Object Recognition Supported by User Interaction for Service Robots

Yasushi Makihara, Masao Takizawa, Yoshiaki Shirai, Jun Miura, and Nobutaka Shimada
Dept. of Computer-Controlled Mechanical Systems, Osaka University
2-1, Yamadaoka, Suita, Osaka 565-0871, JAPAN
{makihara, takizawa, shirai, jun, shimada}@cv.mech.eng.osaka-u.ac.jp

Abstract

This paper describes an interactive vision system for a robot that finds an object specified by a user and brings it to the user. The system first registers object models automatically. When the user specifies an object, the system tries to recognize the object automatically. When the recognition result is shown to the user, the user may provide additional information via speech such as pointing out mistakes, choosing the correct object from multiple candidates, or giving the relative position of the object. Based on the advice, the system tries again to recognize the object. Experiments are described using real-world refrigerator scenes.

1. Introduction

Nowadays, there is a growing necessity of welfare robots in the aging society. This paper describes an interactive vision system for a service robot which can recognize a user-specified object and bring it to the user.

In order to bring a target object, the robot has to know the object position. One may put bar codes on objects. However, such methods are troublesome and not effective for occlusion.

Takahashi et al.[1] developed a small mobile robot system which brings daily objects using vision. It is sometimes difficult to automatically recognize partial occluded objects and adapt to the change of the lighting condition. In such cases, we need to use additional information.

There are researches on combining visual information with verbal information. Some methods[2][3] generate a scene explanation based on the visual recognition. Watanabe et al.[4] proposed a system to recognize flowers and fruits in a botanical encyclopedia using explanation texts attached to each figure. Other methods use the user's advice[5][6]. These methods search for the position where the image features are most consistent with the user's advice. However, the methods do not recover from recognition error using the verbal information.

Takahashi et al.[7] proposed a welfare robot which recognizes apples and books through verbal and gestural interaction. The robot recognizes target objects at the place pointed by a user with gestural interaction. If the robot extracts multiple object candidates, it lets the user choose the correct object from them via speech. However, they do not have recovery method when no object candidates are extracted.

In our research, the robot first recognizes cans, bottles, and PET bottles in the refrigerator as automatically as possible. Then, if the robot fails in recognition, it tries again to recognize the objects with user interaction via speech.

The outline of this paper is as follows. In section 2, we describe the registration of object models for recognition. Next, the automatic object recognition based on the object models is described in section 3. Then, the recognition supported by user interaction is described in section 4. Lastly, we give a conclusion of this paper in section 5. Note that manipulation and locomotion are not dealt with in this paper (see [8]).

2. Registration of object models

In order to recognize objects, the system registers object models automatically in advance. The model of an object has to contain enough features to recognize the object from any directions. The features consist of the actual size of an object, representative colors, and secondary features such as the color, the position, and the size of uniform color regions. We propose a strategy to register a minimum set of features for discriminating the object. For example, if there are no objects with the same representative colors, the system registers only the representative colors. Every time an object with the same feature is registered, the system adds distinguishable features. In experiments, the system registered 40 cans, 3 bottles, and 21 PET bottles.

2.1. Construction of a developed image

In order to make those object models, we use one image made by developing the surface texture of the object into a rectangle plane. We call it "developed image". The approximate developed image is constructed in the following steps. First, the system takes images of the object from each direction (see Figure 1(a)). Next, the system extracts each piece image (see Figure 1(b)). Lastly, the system merges ad-

jacent piece images by image mosaicing and constructs the developed image (see Figure 1(c)). Note that the system can make developed images of any objects except self-occluded objects.

(a) Original image (b) Piece image (c) Developed image

top : can, bottom : square type PET bottle

Figure 1. Procedure for constructing a developed image

2.2. Registration of features of objects

The size of an object consists of its width and height and is computed from the developed image. If the width of an object changes depending on the viewing directions, the range of the width is registered.

The representative colors and the secondary features depend on the viewing direction. We determine the intervals of directions where the similar feature values are seen. For example, the developed image in Figure 2(a) is divided into the interval I_1 of white and the interval I_2 of blue.

If an interval is not distinguishable from those of the other objects, the interval is further divided into multiple intervals according to secondary features if distinct secondary features are extracted in the interval. We continue this process until all the intervals are distinguishable from other objects.

For example, suppose that object A with two intervals is already registered and that object B with the same feature (blue) is now added (see Table 1(a)). Because the system cannot distinguish interval (2) of A from interval (1) of B, it extracts secondary features in developed images in the intervals. The result is shown in Table 1(b).

Table 1. Addition of secondary features

((interval index), representative color, [secondary feature])

(a) Before adding distinguishable features

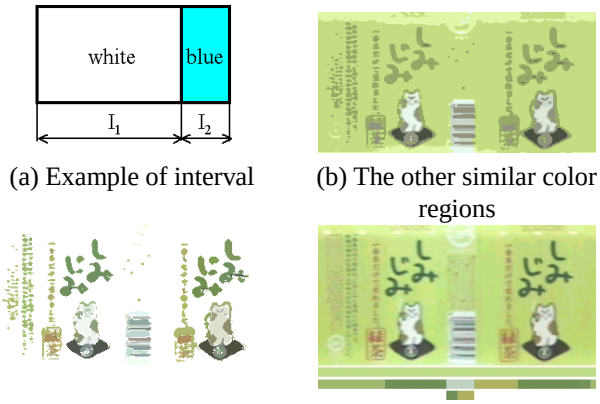
A	((1), white), ((2), blue)
B	((1), blue)

(b) After adding distinguishable features

A	((1), white), ((2), blue, [white])
B	((1), blue, [black]), ((2), blue, [gray])

The system extracts those direction-dependent features for every direction. The system first segments the developed image into uniform color regions [9] [10] (see Figure 2(b)). For each direction, the system reconstructs a piece image of uniform regions and registers the representative color as the largest region in the image.

Then the system extracts the other uniform color regions (see Figure 2(c)). For each direction, the system reconstructs a piece image in the same way and registers the secondary feature as the color, the position, and the size of the largest region in the image. If more than one secondary features are needed, the system extracts a necessary number of regions according to priority of the size.



(a) Example of interval (b) The other similar color regions

(c) Segmented image (d) Features

In (d), top line: representative color, middle line: color of secondary feature 1, bot. line: color of secondary feature 2

Figure 2. Exraction of features

3. Automatic object recognition

We consider two cases for automatic object recognition: (1) when the name of an object is specified, the system tries to recognize it and (2) when a user asks what objects are in the refrigerator, the system tries to recognize all objects. In both cases, the system recognizes objects in the following steps:

1. Extract candidate regions for objects.
2. Verify or determine object types (can, bottle, PET bottle).
3. Match extracted regions to object models.

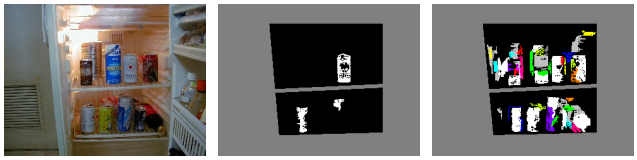
3.1. Extraction of candidate regions

When the name of an object is specified, the system retrieves the representative color of the object and extracts regions with that color. Such a color-based extraction may extract regions other than objects, so the system eliminates unsuitable regions for objects. The resultant candidate regions are shown in Figure 3(b).

When the system tries to recognize all objects in the refrigerator, it segments whole area in the refrigerator into uniform color regions. After eliminating unsuitable regions for objects, the resultant candidate regions are extracted (see Figure 3(c)).

3.2. Recognition of object types

In this section, we describe the object type models and recognition method for those types.



(a) (b) (c)

(a) Original image, (b) Candidate regions for specified object, (c) Candidate regions for all objects

Figure 3. Candidate regions

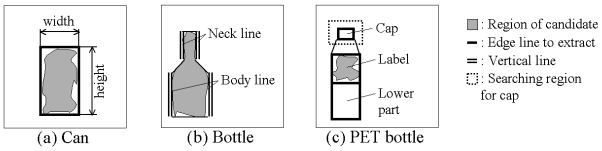
A can is regarded as a rectangle in the image. If four edges around the candidate region (see Figure 4(a)) are extracted and the aspect ratio of the rectangle is approximately 2.0, it is given a high evaluation value as a can.

A bottle has two pairs of vertical lines of the neck and the body (see Figure 4(b)). If lines corresponding to the neck and the body are extracted, the region is given a high evaluation value as a bottle.

A PET bottle consists of three parts: the cap, the label, and the lower part (see Figure 4(c)). After a label is extracted, if edges corresponding to a cap and a lower part are extracted, it is given a high evaluation value as a PET bottle.

When the name of an object is specified, the system verifies the object type based on the evaluation value of the region. When the system tries to recognize all objects, it regards the region as the object type with the highest evaluation value.

Examples of each object type's region are shown in Figure 5.



(a) Can (b) Bottle (c) PET bottle

Figure 4. Models of each object type

(a) Can (b) Bottle (c) PET bottle

Figure 5. Region of each object type

3.3. Matching to object models

When the name of an object is specified, the system first checks the object size. Note that the system can compute the object size in the image with the actual size stored in the object model because it knows the approximate distance to the refrigerator. Then the system retrieves intervals whose representative color matches to that of the extracted region. If no secondary features are registered, the recognition procedure is finished. Otherwise the system extracts secondary features for each interval from the extracted region. If the all the secondary features of one of the intervals are extracted, the region is regarded as the object viewed from the direction corresponding to the interval. Otherwise the system regards the regions as the object candidates for which the most features are extracted.

When the system tries to recognize all objects in the refrigerator, the system compare the region with all the object models and recognizes it in the same way.

4. Recognition supported by dialog

When the system cannot recognize the target object automatically, it uses dialog with the user to obtain additional information.

Case 1: Multiple (n) objects are found.

The system asks the user "I have found n objects. Which object shall I bring?" After the correct object is specified, the system suspects whether the others might be different objects (let one of them be A). In order to avoid the same mistake, the system asks the user, "By the way, are the other objects also target objects?" If the user points out a mistake, the system notices that the color of A shifts to that of the target object and corrects the color range of the object models (see sec. 4.2).

Case 2: No objects are found, and candidate regions are found.

The system considers that the target object overlaps with another object of the same color (see Figure 6(a)). Because the system extracts two overlapping objects as one connected region (see Figure 6(b)) and the region is too large for the object, the system does not interpret the region as the target object but as a candidate region. Then the system shows the region to the user and asks, "I have no confidence, but is this the target object?" The user may say, "No. Two objects overlap." or "Take the front object." Using these advices, the system tries to recognize each object (see sec. 4.1 for detail).

Case 3: No candidate regions are found.

The system considers the following two situations.

Case 3-1: The target object is partially occluded by another object of different color.

The system asks the user the approximate position of the target object such as "I have not found it. Where is it?" Then the user may answer, "The target object is behind object A" Using this advice, the system tries to recognize the occluded object (see sec. 4.1 for detail).

Case 3-2: The color of the target object is shifted due to the change of the lighting condition.

The system asks the user the approximate position of the target object (let it be A) in the same way. After the user answers "At the top shelf." or "Left side of the bottom shelf.", the system tries again to extract the object in the given area by extending the range of the representative color. After the

system extracts the region, it corrects the color range of the target object (see sec. 4.2).

We use ViaVoice by IBM for speech recognition. More detail dialog cases are described in [12].

4.1. Recognition of the object partially occluded by another object

When the target object is partially occluded by another object of the same color, the system first tries to extract the region of the occluding object. Because the bottom boundary of the occluding object is projected at the lower position in the image (see Figure 6(a)), it is determined by the slope of the line which is fitting to the bottom boundaries (see Figure 6(c)). According to the recognized configuration, the system determines the rest of the occluding boundaries lines (see Figure 6(d)).

If the occluded object is specified, based on the known boundary lines, the system predicts the other boundary lines to estimate the contours of the object. Then the system regards the regions surrounded by them as the occluded object (see Figure 6(d)).

(a) (b) (c) (d)

(a) Original image, (b) Result of automatic recognition, (c) Recognition of the object configuration, (d)Recognition result (solid line: occluding object, broken line: occluded object)

Figure 6. Recognition of two overlapping objects of the same color

When the target object is partially occluded by another object of different color, the system first detects the occluding object (see Figure 7(b)). Secondly, the system searches the both sides of the occluding object for the region of the representative color of the target object and extracts a vertical edge (see Figure 7(c)). Lastly, the system extracts the other edges and regards the regions surrounded by them as the target object in the same way (see Figure 7(d)).

(a) (b) (c) (d)

(a) Original image, (b) Occluding object, (c) Extracted vertical edge, (d) Recognition result (broken lines: occluded edges)

Figure 7. Recognition of the occluded object of the different color

4.2. Learning by dialog

In Case 1 and Case 3-1, the reason of the recognition error is that the color range of the feature of A (let it be R_A) is shifted to a range (R_I) which is not overlapping with R_A. Therefore the system needs to modify R_A.

The system first modifies R_A to include R_I. If R_A appears in other intervals, they are similarly modified. If this change causes confusion with other objects, the necessary modification is performed as described in sec. 2.2.

5. Conclusion

We propose a method to register object models so that object may be recognized from any directions. The system recognizes objects automatically using those models. If the system fails in recognition, it interacts with a user. Based on the user support such as pointing out mistakes and choosing the correct object from multiple objects, the system works again. Lastly, the effectiveness of using dialog is presented with examples such as recognition of occluded objects and learning new features of the objects.

Future works are as follows.

- Recognition of various kind of objects
- Recognition of objects in various places such as a cupboard
- Interaction with a user without a display

References

[1] Y. Takahashi et al., "Development of the mobile robot system to aid the daily life for physically handicapped", *Proc. of ICMA2000*, pp. 549-554, 2000.

[2] K. Fujii et al., "A Method of Generating a Spot-Guidance for Human Navigation", *Trans. of IEICE D-II*, Vol. J82-DII, No. 11, pp. 2026-2034, 1999 (in Japanese).

[3] M. Iwata et al., "Linguisitic Expressions of Picture Information Considering Connection between Pictures", *Trans. of IEICE D-II*, Vol. J84-DII, No. 2, pp. 337-350, 2001 (in Japanese).

[4] Y. Watanabe et al., "Image Analysis Using Natural Language Information Extracted from Explanation Text", *Proc. of MIRU'96*, Vol. 2, pp. 271-276, 1996 (in Japanese).

[5] S. Wachsmuth et al., "Connecting Concepts from Vision and Speech Processing", *Workshop on Integration of Speech and Image Understanding*, 1999.

[6] U. Ahlrichs et al., "Knowledge Based Image and Speech Analysis for Service Robots", *Workshop on Integration of Speech and Image Understanding*, 1999.

[7] T. Takahashi et al., "Human-Robot Interface by verbal and Nonverbal Communication", *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 924-929, 1998.

[8] Y. Makihara et al., "An Assistant Robot Acting by Occasional Dialog -Object Recognition and Manipulation Using Dialog with User-", *Proc. of ROBOMEC'01*, 2001 (in Japanese).

[9] Y. Shirai *Three-Dimensional Computer Vision*, Springer-Verlag, pp. 62-68, 1987.

[10] A. Okamoto et al., "Integration of Color and Range Data for Three-Dimensional Scene Description", *IEICE Trans. Inf. and Syst.*, Vol. E76-D, No. 4, pp. 501-506, 1993.

[11] The Color Science Association of Japan, *Handbook of Color Science*, University of Tokyo press, 1989 (in Japanese).

[12] Y. Makihara et al., "Object Recognition Supported by User Interaction for Service Robots", *Proc. of 5th ACCV*, Vol. 2, pp. 719-724, 2002.