Proc. 14th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp. 214-219, Beppu, Japan, Jan. 24-25, 2008.

Vision Planning for Object Search using Multiple Visual Features

Akihide Shibata† and Jun Miura‡

† Department of Mechanical Engineering, Osaka University

‡ Department of Information and Computer Sciences, Toyohashi University of Technology

Abstract This paper describes a vision planning method for object search using multiple visual features. To realize an efficient search, it is important to select appropriate search actions (i.e., features to use, fixation points, and resolution) depending on the status of the search. We first define similarity measures between the target object image and image regions in input images, and built similarity distribution models for the target object and the background. We then use the distribution for estimating the detection probability of the target in each location in the image for each search action. The planning method repeatedly selects the best search action which maximizes the expectation of the detection probability. We also develop a method of estimating the distribution models for the background from relatively simple image features. We compare the proposed method with methods with fixed search strategies to show its effectiveness.

Keywords: Vision planning, Visual search, Multiple visual features, Similarity distribution models

1 Introduction

Visual object search is one of the growing research areas in computer vision, especially for security and surveillance purposes. We usually select viewpoints, focus of attention, or features to use adaptively depending on the situation, in order to find or to carefully examine a specific object. Since the processing cost of computer vision is generally high, such a vision planning will be important.

Vision planning under uncertainty of visual information is an important area of research. Information-theoretic approaches [2] are useful when the goal of vision is object identification. In this approach, we have a probabilistic distribution on object identities, and update the distribution as an observation is carried out. A usual strategy is to repeatedly select the most promising observation by using the minimum expected entropy criterion [3, 7]. Rimey [6] dealt with a vision planning for recognition of a complicated scene. He proposed to represent the relationship between objects using a Bayesian network and to select the best observation considering the planning cost. The vision planning methods in these works deal with the problem of selecting viewpoints or focus of attention, but do not deal with the feature selection problem.

Automatic generation of recognition programs [5] is another type of vision planning. This work dealt with the problem of localizing known objects using vision. The proposed method generates a decision tree for determining which image features to use based on a sensor model and the analysis of possible object states. The method does not deal with object identification.

Many methods have been proposed for object search in an image. A fast method [9] is proposed which uses the *active search* technique for color histograms. Object search methods using a single visual feature are not, however, applicable to the cases where that feature is not effective in discriminating objects or there are many objects with similar values for that feature.

Using multiple visual feature is, therefore, important for robust object recognition. A Bayesian inference-based object extraction method [8] was proposed which uses probabilistic distributions of target objects on three visual features (color, position, size) and determines the region with high probability of object existence. Although any combination of features can, in principle, be used in this approach, it is inefficient to use all such features all the time; only a small number of features may be effective in a specific situation.

This paper therefore proposes a vision planning method which adaptively chooses effective visual features and observation conditions based on the expected object detection probability. This probability is calculated from the current probabilistic distribution of object existence and knowledge of object detectability measures for every combination of visual feature and resolution. Since the object detectability of a feature depends both on the target object and the background, and since the background may not be known in advance, we also propose a method of estimating background models from a simple visual feature of the image. We will show that the proposed method outperforms other methods which uses fixed search strategies.

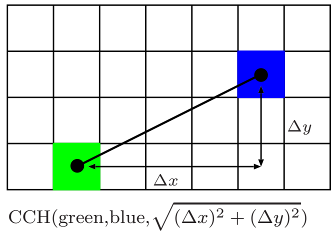


Fig. 1: Color cooccurrence histogram.

2 Visual Features Used for Object Search

We use three visual features: color histogram, color cooccurrence histogram, and edge pattern. The following subsection explains these features in detail.

2.1 Color Histogram

Color histogram is robust to shape changes of objects and low-cost, and are therefore widely used. Since the similarity on color histogram changes spatially gradually, it is not well suited for accurately localizing objects.

We use the HSV color space and divide the space into 38 bins as follows. We first set the black region where V is small and set the white-gray region where S is small. The rest part of the space is divided into 12 in the H axis and into 3 in the S axis. The similarity of the color histogram is calculated by the following *intersection*:

$$S_{MI} = \sum_{i=1}^{38} \min(M_i, I_i), \tag{1}$$

where M and I are the normalized histogram of a model and an input image, respectively.

2.2 Color Cooccurrence Histogram

Color Cooccurrence Histogram (CCH) [1] is a three-dimensional histogram, axes of which are two colors and the distance between two pixel positions (see Fig. 1). This feature can discriminate two objects with different color patterns even if the objects have similar color histograms. We use the same division of the color space and the bins as the ones used for the color histograms. The neighborhood for calculating CCH is the 17x17-pixel square centered at the point under consideration.

2.3 Edge Pattern

Edge patterns represent shapes and surface textures of objects. A target object deforms depending on the viewpoint, but we consider here only the scaling and the rotation in the image, by assuming a viewpoint looking down a region for search as in the case of searching for an object on the table. More concretely, we consider eight orientations and two or three scalings of each object.

The *Hausdorff distance* defined below [4] is often used as a matching criterion:

$$h(M, I) = \max_{m \in M} \min_{i \in I} ||m - i||, \tag{2}$$

where M and I are edge patterns of a model and an input image. This expression calculates the maximum distance among those of all possible pairs of edges in both images. By not using the sum of the distances but using the maximum distance, a more globally matched pattern is selected. The original Hausdorff distance defined above is, however, weak to noise or partial occlusions which causes some of edges not to have corresponding edges in their neighborhoods. So the following modified expression is used in this paper:

$$h^f(M, I) = \underset{m \in M}{f\text{th}} \min_{i \in I} ||m - i||, \tag{3}$$

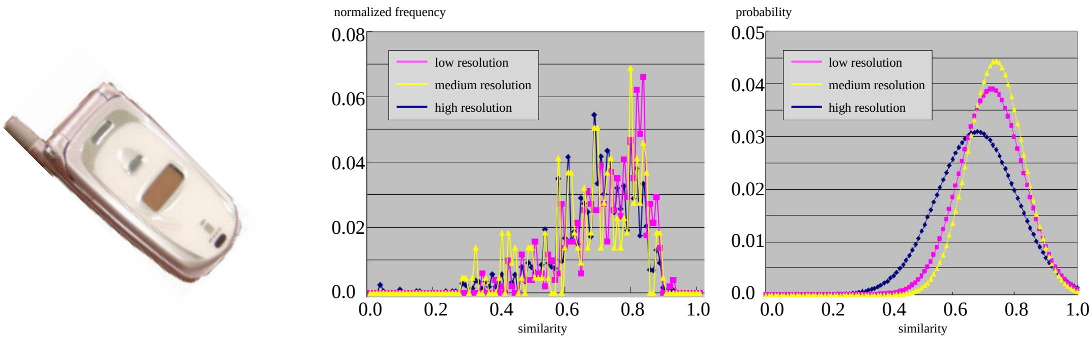
where fth indicates the value which divides the ascending ordered list of the distances to two groups with the ratios of f and $1 - f$ $(0 \leq f \leq 1)$. This distance being some value means that the part of edges with ratio f have the corresponding edges within the distance of that value. In calculating the similarity of edge patterns, we set some threshold and calculate the ratio f of edges which have the corresponding edges within the threshold distance.

2.4 Object Search using Multiple Visual Features

Fig. 2 shows an example of object search using the three features. The top row of the figure shows the input image, the reduced color-image (38 colors), and the extracted edges. The bottom low shows the process of reducing the candidate regions for the object using the three features in the order of color histogram, CCH, and edge pattern. This order is effective in this case, but is not always the best.

3 Vision Planning in Object Search

In object search using multiple visual features, the best strategy will depend on the situation. We therefore proposes a method of selecting effective visual features and



(a) Model image (b) Distributions of similarities (c) Fitted normal distributions

Fig. 3: An example of similarity distribution model (SDM) for color histogram.

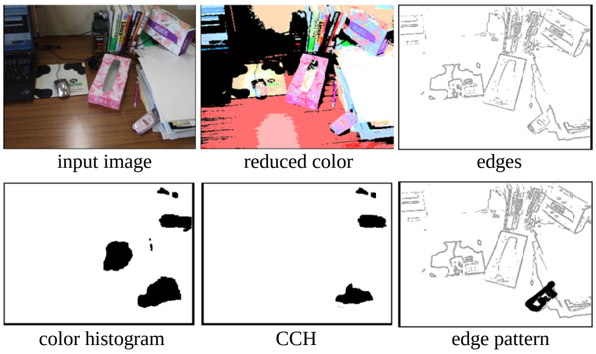


Fig. 2: Object search using multiple features.

observation parameters adaptively to the model and the input image.

The problem setting we consider in this paper is as follows. We use the three visual features mentioned above. We use a high-resolution digital camera to take images in advance and emulate the zoom effect by re-sampling the images. We consider three resolution levels (low, medium, high); for CCH, we do not use low resolution because of the limitation of the size of neighborhood. The search area is limited on a table and the whole area is visible in low resolution. We assume that the angle between the viewing direction and the table top is almost constant and that the approximate size of the objects in the image is known in each resolution, thereby limiting the size of search window. The search finishes when all candidate regions have been examined.

3.1 Similarity Distribution Model

We search for a target object using the similarity of features between the model image of the target and input images. To select an appropriate feature in each search step, we need to assess the effectiveness of each feature on object detection, we characterize each feature by its *similarity distribution model* (SDM). An SDM models the distribution of the similarity measure for a feature between the model and the corresponding object in input images at some resolution. Fig. 3 shows an example of object model for color histogram.

The object detection power of a feature depends not only on how a target object appears in the image but also on how the background looks like; we know that it is very difficult to find an object in a similar background. We, therefore, model the characteristics of the background, also in the form of SDM. The problem, however, arises that the background cannot be known in advance and a concrete background SDM cannot be prepared. We thus develop a method of predicting background SDM from a relatively simple image feature.

The details of the acquisition of SDMs will be described in Section 4.

3.2 Detection Probability

We select recognition strategies based on the predicted probability that the target object is found, called *detection probability*. The detection probability is defined using SDMs. Let ω be an observation parameter, including visual feature to use and observation conditions such as zoom and focus of attention, M be the target object, and $P(M|s;\omega)$ be the probability that object M exists if similarity s is obtained with parameter ω. Then the detection probability $F(\omega)$ is defined as:

$$F(\omega) = \int_0^1 f(P(M|s;\omega))P(s;\omega)ds \quad (4)$$

$$f(P) = \begin{cases} 1 & (P > \text{threshold}) \\ 0 & (P \leq \text{threshold}) \end{cases} \quad (5)$$

$$P(M|s;\omega) = P(s|M;\omega)P(M)/P(s;\omega) \quad (6)$$

$$P(s;\omega) = P(s|M;\omega)P(M) + P(s|\overline{M};\omega)P(\overline{M}) \quad (7)$$

where $P(M)$ is the prior probability, $P(\overline{M}) = 1 - P(M)$, $P(s|M;\omega)$ and $P(s|\overline{M};\omega)$ are the object and the background SDM, respectively.

3.3 Planning the Best Next Observation Parameter

The planning method determines the best next observation parameter ω^* using the detection probability. Since the observation result is not deterministic in general, we, in principle, need to devise a contingency plan, which has a corresponding action for each possible outcome to optimize the object search. However, such optimization is very costly. So we adopt a strategy of maximizing the utility increase per unit time. The next observation parameter is thus determined by:

$$\omega^* = \arg\max_{\omega} \left[\Delta F(\omega)/\mathrm{cost}(\omega) \right], \tag{8}$$

where $cost(\omega)$ is the expected execution time of observation with parameter ω and $\Delta F(\omega)$ is set to $F(\omega)$ because the target object has not been found until now.

The concrete steps for the object search with planning is as follows.

1. Extract candidate regions using color histogram in the low-resolution image by active search [9]. Only the extracted regions are processed in the later steps.
2. Determine ω^* among unused parameters for each candidate region, and for the region which has the highest detectability $F(\omega^*)$, execute the corresponding observation with ω^*.
3. Examine if the target object is found after the observation. If it is found, verify the region using the edge pattern in the high-resolution image for determining the object pose. If the probability of object existence is very low, delete the region. Otherwise, update the probability.
4. Repeat steps 2 and 3 until all candidate regions are processed.

Each candidate region is composed of a set of representative points of windows or edge patterns. Similarity s of a region is calculated as the average of similarity values of the pixels in the region.

The condition that an object is found is that the posterior probability $P(M|s;\omega)$ is larger than a threshold (currently, 0.8) *and* the size of the region is less than twice the predicted size of the target object under the current resolution. The condition that an object is considered not to exist is that the posterior probability is very low (currently, less than 0.01) *or* the size of the region is much smaller than predicted (currently, less than 10% of the predicted) with a low posterior probability (currently, less than 0.1).

4 Acquisition of Similarity Distribution Models

This section explains how to obtain similarity distribution models (SDMs) for target objects and backgrounds. Since the SDMs depend on the target object, we use the data for the object at the left of Fig. 3 as examples.

4.1 SDM for Target Object

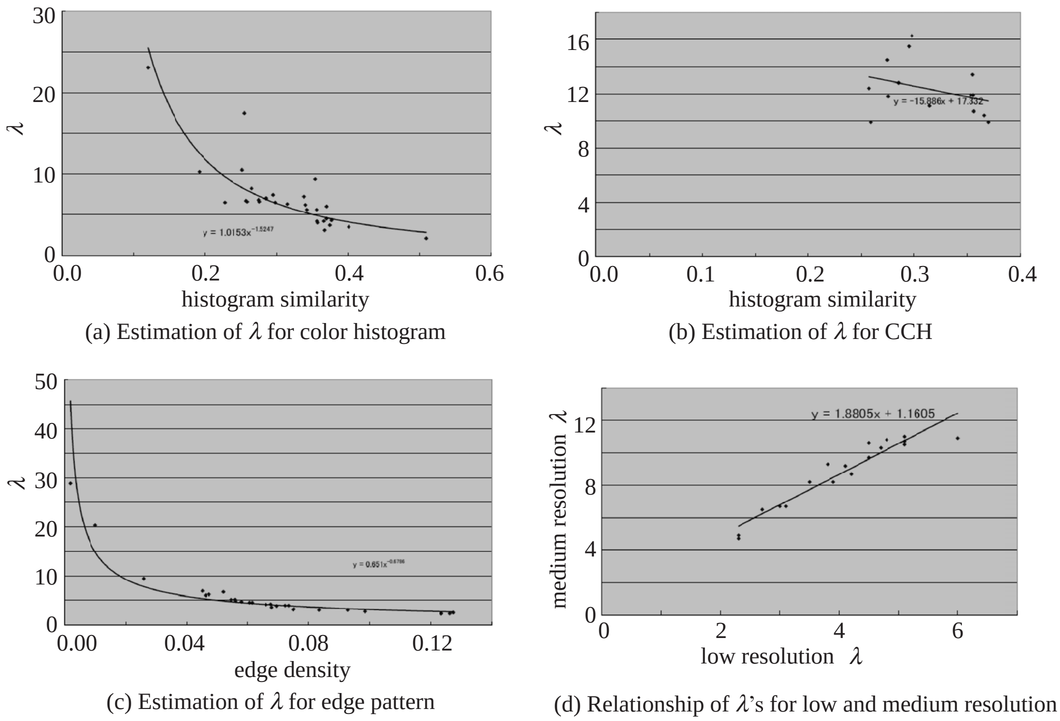
We collect sample images of the target object by observing the object under various conditions and by segmenting the object regions manually. We then calculate the similarity values for the sample images and the model image for each feature. We finally construct SDMs by fitting a normal distribution to each data set. Fig. 3 shows the sample sets and the obtained SDMs for color histogram and the three image resolutions. The result shows that the similarity increases as the resolution becomes higher.

4.2 SDM for Background

The similarity between the target object and the background is usually low. We model the SDM for the background, which is the distribution of the similarity between the background and the current target object, using an exponential probability; $P(s) = \lambda \exp(-\lambda s)$. The problem here is that SDMs are not constructed when starting the search. One possible way to cope with this is to construct *general* SDMs in advance by analyzing the similarity values for various backgrounds. This way is, however, inappropriate because we need to predict the detection power under some specific situation currently under consideration. We therefore develop a method of predicting background SDMs from a relatively simple image feature. We use the color histogram of the entire image for predicting SDMs for both color histogram and CCH, and use the edge density for SDMs for edge patterns.

Fig. 4(a) shows the relationship between the image features (i.e., similarity measure for the color histogram of the entire image) and the estimated distribution parameter λ's; the approximation result of the data using a power function is also shown. Fig. 4(b) shows the result of a similar process for CCH; we use a linear approximation instead of the power function for CCH. For these two features, we use the same λ for every resolution because the change of λ is insignificant over the resolutions.

Fig. 4(c) shows the result for the edge pattern. In this case, parameter λ largely changes according to the resolution because a higher resolution reveals finer textures. We thus decided to use different λ's for different resolutions, and to estimate λ's for the medium and the high resolution from λ for the low resolution, which can be estimated at the beginning of the object search in the low resolution. Fig. 4(d) shows the relationship between λ's for the medium and the low resolution and a fitted line to be used for the estimation.



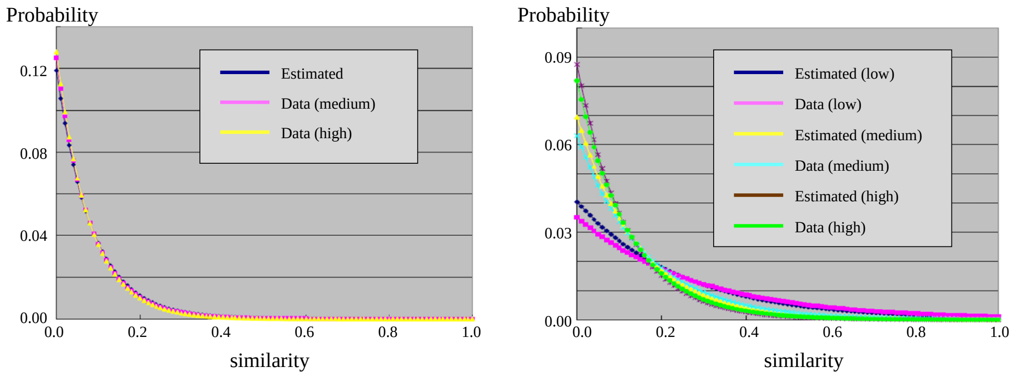
(a) Estimation of λ for color histogram

(b) Estimation of λ for CCH

(c) Estimation of λ for edge pattern

(d) Relationship of λ's for low and medium resolution

Fig. 4: Modeling SDMs for background.



(a) CCH

(b) Edge pattern

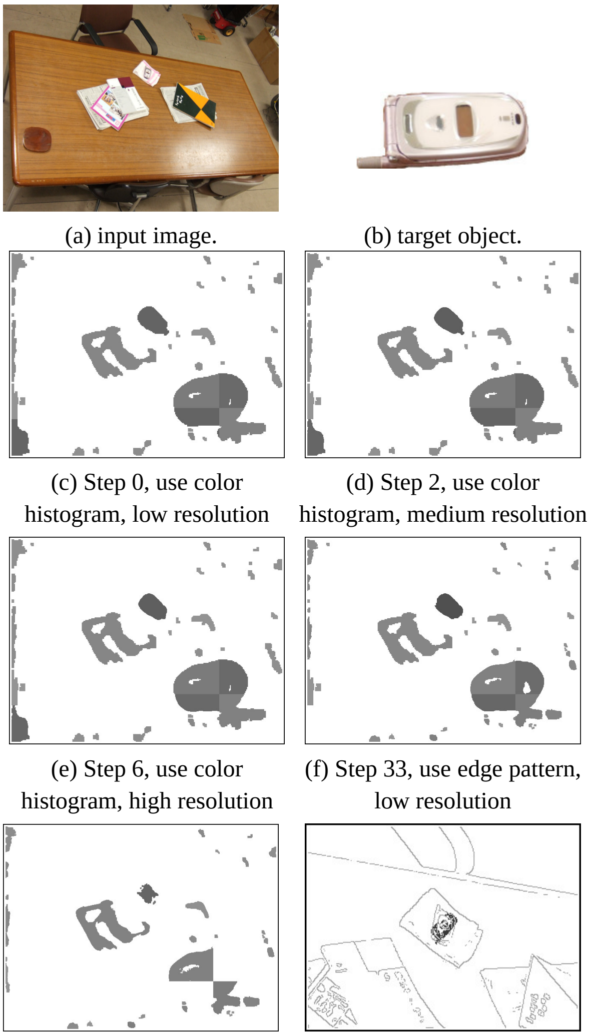
Fig. 5: Evaluation of background SDM prediction.

Fig. 5 shows some results of estimating the SDMs for the background from the simple image features. The results show that the SDMs are estimated reasonably well.

5 Experimental Results

Fig. 6 shows an example of object search. In the figure, (a) shows the input image and (b) shows the target object. The initial candidate regions are shown in (c). The intensity of each region indicates the probability of existence of the target object (darker is higher). The steps of search are as follows. The system first searches the bottom left region and then the region including the target (step 2, (d)). The probability has reached 0.97 at this state, but the size of the region is too large to consider the target is found. After this, the system performs the search with the color histogram in the high resolution (step 6, (e)) and then with the edge pattern in the low resolution (step 33, (f)), and finally finds the target with the edge pattern in the high resolution (step 56, (g)). (h) shows the detected target with a pose estimate.

We then compare the proposed method with others which use fixed object search strategies. Table 1 shows the order of pairs of feature and resolution for each fixed method. Table 2 shows the comparison result. The performance of fixed methods depend on the scene, while the proposed method shows the best performance for all scenes by adaptively selecting appropriate strategies.



(a) input image. (b) target object.

(c) Step 0, use color histogram, low resolution (d) Step 2, use color histogram, medium resolution

(e) Step 6, use color histogram, high resolution (f) Step 33, use edge pattern, low resolution

(g) Step 56, use edge pattern, high resolution (h) object found.

Fig. 6: An experimental result.

6 Conclusion and Future Works

This paper has described a vision planning method for object search using multiple visual features. The method determines the feature and the image resolution to use based on the target detection probability. We also proposed a method of estimating the probabilistic models of similarity measures not only for target objects but also for the background. The experimental results show that the proposed method outperforms the methods with fixed search strategies.

A future work is to investigate the use of other visual features which will be necessary for object search for various objects. Another future work is to deal with the vision planning in a wider area like a room or a floor. In this case, a camera moves around on a mobile platform and the planning method needs to be extended to consider the cost of camera movements.

Table 1: The search strategies of fixed methods.

method	Order of search steps
set1	$1 \rightarrow 2 \rightarrow 3 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 8 \rightarrow 9$
set2	$1 \rightarrow 7 \rightarrow 2 \rightarrow 5 \rightarrow 8 \rightarrow 3 \rightarrow 6 \rightarrow 9$
set3	$1 \rightarrow 5 \rightarrow 9$
set4	$1 \rightarrow 9$

color histogram: low (1), medium (2), high (3)
CCH: medium (5), high (6)
edge pattern: low (7), medium (8), high(9)

Table 2: Comparison of planning-based method with fixed methods (in seconds)

scene ID	planning	set1	set2	set3	set4
1	31.9	34.4	49.3	55.8	52.8
2	47.2	50.1	58.0	59.4	56.9
3	39.4	41.5	42.6	49.9	52.6
4	37.9	54.0	39.5	54.7	56.8
5	37.2	49.5	41.9	52.2	47.7

References

[1] P. Chang and J. Krumm. Object Recognition with Color Cooccurrence Histograms. In *1999 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 498–504, 1999.

[2] J.J. Denzlera and C.M. Brown. Information Theoretic Sensor Data Selection for Active Object Recognition and State Estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 2, pp. 145–157, 2002.

[3] S.A. Hutchinson and A.C. Kak. Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities. *IEEE Trans. on Robotics and Automat.*, Vol. 5, No. 6, pp. 765–783, 1989.

[4] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing Images using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850–863, 1993.

[5] K. Ikeuchi and T. Kanade. Towards Automatic Generation of Object Recognition Programs. *Proceedings of IEEE*, Vol. 76, No. 8, pp. 1016–1035, 1988.

[6] R.D. Rimey. Where to Look Next using a Bayes Net: An Overview. In *Proceedings of 1992 DARPA Image Understanding Workshop*, pp. 927–932, 1992.

[7] H. Takizawa, Y. Shirai, Y. Kuno, and J. Miura. Rocognition of Intersection Scene by Attentive Observation for a Mobile Robot. In *Proceedings of 1996 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1648–1654, 1996.

[8] Y. Taniguchi and Y. Shirai. Evidence-Based Scene Interpretation Considering Subjective Certainty of Recognition. In *Proceedings of 1998 Asian Conf. on Computer Vision*, pp. 432–439, 1998.

[9] V.V. Vinod and H. Murase. Focused Color Intersection with Efficient Searching for Object Extraction. *Pattern Recognition*, Vol. 30, No. 10, pp. 1787–1797, 1997.