Recognition of multiple objects based on global image consistency

Manabu HASHIMOTO Kazuhiko SUMI Teruo USAMI
Industrial Electronics & Systems Laboratory,
Mitsubishi Electric Corp.

Shuji NAKATA
Department of Manufacturing Science, Osaka University

Abstract

In this paper, we describe an algorithm designed to recognize a large number of objects simultaneously. When many objects are in contact with each other, the conventional model-matching method frequently causes false results because of the difficulty involved in segmenting objects one by one. We propose to solve this problem by employing global consistency between a scene and an acquired image instead of model-to-image consistency. The algorithm is based on a hypothesis generation and verification strategy. The solution is obtained by selecting the most rational hypothesis of the scene from the generated hypotheses. We have employed global consistency as the criterion to estimate rationality. We have also applied a Genetic algorithm as a search method for high-speed processing. Experiments indicate that the algorithm is practical for robot tasks.

1 Introduction

In shipment stations of warehouses and factories, automated robot systems have been needed in order to handle loads. However, to realize such a system, it is necessary to develop a vision system which recognizes randomly stacked objects. In this paper, we describe an algorithm capable of recognizing multiple hexahedral objects simultaneously. The main features of this situation are as follows:

1. Objects frequently are in contact with neighboring objects.
2. A variety of textures and designs can be found on the surface of objects.
3. Processing speed must be reasonably fast for utilization in a practical system.

An example of the three-dimensional scene which we are concerned with in this research is illustrated in Figure 1. A large number of hexahedral objects are stacked randomly and they are in contact with each other. Various characters and designs are often printed on the surface of the objects.

BMVC 1999 doi:10.5244/C.13.15

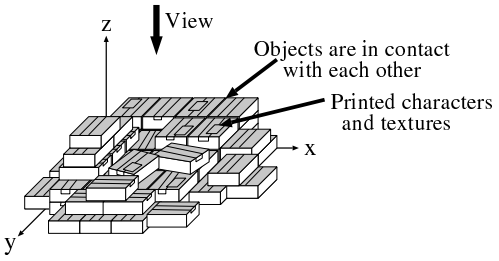


Figure 1: An example of a 3-D scene consisting of many hexahedrons.

Mainly two 3-D recognition methods have been studied[1]-[6]: the model search and a method that is based on the "hypothesis generation and verification" strategy, which is an extended version of the model search. Various aspects of images such as range, contour, or color were employed in these studies to verify hypotheses (i.e., objects). All of these approaches have adopted model-to-image matching in order to evaluate their rationality. Each object hypothesis is matched to a model and the degree of similarity is calculated. The model-matching is certainly useful; however, it is difficult to detect whether an object is an actual one or not when it is in contact with neighboring objects. The reason is that even false objects can appear to be extremely similar to the model because they cannot be properly segmented from other objects. As mentioned above, this problem frequently occurs with objects which are in contact with neighboring objects, so this problem is very important.

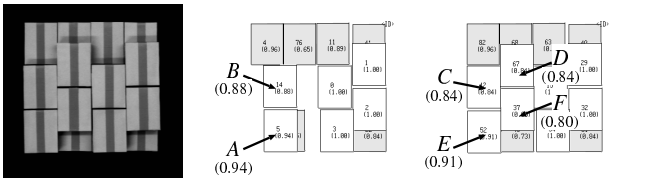
The basic idea we would like to propose in order to solve this problem is to employ global consistency instead of model-to-image consistency. The hypothesis is defined as a scene, not an object, and global consistency is the similarity calculated from a global viewpoint between a scene and an acquired image. Although the true scene is completely consistent with the acquired image, false scenes are expected to be inconsistent. Therefore, by calculating global consistency, we can evaluate the rationality of scene hypotheses. We propose a new recognition algorithm by applying this idea to a hypothesis generation and verification strategy.

The algorithm we propose here consists of three steps. The first step is to extract object candidates from an image, the second step is to generate scene hypotheses by combining the candidates, and the final step is to determine the most consistent hypothesis with the acquired image. A range image is used in the final step, because it is robust in varying lighting conditions. In order to achieve practical processing speed, we also propose an efficient search method using a Genetic algorithm.

In Section 2, we describe problems with the conventional model-matching method and the advantages of global image matching. In Section 3, an object recognition algorithm using global consistency is introduced. In Section 4, several experimental results are shown and the practicality of the algorithm is demonstrated. The final section provides a summary of the paper.

2 Model matching and global image matching

Figure 2 shows the recognition results achieved by template matching[8], a kind of conventional model-matching method. (a) is an original image, (b) is a mis-recognized result, and (c) is the true result which was generated artificially. The numbers in parentheses indicate the similarity factor of each object region which was calculated by matching to the model data. Although both A and B are false, they are significantly similar to the model. On the other hand, objects C, D, E, and F are true, but their degree of similarity to the model is lower than that of the false objects. This method uses contour images; however, even if gray scale or range images are employed, this problem cannot be solved as long as model-matching is used due to unreliable consistency.



(a) Original image (b) False results by model-matching. (c) Artificial true results

Figure 2: False recognition by the model-matching method and artificial true results. White rectangles are in the upper layer, and gray are in the lower layer.

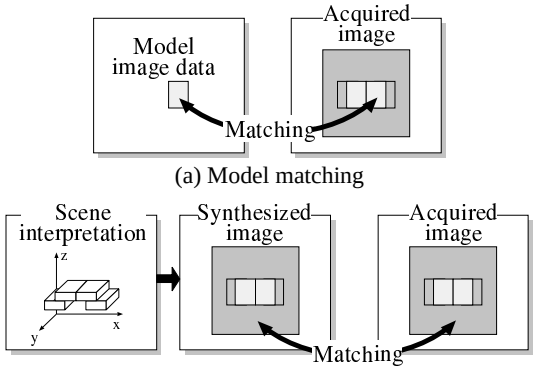


Figure 3: Model-matching and global image matching.

Investigations of why mis-recognition occurs have found that the false objects have high model consistency as a result of parts of neighboring objects being included as part of their image. However, when this scene is analyzed from a global viewpoint, the true scene is completely consistent with the acquired image

and the false scene is inconsistent. Therefore, we are proposing a new object recognition algorithm based on global image matching of a scene hypothesis to an input image. A schematic diagram of the differences between model-matching and global image matching is illustrated in Figure 3.

3 A recognition algorithm based on global image consistency

In Section 3, we will describe the object recognition algorithm which is based on the ideas presented in Section 2.

3.1 The generation and verification of a scene hypothesis

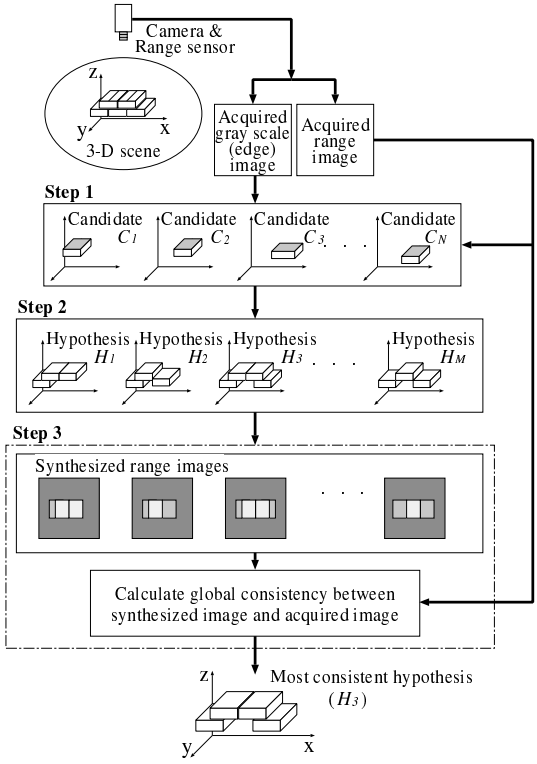


Figure 4: Object recognition algorithm.

The algorithm we propose is shown in Figure 4. The strategy we have utilized consists of the following three steps.

Step 1 (The Extraction of object candidates): First, a large number of straight lines are detected by the Hough transformation method[7] using an edge image. In this way, the boundaries of objects can be easily detected even if they

are in contact with one another. Next, rectangular patterns are extracted by the contour template matching[8] using the crossing points of straight lines. These patterns are stored as the object candidates $C_i(i = 1, 2, \ldots N)$ in 3-D space. The candidates include not only true objects but also false objects.

Step 2 (The generation of a scene hypotheses): By combining several candidates, 3-D scene hypotheses are generated and stored. For example, M scenes $H_i(i = 1, 2, \ldots M)$ are hypothesized from N candidates. Any hypothesis which includes a combination of spatially interfering objects is rejected.

Step 3 (Selection of the best hypothesis): This final step determines which hypothesis is the most rational one. Range images for each hypothesis are synthesized using a 3-D object model. Following this, each image is compared with the acquired range image. The range image is acquired by employing stereo vision with optical pattern projection[9]. It can measure the range of various objects including those with a non-textured surface.

This algorithm is categorized as a hypothesis generation and verification strategy. The approach is based on two ideas. First, the detected object in the first step is an element which constructs the scene hypothesis because it cannot be proved true or false by itself. Second, the hypothesis which has the least inconsistency is the most rational interpretation of the scene.

3.2 Hypothesis verification using global image consistency

The global image consistency S_G is defined as follows:

$$S_G = \frac{1}{n \cdot m} \sum_{j=1}^{m} \sum_{i=1}^{n} f_G(I_i(i,j), I_h(i,j)) \tag{1}$$

$$f_G = \begin{cases} 1 & (\text{ if } |I_i(i,j) - I_h(i,j)| < \varepsilon \ ) \\ 0 & (\text{else}) \end{cases} \tag{2}$$

Where I_i is the acquired range image, I_h is the synthesized range image, and each hypothesis is seen from same viewpoint of I_i. n and m are the size of the region used for image matching. ε is a threshold value. Here, S_G takes the value of 1 when I_h is matched to I_i completely.

Figure 5 illustrates two scene hypotheses, H_1 is "true" and H_2 is "false". H_1 consists of two objects (C_1 and C_2). H_2 has an object (C_3). Only C_3 is false. Here, consistency as calculated by the model-matching method is high for both the true and the false hypothesis. However, if global consistency S_G^1 and S_G^2 are calculated, S_G^2 is found to be lower than S_G^1 because the unmatched region occurs only in the false hypothesis. Therefore, we can verify which hypothesis is the most appropriate interpretation of this scene.

In order to obtain a solution with this method, the correct solution must be present in a set of hypotheses. Thus, the following conditions must be met:

1. The models of all objects in the scene are known.

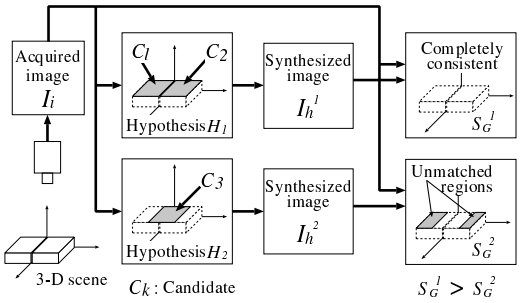


Figure 5: Verification of a scene hypothesis using global image consistency.

2. All visible objects in the scene have to be extracted as candidates even if they are partially occluded by other objects.

Condition 1 can be satisfied in practical applications. Condition 2 is satisfied by using an algorithm which can extract occluded objects from an image, such as contour template matching[8].

3.3 Efficient hypothesis generation and verification utilizing a Genetic algorithm

In the method we have proposed in Section **3.1**, a considerable amount of time is required to find a solution when there is a large number of object candidates because a large number of hypotheses must be verified; therefore, a high-speed algorithm is required. This problem is regarded as a kind of labeling procedure which gives true or false labels to each object candidate. Therefore, we have employed the efficient labeling algorithm which we proposed[10], using a Genetic algorithm (GA)[11].

In the first step of the GA, a chromosome which represents a solution is defined. Then, two good parent chromosomes are selected from the initial set of chromosomes, and new chromosomes are produced by combining parts of the parents using genetic operations. Such a reproduction is iterated in the alternation of generations, and finally the best chromosome is determined.

Figure 6 shows the definition of a chromosome in the GA. The chromosome $A_k(k = 1, 2, \ldots, N)$ is a binary string which has N bits. N is the number of object candidates. Each bit represents the existence of an object candidate. It takes a value of 1 when the candidate truly exists in a scene, and 0 when it is a false object. A 3-D scene can be reconstructed by combining objects with a value of 1.

Mutations and a crossover are used as genetic operations. Figure 7 shows how the crossover works in the algorithm. As this figure shows, the crossover generates a new scene hypothesis (child a) by combining partial scene interpretation P and Q in parents A and B. By clearly relating such a genetic operation to the recognition process, we can effectively utilize the ability to search. The global image consistency factor S_G is used as the fitness value.

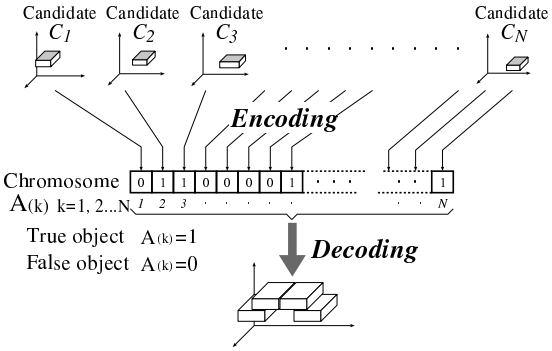


Figure 6: Definition of a chromosome string.

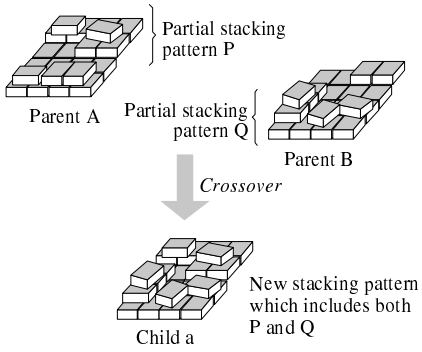


Figure 7: Illustration of how the crossover works in the method.

4 Experiments and discussion

4.1 Recognition reliability

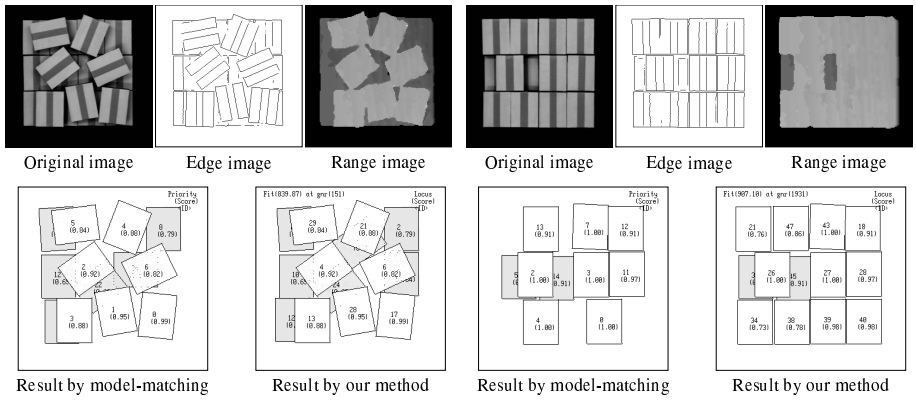
The reliability of the algorithm was examined through 120 real images, which include 1183 hexahedral objects as a total. Objects are stacked in 2 or 3 layers. The false recognition rate and missing rate were calculated, and the results are shown in Table 1. The false recognition rate is 0.9%, and the missing rate is 0.4%. These rates are low enough for use in a practical recognition system. In this case, the processing time of a 300MHz Pentium-II computer is approximately 3 seconds per object. A pick-and-place motion of usual big scale robot for unloading loads takes about 12 seconds, so this recognition time is within practical range.

Two examples of the recognition results are shown in Figure 8. White and gray rectangles indicate the upper and lower layer objects respectively. This experiment shows that the algorithm proposed in this study is better than the model-matching method, even if objects are stacked closely to one another. As mentioned in

Table 1: Recognition error rate. (Total: 120 images)

	Number of objects	Number of mis-recognized objects	
		False recognition	Missing
Top layer	980	2 (0.2%)	5 (0.5%)
Other layer	203	8 (3.9%)	0 (0.0%)
Total	1183	10 (0.9%)	5 (0.4%)

Sec.2, model-matching method can recognize only when objects are arranged in sparsely. Our method is effective not only in such situation but also when objects are arranged closely contacted to each other.



(a) Sparsely placed objects (b) Closely contacted objects

Figure 8: Results of comparing our method with the conventional model-matching method.

4.2 Behavior of the genetic algorithm

In this section, we have analyzed the process of the alternation of generations in order to examine the searching behavior of the Genetic algorithm. Figure 9 shows a tested image and a history of the fitness value in the alternation of generations. The fitness f in this figure is $1000 \cdot S_G$. The number of initial chromosomes is 250, the crossover ratio is 0.1, and the mutation ratio is 0.1. As this figure shows, the mean fitness value in the chromosome pool increases gradually, but the maximum fitness value increases discretely. This step by step change shows that a superior chromosome was contingently generated by the genetic operations. Points (a) to (g) in Figure 9 are changing points.

Figure 10 shows the reproduction process at point (c). The images are gener-

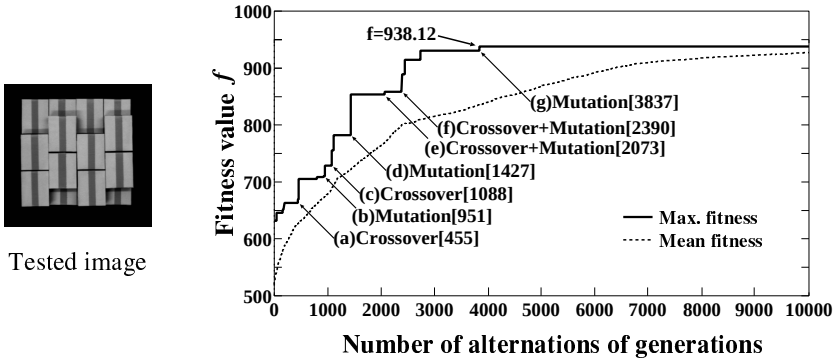


Figure 9: History of the fitness value in the alternations of generations.

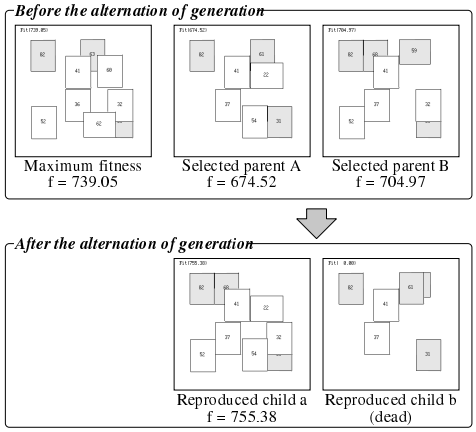


Figure 10: An example of behavior of the GA (generation = 1088).

ated by decoding chromosomes. At generation 1088, point (c), a child a which has a fitness of 755.38, was produced by a crossover operation from parents A (fitness = 674.52) and B (fitness = 704.97). The fitness of a is larger than that of his parents. Another child b has been eliminated as a lethal chromosome, because it includes spatial interference with other objects. At generation 1427, point (d), a new child which has a larger fitness value appeared by mutation. At the point (g) of generation 3837, the best child was produced by mutation. This child was the final result of the experiment.

Through the above analysis, it has been shown that the hypothesis generation and verification method using the genetic algorithm is an effective approach. The reason why the final fitness of 938.12 is less than 1000, the theoretical maximum, is due to the geometric error of the object model and its alignment error.

5 Conclusions

This paper has presented a new object recognition algorithm based on the estimation of global consistency between a scene hypothesis and an acquired image. This method can recognize a large number of objects simultaneously even if segmentation is difficult as a result of some objects being in contact with neighboring objects. To reduce the computational cost of the hypothesis generation and verification process, we applied a genetic algorithm to the search problem. Real image tests showed that the algorithm performs well enough for utilization in practical robot vision systems. In future studies, we will expand this algorithm to more complicated scenes which include objects with a variety of shapes.

References

[1] Bolles, R. C. et al., "3DPO: A Three-dimensional part orientation system," The International Journal of Robotics Research, vol.5, no.3, pp.3-26, 1986.

[2] Grimson, W. E. L. et al., "Localizing Overlapping Parts by Searching the Interpretation Tree," IEEE Trans. on PAMI, vol.9, no.4, pp.469-482, 1987.

[3] Jain, A. K. and Hoffman, R., "Evidence-based recognition of 3-D objects," IEEE Trans. on PAMI, vol.10, no.6, pp.783-802, Nov. 1988.

[4] Flynn, P. J. and Jain, A. K., "BONSAI: 3-D Object recognition using constrained search," Proceedings of the 3rd ICCV, pp.263-267, 1990.

[5] Wheeler, M. D. and Ikeuchi, K., "Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition," IEEE Trans on PAMI, vol.17, no.3, pp.252-265, March 1995.

[6] Yi, J. H., "Model-based 3D object recognition using Bayesian indexing," Computer Vision and Image Understanding, vol.69, no.1, pp.87-105, Jan. 1998.

[7] Duda, R.O. and Hart, P.E., "Use of the Hough transformation to detect lines and curves in pictures," Trans. Comm. ACM, vol.15, no.1, pp.11–15, 1972.

[8] Hashimoto, M., Sumi, K. and Kawato, S., "High speed template matching algorithm using contour information," Proceedings of SPIE Symposium on Electronic Imaging & Science and Technology, vol. 1657, pp.374-385, 1992.

[9] Hashimoto, M., Sumi, K., and Kuroda, S.,S. Kuroda, "Loads recognition by fusion of rough depth image and edge information from intensity image," Proceedings of the 3rd Symposium on Sensing via Image Information, vol. H-3, pp.353-358, June 1997 (in Japanese).

[10] Hashimoto, M., Sumi, K. and Kuroda, S., "Vision System for Depalletizing Robot using Genetic Labeling," IEICE Trans. Inf. & Syst., vol.E78-D, no.12, pp.1552-1558, Dec. 1995.

[11] Goldberg, D. E., "Genetic Algorithms in Search, Optimization, and Machine Learning," Addison Wesley, 1989.