MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

13—8

Pose Detection of 3D Object by Genetic Algorithm

Kimiya Aoki* and Toyohisa Kaneko†
Department of Information and Computer Sciences,
Toyohashi University of Technology‡

Abstract

In this study, an algorithm for detection of the specified object pose is proposed. In input data are range images being accumulated in a process of which a stereovision moves arbitrary. An improvement genetic algorithm (GA) is applied to the model base matching. Since it thought that viewpoint movement brings dynamic environment to GA, to cope with the multi-peak problem and changes of an evaluation function, two concepts, "breed" and "competitive coexistence" are introduced to GA. For performing robust matching, a generation technique of an evaluation space using the 3D morphological operation is proposed. Finally experiments are performed, and the validity of algorithm is shown.

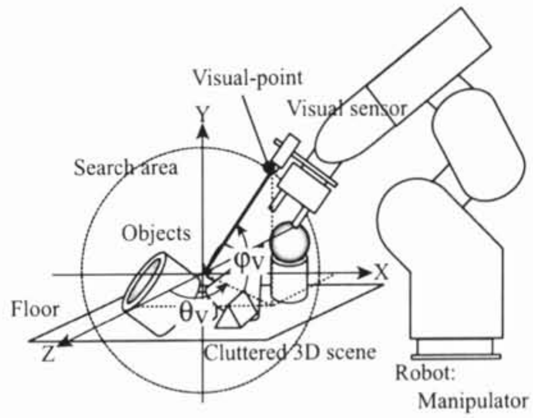


Figure 1: Assumption system

1 Introduction

In this study, a method for detecting a pose of 3D object is proposed, and it is purpose to mount to welfare and home robots. 3D object recognition is an important subject of the computer vision research, and so various methods have been proposed[1]~[5]. Technical subjects in our study are summarized as follows:

1. Input data are range images including omissions of distance information because of a mistake in detecting correspondence point between azimuth difference images.
2. Cluttered 3D scenes in which there are plural objects are search targets.
3. Location and orientation (there are 6 parameters.) of the object specified by a user can be detected.
4. Range images are accumulated in a reconstructed 3D space because a stereovision moves around the search space, it is assumption that our system is mounted to welfare or home robots.

An assumption system is illustrated in Fig.1. The view-point (ϕ_v, θ_v) can move on the sphere covering up the search space.

For above-mentioned technical subjects, we proposed an algorithms having following specialties:

1. The improvement genetic algorithm (GA) is applied to the model base matching (3D to 3D).
2. Complex 3D objects can also be registered as a template by useing a polygonal model having an arbitrary mesh form.
3. In order to perform a robust detection, evaluation space is constructed by using recursive dilation on the 3D morphological operation.
4. The viewpoints where the specified object is most distinctly visible during viewpoint movement are searched.
5. Viewpoint movement brings dynamic environment to GA. Therefore, in order to cope with the multipeak problem and changes of an evaluation function, two new concepts, "breed" and "competitive coexistence" are introduced to GA.

2 Experiment Setup and Data Flow

We have made a system having above-mentioned specialties on an experimental basis. The system validity is shown by experiments with using real images. The experiment equipment is shown in Fig.2. The system is assembled by a 3-eyes stereovision camera, a rotation table, a personal computer and etc.. If the stereovision and the rotation table are controlled about angle and height, a range image similar to the assumption system as shown Fig.1 can be obtained. Stuffing plural objects into a transparent case, a cluttered 3D scene is constructed.

Data-flow of input range images is shown in Fig.3. Input images are converted to orthogonal projection images (Fig.3(c)) using an experimental calibration function, and then input to a voxel space (Fig.3(d)).

*E-mail:aoki@ics.tut.ac.jp
†E-mail:kaneko@ics.tut.ac.jp
‡1-1 Hibarigaoka, Tempaku, Toyohashi-shi, 441-8580 Japan.

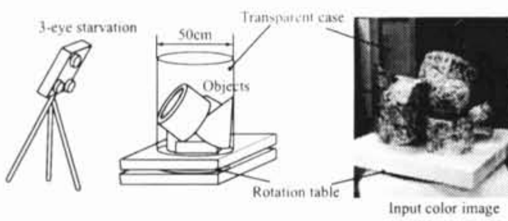


Figure 2: Experiment setup

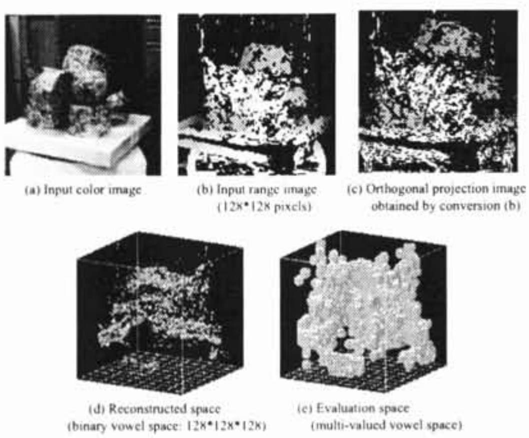


Figure 3: Data flow

Moreover, this binary voxel space is converted to an evaluation space (Fig.3(e)) using the 3D morphological operation(It is 3.2 for details.). The stereovision camera moves around a search space and range images are accumulated in this voxel space. A target may not appear from an initial viewpoint. Therefore, the viewpoint is moved, and it is necessary to perform in the same time viewpoint movement and matching process.

3 Method

3.1 Approach

With huge search space, GA is available. But it has a weakness for the multi-peak problem[6],[7]. Consequently, for diversifying a population, a search space is divided into plural subspaces. The subpopulation acting to detection independently in each subspace is defined as "breed" in this study. An illustration that breed is set up are shown in Fig.4. In actual search process, it is necessary to globally search, until the specified object becomes visible, and focus on the subspace in which there is the target object when it is visible. Therefore, we propose that the number of population in each breed is adjusted corresponding to each breed's fitness changing under the influence of a viewpoint movement. That is to say, the population of a breed having a relatively high fitness increases in number, and it decreases in converse case. Because that breeds struggle for existence, we called this method "competitive coexistence". This new operation improves the efficiency of search. Moreover, it is possible to perform view-point movement and matching process in the same time.

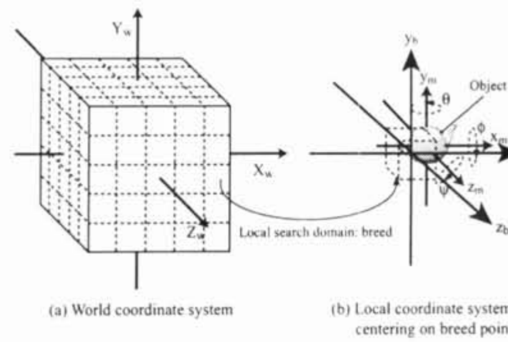


Figure 4: Division of search space

3.2 Definition of Fitness

In using the model-based matching, the evaluation rules of an individual have to be defined. In this study, a fitness level between a model object and a certain region in the re-constructed voxel space is evaluated.

A model template is registered as polygonal apexes data, and it is converted to voxel data when matching process is performed. Making reference to a depth buffer that is used for rendering, the polygonal model is quantized. Thus, a polygonal model having an arbitrary mesh form can be registered independently of an input data format from a vision sensor.

The model template data with arbitrary location and orientation are superimposed on the voxel space re-constructed by input range images. Then, counting the number of overlapping voxels (=volume value 1) is a simple method for evaluation, but using this method; the fitness value does not increase, until the model object adequately matches the re-constructed object corresponding to it in the voxel space. Therefore, the score space based on the re-constructed voxel space is proposed. The score space has the same structure of the voxel space. First, if a voxel in the range data space is 1 as a volume value, the same coordinate voxel in the score space is substituted by maximum score L. Then, other voxels are assigned to the score value with distance from the voxel having L. The score value of each voxel is given by Eq.(1).

$$\begin{gathered} D_i^{'} = (D_{i-1}^{'} \oplus K) \qquad D_i = D_i^{'} - D_{i-1}^{'} \\ i = (1, 2, 3, \ldots, L-1) \\ Score \quad D_0 := L \qquad D_i = (L - i) \end{gathered} \tag{1}$$

These equations indicate recursive dilatation of the 3D morphological operation; where K is a structure element whose size is incremented by 1. Namely, as the dilatation operation is performed recursively, it starts from the voxels having score L, and new voxel sets (D_i) with i times dilatation are substituted ($L - i$) as the score value. Concrete fitness value is given by Eq.(2), using the proposal score space.

$$fitness = \frac{1}{n\,L^2} \sum_{j=0}^{n-1} S_j^2 \tag{2}$$

n is total number of voxels (=volume value 1) constructed with polygon surfaces of the model object, and S_j is score value about each voxel belonging to the model object in the score space.

3.3 Generation of Matching Model

Although the 3D model base matching method deals generally with only six matching parameters concerning translation and rotation, the viewpoints where the specified object is most distinctly visible during viewpoint movement are also searched in our system. When model template is converted to voxel data by reference of a depth buffer, the viewpoint setting is important. In this study, because the viewpoint moves, in order to construct a matching model, it is necessary to link depth buffer data from multiple viewpoints. Since the target object is not all visible from viewpoints, reference viewpoints should be chosen. In this paper, three viewpoints are chosen from last four viewpoints. The one viewpoint that isn't chosen is decided by gene 'v' built in a chromosome.

3.4 GA Operations

A method for coding a phenotype to a genotype is important in GA. In this study, a phenotype has seven parameters indicating translation (x, y, z), rotation $(\phi, \theta, \psi)$and 'v', and a genotype is a bit line composed by 0 and 1. In this study, the elitism method is used for reproduction. The uniform crossover is adopted for crossover operation. In mutation operation, a bit value is turned over (0to1 or 1to0) with mutation rate.

The 3D information on a scene is accumulated because a camera moves in the arbitrary directions from an initial viewpoint. Therefore, a fitness function can be updated on each generation. Moreover, it is a different point from general GA. In this study, a viewpoint moves every T_g generations. The flow of processing is explained according to numerical order in Fig.5.

1. An individual group is generated about every breed and the fitness in an initial viewpoint is calculated.

2. In the breeds in activity, selection, crossover, and mutation processing are performed, and a new individual is generated. A matching model is generated using viewpoint information and the fitness is calculated. The breed with the maximum fitness more than a threshold Th_α fixes a standard viewpoint to the present viewpoint. The lesser breed than the others stops activity at every $T_g/2$ generation.

3. Processing is ended when the set-up search term G generation comes. In this case, it does not mean completing search.

4. When the superior breed into which the maximum fitness exceeds threshold Th_β exists, viewpoint movement is stopped, and search processing is monopolistically performed about this breed.

5. When a search is performed for G_m generations in this superior breed, and the maximum fitness more than threshold Th_γ is obtained, processing is ended as the completion of detection. Viewpoint movement is resumed when it is not able to detect.

6. If this generation is not the cycle of a viewpoint movement, processing flow returns to processing (2). In the case of the generation with viewpoint movement, the range image obtained from a new viewpoint is added to re-construction space, and the score space is updated. Furthermore, every breed's population is updated according to their own maximum fitness. The gene corresponding to parameter v is also updated. Finally, the breed of hibernation is revived. Processing flow returns to processing (2) after the above environmental change.

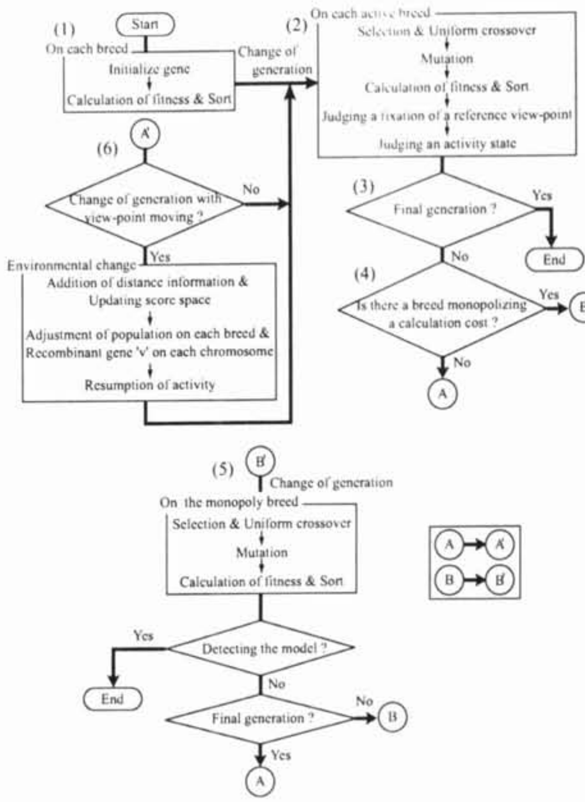


Figure 5: Flwochart

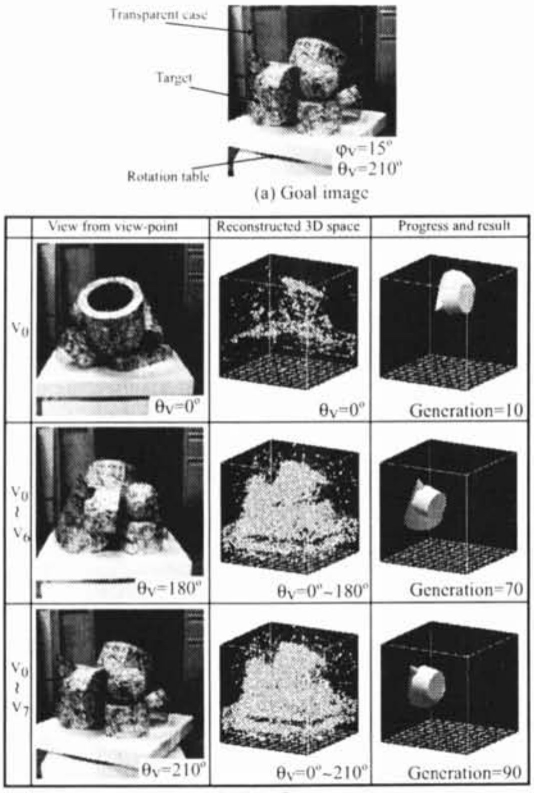
As mentioned above, according to the degree of environmental adaptation, activity and hibernation are set up for every breed. And the number of individuals which can be survived is further adjusted according to the maximum fitness of the breed. The number of individuals of the breed r in generation g which can be survived $(p_{r,g})$ is given by the formula (3).

$$
\begin{aligned}
f\_ratio_{g-c,r} &= \frac{f\_max_{g-c,r} - f\min_{g,all}}{f\_max_{g,all} - f\min_{g,all}} \\
p_{g,r} &= f\_ratio_{g-c,r}(p\_max - p\_min) \\
&+ p\_min
\end{aligned} \tag{3}
$$

$f\_ratio_{g-c,r}$ is the relative ratio of the maximum fitness in the generation $(g-c)$ of breed r. p_{max} and p_{min} are the maximum and minimum numbers of permission individuals, respectively.

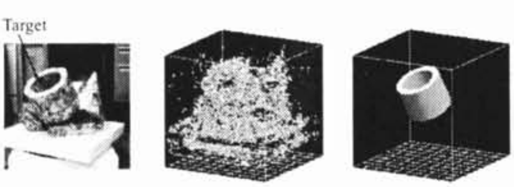
4 Experiment

An example of an experiment result is shown in Fig.6. Stuffing plural objects into a transparent case, a cluttered 3D scene is constructed. In this experiment,



(b) Search progress

Figure 6: Experiment result 1



(a) Goal image (φ_V=15° θ_V=30°) (b) Reconstructed 3D space (θ_V=0°~60°) (b) Result (Generation=45)

Figure 7: Experiment result 2

the stereovision camera was fixed, and the rotation table was rotated 30 degree per 10 generations during searching. The object consisting of two primitive models (cone and cylinder) was the target. In Fig.6, the image of progress and result is a view from a viewpoint from which the target object is most distinctly visible (please compare the image of Fig.6(a) and images on the right line of Fig.6(c)). The parameter used for the experiment is shown in Table 1. Another result is shown in Fig.7. From these figures, it is shown that pose detection become successful by our method even though reconstructed search space includes omissions of distance information.

5 Conclusion

In this paper, an external world recognition system of robots coexisting with humans was discussed. And the algorithms for detecting 3D objects using range information were investigated. In order to construct more robust detection system, GA with the addition of two ideas ("breed" and "competitive coexistence") was

Table 1: Search conditions

Item	Value
Search space $[cm^3]$	50
Reconstructed 3D space [voxel]	128^3
L	5
N (the number of breeds)	27
G,G_m,T_g[$generations$]	150,30,10
Th_α,Th_β,Th_γ	0.6,0.7,0.72
p_{max},p_{min}	50,10
Selection rate (s_{rate})	0.7
Crossover rate (c_{rate})	0.5
Mutation rate (m_{rate})	0.03
Translation range $(x, y, z)[cm]$	$-4.165 \sim 4.165$
Rotation range (ϕ)	$-\pi/2 \sim \pi/2$
Rotation range (θ)	$0 \sim 2\pi$
Rotation range (ψ)	$0 \sim 2\pi$
Chromosome length (sum)[bit]	41
Chromosome length (x, y, z)[bit]	4×3
Chromosome length (ϕ, θ, ψ)[bit]	9×3
Chromosome length (v)[bit]	2

applied to the model-based matching. GA with these ideas was able to cope with the multi-peak problem and a change of the environment. Experiments using the proposal technique with real images were conducted, and the validity of algorithms was shown. It is concluded that the proposed algorithm is suitable for an active robot vision.

Acknowledgments

A part of this work is supported by "Yazaki Memorial Foundation for Science and Technology" and "THE HORI INFORMATION SCIENCE PROMOTION FOUNDATION".

References

[1] M.D.Wheeler and K.Ikeuchi,"Sensor Modeling Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition," IEEE Trans. PAMI, Vol.17, No.3, pp.252-265, Mar. 1995.

[2] A.E.Johnson and M.Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Trans. PAMI, Vol.21, No.5, pp.433-449, 1999.

[3] Y.Sumi, Y.Ishiyama, and F.Tomita, "Real-Time Free-Form Object Tracking Using a Model-Based Approach for Stereo Vision Systems," Trans. IEICE Japan, Vol.J84-D-II, No.8, pp.1693-1700, 2001. (in Japanese)

[4] T.Takeguchi, S.Kaneko, T.Kondo, and S.Igarashi, "Robust Object Recognition Based on Depth Aspect Image Matching," Trans. IEICE Japan, Vol.J84-D-II, No.8, pp.1710-1721, 2001. (in Japanese)

[5] K.Aoki and T.Kaneko, "Detection 3D Object with Color Texture by Genetic Algorithm," Proc. ACCV2002, Vol.1, pp.380-385, Jan. 2002.

[6] S.Tsutsui and Y.Fujimoto, "The fGA: Forking Genetic Algorithm with Blocking and Shrinking Modes," Trans. JSAI Japan, Vol.9, No.5, pp.741-746, 1994. (in Japanese)

[7] D.Whitley, J.R.Beveridge, C.Guerra-Salcedo, and C.Graves,"Messy Genetic Algorithms for Subset Feature Selection," Proc. The 7th International Conference on Genetic Algorithms, Jul. 1997.