Object Recognition Using Sequential Images and Application to Active Vision

Masaki Onishi, Masao Izumi, and Kunio Fukunaga

Department of Computer and Systems Sciences, College of Engineering,
Osaka Prefecture University, 1-1 Gakuen-cho, Sakai, Osaka, 599-8531 Japan

Abstract. In this paper, we propose an object recognition system which integrates features of an object using sequential images. Recently, it has been recognized that an active vision based on the control of camera position is a hopeful approaches to realize a robust object recognition system. Our system integrates features of the unknown object extracted from sequential images while the camera moves to the best position to recognize the object. Integration of features promises a reliable recognition at early step of camera control. The experimental results show the proposed approach promises a reliable and high speed object recognition.

1 Introduction

In the field of computer vision, there have been various kinds of approaches to object recognition [1],[2]. According to these results, it is hard to obtain correct recognition from only one input image, in the case when an input image has occlusion or there are objects which have similar aspects among models. Even though features extracted from one image is not enough to recognize the object, the succeeding sequential images from another viewpoint may make up for lack of features to recognize. Moreover, integration of features from more than two aspect images leads to a reliable recognition system. In this paper, we propose an object recognition system using integration of features obtained from the sequential images at various view points.

In the first, our proposed system recognizes an object using model-based object recognition approach on the basis of a least squares method [3]. In the second, we generate model potential images from 3-D CAD data of the model objects and calculate basic probabilities suitable for an expression of recognition ambiguity as features of the object from these potential images. In the last, using Dempster's rule of combination, it integrates feature which is expressed by basic probabilities calculated from sequential input images at various view points. And we obtain a conclusive recognition result from integrated features.

Besides, it has been recognized that an active vision is one of the hopeful approaches to realize a robust object recognition [4],[5],[6]. Active vision is based on a control of camera position to seek the best camera position to recognize an unknown object. In general, it takes many control steps to move to the best position to recognize an unknown object clearly. Our proposed method could apply to active vision. In the end, our system integrates features of the unknown object extracted from sequential images while the camera moves to the best

position. The proposed method is able to recognize more accurately and fewer moving steps than the conventional method which recognizes without making use of any integration of features. We confirm effectiveness of our method by experimental results.

2 Model-based object recognition

Our system uses three coordinates systems to express 3-D space, world coordinate, camera coordinate and model coordinate system (Fig.1). We adopt a world coordinate system $O^{(w)}$-$x^{(w)}y^{(w)}z^{(w)}$. A camera coordinate system $O^{(c)}$-$x^{(c)}y^{(c)}z^{(c)}$ of which origin is placed on a center point of camera is adopted, such that the axis $O^{(c)}z^{(c)}$ coincides with the optical axis. And to express 3-D models data, a model coordinate system is adopted. Here, we define a camera parameter $\boldsymbol{x}_c$ expressed by a world coordinate system.

$$\boldsymbol{x}_c \equiv [\theta_c \ \varphi_c \ \rho_c \ x_c \ y_c \ z_c]^T. \tag{1}$$

This camera parameter shows the camera position $x_c \ y_c \ z_c$ and camera orientation $\theta_c \ \varphi_c \ \rho_c$ which shows degree of rotating each axis. Similarly, concerning to the model object we define a model parameter $\boldsymbol{x}_m$ expressed by camera coordinate system.

$$\boldsymbol{x}_m \equiv [\theta_m \ \varphi_m \ \rho_m \ x_m \ y_m \ z_m]^T. \tag{2}$$

Next, we discuss the model-based recognition by estimating the similarity between the edge potential of the input image and the edge images of the model objects, based on a least squares method [3]. This method recognizes an object using an input edge image taken by a CCD camera and models data. Here, we define an edge potential image $P(\boldsymbol{X})$. Applying the edge extraction operator to the input image, we compose an edge image $eg(\boldsymbol{X})$, which is obtained by applying the normal function to each of the edge points of the image.

$$P(\boldsymbol{X}_p) = \max_q \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\boldsymbol{X}_p - \boldsymbol{X}_q\|}{2\sigma^2} \right) \cdot eg(\boldsymbol{X}_q) \right], \tag{3}$$
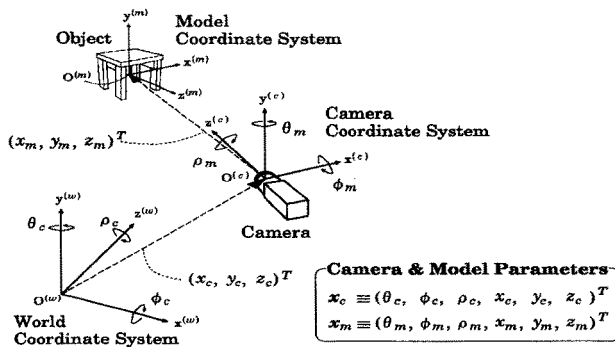
Fig. 1. Camera and model parameters.

(a)Edge image   (b)Edge potential image
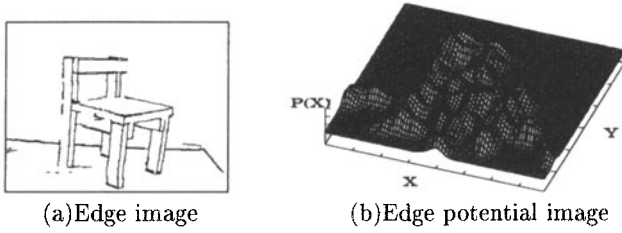
Fig. 2. Example of edge potential image expressed by 3-D.

where $\boldsymbol{X}_p \equiv [X_p \ Y_p]^T$ shows the pixel points, and σ is a standard deviation of a normal distribution. Fig.2 shows an example of edge potential image.

In the similar way, we compose an edge potential image $P_{M_i}(\boldsymbol{X}, \boldsymbol{x}_m)$ for each of the model M_i at the position/orientation $\boldsymbol{x}_m$. This method calculates a residual $S_{M_i}(\boldsymbol{x}_m)$ between an input edge potential image $P_I(\boldsymbol{X})$ and model one $P_{M_i}(\boldsymbol{X}, \boldsymbol{x}_m)$. Suppose that the number of model objects is m, and these model objects are represented by $\mathcal{M} = \{M_1, M_2, \cdots, M_m\}$. Eq.(4) is calculated for each of the model objects M_i, then we decide the candidate object and its position/orientation which has the minimum $S_{M_i}(\hat{\boldsymbol{x}}_m)$ at the optimum position/orientation using those residuals.

$$S_{M_i}(\boldsymbol{x}_m) = \sum_{\boldsymbol{X}_p} \{P_I(\boldsymbol{X}_p) - P_{M_i}(\boldsymbol{X}_p, \boldsymbol{x}_m)\}^2. \tag{4}$$

3 Feature extracted by basic probability

Let us consider basic probabilities for recognizing an object appeared on the input image. Before discussing the basic probabilities, we define a similarity between the input edge image and an edge image of a model object. We assume the number of the model objects is three without loss of generality. In the first, $s_i(i = 1, 2, 3)$ is defined by a similarity between the input image and the image of model M_i. A similarity between the input image and the images of both of model M_i and M_j is given by s_{ij}. And, s_{ijk} is defined by a similarity between the input image and the images of all of the model M_i, M_j and M_k. The relationship of the similarities $\{s_1, s_2, s_3, s_{12}, s_{23}, s_{13}, s_{123}\}$ is shown in Fig.3. Here, we introduce the basic probabilities $\{b_1, b_2, b_3, b_{12}, b_{23}, b_{13}, b_{123}\}$ for the model objects $\{M_1, M_2, M_3\}$. The basic probability $b_i \equiv b(M_i)$ is a rate of feature only belonging to the model M_i, and $b_{ij} \equiv b(M_i, M_j)$ is a rate of feature belonging to both of models M_i and M_j. The common feature among the model M_i, M_j and M_k is given by $b_{ijk} \equiv b(M_i, M_j, M_k)$. The relationship between the similarities and basic probabilities is given by the Fig.3. For example, the similarity s_1, s_{12}, s_{123} is given by following equation.
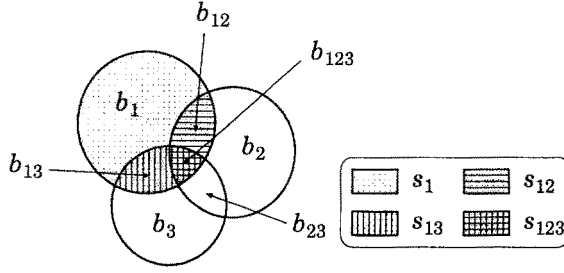
$$s_1 = b_1 + b_{12} + b_{13} + b_{123}, \tag{5}$$

Fig. 3. Basic probability assignment.

$$s_{12} = b_{12} + b_{123}, \tag{6}$$

$$s_{123} = b_{123}. \tag{7}$$

Using the relationship between similarities and basic probabilities, we obtain by

$$\boldsymbol{b}_3 = \frac{1}{T_3}(\boldsymbol{R}_3^{-1}\boldsymbol{s}_3), \tag{8}$$

where, $\boldsymbol{s}_3$ and $\boldsymbol{b}_3$ are the vectors defined by

$$\boldsymbol{s}_3 = [s_1, s_2, s_3, s_{12}, s_{23}, s_{13}, s_{123}], \tag{9}$$

$$\boldsymbol{b}_3 = [b_1, b_2, b_3, b_{12}, b_{23}, b_{13}, b_{123}]. \tag{10}$$

And, the matrix $\boldsymbol{R}_3$ is obtained by Eqs. (5), (6) and (7). In general, an entity R_{ij} is 1 when $\{j_1, \cdots, j_q\}$ of $b_{j_1 j_2 \cdots j_q}$ (j-th entity of $\boldsymbol{b}_m$) includes all of the entities $\{i_1, i_2, \cdots, i_p\}$ of $s_{i_1 i_2 \cdots i_p}$ (i-th entity of $\boldsymbol{s}_m$). The generalized relationship between $\boldsymbol{s}_m$ and $\boldsymbol{b}_m$ is given by the following equation by taking account of Eq.(8).

$$\boldsymbol{b}_m = \frac{1}{T_m}(\boldsymbol{R}_m^{-1}\boldsymbol{s}_m), \tag{11}$$

where, T_m is a normalized constant.

In the next, in order to obtain the similarity, we use an edge potential image which is defined by Eq.(3). And we define similarity s_i between an input and model M_i image by Eq.(12).

$$s_i = \sum \frac{[P_I(\boldsymbol{X}) \cdot P_{M_i}(\boldsymbol{X}, \hat{\boldsymbol{x}}_m)]}{T_P}, \tag{12}$$

$$T_P = \sum [P_I(\boldsymbol{X})^2]. \tag{13}$$

Next, we define similarity s_{ij}, s_{ijk} by the following equation.

$$s_{ij} = \sum \frac{[s_i \cdot P_I(\boldsymbol{X}) \cdot P_{M_j}(\boldsymbol{X}, \hat{\boldsymbol{x}}_m)]}{T_P}, \tag{14}$$

$$s_{ijk} = \sum \frac{[s_{ij} \cdot P_I(\boldsymbol{X}) \cdot P_{M_k}(\boldsymbol{X}, \hat{\boldsymbol{x}}_m)]}{T_P}. \tag{15}$$

Using these similarities, all of the basic probabilities can be obtained by Eq.(11).

4 Integration of features

4.1 Recognition probability

It is hard to recognize by using basic probability itself, because basic probability is assigned to a set of model objects. We introduce recognition probability as a discrimination parameter. Each of recognition probability is obtained by basic probabilities. In the case when a focal element that consists of more than two model objects, we calculate a recognition probability for each model object by distributing the basic probability to each of model object equally.

$$R_{M_i} = \sum_{M_i \in \mathcal{M}_k} \frac{b(\mathcal{M}_k)}{|\mathcal{M}_k|}, \tag{16}$$

where $\mathcal{M}_k$ denotes the set of model objects, and $|\mathcal{M}_k|$ shows a number of model objects in the $\mathcal{M}_k$. The recognition probability R_{M_i} shows a rate how degree the input object is the model M_i. A model with largest recognition probability becomes first candidate for model M_{t_1}.

4.2 Integration of features

Even if one input image have insufficient feature of the object, recognition adopted integration of features extracted from sequential images leads to be reliable one. Here, the feature is represented by the basic probabilities. We introduce the Dempster-Shafer's rule of combination to integrate the features obtained by the plural images [7].

5 Application to active vision

Our method, object recognition using sequential images, is applicable to active vision system. Our active vision is based on the control of camera position which is the best position to discriminate the object [6].

The basic probabilities show the degree that the input object belongs to a model subset. Let us define a degree of recognition-ambiguity concerning to model M_{t_1} by using basic probabilities whose focal elements is more than 2, and one of the focal element is model M_{t_1}. Using these probabilities, we define a degree of recognition-ambiguity A.

$$A = \sum_{(|\mathcal{M}_k| \geq 2) \wedge (M_{t_1} \in \mathcal{M}_k)} b(\mathcal{M}_k). \tag{17}$$

The degree of recognition-ambiguity A is defined by the sum of the basic probabilities whose focal element includes the model candidate M_{t_1} which is most similar to the object on the input image. The degree of recognition-ambiguity

shows possibility of belonging to other model although we suppose the unknown object on the input image to be the model M_{t_1}. This value is getting large if the unknown object is getting similar to the aspect image of other model M_k ($k \neq t_1$). As mentioned above, the principle of controlling the camera parameter x_c is to decrease the degree of recognition-ambiguity A on the another position, and this position promises that the system discriminates the input object more clearly. In general, it take many control steps, however, to move the best position to recognize an unknown object clearly. Here, our proposed method could apply to active vision decreasing the number of control steps, and promising more accurate recognition system.

First, our approach is to seek the better camera position by decreasing A. In the next step, to recognize the object in a short time, we integrate the basic probabilities calculated by sequential input images taken at each steps of camera position.

6 Experimental results and considerations

6.1 Experiment using sequential images

We have confirmed the effectiveness of our method by real scene experiments. In our experiments, we use sample model chairs which have similar aspect each other as shown in Fig.4. We decide the parameters σ (concerning to edge potential) to be $\sigma = 5$ under considering the experimental results. As an example of the experimental conditions, we choose the model chair4 for an input object. Fig.5 shows two input images with occlusion and their recognition probabilities at each images. Ave(1,2) shows the average of recognition probabilities Image1 and Image2. And, DS(1,2) shows integration of these probabilities. We can confirm that recognition probabilities are improved by using sequential images than only one image.

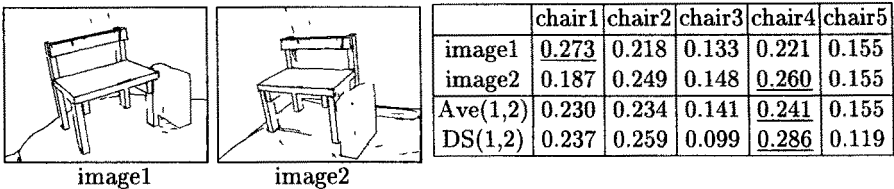chair1   chair2   chair3   chair4   chair5

Fig. 4. Models.

	chair1	chair2	chair3	chair4	chair5
image1	0.273	0.218	0.133	0.221	0.155
image2	0.187	0.249	0.148	0.260	0.155
Ave(1,2)	0.230	0.234	0.141	0.241	0.155
DS(1,2)	0.237	0.259	0.099	0.286	0.119

image1   image2

Fig. 5. Sequential input images and recognition probability R.

STEP0   STEP3   STEP6   Recognition
                        result
(a)Change of input image at each step.

(b)Conventional method.   (c)Proposed method.
Change of evaluation value R at each step.
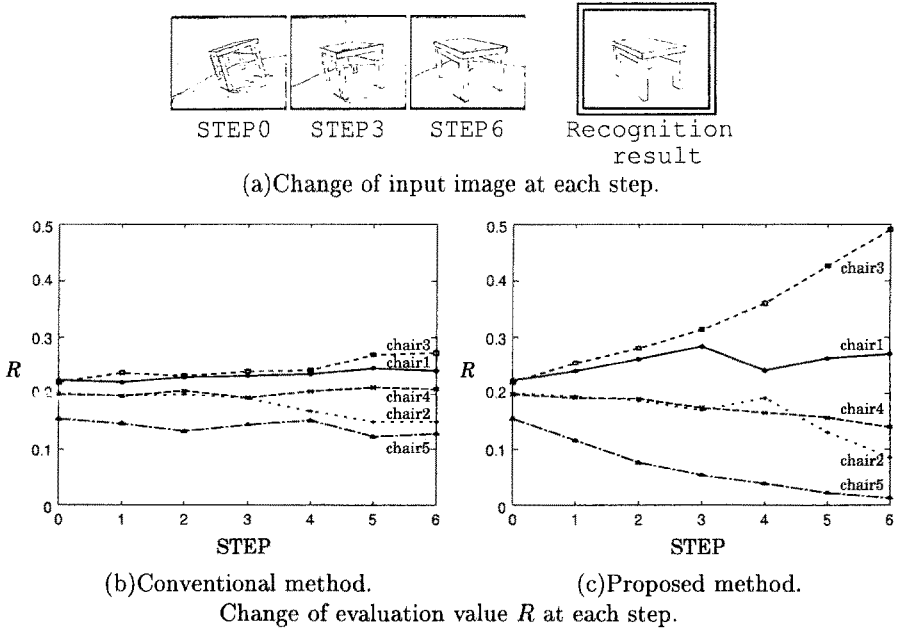
Fig. 6. Experimental result -chair3-.

6.2 Experiment of active vision

Next, we have confirmed the effectiveness of active vision system using our method. We made an experiment by controlling the camera position using robot-arm. As the experimental conditions, we choose the model chair3 as an input unknown object and set the initial position/orientation parameter which makes similar aspect to chair1. Fig.6(a) shows the input edge image at each iteration step, and recognition result at last step. Fig.6(b) shows the change of the recognition probability R concerning to conventional method [6], and (c) shows result of our proposed method. It is hard to decide from Fig.6(b) what number of step recognizes an unknown object clearly. However, we confirm a definite improvement of recognition result from Fig.6(c). These experimental results show that our proposed method can obtain the correct recognition result in early step of movement which is the effect of integrating the features of the object.

6.3 Considerations

In the next, we discuss our proposal approach and experimental results.

1. We express by basic probabilities how similar to models an input object is. And we can reach an conclusive recognition result by integration of basic probabilities calculated from sequential images using Dempster's combination rule. Our approach promise more an accurate result than recognition using only one image.

2. In the case of integrating features from sequential images, we are able to expect accurate result to make up for lack of features. That explains why proposed method apply to active vision. Our active vision system makes an act plan so as to decrease the degree of recognition-ambiguity and controls a camera to move to the optimum position. Our proposed active vision promises a high speed recognition system and an accurate recognition result.

3. A model-based recognition assumes that models database are made preparation in advance. It is effective, however it is not general use, to discriminate and assort industrial parts on the belt conveyor or the industrial products placed in the factory.

4. It is hard to recognize an object using model-based object recognition approach on the basis of a least squares method under complex background in general. That explains why our system need the method which distinguish an object from background as subtract an input image from background image.

7 Conclusions

In this paper, we have proposed a method based on an active vision which integrates features obtained from input sequential images. Even if we could not obtain enough feature from only one input image, a result of recognition can be improved by using integrated features from sequential images, because it makes up the lack of feature each other. Our proposed method carries out a reliable and high speed recognition system.

References

1. Karen B. Sarachik, "The Effect of Gaussian Error in Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 19, No. 4, pp. 289–301, Apr. 1997.
2. David W. Jacobs and Ronen Basri, "3-D to 2-D Recognition with Regions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 547–553, 1997.
3. S. Yamane, M. Izumi and K. Fukunaga, "Model-Based Object Recognition and Its Position and Pose Estimation," *IEICE,* Vol. J-79-D-II, No. 2, pp. 165–173, Feb. 1996 (In Japanese).
4. J. Aloimonos, I. Weiss and A. Bandyopadhyay, "ACTIVE VISION," *Proc. 1st IEEE Int. Conf. on Computer Vision,* pp. 35–54, 1987.
5. T. Akutsu, K. Arakawa and H. Murase, "Active Viewpoint Control for Shape from Occluding Contours," *Proc. 3rd Asian Conf. on Computer Vision,* Vol. I, pp. 321–328, 1998.
6. K. Fukunaga, N. Nishikawa, T. Matsumoto and M. Izumi, "Model-Based Object Recognition Using Camera Control," *Proc. Conf. on Emerging Technologies and Factory Automation,* Vol. 2, pp. 707–711, 1996.
7. Seth. A. Hutchinson and Avinash. C. Kak, "Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities," *IEEE, Trans. on Robotics and Automation,* Vol. 5, No. 6, pp. 765–783, Dec. 1989.