Perspective View of Spatial Relations in 3D Space

Atsushi YAMADA
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, 630-01, JAPAN
phone: +81-7437-2-5241
fax: +81-7437-2-5249
e-mail: yamada@is.aist-nara.ac.jp

Abstract

The objective in this paper is to present a computational model with proper treatment of the projective prepositions. The contribution of the surrounding objects and the view of the observer to the interpretation of the spatial relations is the core problem. At all the previous works, the spatial relations are determined only based on the direction and the proximity between the two (or three) objects in question, and the relation is just for the physical existence of the objects in the world. Incorporating the techniques from the computational geometry and the computer graphics, this paper will show a method of constructing the possible area, and a method of calculating the spatial relation on the perspective view of the observer.

keywords: Natural Language Processing, Spatial Reasoning, Appearance,
Projective Prepositions, Perspective Projection

1 Introduction

People often use verbal expressions to transfer the information about the world. For an AI system that will use verbal expressions as one of the communication medium, connecting visual perception and verbal expressions is an important problem. As we realize the spatial relations among the objects in the world visually, here we concentrate on the spatial relations in 3D space.

In natural language descriptions, the location of some object is mentioned with respect to another object. When describing such relations, people often

use spatial prepositions. Some of these prepositions, such as *at, on, in*, describe topological relations. Others, such as *in front of, behind, to the right of, to the left of*, describe the directional relations. Following Herskovits[Her86], we call the former *topological prepositions*, and the latter *projective prepositions*. As our observation of the world mainly depends on the visual information, the interpretation of the spatial prepositions, especially the projective ones, must be influenced by the characteristics of the view.

This paper argues about the contribution of the surrounding objects and the view of the observer to the interpretation of the spatial relations, and proposes a new method incorporating the technique from computational geometry and computer graphics.

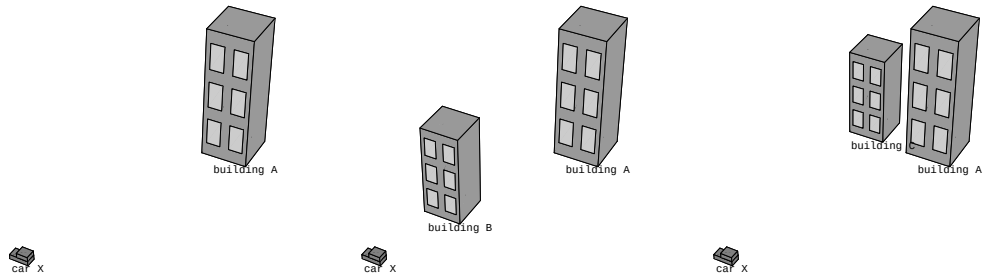
2 Related Work

Several computational models based on the geometric representation are proposed. Here, following Retz-Schmit[RS88], we call the object that is to be located the *primary object*, and the object in relation to which the primary object is located the *reference object*. Retz-Schmit argued about the intrinsic, deictic, and extrinsic use of the projective prepositions. This is the problem how the direction is determined. For example, for the phrase "the ball in front of the car", if the front direction is determined from the orientation of the car itself, it is said intrinsic use. If the direction is determined from the location of the point of view of the observer, it is said deictic use. And if the direction is determined from other outer factors, such as the direction of the motion of the car, it is said extrinsic use. These differences are thought to be qualitative. There also are the quantitative differences. Today, many researchers agree that there are degrees of applicability (vagueness) of spatial prepositions, and it is inappropriate to assume a fixed borderline that separates the area (The exception is Kalita and Badler[KB91], who use the inclusion into the zone that is the outward projection of the extent of the reference object). Not only the direction but also the proximity of the primary object and the reference object is the considered constraint.

The systems CITYTOUR and SOCCER of VITRA project were constructed to reflect these differences, but they are restricted in 2D space. The dimensionality, size and shape of the primary and reference objects are considered in the listener model of the system SOCCER[SS93], but it also works in 2D space.

Gapp[Gap94] proposed a model in 3D space for VITRA project. The evaluation function for the direction and the proximity is defined using spline function in the local coordinate system constructed only from the reference object.

Abella and Kender[AK93] worked on the 2D images, and modeled the



(a) The car is in front of the building A (b) Is the car in front of the building A? (c) Is the car in front of the building A?

Figure 1: Effects of the Surrounding Objects

spatial relations on that image (i.e., 2D plane). They used a set of inequalities and fuzzification using Monte Calro simulation to account for the vagueness.

The system SPRINT[YYI+92] was a system that reconstructs 3D model of the described world. The WIP system[OMT94] is another example that produces depictions of the scene descriptions. Both used the potential field to represent the vagueness, and set some constants according to the extent of the primary and reference object.

3 Effect of Surroundings

At first, let us consider the effect of the surroundings upon the determination of the spatial relations between the primary object and the reference object. At all the previous works, the spatial relations are determined only based on the direction and the proximity between the primary object and the reference object. Is it proper formalization?

For example, if you are talking about a pen on your desk, the ink bottle to the left of the pen should be on your desk, and not on your side table although it is to the left in purely geometric sense. There is a topical region, and in this case, it is the region on the desk.

But there is another kind of limitation of the possible area. Now let us consider the situation in figure 1 (a). There is a car at a place not so far and almost in the front direction of the building A. We may describe the car as "the car in front of the building A". It may not violate the constraints of the direction and the proximity. Then holding the geometric relation between the two, add the building B between the two (figure 1 (b)). Can we still say "the car is in front of the building A"? Or how about the situation like figure 1 (c) where the building C is next to the building A?

These facts do not deny the condition upon the deviation from the ideal direction nor the proximity themselves, but suggest that we must take into account the effects of the surrounding objects besides the geometric relation

between the primary object and the reference object.

The surrounding objects that affect the determination must be salient in some way. Therefore, a small object between the primary object and the reference object may not affect the determination. What kind of object becomes salient is the difficult problem. It may depend on surrounding environment, on many contextual factors and even on the world knowledge. But we may say that the objects that interrupt the view become salient.

Anyway, as in figure 1 (b), there should be no salient object between the primary object and the reference object. Or as in figure 1 (c), the salient object nearby has its own territory. This can be formalized as follows:

> the yonder region of the surrounding salient object, which may be the obstacle that blocks the view between the primary object and the reference object, is not the possible area for the primary object.

When there are several candidate objects surrounding the reference object, not all of them will contribute to the construction of the possible area. For example, the object included in the yonder region of another object will not have any effects. So how can we choose the contributing objects and construct the possible area?

Now as a rough evaluation, let us approximate each object to the center point of the base. The following explanation is for the horizontal 2D space after this approximation. Let the point for the reference object be O, and the set of the candidate points be S, where $P_i \in S$. For O and P_i, we define the line l_i that passes through the point P_i and orthogonal to the line OP_i. Between the regions that the line l_i separates, let us call the region that includes the point O the hither region H_i, and the opposite side the yonder region Y_i. Then the intersection of H_i will become the possible area. Let T be the set of the points that construct the possible area.

At first, the nearest point $P_0 \in T$. Because, if $P_0 \notin T$, there must be P_i whose Y_i includes P_0. This means $\overline{OP_0} > \overline{OP_i}$, but this is a contradiction.

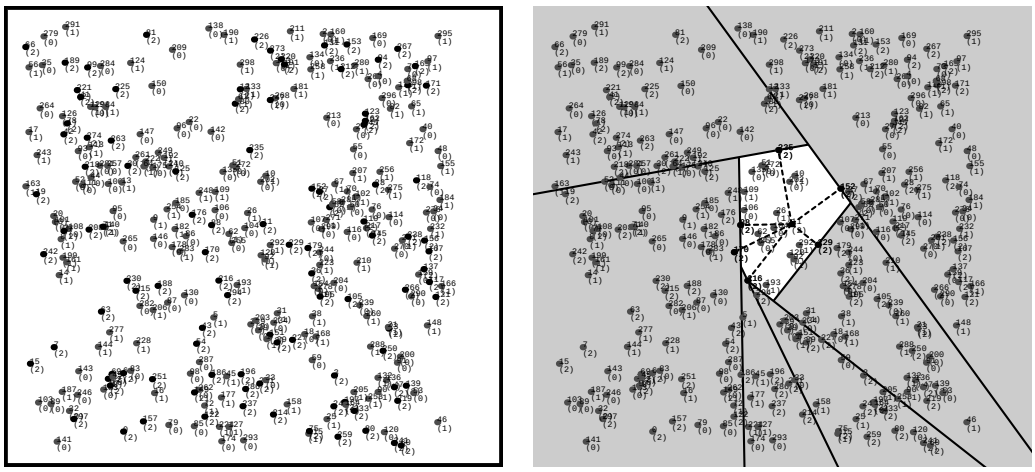
And for the neighboring three points P_{i-1}, P_i, P_{i+1}, if $P_{i-1}, P_{i+1} \in T$ and P_i is inside the circle that passes through O, P_{i-1}, P_{i+1} (we call this condition the *inside circle condition*), $P_i \in T$.

Then the calculating algorithm is as follows:

step 0 for given O, S, make a polar coordinate system (r, θ) whose origin is O, and sort the elements of S by the angle θ. Let the point whose r is minimum be P_0, and make a circular list whose head is P_0.

step 1 $i = 1, T = S$.

step 2 if T includes less than three points, then end.



(a) sample world (b) result of calculation

Figure 2: Construction of the Possible Area

step 3 get neighboring three points P_{i-1}, P_i, P_{i+1} from the circular list.

step 4 if P_i satisfies the inside circle condition for the neighboring points, then goto step 5, else goto step 6.

step 5 if $P_{i+1} = P_0$, then end, else $i = i + 1$, goto step 2.

step 6 remove P_i from T. if $i = 1$, then goto step 2, else $i = i - 1$, goto step 2.

The result of the calculation using this algorithm for a sample world is shown in figure 2. In this figure, for each point, the number above is the point number, and the number below in parentheses shows the object type. Let the point 11 be the reference point, and the salient object be of type 2, the algorithm was applied. The calculated possible area consists of six points, point 229, 216, 170, 88, 235, and 152.

One of the more realistic examples is shown in figure 3. In this case, the possible area for "in front of the tower" is constructed. When the primary object is inside this area, we can argue the degree of the applicability of that relation considering the deviation from the ideal direction and the proximity between the primary object and the reference object.

4 Model of Appearance

4.1 Simple Model of Visibility

As our observation of the world mainly depends on the visual information, whether the object in question is visible or obscured is the essential problem.

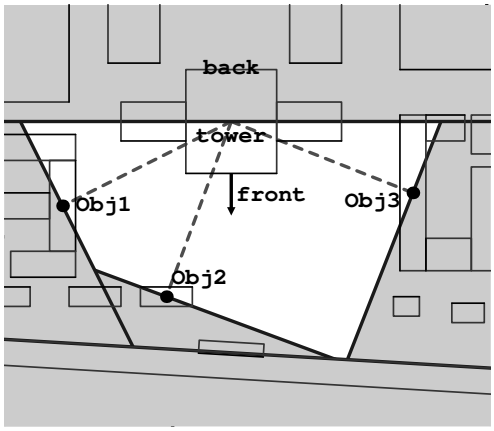


Figure 3: The Possible Area for "in front of"

Usually, the reference object may be visible from the place of the primary object in the intrinsic or extrinsic use, and both objects may be visible from the place of the observer in the deictic use. The former case was argued constructing the possible area in the last section, so let us consider the modeling of the visibility from the observer in this section.

The simplest approach is as follows:

1. check if the primary object satisfies the geometric constraints (direction and proximity) with respect to the reference object.

2. check if the primary object is visible from the location of the observer.

Here, how can we check the visibility? The visibility can be regarded as the condition that there is no obstacle between the eye point of the observer and the focal point on the primary object. Therefore, one solution is:

1. select some representative points on the primary object as focal points (e.g., center of gravity),

2. construct view lines between the eye point of the observer and the focal points,

3. check if the view lines pass through other objects.

This method works well in many cases. But the selection of the representative points is difficult, and the degree of the visibility cannot be handled.

4.2 Construction of Perspective View

There is another problem when we aim at the proper treatment of the spatial relations in 3D space. All the previous approaches are based on the physical

existence of the objects in the world. The determination of the spatial relations was done only for the geometric relation among the actual locations in 3D space. But some of the spatial expressions refer to not the existence but the appearance of the objects. It is typical in the deictic use. For example, one may say "the building to the left of the tower" (in deictic sense) ignoring the difference of the distance from her/him (depth of the view), where the building can be seen to the left of the tower. "The north star to the left of the top of the mountain" [Her86] is one of the extreme examples. In this case, no one thinks of the geometric relation between the physical existence. Herskovits proposed the projection to the infinity plane for this example.

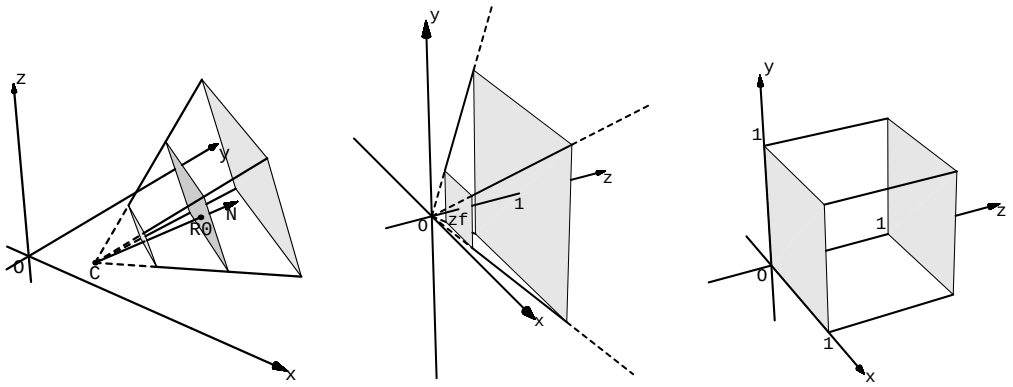
Let us consider this projection more carefully. Mere omission of the depth dimension from the observer does not work. It does not reflect the view of the observer at all. The characteristic that the farther an object is, the smaller it seems, which is called perspective foreshortening, is an important one. This brings us not to the parallel projection but to the perspective projection. In other words, as a computational model, we will simulate the view of the observer and interpret the locative expression on that view plane. We can use the techniques developed in the computer graphics[PK86] for this purpose. But we do not need the full power of the computer graphics. We just want the clue for the determination of the spatial relations, and the realistic view that the researches in the computer graphics pursue is not necessary. So we use the bounding box representation for each object, and adopt the technique originated in the computer graphics. The problem is now to construct the perspective view of the bounding boxes given the point of the observer and the reference object.

Let the world coordinate system be right-handed, where positive direction of z-axis is up (defined from the direction of the gravity) and the view plane coordinate system be left-handed, where viewpoint C is the origin, and the z-axis directs the view. We are using homogeneous coordinate for the sake of simplicity of the calculation.

Now, let the point of the observer be C and the location of the reference object be R_0 in the world coordinate system. The view plane normal vector $\mathbf{N}$ is a unit vector from C to R_0. As a preparation, the bounding box aligned to the vector $\mathbf{N}$ is calculated for each object.

Then, the perspective viewing transformation can be constructed in three steps:

1. transformation from the world coordinate system (figure 4 (a)) to the normalized viewing coordinate system (figure 4 (b)): N_{per}
2. clipping the outside the view: CL
3. normalized perspective to parallel transformation that makes normalized display space (figure 4 (c)): NT_p



(a) world coordinate system (b) normalized viewing coordinate system (c) normalized display space

Figure 4: Perspective Viewing Transformation

world coordinate system to the normalized viewing coordinate system N_{per} is constructed of 6 transformations:

$$N_{per} = S_{S_x,S_y,S_z} \cdot Sh \cdot T_{RL} \cdot R_{\theta,\mathbf{K}} \cdot A_{\mathbf{N},-\mathbf{K}} \cdot T_{-C}$$

T_{-C} is for the translation of $C(a,b,c)$ to the origin

$$T_{-C} = \begin{pmatrix} 1 & 0 & 0 & -a \\ 0 & 1 & 0 & -b \\ 0 & 0 & 1 & -c \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$A_{\mathbf{N},-\mathbf{K}}$ transfers vector $\mathbf{N} = n_1\mathbf{I} + n_2\mathbf{J} + n_3\mathbf{K}$ to vector $-\mathbf{K}$

$$A_{\mathbf{N},-\mathbf{K}} = \begin{pmatrix} \frac{\lambda}{|\mathbf{N}|} & -\frac{n_1 n_2}{\lambda|\mathbf{N}|} & -\frac{n_1 n_3}{\lambda|\mathbf{N}|} & 0 \\ 0 & -\frac{n_3}{\lambda} & \frac{n_2}{\lambda} & 0 \\ -\frac{n_1}{|\mathbf{N}|} & -\frac{n_2}{|\mathbf{N}|} & -\frac{n_3}{|\mathbf{N}|} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $|\mathbf{N}| = \sqrt{{n_1}^2 + {n_2}^2 + {n_3}^2}$, $\lambda = \sqrt{{n_2}^2 + {n_3}^2}$. If $n_2 = n_3 = 0$ then

$$A_{\mathbf{N},-\mathbf{K}} = \begin{pmatrix} 0 & 0 & -\frac{n_1}{|n_1|} & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{n_1}{|n_1|} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$R_{\theta,\mathbf{K}}$ is a rotation around $\mathbf{K}$ which transfers vector $\mathbf{I}'_p$ to vector $\mathbf{I}$, where vector $\mathbf{I}'_p$ is the transformation of vector $\mathbf{I}_p$. Vector $\mathbf{I}_p$ makes the original view plane coordinate system. These vectors are calculated as follows:

$$\mathbf{J}_q = \frac{\mathbf{U}_q}{|\mathbf{U}_q|}$$

where $\mathbf{U}_q = \mathbf{U} - (\mathbf{N} \cdot \mathbf{U})\mathbf{N}$ and $\mathbf{U}$ is up vector $(0, 0, 1)$. Then

$$\mathbf{I}_p = \frac{\mathbf{N} \times \mathbf{J}_q}{|\mathbf{N} \times \mathbf{J}_q|}$$

and

$$\mathbf{I}'_p = A_{\mathbf{N},-\mathbf{K}} \cdot \mathbf{I}_p = a'_p\mathbf{I} + b'_p\mathbf{J}$$

Here,

$$R_{\theta,\mathbf{K}} = \begin{pmatrix} \frac{a'_p}{\sqrt{a'^2_p + b'^2_p}} & \frac{b'_p}{\sqrt{a'^2_p + b'^2_p}} & 0 & 0 \\ -\frac{b'_p}{\sqrt{a'^2_p + b'^2_p}} & \frac{a'_p}{\sqrt{a'^2_p + b'^2_p}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

T_{RL} is the transformation from the right-handed to the left-handed

$$T_{RL} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Sh is the shearing to transfer the center line of view volume to z-axis. Now, let the lower left corner of the original view plane be $L(p_{min}, q_{min})$, and the upper right corner of the original view plane be $R(p_{max}, q_{max})$, then the center point $C_w = \left(\frac{p_{min}+p_{max}}{2}, \frac{q_{min}+q_{max}}{2}\right) = (w_p, w_q)$. In the world coordinate system, the view plane was defined by origin $R_0(x_0, y_0, z_0)$, vector $\mathbf{I}_p = (a_p, b_p, c_p)$, and $\mathbf{J}_q = (a_q, b_q, c_q)$. Therefore

$$\overrightarrow{CC}_w = (x_0 + w_p a_p + w_q a_q - a)\mathbf{I} + (y_0 + w_p b_p + w_q b_q - b)\mathbf{J} + (z_0 + w_p c_p + w_q c_q - c)\mathbf{K}$$

and the transformation of $\overrightarrow{CC}_w$ is

$$(\overrightarrow{CC}_w)' = T_{RL} \cdot R_{\theta,K} \cdot A_{N,-K} \cdot \overrightarrow{CC}_w = c'_x\mathbf{I} + c'_y\mathbf{J} + c'_z\mathbf{K}$$

and

$$Sh = \begin{pmatrix} 1 & 0 & -\frac{c'_x}{c'_z} & 0 \\ 0 & 1 & -\frac{c'_y}{c'_z} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

S_{S_x,S_y,S_z} is a scaling that makes the back clipping plane be $z = 1$ and window boundaries be $x = z, x = -z, y = z, y = -z$. Let $l = p_{max} - p_{min}$, $w =$

$q_{max} - q_{min}$, and the front distance (between the front clipping plane and R_0) be f, the back distance (between the back clipping plane and R_0) be b. Then

$$S_{S_x,S_y,S_z} = \begin{pmatrix} \frac{2c'_z}{l(c'_z+b)} & 0 & 0 & 0 \\ 0 & \frac{2c'_z}{w(c'_z+b)} & 0 & 0 \\ 0 & 0 & \frac{1}{c'_z+b} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Here $z_f = \frac{c'_z+f}{c'_z+b}$.

clipping the outside of the view If the transformed point (x, y, z) satisfies following condition, then the point is inside the view volume:

$$z - y \geq 0 \wedge y + z \geq 0 \wedge z - x \geq 0 \wedge x + z \geq 0 \wedge 1 - z \geq 0 \wedge z - z_f \geq 0$$

normalized perspective to parallel transformation NT_p transforms the normalized viewing coordinate system to the normalized display space $(0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1)$

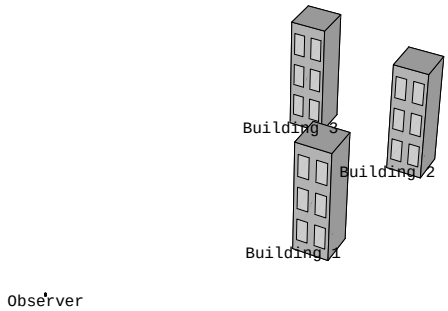
$$NT_p = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{1-z_f} & -\frac{z_f}{1-z_f} \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

4.3 Spatial Relations in Perspective View

The above method is implemented and tested on Sun SPARC station using Common Lisp and *Mathematica*. Now let us apply this method to the sample world A shown in figure 5. Here, which one is "the building to the left of the building 1"? In this case the building 1 becomes the reference object, and the result of the application of the perspective projection to the center of gravity of each object is shown in figure 6.

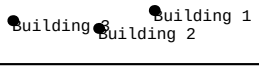
Though the building 2 is not physically to the left of the building 1 in any sense in this world, it is determined to the left of the building 1 on this projected plane because it satisfies the constraints on direction and proximity. The degree of the plausibility can be measured using potential field[YYI+92].

The advantage of this method is that the visibility of the objects can be represented easily. For this purpose, we adopt the 2D occupancy array representation. Occupancy array is an array representing space, whose cells are labeled with the object filling the part of the space. As we have approximated each object with the bounding box aligned to the vector **N**, the front side of the bounding box may become the visible part, and all the z-coordinate (depth) values of that side are the same. Having transformed the



name	location	extent	orientation
Observer	(0, 0, 0)		
Building 1	(150, 150, 0)	(30, 30, 100)	(0, -1, 0)
Building 2	(170, 280, 0)	(30, 30, 100)	(0, -1, 0)
Building 3	(70, 300, 0)	(30, 30, 100)	(0, -1, 0)

Figure 5: Sample World A



name	projected center of gravity
Building 1	(0.5, 0.117851, 0.888889)
Building 2	(0.377778, 0.0785674, 0.962963)
Building 3	(0.189189, 0.095555, 0.930931)

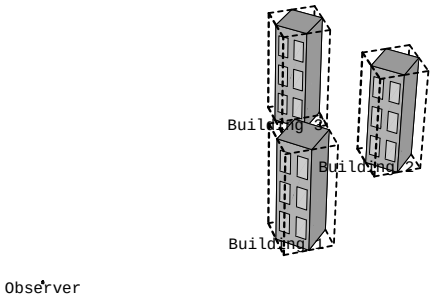
Figure 6: Perspective Projection of the Sample World A (Center of Gravity)

world coordinate system into the normalized display space, simply comparing z-value gives us the information about which one is nearer to the observer than others. The bounding box representation of the sample world A and the calculated constructed array is shown in figure 7. In this case, comparing z-value, the building 1 is the nearest, next one is the building 3, and the building 2 is the farthest. As the actual dimensions of all the buildings are the same, the extent of the appearance in the occupancy array reflects this fact (i.e., the farther object is smaller).

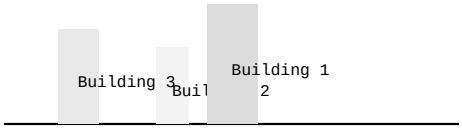
If some object overlaps other object, the pixel will be occupied with the hither object. Therefore in the case of figure 8, the building 2 obscures the building 3.

We can handle the partially visible objects with this representation. In the case of figure 9, the building 3 is partially visible. Furthermore, calculating the ratio of the visible area in the array to the area occupied including the hidden part, we can also handle the degree of the visibility.

Figure 10 shows another example for the calculation of the visibility. The



(a) Bounding Box



name	projection of the front 4 vertices	
Building 1	(0.444444, 0., 0.864198)	(0.555556, 0., 0.864198)
	(0.555556, 0.261891, 0.864198)	(0.444444, 0.261891, 0.864198)
Building 2	(0.333333, 0., 0.952381)	(0.404762, 0., 0.952381)
	(0.404762, 0.168359, 0.952381)	(0.333333, 0.168359, 0.952381)
Building 3	(0.117647, 0., 0.915033)	(0.205882, 0., 0.915033)
	(0.205882, 0.207973, 0.915033)	(0.117647, 0.207973, 0.915033)

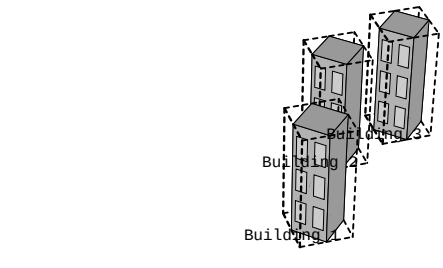
(b) Occupancy Array

Figure 7: Perspective Projection of the Sample World A (Occupancy Array)

parameters of each object in the sample world D was given at random.

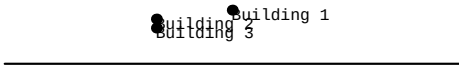
The future problems are as follows:

- The bounding box representation is good as a first approximation and easy to handle, but it does not reflect the real shape of the objects.
- Probably not all the deictic use refers the appearance. The switching between the interpretation on physical location in 3D space and the interpretation on appearance is another problem.
- We have argued the determination of the spatial relations from the given 3D world model to the 2D appearance. Reversely, how can we construct the world model so that it satisfies the given expressions? This becomes the problem when we take a task like depiction from natural language descriptions.



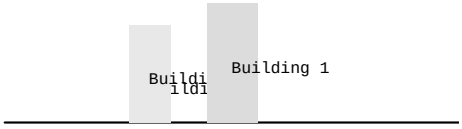
name	location	extent	orientation
Observer	(0, 0, 0)		
Building 1	(150, 150, 0)	(30, 30, 100)	(0, -1, 0)
Building 2	(120, 240, 0)	(30, 30, 100)	(0, -1, 0)
Building 3	(150, 300, 0)	(30, 30, 100)	(0, -1, 0)

(a) Sample World B (with Bounding Box)



name	projected center of gravity
Building 1	(0.5, 0.117851, 0.888889)
Building 2	(0.333333, 0.0982093, 0.925926)
Building 3	(0.333333, 0.0785674, 0.962963)

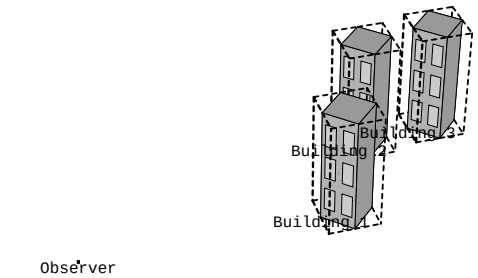
(b) Projected Center of Gravities



name	projection of the front 4 vertices	
Building 1	(0.444444, 0., 0.864198)	(0.555556, 0., 0.864198)
	(0.555556, 0.261891, 0.864198)	(0.444444, 0.261891, 0.864198)
Building 2	(0.272727, 0., 0.909091)	(0.363636, 0., 0.909091)
	(0.363636, 0.214275, 0.909091)	(0.272727, 0.214275, 0.909091)
Building 3	(0.285714, 0., 0.952381)	(0.357143, 0., 0.952381)
	(0.357143, 0.168359, 0.952381)	(0.285714, 0.168359, 0.952381)

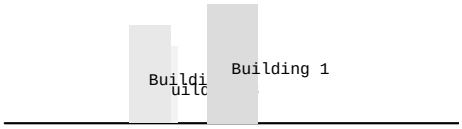
(c) Occupancy Array

Figure 8: Perspective Projection of the Sample World B



name	location	extent	orientation
Observer	(0, 0, 0)		
Building 1	(150, 150, 0)	(30, 30, 100)	(0, -1, 0)
Building 2	(120, 240, 0)	(30, 30, 100)	(0, -1, 0)
Building 3	(160, 290, 0)	(30, 30, 100)	(0, -1, 0)

(a) Sample World C (with Bounding Box)



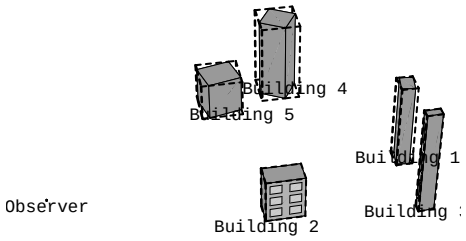
name	projection of the front 4 vertices	
Building 1	(0.444444, 0., 0.864198)	(0.555556, 0., 0.864198)
	(0.555556, 0.261891, 0.864198)	(0.444444, 0.261891, 0.864198)
Building 2	(0.272727, 0., 0.909091)	(0.363636, 0., 0.909091)
	(0.363636, 0.214275, 0.909091)	(0.272727, 0.214275, 0.909091)
Building 3	(0.309524, 0., 0.952381)	(0.380952, 0., 0.952381)
	(0.380952, 0.168359, 0.952381)	(0.309524, 0.168359, 0.952381)

(b) Occupancy Array

name	own area	visible area	visibility ratio
Building 1	0.029099	0.029099	1.
Building 2	0.0194795	0.0194795	1.
Building 3	0.0120256	0.0029153	0.242424

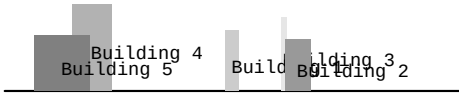
(c) Degree of Visibility

Figure 9: Perspective Projection of the Sample World C



name	location	extent	orientation
Observer	(0, 0, 0)		
Building 1	(347, 285, 0)	(24, 17, 116)	(-0.50, 0.87, 0)
Building 2	(263, 121, 0)	(28, 48, 58)	(0.65, -0.76, 0)
Building 3	(411, 209, 0)	(19, 10, 142)	(-0.72, -0.69, 0)
Building 4	(133, 347, 0)	(38, 37, 113)	(-0.15, 0.99, 0)
Building 5	(84, 274, 0)	(46, 42, 52)	(0.60, 0.80, 0)

(a) sample world D (with Bounding Box)



name	own area	visible area	visibility ratio
Building 1	0.00398726	0.00398726	1.
Building 2	0.00639927	0.00639927	1.
Building 3	0.002091	0.00163951	0.784078
Building 4	0.0164875	0.0116986	0.70954
Building 5	0.015058	0.015058	1.

(b) Degree of Visibility

Figure 10: Perspective Projection of the Sample World D

5 Conclusion

In this paper, we discussed the computational model with proper treatment of the projective prepositions. The importance of the surroundings and the view of the observer was pointed out. At all the previous works, the spatial relations are determined only based on the direction and the proximity between the primary object and the reference object, and on the physical existence of the objects in the world. Incorporating the techniques from the computational geometry and the computer graphics, we proposed a method of constructing the possible area, and a method of calculating the spatial relation on the perspective view of the observer. These methods were applied to some examples to demonstrate the usefulness and the sufficiency. The reflection of the real shape of the objects and other contextual factors to determine the use of the prepositions are the future problems.

References

[AK93] A. Abella and J. R. Kender. Qualitatively describing objects using spatial prepositions. In *Proc. AAAI-93*, pages 536–540, 1993.

[Gap94] Klaus-Peter Gapp. Basic meanings of spatial relations: Computation and evaluation in 3D space. In *Proc. AAAI-94*, pages 1393–1398, 1994.

[Her86] A. Herskovits. *Language and Spatial Cognition.* Cambridge University Press, 1986.

[KB91] J. K. Kalita and N. I. Badler. Interpreting prepositions physically. In *Proc. AAAI-91*, pages 105–110, 1991.

[OMT94] P. Olivier, T. Maeda, and J. Tsujii. Automatic depiction of spatial descriptions. In *Proc. AAAI-94*, pages 1405–1410, 1994.

[PK86] R. A. Plastock and G. Kalley. *Theory and Problems of Computer Graphics.* McGraw-Hill, Inc., 1986.

[RS88] G. Retz-Schmidt. Various views on spatial prepositions. *AI Magazine*, 9(2):95–105, 1988.

[SS93] J. R. J. Schirra and E. Stopp. ANTLIMA — a listener model with mental images. In *Proc. IJCAI-93*, pages 175–180, 1993.

[YYI+92] A. Yamada, T. Yamamoto, H. Ikeda, T. Nishida, and S. Doshita. Reconstructing spatial image from natural language texts. In *Proc. COLING-92*, pages 1279–1283, 1992.