology of Computer Vision, Winston, P.H.(ed.), McGraw-Hill, 1975, pp. 211-277.
- [9] Evans, T. G. "A Program for the Solution of Geometric Analogy Intelligence-Test Questions." In Semantic Information Processing, Minsky, M. (ed.), MIT Press, 1968, pp. 271-353.
- [10] Badler, N.I. "Temporal Scene Analysis : Conceptual Descriptions of Object Movements." Tech. Rep. University of Toronto No. 80(1975) .