Use of Reconstructed Spatial Image in Natural Language Understanding Process

Atsushi Yamada and **Tadashi Yamamoto** and **Hisashi Ikeda**
Toyoaki Nishida and **Shuji Doshita**
Department of Information Science, Faculty of Engineering, Kyoto University
Sakyo-ku, Kyoto 606-01, Japan
e-mail: yamada@kuis.kyoto-u.ac.jp

Abstract

This paper describes the use and the benefit of the spatial image of the world in natural language understanding process. The actual or purely imaginary image of the world helps us to understand the natural language texts. In order to treat the image of the described world, the authors use a geometric representation and try to reconstruct a geometric model of the global scene from the scenic descriptions (in Japanese) drawing a space. An experimental computer program SPRINT is made to reconstruct a model. SPRINT extracts the qualitative spatial constraints from the text and represents them by the numerical constraints on spatial attributes of the described entities in the world. This makes it possible to express the vagueness of the spatial concepts, to accumulate fragmentary information on the memory, and to derive the maximally plausible model from a chunk of such information. In this process, the view of the observer and its transition is reflected. One can hardly treat the view without such geometric representations. The visual disappearance of the spatial entities is also discussed with respect to the view of the observer. By constructing a geometric representation of the world, these phenomena are reviewed.

Introduction

This paper concentrates on the understanding process of the verbal expressions concerning about space, and the use of the spatial image in that process. 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 kind of information. The natural language descriptions contain some of them, but they are not enough. Among them, information about the configuration of the world in one's image plays an important role.

So we decide to use a geometric representation as a world model, and try to reconstruct a geometric model of the global scene from the spatial descriptions in Japanese by an experimental computer program SPRINT (for "SPatial Representation INTerpreter"), which takes spatial descriptions written in Japanese as input, reconstructs a 3-dimensional geometric model of the world, and outputs the corresponding image on the graphic display.

In this paper, we describe our basic approach and discuss how the reconstructed image is used in the natural language understanding process, especially for the descriptions of the view of the observer.

Approach to the Image Reconstruction

The essence of our approach to the image reconstruction 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 geometric 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 numerical 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, i.e. 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. 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 qualitative constraints as the numerical constraints among the entity parameters, and calculate the parameter values.

This process is shown in figure 1.

The numerical constraints are represented as the combination of the primitive constraints. The potential energy function is one of such primitives, and this

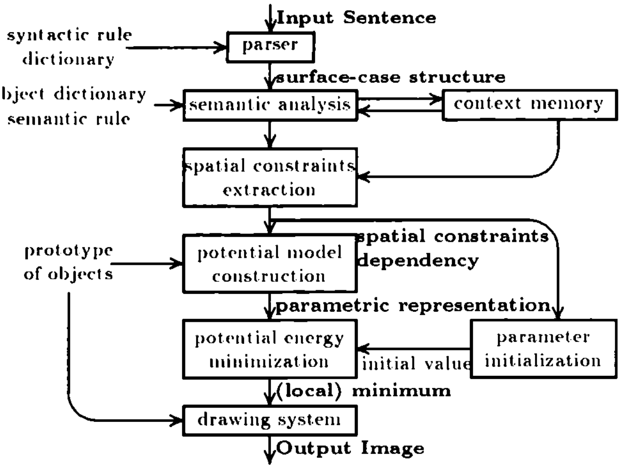


Figure 1: The Overview of SPRINT

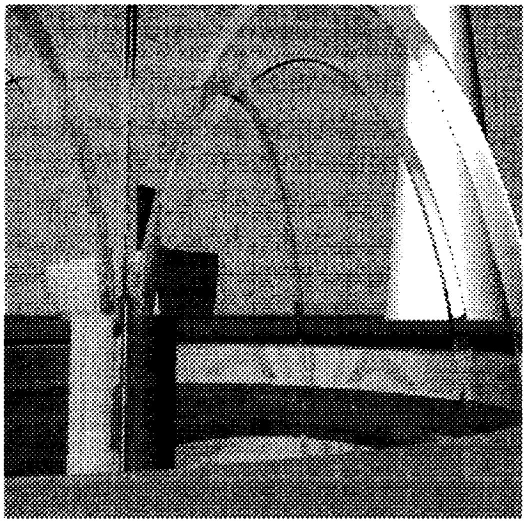


Figure 2: The Output Image of the Final View (with Ray-traced View from the Place of the Observer)

is an efficient method to treat the vagueness in the constraints. Other primitives are the topological constraints and the regions. The potential energy function provides a means for accumulating from fragmentary information. (The details of this process are reported in [Yamada *et al.*, 1988] [Yamada *et al.*, 1990].)

Example of the Reconstruction

Suppose that the following sentences are the inputs to SPRINT.

> 「山下公園の中央には噴水がある。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。氷川丸の右方にはマリンタワーがたっている。」
> (There is a fountain at the center of the Yamashita Park. From that place, you can see Hikawa-maru (a ship) beyond the fence of the park. 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 the spatial constraints. The extracted constraints in this example are about the relation between the "Park" and the "Fountain", among the "Fountain", the "Fence" and the "Ship", and so forth.

Then SPRINT calculates the entity parameter values based on these constraints using potential energy functions. For example, in order to calculate the location of the "Ship", SPRINT uses a directional potential function which represents a direction toward the "Fence" from the "Fountain" and a inhibited region which prohibit the hither side of the "Fence".

Finally SPRINT draw a world image on the graphic display. The final view from the place near the "Fountain" toward the "Ship" and the "Tower" is shown in figure 2. It is a likely view of the observer.

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. For example, 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. 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 that view.

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 views in details.

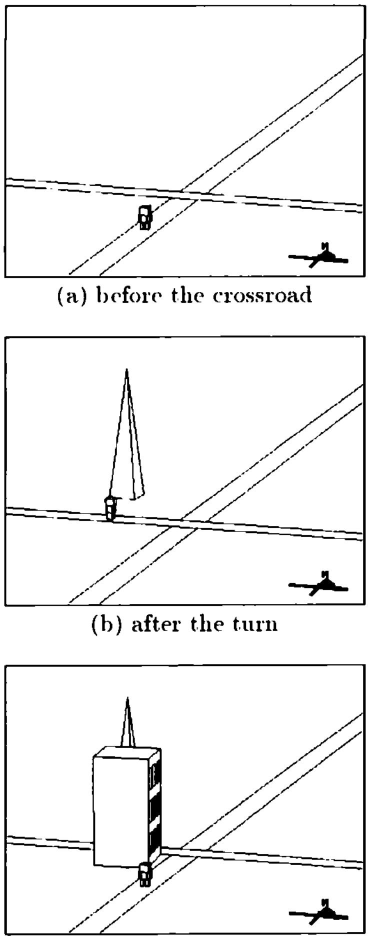
At first, we define the relation about "see" as follows:

> "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 come from this definition.

If the eye point has its own direction, the constraint on the direction becomes a relative one. 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 on the direction of the view depends on the direction of the eye of the observer, and as the observer get across the crossroad and no other information is obtained, 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, each 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. 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 3 (a),(b). In this case, the direction "to the right" is calculated from the last direction of the eye, and it is the same with that of the transfer after the



(a) before the crossroad

(b) after the turn

(c) with the added obstacle

Figure 3: The View Interpretation of the Transfer with the Intermediate Change

turn, which is calculated from the direction before the turn.

Though this interpretation satisfies the constraints in the sentence, 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, 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 at that time.

Anyway if you do not want to see a tower before the crossroad, one of the solutions to this 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 before the crossroad 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 3 (c).

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

Analysis of the Visual Disappearance

In the last section, we find some 'invisible' situations must be considered with respect to the image of the world. In this section we argue about the visual disappearance which occurs by a visible obstacle blocking the view of the observer.

An obstacle can block the view of the observer by

(a) transfer of the eye point

(b) transfer of the aimed object

(c) transfer (or appearance) of the obstacle.

In the understanding process of the visual disappearance expressions, one must determine which transfer has occurred and reconstruct the world image to confirm the phenomena. In each case, the model of the world after the disappearance must satisfy the constraint that the view of the observer cross some obstacle.

Let us consider the following example.

「通りから左前方に木が見えた。通りを進むと、木はビルに隠れた。」
(You saw a tree from the street at the left hand and front of you. As you walked along the street, the tree hid behind a building.)

In this case, the observer moves and you also know trees and buildings won't move, so this is the case of the eye point transfer. If you don't know about the building before you are told the second sentence, your image about the world is like figure 4 (a). Told the second sentence, your image about the world becomes like figure 4 (b). The reconfirmation about the visibility is required at this time, because you know the building was there when you saw the tree. This reconfirmation is done by making an image like figure 4 (c).

The following example is the case of the transfer of the aimed object.

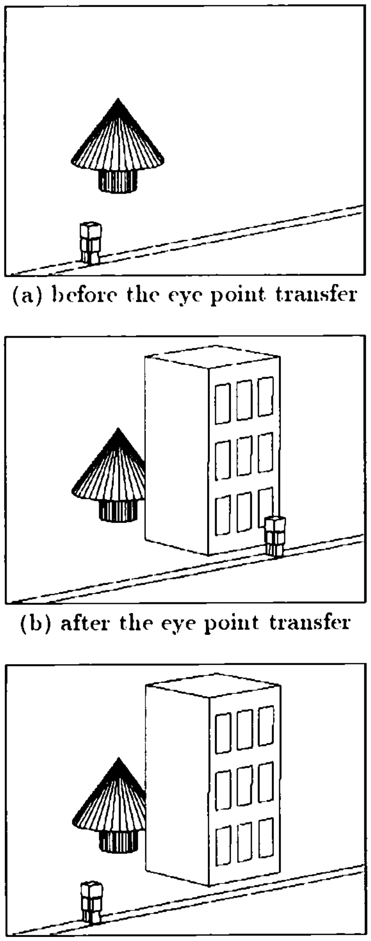
「山の左方に月がでていた。月が山に沈んだ。」
(You saw a moon to the left of the mountain. The moon sank below the mountain.)

The obstacle is the mountain, and the moon is moved behind the mountain so as not to be seen from the observer.

「私の家からはタワーが見えた。ビルが建つと、タワーは見えなくなった。」
(You could see a tower from my house. As a new building was built, you cannot see the tower now.)

is the example of the appearance of an obstacle. The newly built building is put so as to block the view in the model.

If the verbal expression only tells the aimed object and the obstacle like in the expression 「AがBに隠れる」 (A hides behind B), it is difficult to infer the cause of the disappearance. In general, A is a disappeared entity, which was put the aim point, and B is a obstacle, which block the view of the observer. But the cause of the disappearance is influenced by the movability of the



(a) before the eye point transfer

(b) after the eye point transfer

(c) reconfirmation of visibility before the transfer

Figure 4: Visual Disappearance with Eye Point Transfer

entities. For example, as the mountain won't move and the sun is movable, in the interpretation of the sentence 「太陽が山に隠れる」 (the sun hides behind the mountain), the sun as the aimed object is likely to move and the visual disappearance occurs. If 「富士山が雲に隠れる」 (Mt.Fuji hides behind the clouds), the clouds is much more movable, so the clouds as the obstacle move to produce a visual disappearance. In the sentence 「大文字山がビルに隠れる」 (Mt.Daimonji hides behind the building), the mountain and the building are not movable in general. So in many cases, this visual disappearance occurs because of the transfer of the eye point of the observer. But if the situation negates the eye point transfer just like with the phrase 「私の家から」 (from my house), as the building is more movable than the mountain, the building as an obstacle is newly built or extended and the visual disappearance occurs.

In these cases, if one has a image of the world, it can be used as a basis of the inference.

Related Work

From the pure linguistic point of view, A. Herskovits [Herskovits, 1986] 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 [Novak Jr., 1977]. There also is a work by D. Waltz[Waltz, 1981]. 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.

To use a geometric representation as a world model, we can retrieve information that is not mentioned explicitly in the sentences. 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 an accumulator of a chunk of information.

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 and the use of the geometric representation contains a lot of hard issues. 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.
- Degree of the visibility.
  We use very simple model for the view of the observer. It luck the degree of the visibility. For example, with the current model we cannot interpret the expressions like "you can see it clearly/dimly" or "you can see the whole/part of it." How to incorporate these information becomes the problem.

We are now considering the pragmatic use of the verbal expression in the world model. As the use of the geometric model of the world, the research on the denotation with the spatial information is now in progress.

References

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

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

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

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

Yamada, A.; Nishida, T.; and Doshita, S. 1990. Qualitative interpreter for the sense teaching of the spatial descriptions. In *Proc. International Conference on ARCE.* 255–260.