



Title	Planning of observation and motion for interpretation of road intersection scenes considering uncertainty
Author(s)	Takizawa, Hotaka; Shirai, Yoshiaki; Miura, Jun et al.
Citation	IEEE International Conference on Intelligent Robots and Systems. 1998, 1, p. 520-525
Version Type	VoR
URL	https://hdl.handle.net/11094/14062
rights	c1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE..
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Planning of Observation and Motion for Interpretation of Road Intersection Scenes Considering Uncertainty

Hotaka Takizawa, Yoshiaki Shirai, Jun Miura and Yoshinori Kuno

Osaka University

2-1, Yamadaoka, Suita, Osaka 565-0871, JAPAN

E-mail: {takizawa, shirai, jun, kuno}@cv.mech.eng.osaka-u.ac.jp

Abstract

This paper describes a method of planning of observation and motion for a mobile robot to interpret a road intersection scene and to reach the intersection as soon as possible considering the uncertainty of interpretation. From a monocular color image, candidate regions are extracted for objects which are related to the intersection type. The probabilities of the regions coming from the objects are calculated from probabilistic models of the objects. From the probabilities of the regions and the relation between the intersection type and the objects, the current probability of intersection types is calculated. If the intersection type is ambiguous, the robot plans the observation and motion which minimize the expectation of the cost(time) to complete the task. The robot takes these actions and tries to determine the intersection type again. This process is iterated until the task is completed. The experimental result is shown for an actual intersection scene.

1 Introduction

It is an important issue for a mobile robot to make a road map. The robot moves around a town and interprets the types of encountered intersections to use them as landmarks. In order to make the map efficiently, the robot has to interpret each intersection and pass the intersection as soon as possible. It is not easy to interpret the intersection unambiguously because of the following three reasons: (1) objects which are related to intersection types vary in their attributes such as color, shape and size; (2) intersection scenes include various objects whose attributes are similar to one another; (3) enough sensor information is not always available. If the interpretation is ambiguous, the robot repeats the observation while approaching the intersection.

Fig. 1 shows a typical situation, where the robot observes the intersection scene at the first viewpoint(a) and tries to determine the intersection type. Since the left branching road and the straight road cannot be observed in the image, the intersection type is ambiguous. Then, the robot determines which parts it observes attentively after approaching to the intersection. Since the left branching road is wide and the straight road is narrow, the robot observes the left branching road at b and the straight road at c because

the observation of the straight road at b probably results in failure.

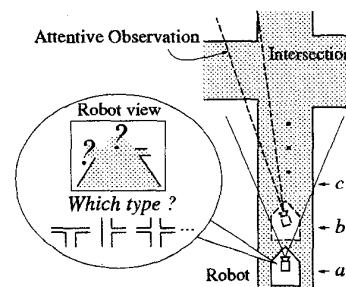


Figure 1: A robot determining type of intersection.

Cameron et al.[1] proposed a method of selecting the observation and motion to determine the position and direction of objects considering the uncertainty of recognition neglecting the cost of observation and motion. Rimey et al.[2] proposed a method of controlling a camera for determination of class and position of objects to minimize the expectation of the cost for determination considering the uncertainty of recognition. This method does not consider the position which the camera reaches finally.

In this paper, we propose a method of planning of observation and motion which minimize the expectation of the cost to interpret a intersection scene and to reach the intersection considering the uncertainty of interpretation. First, regions considered to come from objects are extracted from an observed color image. Then, probabilities of the regions coming from the objects are calculated from probabilistic models of the objects. From these probabilities and the relation between the intersection type and the objects, the probability of intersection types is calculated. If the intersection type is ambiguous, the robot determines the optimal observation and motion to complete the task. The robot takes these actions and tries to determine the intersection type again. This process is iterated until the task is completed.

2 Intersection Model

The types of intersections are characterized by the number and direction of the branching roads (see Fig. 2). The intersections have particular objects such as curve mirror whose classes and positions are determined by the intersection types. In this paper, we divide an intersection scene into eight areas (see Fig. 3). We suppose that each area has only one object and that the object classes depend only on the intersection types. The objects are described by several attributes such as color, shape and size. We suppose that each attribute is independent of other attributes and that each attribute depends only on the corresponding object. The relation between the intersection types, the objects and their attributes are represented as shown in Fig. 2. The relation is determined experimentally.

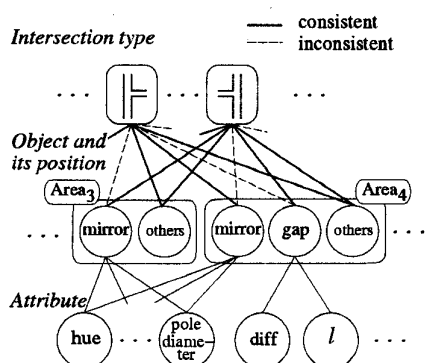


Figure 2: Model of intersection scene. The consistent lines indicate that the classes and areas of the objects are consistent with the intersection types; the inconsistent lines indicate that they are inconsistent with the intersection types.



Figure 3: Eight areas into which an intersection scene is divided

2.1 Object Model

In real intersection scenes, objects vary in their attributes within certain ranges. We represent the variation by describing each attribute as a probability density function. Fig. 4 shows example probability distribution that an object o has an attribute value a . The distribution used here is determined experimentally. The positions of objects in each intersection type

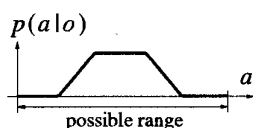
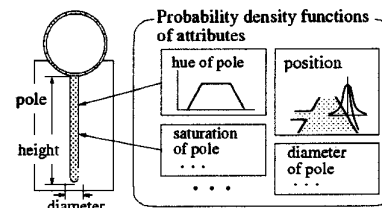
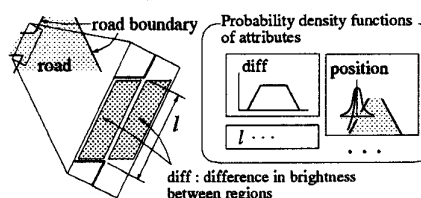


Figure 4: Probability distribution that an object o has an attribute value a .

also vary within certain ranges. We describe their positions by normal distributions. The models of two objects are shown in Fig. 5.



(a) Mirror at right side



(b) Gap of left road boundary

Figure 5: Models of objects

3 Observation Model

Objects are found in the observed color image as regions with several attributes such as color, shape and size. For example, a curve mirror is observed as a thin rectangle region whose color is orange.

If the objects are near to the robot, the regions are large in the image and it is reliable that the attributes of the regions are equivalent to those of the objects. On the other hand, if the objects are far from the robot or the direction of the view of the camera is not relevant, the regions are small and unreliable.

Fig. 6 shows an example case that the reliabilities of the regions depend on the distance from the robot to an object. In Fig. 6(a), a curve mirror which is far from the robot is observed in the white square. Fig. 6(b) shows the pixels whose colors are orange in the square area of Fig. 6(a) (the white superimposed pixels show such orange pixels). In this figure, there are few orange pixels and no reliable regions can be found. On the other hand, in Fig. 6(c), the curve mirror is near and there are many orange pixels (see Fig. 6(d)). In this figure, a large and reliable region is found.

The large and reliable regions are extracted from the image by using the knowledge of the object (see [3]). They are used as evidence for interpretation of the intersection types. On the other hand, the small regions are ignored. In this case, the facts that no regions are extracted are used as evidence.

We suppose that there is an ideal image plane which has an infinite resolution and no noise, and that ob-

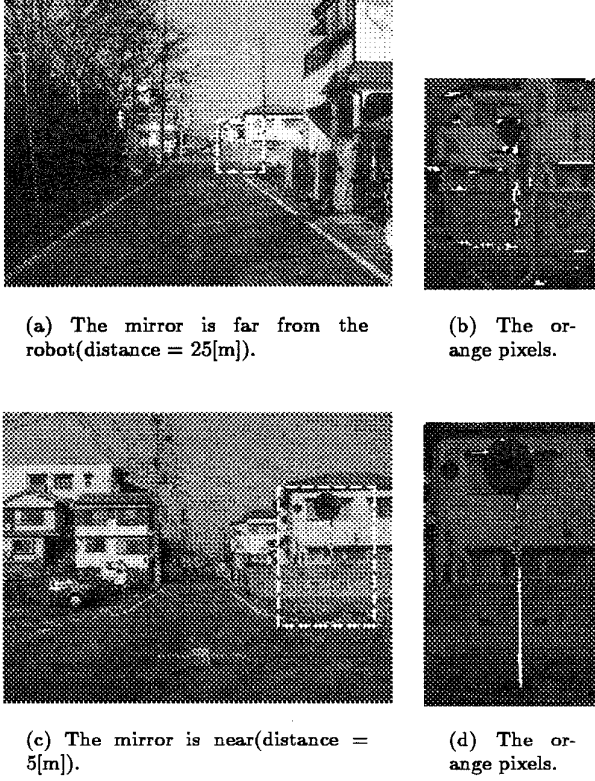


Figure 6: Relation between the orange regions extracted from the image and the distance from the robot to the mirror.

jects are observed as regions on the ideal image(*true regions* denote such regions). If the robot is near enough to the objects, the true regions are large and reliable. If the robot is far, they are small and unreliable. The large true regions are extracted from the image and used as evidence for interpretation(\mathcal{E}_1 denotes such evidence). The small regions are ignored and the facts that no regions are extracted are used as evidence(\mathcal{E}_2). In this section, we describe such observation model by using the probabilistic theory considering the region size which depends on the object size, the distance from the robot to the objects, and so on.

Let $P(e|o, \omega)$ denote the likelihood that the robot obtains the evidence $e (\in \mathcal{E}_1 \text{ or } \mathcal{E}_2)$ by observing the object o with observation parameters ω (the position, view direction and focal length of the camera).

By using the true region r' , $P(e|o, \omega)$ is represented as follows:

$$P(e|o, \omega) = \int_{r' \in R'} P(e|r', o, \omega) p(r'|o, \omega) dr'. \quad (1)$$

where R' is the possible set of the true regions.

$p(r'|o, \omega)$ represents the probability that the true region r' comes from the object o with the observation parameters ω . The true region is represented by a set of the attributes a'_1, a'_2, \dots . Since these attributes are supposed to be independent,

$$p(r'|o, \omega) = p(a'_1|o, \omega) p(a'_2|o, \omega) \dots \quad (2)$$

Each $p(a'|o, \omega)$ is calculated by using the attribute probability distribution shown in Fig. 4. In Fig. 7, two probability distributions for attributes a' and s' (s' represents the size of the true region especially) are shown. The probability distribution $p(r'|o, \omega)$ made up from these attribute probability distributions is also shown in this figure.

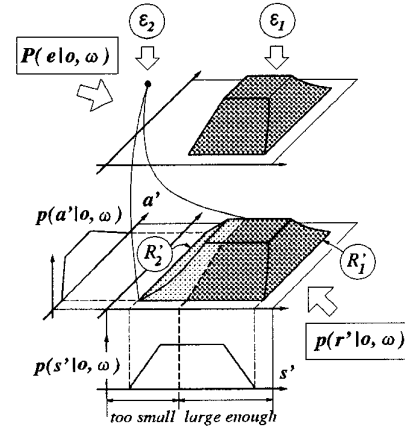


Figure 7: Relation between true region and evidence

The set of the true regions consists of two sets. One is a set of the regions which are large enough in the ideal image(\mathcal{R}'_1 denotes such set), and the other is a set of the regions which are small(\mathcal{R}'_2). In Fig. 7, the set of the regions whose s' is *large enough* is \mathcal{R}'_1 , the set of the regions whose s' is *too small* is \mathcal{R}'_2 .

$P(e|r', o, \omega)$ represents the probability that the evidence e comes from the true region r' . This probability depends on whether the true region belongs to \mathcal{R}'_1 or \mathcal{R}'_2 .

- If r' belongs to \mathcal{R}'_1 , the robot can obtain a region whose attributes are equivalent to those of the true region.
- If r' belongs to \mathcal{R}'_2 , no region is extracted. In this case, the robot obtains the evidence belonging to \mathcal{E}_2 .

In Fig. 7, the relation between the true regions and the obtained evidence is shown.

By using \mathcal{R}'_1 and \mathcal{R}'_2 , Eq. (1) is represented as follows:

$$P(e|o, \omega) = \int_{r' \in \mathcal{R}'_1} + \int_{r' \in \mathcal{R}'_2}. \quad (3)$$

In the first term,

$$P(e|r', o, \omega) = \begin{cases} 1 & (e \in \mathcal{E}_1 \text{ and } e = d(r')) \\ 0 & (\text{otherwise}) \end{cases} \quad (4)$$

where $d(r')$ represents the evidence whose attribute values are equivalent to the discrete values of the corresponding attributes of the true region r' . Thus,

$$\text{first term} = \begin{cases} \int_{r'=d^{-1}(e)} p(r'|o, \omega) dr' & (e \in \mathcal{E}_1) \\ 0 & (e \in \mathcal{E}_2). \end{cases} \quad (5)$$

In the second term,

$$P(e|r', o, \omega) = \begin{cases} 0 & (e \in \mathcal{E}_1) \\ 1 & (e \in \mathcal{E}_2). \end{cases} \quad (6)$$

Thus,

$$\text{second term} = \begin{cases} 0 & (e \in \mathcal{E}_1) \\ \int_{r' \in \mathcal{R}_2} p(r'|o, \omega) dr' & (e \in \mathcal{E}_2). \end{cases} \quad (7)$$

We calculate the likelihood($P(e|o, \omega)$) by using Eq. (1) ~ Eq. (7).

From Fig. 6(a), no regions are extracted and the likelihood that no regions are obtained is 1. From Fig. 6(c), an orange region is extracted and the likelihood that the region is obtained is 3.47×10^{-10} .

4 Interpretation of Intersection Scene

Let $e^k = (e_1^k, e_2^k, \dots)$ denote the pieces of current evidence, ω^k the current observation parameters (assumed to be known) and $E^{k-1} = \{e^{k-1}, \omega^{k-1}, \dots, e^1, \omega^1\}$ the pieces of the past evidence with the past observation parameters. By using e^k, ω^k and E^{k-1} , the robot calculates the conditional probability of the hypothesis that the current intersection type is $I_i \in \mathcal{I}$ as follows:

$$P(I_i|e^k, \omega^k, E^{k-1}) = \beta P(I_i|\omega^k, E^{k-1}) P(e^k|I_i, \omega^k, E^{k-1}), \quad (8)$$

where $\beta = [P(e^k|\omega^k, E^{k-1})]^{-1}$.

Since the intersection type is independent of the observation parameters,

$$P(I_i|\omega^k, E^{k-1}) = P(I_i|E^{k-1}). \quad (9)$$

$P(I_i|E^0)$ is set to a prior probability distribution of the intersection type.

Let $o_a(c_j)$ denote an object whose area is represented by the suffix a and class is represented by the suffix j . From the objects which are in the field of view of the camera with the observation parameters ω^k , we make various combinations $o = (o_1(c_{j1}), o_2(c_{j2}), \dots)$, where $j1 = 1 \sim J1; j1 = 0 \sim J2; \dots$. From the possible combinations, we make an exclusive and exhaustive set $\mathcal{O}(\omega^k)$ (see [3]).

By using $\mathcal{O}(\omega^k)$, $P(e^k|I_i, \omega^k, E^{k-1})$ is represented as follows:

$$P(e^k|I_i, \omega^k, E^{k-1}) = \sum_{o \in \mathcal{O}(\omega^k)} P(e|o, I_i, \omega^k, E^{k-1}) P(o|I_i, \omega^k, E^{k-1}). \quad (10)$$

Since each evidence depends only on the corresponding object and the observation parameters,

$$P(e^k|o, I_i, \omega^k, E^{k-1}) = P(e^k|o, \omega^k). \quad (11)$$

Since each evidence is independent of other evidence,

$$P(e^k|o, \omega^k) = \prod_a P(e_a^k|o_a(c_j), \omega^k). \quad (12)$$

$P(e_a^k|o_a(c_j), \omega)$ is calculated as described in Sec. 3.

Since the object at an area is independent of the objects at the other area and each object depends only on the intersection type,

$$P(o|I_i, \omega^k, E^{k-1}) = \prod_a P(o_a(c_j)|I_i, \omega^k, E^{k-1}) = \prod_a P(o_a(c_j)|I_i). \quad (13)$$

The probability $P(o_a(c_j)|I_i)$ is defined as described in [3].

The probability distribution of the intersection types is calculated by using Eq. (8) ~ Eq. (13). If the intersection type is determined from the probability distribution unambiguously, the robot adopts the best hypothesis of the intersection type.

5 Observation and Motion to Resolve Ambiguity of Interpretation

In real intersection scenes, there are several cases that the robot cannot find the objects even if the objects exist. These reasons are as follows: (1) the sizes of the objects are small; (2) the viewpoint is too far from the objects to observe them; (3) the robot observes the object in the bad direction; (4) the objects are occluded by other objects such as persons, bicycles and cars. In such cases, the interpretation of intersection types becomes ambiguous.

To overcome such situations, first, the robot determine the next viewpoint and the next observed area so that it can determine the intersection type and reach the intersection as soon as possible. Then, the robot moves to the next viewpoint and observes the area attentively. These processes are iterated until the robot completes the task. In order to complete the task as soon as possible, the robot has to determine the optimal area and viewpoint. In this section, we describe a method of searching for the observation and motion which minimize the expectation of the cost to complete the task.

Now suppose that the robot has the pieces of evidence with the observation parameters E^k . If the

robot repeats the optimal observation and motion from the current state, the robot can complete the task in a minimum cost. Let $C(E^k)$ denote such cost. The cost is zero if the task is completed.

Let ω_α^{k+1} denote the observation parameters achieved from the ω^k by an action (motion and observation) α , and $\hat{E}(\omega_\alpha^{k+1})$ denote a set of possible evidence predicted to be obtained with ω_α^{k+1} . If the robot takes the action α , the expectation of the cost is represented as follows:

$$\begin{aligned} C_\alpha(E^k) &= \sum_{\hat{e}^{k+1} \in \hat{E}(\omega_\alpha^{k+1})} C(\hat{e}^{k+1}, \omega_\alpha^{k+1}, E^k) P(\hat{e}^{k+1} | \omega_\alpha^{k+1}, E^k) \\ &+ T(\alpha), \end{aligned} \quad (14)$$

where $T(\alpha)$ is the time for taking the action.

By using $O(\omega_\alpha^{k+1})$, $P(\hat{e}^{k+1} | \omega_\alpha^{k+1}, E^k)$ is represented as follows:

$$\begin{aligned} P(\hat{e}^{k+1} | \omega_\alpha^{k+1}, E^k) &= \sum_{o \in O(\omega_\alpha^{k+1})} P(\hat{e}^{k+1} | o, \omega_\alpha^{k+1}, E^k) P(o | \omega_\alpha^{k+1}, E^k). \end{aligned} \quad (15)$$

Since the evidence depends only on both the objects and the observation parameters,

$$P(\hat{e}^{k+1} | o, \omega_\alpha^{k+1}, E^k) = P(\hat{e}^{k+1} | o, \omega_\alpha^{k+1}) \quad (16)$$

$P(\hat{e}^{k+1} | o, \omega_\alpha^{k+1})$ is calculated as described in Sec. 3.

By using the set of the intersection types $I_i \in \mathcal{I}$, $P(o | \omega_\alpha^{k+1}, E^k)$ is represented as follows:

$$P(o | \omega_\alpha^{k+1}, E^k) = \sum_i P(o | I_i, \omega_\alpha^{k+1}, E^k) P(I_i | \omega_\alpha^{k+1}, E^k). \quad (17)$$

Since the objects depend only on the intersection types, $P(o | I_i, \omega_\alpha^{k+1}, E^k) = P(o | I_i)$. And since the intersection type is independent of the observation parameters, $P(I_i | \omega_\alpha^{k+1}, E^k) = P(I_i | E^k)$.

The best action is represented as follows:

$$\alpha^* = \arg \min_{\alpha} C_\alpha(E^k). \quad (18)$$

If α^* is selected, $C(E^k)$ is updated as follows:

$$C(E^k) = C_{\alpha^*}(E^k). \quad (19)$$

The robot moves to the next viewpoint and observes the next area by taking the action α^* . Then it obtains a new evidence and tries to interpret the intersection types from the evidence. This process is repeated until it determines the intersection type and reaches the intersection.

6 Experimental Result

The experiment is made on the following assumption.

1. The advance speed is constant (20[km/h]).
2. The time needed for scene interpretation and action determination is constant (1.5[sec]).
3. The robot repeats the observation at the interval of the time.
4. If the robot is close to the intersection, it can determine whether or not there is a branching road.

Fig. 8 shows an image of a sample intersection scene. In the position indicated by (a), a region considered to be a gap whose area is $Area_6$ is extracted. From the region, the distance between the robot and the intersection is calculated to be 28[m]. The likelihood of the region is 5.00×10^{-7} . Fig. 9 shows the probability distribution of the intersection types calculated from the region. From the distribution, the robot cannot determine the intersection type.

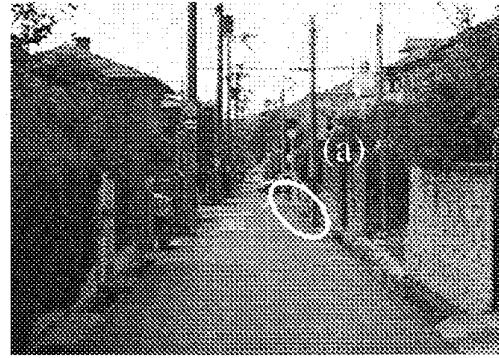


Figure 8: Sample road intersection scene

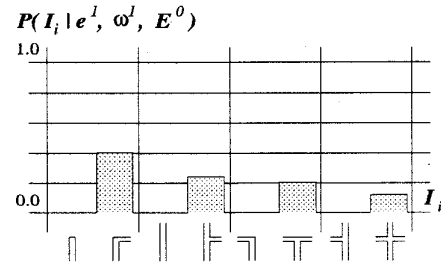


Figure 9: Probability of intersection types for Fig. 8.

The optimal action is to advance eight meters and observe $Area_4$ attentively. The expectation of the cost is 9.02[sec]. Fig. 10 shows a newly obtained image. In the position indicated by (a), a region considered to be a gap whose area is $Area_4$ is extracted (the likelihood is

2.07×10^{-7}). Fig. 11 shows the probability distribution of the intersection types calculated from the region. From the distribution, the robot cannot determine the intersection type yet.

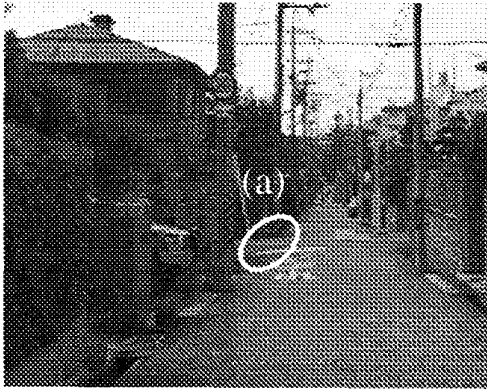


Figure 10: New image obtained by attentive observation

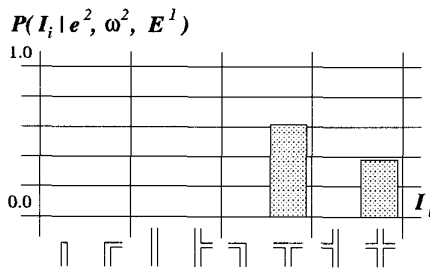


Figure 11: Probability of intersection types for Fig. 10.

The optimal action is to advance eight meters more and observe $Area_2$ attentively (the expectation of the cost is 5.83[sec]). Fig. 12 shows a newly obtained image. In the position indicated by (a), a region considered to be a straight road whose area is $Area_2$ is extracted (the likelihood is 2.33×10^{-9}). Fig. 13 shows the probability distribution of the intersection types calculated from the region. From the distribution, the robot can determine that the intersection type is a cross type. And the robot reaches the intersection while it calculates the probability distribution.

7 Conclusion

This paper describes a method of planning of observation and motion for a mobile robot to interpret a road intersection scene and to reach the intersection as soon as possible considering the uncertainty of interpretation.

From a monocular color image, candidate regions are extracted for objects such as curve mirror. The likelihoods of the regions are calculated from the probabilistic models. From the probabilities of the regions

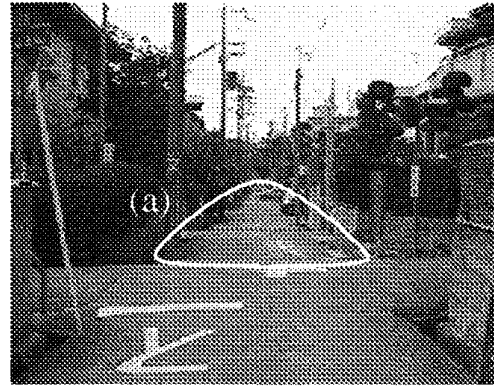


Figure 12: New image obtained by attentive observation

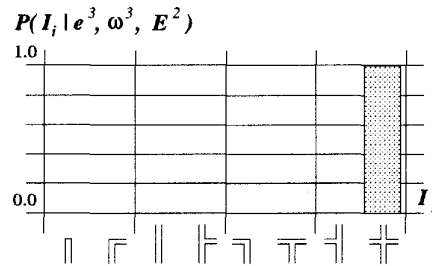


Figure 13: Probability of intersection types for Fig. 12.

and the relation between the objects and the intersection types, the current probability of intersection types is calculated. If the intersection type is determined unambiguously, the robot adopts the intersection type and goes to the intersection.

Otherwise, the robot determines the observation and motion which minimize the expectation of the cost to complete the task. The robot takes these actions and tries to determine the intersection type again. This process is repeated until the task is completed.

The experimental result is shown for an actual intersection scene.

References

- [1] Alec Cameron and Hugh Durrant-Whyte. "A Bayesian Approach to Optimal Sensor Placement". *The International Journal of Robotics Research*, Vol. 9, No. 5, pp. 70-88, 1990.
- [2] Raymond D. Rimey and Christopher M. Brown. "Control of Selective Perception Using Bayes Nets and Decision Theory". *International Journal of Computer Vision*, Vol. 12, No. 2/3, pp. 173-207, 1994.
- [3] Hotaka Takizawa, Yoshiaki Shirai, Yoshinori Kuno, and Jun Miura. "Recognition of Intersection Scene by Attentive Observation for a Mobile Robot". In *Proc. Int. Conf. on Intelligent Robots and Systems*, Vol. 3, pp. 1648-1654, 1996.