Reconstructing Spatial Image from Natural Language Texts

Atsushi Yamada Tadashi Yamamoto Hisashi Ikeda
Toyoaki Nishida Shuji Doshita
Department of Information Science, Faculty of Engineering, Kyoto University
Sakyo-ku, Kyoto 606-01, Japan
phone: +81-75-753-4952
email: yamada@kuis.kyoto-u.ac.jp

Abstract

This paper describes the understanding process of the spatial descriptions in Japanese. In order to understand the described world, the authors try to reconstruct the geometric model of the global scene from the scenic descriptions drawing a space. It is done by an experimental computer program SPRINT, which takes natural language texts and produces a model of the described world. To reconstruct the model, the authors extract the qualitative spatial constraints from the text, and represent them as the numerical constraints on the spatial attributes of the entities. This makes it possible to express the vagueness of the spatial concepts and to derive the maximally plausible interpretation from a chunk of information accumulated as the constraints. The interpretation reflects the temporary belief about the world.

1 Introduction

This paper concentrates on the understanding process of the verbal expressions concerning about space. One can easily imagine the described world from the verbal expressions. We regard the interpretation of descriptions as an active process, that is the process of reconstruction of a situation which the speaker intended. In this process, one will use many kinds of information. The natural language descriptions contain some of them, and it is very important to extract and use them, but they are not enough. Among them, information about the configuration of the world in one's image plays an important role.

We have made an experimental computer program SPRINT (for "SPatial Representation INTerpreter"), which takes spatial descriptions written in Japanese, reconstructs 3-dimensional model of the world, and outputs the corresponding image on the graphic display.

In this paper, We describe the overview of this system.

2 The Approach

The essence of our approach is as follows:

- Meaning of the natural language expressions as the constraints among the spatial entities
- Image representation of the world as a collection of the parameterized entities
- Interpreting the qualitative relations as the numerical constraints among the parameters
- Potential energy functions for the vague constraints
- Extracting the procedure of the reconstruction from the natural language expressions
- Successive refinement and modification of the world model.

We regard the world as an assembly of the spatial entities, and represent each entity as the combination of its prototype and the real values of its parameters. We prepare the graphic objects corresponding to the prototypes. Each graphic object is represented by the parameters prescribing the details of it. The parameters prescribe its location, orientation, and extent.

Now the task becomes to generate the graphic objects corresponding to the described entities and to determine the parameter values prescribing them.

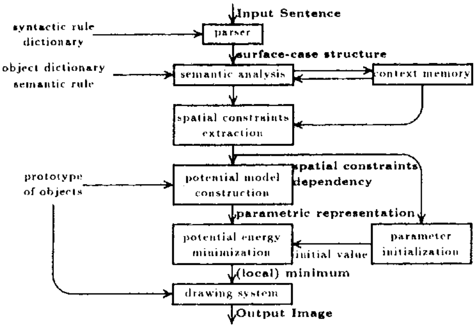

Figure 1: The Overview of the Experimental System SPRINT

It is difficult to determine the parameter values directly from the natural language descriptions, because of the partiality of the information and the vagueness about the spatial relations among the entities. So, at first, we extract such information as the qualitative spatial constraints among the spatial attributes of the entities, and then, interpret these constraints and calculate the parameter values. This process is shown in figure 1.

Given a text, SPRINT makes a surface case structure using the lexical information. Each entity is described as a noun. Next, SPRINT extracts spatial constraints about the entities by analyzing the related words in the case structure. At this time, SPRINT also extracts the sequence of the information references from the lexical information as dependencies, which are used as cues in the calculation of the parameters.

At the next stage, the extracted qualitative constraints are interpreted as the numerical constraints among the entity parameters. These numerical constraints are represented as the combination of the primitive constraints. The potential energy function is one of such primitives, and this is an efficient method to treat the vagueness in the constraints. Other primitives are the topological constraints and the regions. The potential energy function is a kind of the cost functions which takes all related parameters and output the cost. The less the value of the potential energy function, the more credit the combination of the geometric parameters gains. Using the gradient descendent method, the solution with minimum cost is calculated. The potential energy function provides a means for accumulating from fragmentary information. (The basic idea of the potential energy function is reported in [4].)

3 The Example

Suppose that the following sentences are the inputs to SPRINT.

(1) 「山下公園の中央には噴水がある。」(There is a fountain at the center of the Yamashita Park.)

(2) 「噴水のところから公園の柵の向こうに氷川丸を見ることができる。」(From that place, you can see Hikawa-maru (a ship) beyond the fence of the park.)

(3) 「氷川丸の右方にはマリンタワーがたっている。」(There is a marine tower to the right hand of Hikawa-maru.)

From these sentences, SPRINT gets the surface case structures and interprets each connection in the structures to extract spatial constraints. The extracted constraints in this example is shown in table 1.

Then SPRINT calculates the entity parameter values based on these constraints using potential energy functions. The example of the potential energy function is shown in figure 2. This is one which is used to calculate the location of the ship. In this figure, the line represents the edge of the park, and the thither side of the line means the inside of the park. Finally SPRINT draw a world image on the graphic display. This is shown in figure 3.

4 The Analysis of the View

In the last example, the treatment of the view is very important. Usually an observer sees the world and notices how the world is. If you did not know which direction the observer sees, you would not determine the direction "to the right" and could not imagine where the tower is. Another way to determine the direction "to the

Table 1: The Extracted Spatial Constraints

- The Spatial Entities in the World

  Yamashita Park (a park), a fountain, a fence, Hikawa-maru (a ship), a marine tower (a tower)

- The Relations among the Entities

  location(Fountain)=near(center(Park)), inside(region(Park))

  location(Eye-point)=near(location(Fountain)), outside(region(Fountain))

  location(Fence)=at(boundary(region(Park)))

  orientation(View)=from(location(Eye-point), to(location(Fence))

  location(Aim-point)=hinter(Fence, Eye-point), location(Ship)

  location(Tower)=right(Ship, Eye-point)

- The Used Knowledge about the Entities

  Park = Two-dimensional region (a kind of Ground)

  Fountain = Three-dimensional object with Water

  Fence = Three-dimensional object at the boundary of the two-dimensional region

  Ship = Three-dimensional object (a kind of Vehicle on Water)

  Tower = Three-dimensional object (a kind of Building)

- The General World Knowledge

  Any two objects cannot occupy the same place at the same time.

  Every object is under the law of gravitation (i.e. it must supported unless a special reason).

  Every object has its plausible extent.

  There exists a distance scale according to the extent.

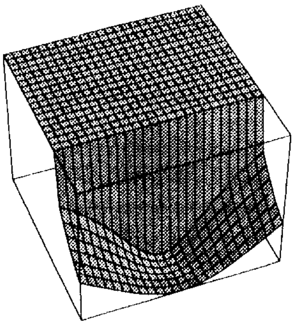

Figure 2: The Example of the Potential Energy Function

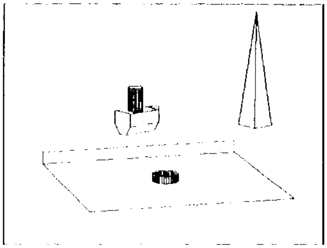

Figure 3: The Output Image of the Interpretation

Table 2: The Basic Constraints from the Eye Point

- constrain the eye point by the point of "observation"
  constrain the aim point by the aimed entity
- constrain the view by the eye point
- constrain the view by the aim point
- constrain the view by the eye direction

right" is to calculate it only from the orientation of the ship, but we do not think it is usual. This means that the spatial image reflects the history of the inference, and the constructed image is used again to understand the next sentence.

So SPRINT also has to

- pursue of the eye point of the observer.
- set the view of the observer from the eye point,
- infer the spatial configuration from the view.

For this sake, we modeled the view of the observer as one of the spatial entities, which has the eye point, the aim point, and the eye direction. In this section, we analyze the descriptions about view in details.

At first, we define the relation about "see" as follows:

> "There is no visible obstacles between the eye point and the aimed entity."

The constraints about the eye point, eye direction, and the aim point comes from this definition.

The simplest case is shown in table 2. For example, there are 5 constraints to the sentence 「駅前広場のところから北の方にタワーが見

える」 (You can see a tower north from the station square.)

- **(A)** location(Eye-point1)=Station Square
- **(B)** location(Aim-point1)=Tower
- **(C)** start-point(View1)=Eye-point1
- **(D)** end-point(View1)=Aim-point1
- **(E)** direction(View1)=North

If the eye point has its own direction, the 5th constraint in 2 becomes a relative one based on the direction of the eye point. For example, to the sentence 「十字路を越えると、右手にタワーが見える」 (If you get across the crossroad, You can see a tower to the right hand.) the constraint (E) above becomes

(E′) direction(view1)=to-the-right(view-point2)

which means the direction "to the right" is determined by the direction of the eye of the observer. In this case the observer get across the crossroad and no other information is obtained, so the direction of the eye is determined as the same as that of the transfer of the observer.

There are the cases where the direction of the eye changes among the transfer. In such cases, the last eye direction must be calculated according to the intermediate changes. So the change point is put, and it mediates the change of the direction of the transfer. The necessary constraints are as follows:

- constraint about the change point
- constraint about the transfer before the change
- constraint about the transfer after the change

For example, there are 5 constraints for the sentence 「十字路を左折する」 (turn left at the crossroad).

- **(A)** location(Change-point1)=Crossroad
- **(B)** start-point(Transfer-vector1)=the last Eye-point
- **(C)** end-point(Transfer-vector1)=Change-point1
- **(D)** start-point(Transfer-vector2)=Change-point1
- **(E)** direction(Transfer-vector2)=to-the-left(Transfer-vector1)

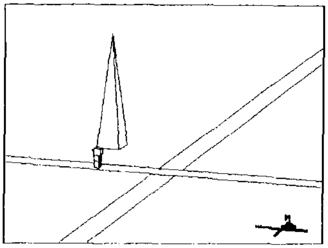

Figure 4: The View Interpretation of the Transfer with the Intermediate Change

The direction of the eye after the change is same as the direction of the transfer-vector2. For example, the sentence 「十字路を左折すると、右手にタワーが見える」 (If you turn left at the crossroad, you can see a tower to the right hand.) is interpreted as in figure 4. In this case, the direction "to the right" is calculated from the last direction of the eye.

This interpretation satisfies the constraints in the sentence, however, one may think this is not the same as he/she imagine because in this interpretation the observer can see the tower even before the crossroad. The sentence "If you turn left ..." seems to imply that "until you turn left at the crossroad, you cannot see a tower yet," and this is not in the case of the logical sense. Of course this is not always true. Suppose the situation where you see a tower now and are told the last sentence (probably in English you say not "a tower" but "the tower"), this will be the case of the integration of the several views. So the additional pragmatic constraints are strongly influenced by the purpose of the utterance.

Anyway if you do not want to see a tower before the crossroad, one of the solutions to this problem is like this: put some obstacle on the view of the observer before the crossroad, that means put it between the point of the observer and the tower. In this case, till the observer turn at the corner, there is no way to know the location of the tower, so no way to put the obstacle. The interpretation according to this solution is shown in figure 5.

One of the other solutions is that you know

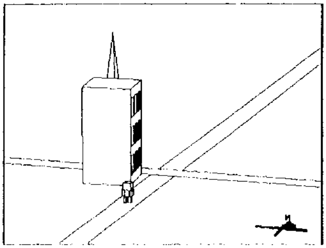

Figure 5: The View Interpretation with the Added Obstacle

there is buildings or something along a street and use them as a obstacle.

This kind of 'invisible' situation must be discussed with respect to the read world and the daily language use.

5 Related Work

From the pure linguistic point of view, A. Herskovits [1] analyzed locative expressions in English. As for constructing a computer model, conventional logic falls short of our purpose. Among the formulations based purely on conventional logic, most typical is slot-filler representation such as a formulation by Gordon Novak Jr [2]. There also is a work by D. Waltz[3]. It is however hard to draw logical conclusion out of a set of axioms which may involve predicates vague and to get a reusable model of the world configuration.

Our approach allows both continuous and discontinuous functions to represent spatial constraints, so that the probability changes either continuously and discontinuously.

It also works as a chunk of the information. Though it seems that our approach is rather subjective, it seems impossible to construct a model for the world without some kind of subjective.

6 Conclusions

We have presented an experimental computer program which produces 3-dimensional image as an interpretation of the given natural language texts. The area of space-language relationship contains a lot of hard issues, and some problems related to this work are mentioned below.

- presentation of the image.
  Our program makes a internal 3-dimensional model of the world, but the presentation on the screen is now manually done, which means that the camera position for the computer graphics is manually decided (it is usually a bird's-eye view). How to present the internal configuration as an image is a further problem.
- integration of the initial image.
  If all the model is constructed based on the verbal information, how to give the initial values of the parameters effectively becomes the problem. If the the reconstruction begins with an initial image, the integration of that image and the verbal information is the other problem. (Probably the initial image is also vague.)

We are now considering the pragmatic use of the verbal expression in the world model, and making a model of the visual disappearance.

References

[1] A. Herskovits. *Language and Spatial Cognition.* Cambridge University Press, 1986.

[2] G.S. Novak Jr. Representations of knowledge in a program for solving physics problems. In *Proc. IJCAI-77*, pages 286 291, 1977.

[3] D.L. Waltz. Towards a detailed model of processing for language describing the physical world. In *Proc. IJCAI-81*, pages 1 6, 1981.

[4] A. Yamada, T. Nishida, and S. Doshita. Figuring out most plausible interpretation from spatial descriptions. In *Proc. COLING-88*, pages 764 769, 1988.