ACCV2002: The 5th Asian Conference on Computer Vision, 23—25 January 2002, Melbourne, Australia.

Detecting 3D Object with Color Texture by Genetic Algorithm

Kimiya Aoki and Toyohisa Kaneko
Department of Information and Computer Sciences
Toyohashi University of Technology
aoki@ics.tut.ac.jp

Abstract

Development of welfare and home robots that replace man in aging society and perform housekeeping and care is important. It focused on the external world recognition system of robots coexisting with man, and this research considered recognition of 3-dimensional object using distance information and texture information as introductory part.

Various methods for recognition of 3D object have been proposed from the former, in this study, in order to construct more robust recognition system, we tried to apply the genetic algorithm (GA) to the model-based matching. Moreover, we introduced two concepts, "Breed" and "Competitive coexistence", into GA in order to fit a robot vision. And a robust method using range and color texture information for evaluating fitness on the model-based matching was proposed. GA added these ideas is able to cope with the multi-peak problem and change of environment (=fitness function) on search. Therefore, it is possible for the proposal system for detecting 3D object to deal with complexity of the searched object shape and texture, and a lack of range information obtained by its visual sensor. Finally, simulation was performed about the proposal technique and the validity of algorithms was shown.

1.Introduction

Development of welfare and home robots that perform care and easy housekeeping tasks for men is an important subject in aging society. Such robots are required to equip with the mounting ability for self-driving, recognizing, and picking up 3D objects, that is to say, autonomous ability [1,2]. Robots have spread in the industrial field from past to present, the work environment of industrial robots is extremely improved in lighting, positioning of parts (= recognition targets) and etc., so robots are able to recognize the external world comparatively easily using their visual sensor [3,4,5]. For example, recognition targets are of comparatively easy shape consisting of planes and simple curved surfaces in case of industrial robots. But welfare robots have to deal with common household goods which tend to have more complex shapes and colorful textures such as a brand name and trade mark. The technique for making recognition task easy by marking on detection symmetry was also proposed [6].

In this study, the external world recognition system of robots coexisting with human was focused, and robust

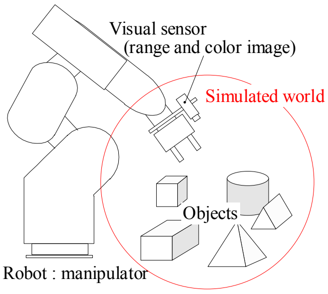


Figure 1. Assumption system

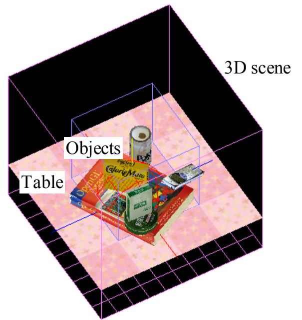


Figure 2. An example of a test scene

algorithm for detection 3D objects using range and texture (color) information was investigated.

This research is concerned with the problem of extraction and re-construction of 3D objects with form matching. Various methods to re-construct solid with accounting 3D structure using edges and apexes information obtained by 2D analysis to image have been investigated [7-11]. However, many of past researches have considered 3D objects of basic shapes such as a cube, pillar, triangular pyramid and so on. Because ideal lighting and photographic conditions are assumed or prepared, the input image is clear and has good contrast (edges, apexes and planes are detected by simple image processing easily). Moreover, form-matching methods on 2D have been studied and applied to part detection and others in the industrial field [12], but form-matching methods on 3D have few reports. Consequently, in this study, the form-matching method for recognizing 3D objects with color texture on disorderly background will be investigated.

Concretely, the genetic algorithm (GA) and model-based matching are applied to the form-matching method [13-15]. We propose adding two ideas called "*Breed*" and "*Competitive coexistence*" to general GA operations, and the robust method using range and color-texture information for evaluating fitness on the model base matching [16].

2.Outline of Processing

The system is illustrated in Fig.1. It is assumed that the robot has a certain vision sensor, which can acquire distance and color information. That is to say, in this paper, the sensor part of the robot and all 3D objects in the external world scene are built in a computer virtually. An example of a test scene is shown in Fig.2. It searches the position and posture of the specified object in the space

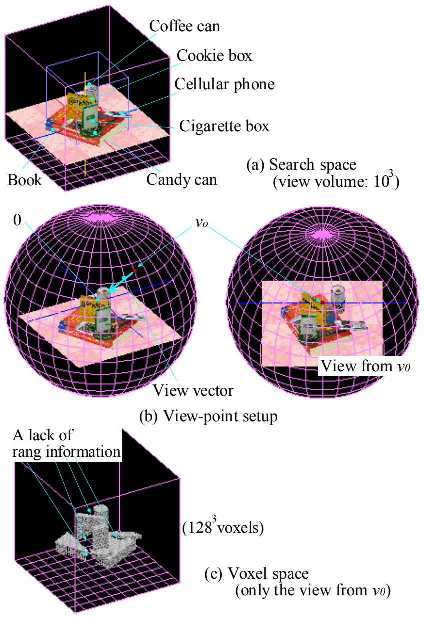


Figure 3. Relationship of spaces

shown in Fig.2. Of course, required information, which can be used in the case of detection, is distance value from a view-point to object surfaces. For example, data occluded by another object cannot be used.

The method proposed in this paper is the model-based matching. There are six matching parameters, which are translation (x,y,z) and rotation (ψ,θ,ϕ). The 3D shape and texture of the model are employed for matching directly in the search space, so robust recognition under a lack of distance information and noises is expected. But it is not surprising that the search space (=parameter range) becomes very huge. Then, GA that is a method of optimization requires some adaptation.

3.Problem Statement

An example of the search space is shown in Fig.3(a). The size of view volume is 10^3, and in this example, six objects such as a book, cigarette box, cookie box, cellular phone, coffee can, candy can, are located with arbitrary orientation. The problem is to identify their location and orientation. The search process uses the distance from a view-point to surfaces and also color or textures. If the view-point is v_0 and the view vector is zero point direction in Fig.3(b), a range image is obtained as shown in Fig.3(c). The range image is then converted and substituted into the voxel space. The size of this voxel space is 128^3 cells. The voxels corresponding to the surface of objects in the external world have volume value *1*, whereas other voxel's value is *0*. (It is to be noted that this voxel space is binary.) Moreover, for voxel's value of *1*, RGB color is kept. That is to say, this voxel space corresponds to a robot's memory where the range information acquired by a vision sensor is re-constructed and converted to the 3D space. So search processing is carried out in the voxel space.

Now, it is important for re-construction of the 3D space to move a view-point, and amount of information in the voxel space have an influence on success of search. From Fig.3(c), it is confirmed that objects on counter side from the view-point and covered by other object lack the surface information. If the view-point is moved around the search space enough for re-construction of the 3D space, more complete voxel space will be constructed. And then, there are two kinds of strategy in search process. One is a method of search using the complete voxel space obtained by fully moving the view-point. Another is a method of performing view-point movement and matching processing in parallel. The latter strategy is natural, if search of human is considered; though more advanced algorithm is needed. In this paper, the latter strategy was examined.

4.Definition of Fitness

4.1 Basic Definition

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

First of all, polygon surfaces of the model object are converted to voxel data with the same way in the search space. This model object data with arbitrary location and orientation are superimposed on the voxel space (=robot's memory). And then, counting the number of overlapping voxels (=volume value *1*) is a simple method for evaluation, but using this method, fitness value does not increase until the model object matches the re-constructed object corresponding to it in the voxel space. Therefore, the score space based on the voxel space is constructed. The score space is the same structure of the voxel space constructed from range information. It is a voxel space and its size is 128^3. First, if a voxel of range data space has *1* as volume value, the same coordinate voxel in the score space is substituted maximum score *L*. Then, surrounding voxels are substituted score value with distance from a voxel having *L*. The score value of each voxel is given by Eq. (1).

$$D'_i = (D'_{i-1} \oplus K) \quad D_i = D'_i - D'_{i-1} \quad i=(1,2,\ldots,L-1) \qquad (1)$$
$$Score \quad D_0 := L \quad D_i := (L-i)$$

These equations indicate recursive dilatation of the 3D morphological operation [17,18]; where *K* is a structure element whose size is incremented by *1*. Namely, as the dilatation operation is performed recursively, it starts voxels with score *L*, and new voxel sets (D_i) with *i* times dilatation are substituted ($L-i$) as the score value. An example of the score space is shown in Fig.4.

Concrete fitness value is given by Eq.(2), using the proposal score space.

$$fitness = \sum_{j=0}^{n-1} S_j \Big/ L \cdot n \qquad (2)$$

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 (=volume value *1*) belonging to the model object in the score space.

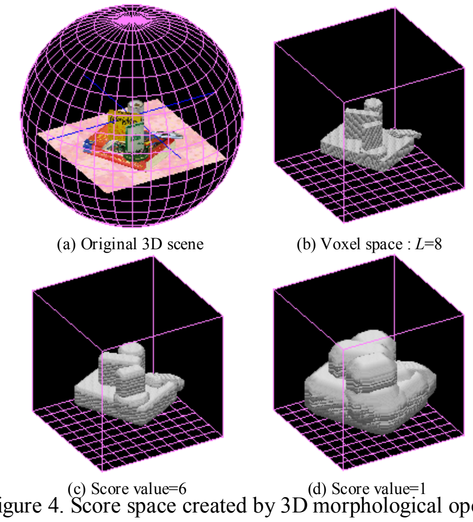


Figure 4. Score space created by 3D morphological operation

Using the proposal score space to evaluate in time of search permits that fitness value increases in a degree, if the model object misses slightly, and then robust evaluation is performed.

When an object is symmetrical such as a cuboid, the above method is not applicable to the identification of object direction because shape information alone is used. Next, the score space is expanded by color-texture information.

4.2Utilization of Color-texture Information

First, RGB color data are converted to HSV color data. Hue datum that is robust to variation of lighting is utilized for evaluation of matching rate. Hue-Saturation plane is divided into six equal part (these are 60° each) and labeled '0～5'. In low saturation area, '6'(white) or '7'(black) label are given according to intensity value. Then, the voxel space obtained by range information is divided into eight voxel spaces with same structure. These voxel spaces correspond to eight color labels respectively. And each color voxel space is converted color score space by Eq.(1). An example of the color score space is shown in Fig.5 (about score value = 6).

When matching rate is calculated, then model voxel data are divided using the color-label table and superimposed on the color score space respectively. Concrete fitness value is given by Eq.(3).

$$fitness = \sum_{c=0}^{7}\sum_{j=0}^{n_c-1} S_{c,j} \Big/ L \cdot \sum_{c=0}^{7} n_c \qquad (3)$$

Where n_c is total number of voxels (=volume value *1*) constructed with polygon surfaces of the model object about each color-label *c*, and Sc,j is score value about each voxel (=volume value *1*) belonging to the model object in the color score space labeled *c*.

Using the proposal color score space permits that shape and color texture of a model object are evaluated at the same time.

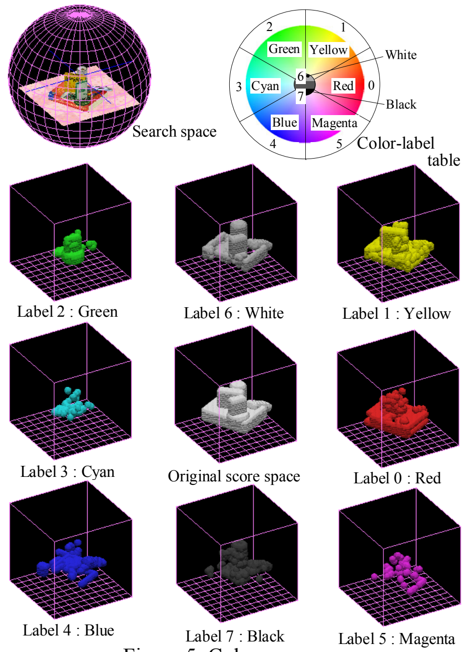


Figure 5. Color score space

5.Coding of Matching Parameter

A method coding phenotype to genotype is important in GA. In this study, phenotype is six parameters indicating translation (*x,y,z*) and rotation (ψ,θ,ϕ), and the genotype is a bit line composed by *0* and *1* generally. These parameters are coded and this bit line is defined as genotype for location and orientation of a searched model object. Figure 6 shows relationship between phenotype and genotype.

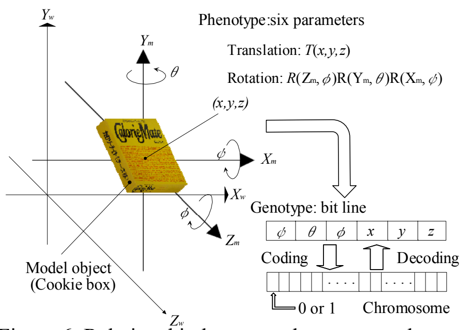


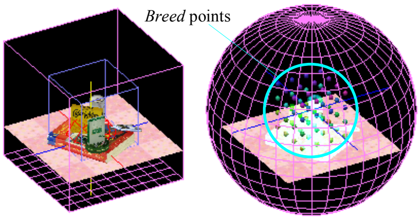
Figure 6. Relationship between phenotype and genotype

6.Rule of Reproduction, Crossover and Mutation

As mentioned above, parameters for translating and rotating a model object are defined as an individual on GA. In this study, the elitism method is used for reproduction. In a generation, individuals are arranged according to the fitness value, and these of low ranking are eliminated. Then, a couple is chosen from higher rank at random, and new individual is created by crossover operation. The number of the population is kept. Using this reproduction method, a reasonable solution is detected earlier.

The uniform crossover is adopted for crossover operation.

In mutation operation a bit value is turned over *0* to *1*, or *1* to *0* by mutation rate.



(a) Search space (b) Divided search space (Arranged *breed* points)

Figure 7. Division of the search space by *breeds*

7.Definition of "*Breed*"

GA is a useful search technique, but it has a weakness in the multi-peak problem [19,20]. Also in this research, it shows the multi-peaks especially in search of location. Then, the search space is divided and the individual groups in which GA operation is performed independently are defined for every division domain. This definition is called "*Breed*". Specifically, as shown in Fig.7, *r_n* cubes divide the search space, and the *breed* point (=base point for local search by each *breed*) is installed in the center respectively. The individual group belonging to each *breed* searches only the local domain centering on the *breed* point. Coordinates conversion of a model object are given by Eq. (4).

$$P' = (T_r TR)P \qquad (4)$$

Where P is the vector of polygon apexes belonging to a model object, and T_r is the matrix for translating from zero point to each *breed* point, and T and R is the conversion matrix (translation and rotation) indicated parameters by GA search.

As the feature, GA parameters different for every *breed* could be set up; the number of populations in the *breed* can be adjusted this time. That is, the total number of populations in the search space is set constant, and the *breed* scramble for the number of survival populations. The number of populations in a generation with each *breed* is given by Eq. (5) as $pop_{g,r}$.

$$pop_{g,r} = f\_ave_{g-c} \cdot \frac{pop_{world}}{\sum_{r=0}^{r\_n-1} f\_ave_{g-c,r}} \qquad (5)$$

Where $f\_ave_{g,r}$ is the average fitness of *breed* r at generation g, and pop_{world} is the total number of populations. This corresponds to the survival struggle between *breeds*. It enables to calculate an optimum solution without getting trapped into local minimum, introducing "*Breed*" and its competition processing.

8.Experiment

The simulation of 3D object detection was performed using the algorithms described in the above.

8.1Flow of Processing

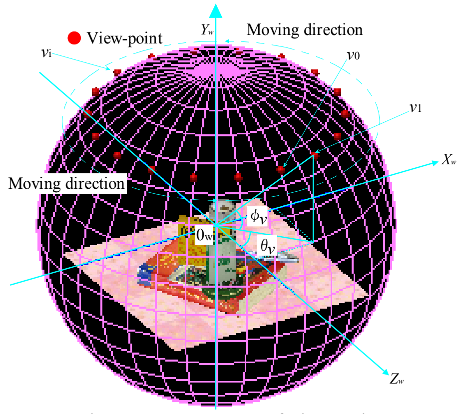


Figure 8. Movement of view-point

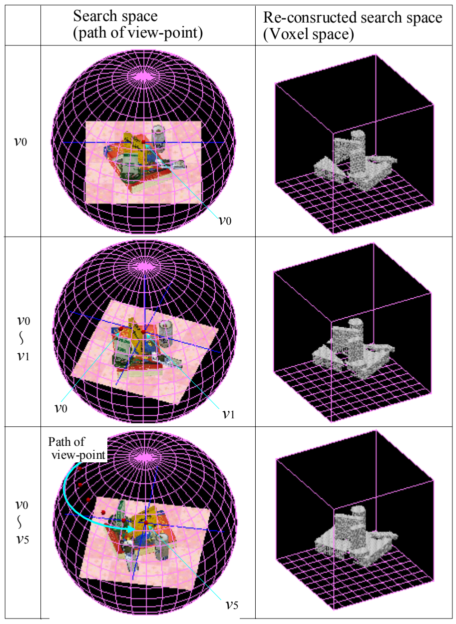


Figure 9. Relationship between movement of view-point and voxel space

In the search space, *breed* points are set up first. Matching processing is performed in parallel with moving a view-point with 20° about θ_v (ϕ_v is kept const = 45°) per five generations as shown in Fig.8. The view-point moves on the wire sphere warping the search space. At this time the direction of view vector is from view-point to zero point. Figure 9 shows the relationship between location of view-point and a re-constructed voxel space. At first, the 3D scene is re-constructed in the voxel space using range information acquirable in the view-point v_0, and the score space is created and matching processing is carried out. Then, at five generations passed, view-point is moved to v_1, and range information acquired from there is added to the established voxel space, and then the score space is newly generated. Of course, voxel data of model object used in matching process is obtained by the same view-point path. This method is more natural and close to the human recognition processing. For example, in case of searching the cookie box, it is impossible to detect the cookie box

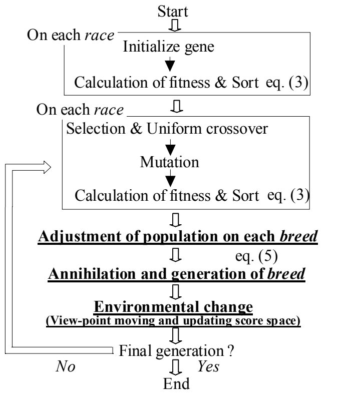


Figure 10. Flowchart

hidden by other objects seen at first view-point, but by moving view-point and getting new view, it is possible to recognize the box with ease.

8.2Matching Processing

Figure 10 shows a flowchart of GA operations. The notable point is updating score space in process of search. This means introduction of the environmental change in GA and change of evaluation space. That is to say, the *breed* with low average fitness value in early stages of search, because 3D information is not enough, might flourish (=population increasing) with updating score space arising from movement of a view-point. Conversely, the *breed* with high average fitness value in early stages of search might decline relatively. Moreover, a *breed* with low degree of fitness average in comparison with other *breeds* is annihilated until the view-point moves and the score space updates. As mentioned above, this means introduction the operation of "*Competitive coexistence*" to general GA. Using this new operation, it is possible to perform moving view-point and matching a model object in parallel. In addition, the various parameters used for the experiment are shown in table 1.

Table 1. Search conditions

Item	Value	Item	Value
○Search space	10.000^3 view volume	○ Updating viewpoint	20° each 5 generation
○ Voxel space	128^3 voxels	○ Adjustment of population	each 5 generations
○ Score space	128^3 voxels	○Translation range (x,y,z)	-0.625～ 0.625
○ L	8	○ Rotation range	
○ *Breed*	64	▹ ϕ	$-\pi/2$～$\pi/2$
○ Max.population of *the world*	1920	▹ θ	0～2π
○Population of *breed*	10～100	▹ ϕ	0～2π
○Max.generation	100	○ Chromosome length	39 bits
○ Selection rate	0.9	▹ (x,y,z)	4*3 bits
○Crossover rate	0.5	▹ (ϕ,θ,ϕ)	9*3 bits
○ Mutation rate	0.03		

8.3Result

A search result of the cigarette box is shown in Fig.11 and Fig.12 as an example. In early stages of search, *breeds* belong to false domain (=extra aria of the cigarette box) increase fitness value because the information about cigarette box is lacking, but accumulation of 3D information with moving view-point brings high fitness value to the right *breed* afterward.

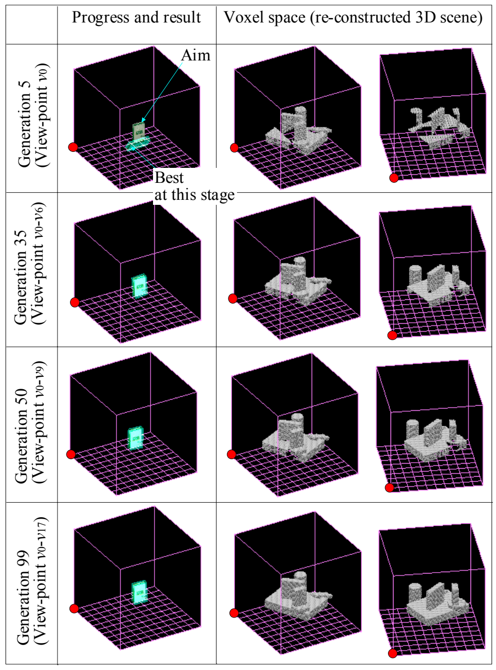
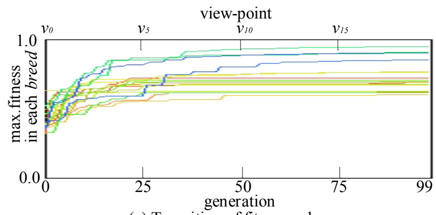


Figure 11. Search result (Cigarette box)



(a) Transition of fitness value

	x	y	z	ϕ°	θ°	ϕ°
aim	-1.100	-0.300	2.300	90.000	357.000	0.000
result	-1.083	-0.333	2.333	88.027	356.686	0.000

Max.fitness *breed*'s population in the final generation =38 (first value=30)

(b) Search result of the cigarette box

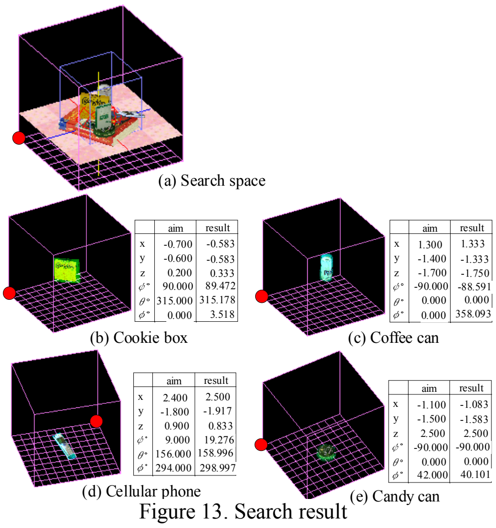
Figure 12. Data of search result(about fig.11)

And search results of other object are shown in Fig.13. The search experiment about each object was carried out serially, and the good results were obtained in almost all the cases. From these figures, it is confirmed that proposal algorithms are effective for the robot vision system.

9.Conclusion and Future Works

In this paper, the external world recognition system of robots coexisting with humans is discussed. And as introductory research, algorithms for detecting 3D object using range and color texture information are investigated.

In order to construct more robust detection system, GA that was added two ideas called "*Breed*" and "*Competitive coexistence*" was applied to the model-based matching. GA with these ideas was able to cope with the multi-peak problem and change of environment. So this algorithm is suitable for an active robot vision. Using the proposal color score space permits that shape and color texture of the



(a) Search space

(b) Cookie box

	aim	result
x	-0.700	-0.583
y	-0.600	-0.583
z	0.200	0.333
φ°	90.000	89.472
θ°	315.000	315.178
φ°	0.000	3.518

(c) Coffee can

	aim	result
x	1.300	1.333
y	-1.400	-1.333
z	-1.700	-1.750
φ°	-90.000	-88.591
θ°	0.000	0.000
φ°	0.000	358.093

(d) Cellular phone

	aim	result
x	2.400	2.500
y	-1.800	-1.917
z	0.900	0.833
φ°	9.000	19.276
θ°	156.000	158.996
φ°	294.000	298.997

(e) Candy can

	aim	result
x	-1.100	-1.083
y	-1.500	-1.583
z	2.500	2.500
φ°	-90.000	-90.000
θ°	0.000	0.000
φ°	42.000	40.101

Figure 13. Search result

model object are evaluated at the same time.

Finally, future woks are shown as follows:

(1) More quantitative evaluation of the proposal system.
(2) Optimization of GA parameters.
(3) Examination of an installation method of the *breed* point.
(4) Optimization of the strategy of view-point movement in an active vision.
(5) Application in a real image.

The technique stated in this paper is only the introductory part of research of a robust robot vision. Therefore, it is due to continue the study.

Acknowledgments

A part of this work is supported by “Yazaki Memorial Foundation for Science and Technology” and “The Hori Information and science promotion foundation”.

References

[1] K.sugihara, “Some Location Problem for Robot Navigation Using a Single Camera,” Proc. Graphics and Image, 43, 112, 1988.

[2] S.Kotani, H.Mori and N.Kiyohiro, “Development of the Robotic Travel aid “HITOMI”,” Robotics and Autonomous System, 17, 119, 1996.

[3] K.Nakano, S.Kanno and Y.Watanabe, “Recognition Assembly Parts Using Geometric Models,” Bull. Japan Society of Precision Engg., Vol.24, No.4, 1990.

[4] N.Nishikawa, M.Onishi, T.Matsumoto, M.Izumi and K.Fukunaga, “Object Recognition Based on Camera Control,” T.IEE Japan, Vol.118-C, No.2, pp.210-216, 1998.

[5] A.J.Sanchez and J.M.Maritinez, “Robot-arm Pick Place Behavior Programming System Using Visual Perception,” Porc. 15th ICPR, Vol.4, pp.507-510, 2000.

[6] S.Aoyama, H.Kinomoto, S.Ieuji, M.Takano, “Recognition and Handling of Tableware by Using a Robot Based on RECS Concept,” IEEJ(Japan), Vol.120-C,No.5,pp.615-623, 2000

[7] M.Oshima and Y.Shirai, “Object recognition using three-dimensional information,” IEEE Trans., Vol.PAMI-5, No.4, pp.353-361, 1983.

[8] T.Ozaki, K.Sato and S.Inokuchi, “Rule-driven processing recognition from range image,” Proc. 9th ICPR, pp.804-807, 1988.

[9] S.Furukawa, S.Mukai and M.Kuroda, “A Convex Hull Algorithm and Its Application to Shape Comparison of 3D Objects,” Proc. IFIP Conference on Computer Applications in Production and Engineering, 5, pp.111-120, 1996.

[10] B.J.A.Kruse and R.Bushoten, “Probabilistic localization by appearnce models and active vision,” Proc. IEEE ICRA, pp.2255-2260, 1999.

[11] Y.Mae, T.Umetani, T.Arai and K.Inoue, “Object Recognition using Appearance Models Accumulated into Environment,” Proc. 15th ICPR, Vol.4, pp845-848, 2000.

[12] H.Kawanishi and M.Habiwara, “A Shape Detection Method Using Improved Genetic Algorithm,” IEEE International Conf. on Syst. Man and Cybern., p235, 1995.

[13] L.Davis, “Handbook of Genetic Algorithms,” Van Nostrand Reinhold, 1991.

[14] R.C.Bolles, P.Horaud and M.J.Hannah, “3DPO: A three-diensional part orientation system,” Proc. 8th IJCAI, pp.1116-1112, 1983.

[15] T.Fujita, K.Sato and S.Inokuchi, “Range image processing for bin-picking of curved object,” Proc. IAPR Workshop on CV, pp.451-545, 1988.

[16] T.Yomo, About “Comettive coexistence”, http://www.bio.eng.osaka-u.ac.jp/urabe/mainnormal.html

[17] S.Jean ed, “Image Analysis and Mathematical Morphology,2,” Theoretical Advances, Academic Press, 1988.

[18] Giardina, Charles R. and Dougherty, Edward R., “Morphological Methods in Image and Signal Processing,” Prentice Hall, Englewood Cliffs, 1988.

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

[20] F.Toyama, K.Shoji and J.Miyamichi, “A New Fitness Function for Shape Detection Using Genetic Algorithm,” ICICE(Japan), D-II, Vol.J82-D-II, No.8, pp.1262-1268, 1999.