

# 3-D Object Recognition Based on Integration of Range Image and Gray-scale Image

Manabu Hashimoto and Kazuhiko Sumi  
Industrial Electronics and Systems Laboratory  
Mitsubishi Electric Corporation  
Amagasaki-city, 661-8661 Japan  
hasimoto@sen.sdl.melco.co.jp

## Abstract

In this paper, we propose a practical object recognition system which consists of two functional modules. The first is object extraction module using a range image, and the second is a precise position measurement module using a gray-scale intensity image. Both high-reliability and high-accuracy can be achieved by effective image integration. We also propose an idea of stereo vision with random-dot pattern projection as an effective way to obtain a range image. This method enables reliable stereo matching, even for objects with no texture. Through an experiment with real images, we have demonstrated that our system has 99.8% recognition reliability and processing time is approximately 5 seconds per image; as a result, the system can be applied to practical industrial robot vision.

## 1 Introduction

Recently, in the field of factory automation, intelligent robot systems are needed for the automated handling of stacked loads at warehouses or shipment stations. Because loads are placed randomly on pallets, a three-dimensional vision system is indispensable if robots are to recognize the position of each individual object.

As for 3-D object recognition system, two types of methods have been studied. The first is based on an image analysis using gray-scale images [1-3]. The second method is based on range images [4-6]. The method based on gray-scale images requires only inexpensive CCD camera(s) that are capable of providing high-resolution images. Therefore, accurate image measurement with simple hardware is the greatest advantage of this method. However, when actual complex scenes, which consist of multiple objects, such as stacked loads, are encountered, it is difficult to match 2-D images to object models because the system faces difficulties in segmenting one object from the background or other objects.

In contrast, the method based on range images can measure the 3-D shapes of objects directly, so recognition reliability is higher than that of the gray-scale image based method. One drawback to this method is that in order to acquire dense range images, a laser range finder [7, 8] that uses an electronic LCD shutter or a laser scan range finder is required. Since this equipment contains optical devices, such as LCD shutters or laser scanning mirrors for projecting variable light patterns, the hardware structure is both complicated and expensive. In addition, range finders often take a long time to acquire data compared with normal CCD cameras. This is a major reason that 3-D sensors have not been used in various practical applications. Recently, an integration method that uses both gray-scale image and range image has been studied [9-12]. However, a system that satisfies the need for object segmentation, precise positioning, and simple hardware configuration, has not been proposed yet.

In this paper, we propose a recognition method for practical, reliable, and accurate 3-D robot vision system that can be operated with simple equipment. The algorithm we propose consists of two recognition modules, a range image processing module and a gray-scale image processing module. The role of the range image processing module is to extract object candidates and detect their rough position. The role of the gray-scale image processing module is to determine whether each candidate has been correctly identified and to measure its position precisely.

In the object extraction, the stable segmentation of each object is made possible by utilizing the discontinuity of range data of object edges. Because the accuracy of the positioning is guaranteed by the gray-scale image processing finally, a low-resolution range image is adequate for this purpose.

On the other hand, in the object positioning, the algorithm searches for the precise position of objects using a contour model in an edge image that has been generated from the gray-scale image. Because the rough position of each object has been detected by the range image processing module, it is possible to limit the region of pattern search. Such a pattern search in a limited region enables the suppression of mismatches by false edges caused by printed characters or designs on the object. As a result, processing time is reduced and reliable pattern matching is achieved.

In order to acquire a low-resolution range image efficiently, we also propose a new stereo vision system. In this system, an optical pattern projector, which is equipped between two stereo cameras, projects a random-dot pattern onto the surface of a group of stacked objects. This random-dot pattern makes it easy to search for corresponding patterns between the stereo images. Because the aim of projecting the random-dot pattern is to place a unique texture onto the object, the projection pattern does not have to be changed when taking measurements. Therefore, we can realize the pattern projector with very simple hardware.

In Section 2, the vision system for an object handling robot is described, and a basic approach to a recognition algorithm, which consists of a range image processing module and a gray-scale image processing module, is proposed. In Section 3, as a range image processing module, the stereo vision with random-dot projection and an object extraction method are proposed. Following this, as a gray-scale image processing module, a precise object positioning method using the contour pattern matching technique is described in Section 4. In Section 5, through two experiments using real images, the performance of a robot vision system is introduced. The final section provides a summary of this paper.

## 2 Object Recognition by Image Integration

### 2.1 Subjects of Vision System for Object Handling Robot

Figure 1 shows an example of stacked loads on a “pallet”. As shown in this figure, the objects to be recognized are box-shaped or bag-shaped. Their sizes are approximately 0.1 to 1 meter, and the objects are stacked randomly on a pallet. There is one kind of object in each scene, and 5 to 10 objects are in each layer. In order to realize automated object handling robot, two translation parameters in an x-y plane and a rotation parameter around the z-axis must be measured.

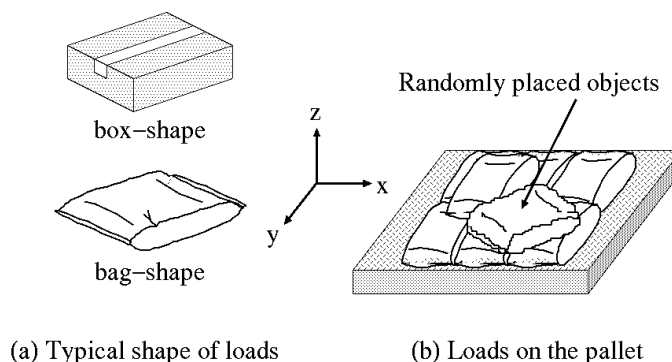


Figure 1. An example of stacked loads.

The optical equipment used for image acquisition must be placed out of range so that it cannot be reached by a robot hand. In this research, we set the optical equipment approximately 3 meters above the floor in order to avoid interference with the robot hand. The required measurement accuracy for robot handling is 5mm along x-y direction in the case of box-shaped objects. However, the required accuracy for range measurement is 5cm, because the vacuum robot hand that picks up an object has 5cm margins.

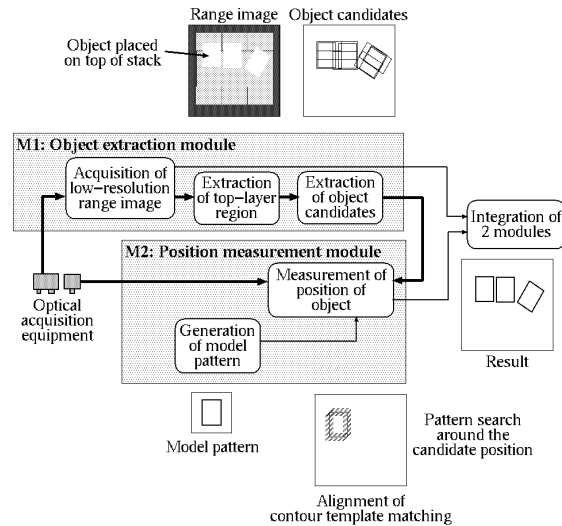
### 2.2 The Structure of the Recognition Algorithm

Figure 2 shows the module structure of the recognition system we propose. This system consists of two functional modules; one is an object extraction module using a range image (M1), and the other is a position measurement module using a gray-scale image (M2).

First, a range image and a gray-scale image are acquired, the objects of the highest layer are extracted and stored by the module M1, and their rough position is detected. Since both true and false objects may be extracted, they are stored as object candidates. The resolution depends on the object's size, required accuracy of rough object extraction, and size of field of view. In this research, we have determined that resolution of the range image for stable object extraction is 150 by 150 pixels. In the range image, a depth discontinuity of the outline of object candidates appears clearly, so the range image enables the stable segmentation and extraction of objects to take place.

Next, an edge image is generated from the gray-scale image in the position measurement module (M2). The contour model of the object is then searched for in the

edge image using the template matching technique. In this way, the precise position of the object can be detected. Here, the model pattern is searched in a limited area near the rough candidate point detected by the module M1. Because it is possible to suppress the influence of false edge patterns caused by printed characters or textures on the surfaces of objects, the computational cost of the template matching can be reduced, and reliable positioning can be realized. A VGA resolution CCD camera can be utilized as the input device to obtain the gray-scale image.



**Figure 2. Module structure of the recognition system using a range image and a gray-scale image.**

## 3 Object Extraction using a Range Image

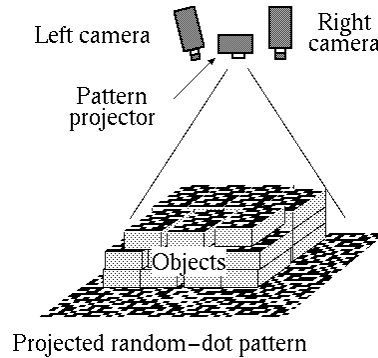
### 3.1 Stereo Vision with Random-dot Pattern Projection

We have chosen to use the binocular stereo method for obtaining a range image because of its simple hardware configuration. In addition, the block matching technique has been adopted in order to determine the corresponding points in a search space. This technique is appropriate because the distribution of range data in the matching-blocks are expected to be uniform because the upper surfaces of the objects are planar, and the block matching process is simple and suitable for high-speed processing.

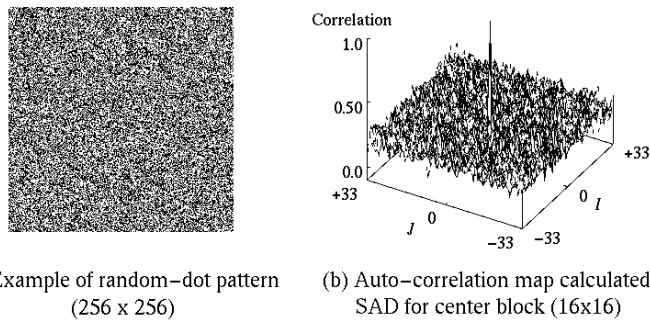
One requirement of the stereo algorithm is high-contrast texture on the object surfaces that allows for a reliable correspondence between the images. However, both textured and no-textured objects are found in actual material handling situations. To overcome this problem, we propose the adoption of stereo vision with a pattern projection equipment that projects a random-dot optical pattern onto the object as shown in Figure 3. The primary advantage of projecting a random-dot pattern is that it is stochastically unique everywhere, and the corresponding points can be detected reliably.

A variety of optical patterns has been proposed in past studies. For example, random size circle pattern [13], binary checker-pattern [14], spatial sine modulated pattern [15], and colour coding pattern [16] have been studied, but these methods require the

projected pattern to be extracted via pre-processing or a complicated pattern projector to generate specified optical patterns. In our method, this kind of pre-processing is unnecessary and the projector used for generating a fixed pattern can be provided easily. Figure 4 shows an example of a random-dot pattern (256x256 pixels) used in our system and a distribution map of auto-correlation values calculated using a 16x16 pixels region in the central part of the pattern. As shown in this figure, the auto-correlation value is very small except in the origin of the map, and it demonstrates that this random-dot is statistically unique.



**Figure 3. Stereo vision with random-dot pattern projection.**



**Figure 4. Auto-correlation map of designed random-dot pattern (256x256).**

We can summarize the advantages of our method as follows.

- (1) The projection of a random-dot pattern makes it possible to achieve reliable stereo matching even when no texture is present on the surfaces of objects.
- (2) The random-dot pattern projector is simple because it is not necessary to change patterns when making measurement.
- (3) The block matching technique for calculating disparities can be carried out with low computational cost.

### 3.2 Object Extraction from a Range Image

Object candidates are extracted with a template matching technique using a binarized range image. The template describing the shape of an object from the top view can be generated automatically. In order to allow for perturbations in the size of objects, additional templates, whose size has been changed 1 or 2% from the original size, have been prepared.

## 4 Precise Positioning using a Gray-scale Image

In this section, we will describe the precise object positioning method using a gray-scale image. For each candidate extracted from the range image, the precise position is measured with the contour template matching technique [17] using a high-resolution(VGA) gray-scale image. A contour template matching algorithm is shown in Figure 5.

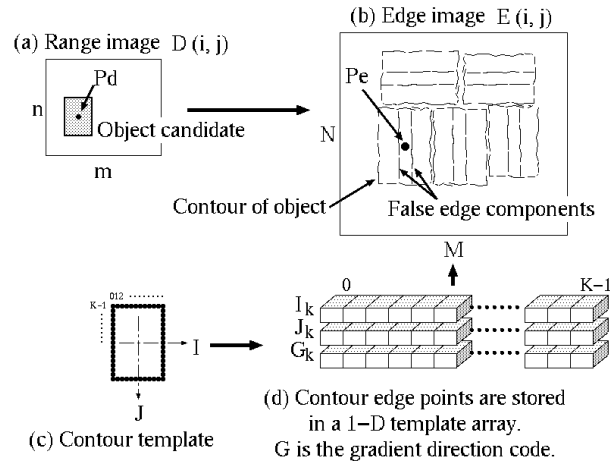


Figure 5. The contour template matching method using an edge image.

In this figure, the contour template (d) has a one-dimensional data structure  $(I_k, J_k, G_k)$ , which stores contour point information (c) about object shape.  $I_k, J_k$  are the coordinates of  $K$  contour points, and  $G_k$  is the direction code number which indicates the direction of the maximum gradient of each pixel as eight numbers defined every 45 degrees. The contour template can be generated automatically using geometrical data of the object model.

An edge image,  $E(i, j)$  is generated by dilating original contour pixels detected from the acquired gray-scale image in order to allow the perturbation of the size or shape of objects. In addition,  $E$  provides the same information about the gradient direction as the template data.  $P_e = (i_e, j_e)$  is the start point of the pattern search, which is calculated using the position of object candidates  $P_d$ . The best match point is detected as the maximum value in a similarity map  $S(d_x, d_y, d_\theta)$ , which is calculated by moving the template on the edge image.  $S$  is defined as equations (1) and (2). Where,  $d_x, d_y, d_\theta$  are parameters of the matching position in translation and rotation, and  $\delta$  indicates a similarity factor of the direction code between the template and the edge image.  $\delta$  takes a value of 1 when the direction code is similar, and a value of 0 when it is dissimilar.

$$S(d_x, d_y, d_\theta) = \sum_{k=0}^{K-1} E(d_x + i_k, d_y + j_k) \cdot \delta \quad (1)$$

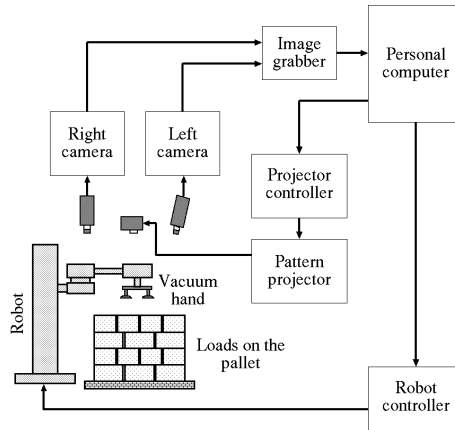
$$\begin{cases} i_k = I_k \cos d_\theta - J_k \sin d_\theta + i_e \\ j_k = I_k \sin d_\theta + J_k \cos d_\theta + j_e \end{cases} \quad (2)$$

As mentioned above, high-speed template matching is achieved by utilizing only contour pixels in a 1-D template. This approach is suitable for accurate pattern alignment because the contours of objects provide useful information for determining precise position. Moreover, it is possible to decide whether each object candidate is true or false by comparing the maximum similarity value arrived at through template matching with an appropriate threshold value. If we attempted to extract object candidates by template matching using a gray-scale image instead of a range image, many false objects would be extracted due to disturbances caused by false edges as shown in Figure 5(b). However, in cases in which the rough position of object candidates has already been detected using another method like our algorithm, the matching reliability can be increased because the area of the pattern search can be limited effectively.

## 5 Experiments and Discussion

### 5.1 Vision System for Object Handling Robots

Figure 6 shows an object handling robot system using a vision system based on the recognition algorithm we have proposed in this paper. The system consists of two cameras, a pattern projector, an image grabber, a personal computer, and a robot system.



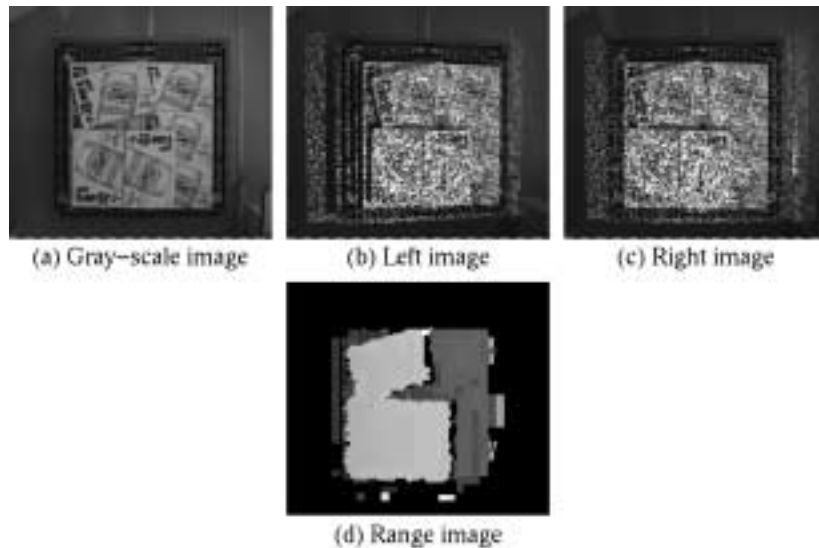
**Figure 6. The structure of a vision system for an object handling robot.**

The pattern projector includes a metal-halide lamp and a glass plate upon which a random-dot pattern is printed. A pair of stereo images is acquired and input into the computer, and a range image is generated by software processing. The camera on the

right is also used to obtain a gray-scale image. Because a range image and a gray-scale image have the same view direction, the coordinate systems between them can be easily transformed. Finally, recognition results are transferred to the robot controller.

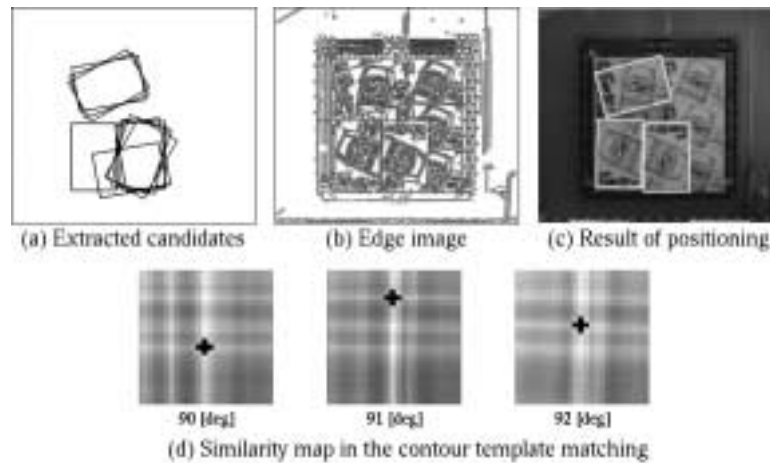
## 5.2 The Estimation of System Performance

Figure 7 shows examples of a gray-scale image, stereo images with random-dot projection, and a range image computed by the stereo images. The size of the matching block in this experiment is 16x16 pixels. There are unstable regions of range data near the slanted contours because the matching block is rectangular. However, this does not present a serious problem for the purpose of the detecting rough position of the candidates. Absolute measurement error was estimated as  $\pm 10.7mm$  in the x-direction,  $\pm 16.1mm$  in the y-direction, and  $\pm 32.8mm$  in z, when the distance between the camera and the object was  $3.2m$ , and the distance between the two cameras was  $665mm$ .



**Figure 7. Examples of input images and a range image.**

The final recognition results for the box-shaped object in Figure 7 are shown in Figure 8. Figure (a) shows all of the object candidates that were extracted by the range image processing module, (b) is an edge image used for precise positioning, and (c) is the recognition results. Three objects on the highest layer were completely recognized. In figure (d), examples of similarity maps generated by contour template matching are shown for  $d_\theta = 90^\circ, 91^\circ, 92^\circ$ . The black cross mark indicates the peak position which takes the maximum value in each similarity map. In this experiment, the final best match point was detected as the maximum value in the case of  $d_\theta = 91^\circ$ , and it was the correct result.



**Figure 8. Recognition results for box-shaped objects.**

Next, recognition reliability was evaluated using 1048 real images. The size of the object was 60x40cm, and it was 15cm thick. We estimated whether the error between the recognized position and the true position was within 3cm or not. If so, it was considered a successful recognition result. In this experiment, the recognition success rate was 99.8%, which is sufficient for industrial use.

In terms of processing time, the experimental results showed that the system took approximately 5 seconds to recognize a scene. In this case, it took 4.2 seconds for generation of a range image, 0.7 seconds for candidate extraction, and 0.1 seconds for accurate positioning using a gray-scale image. Generally, robot motion such as picking and placing tasks takes 10 seconds even if it does not have vision systems. Because 4 seconds of the processing time for object recognition can take place during the robot motion, it can be said that this processing time is short enough for practical application.

## 6 Conclusions

We have proposed a new practical object recognition algorithm that consists of an object extraction module using a range image and a position measurement module using a gray-scale image. By sharing the roles of range image and gray-scale image clearly, a method for obtaining and processing the range image can be simplified. Accurate position measurement is achieved using the contour template matching technique, so final accuracy of the system can be guaranteed.

We also proposed a stereo vision system which involves projecting an optical random-dot pattern for reliable stereo matching. The optical equipment is made up of two cameras and a simple pattern projector. It has been demonstrated through real image tests that the recognition success rate of the system is 99.8% and processing time is approximately 5 seconds. This performance shows that the system can be applied to industrial robot systems as the practical vision system.

## References

- [1] D.G.Lowe, "Visual Recognition from Spatial Correspondence and Perceptual Organization", Proceedings of 9th IJCAI, pp.953-959, 1985.
- [2] M.Hashimoto, K.Sumii, and S.Kuroda, "Vision System for Depalletizing Robot using Genetic Labeling", IEICE Trans.on Inf.and Syst., Vol.E78-D, No.12, pp.1552-1558, 1995.
- [3] J.Edwards and H.Murase, "Appearance Matching of Occluded Objects using Coarse-to-Fine Adaptive Masks", Proceedings of CVPR, pp.533-539, 1997.
- [4] R.C.Bolles and P.Horand, "3DPO: A Three-Dimensional Part Orientation System", International Journal of Robotics Research, Vol.5, No.3, pp.3-26, 1986.
- [5] P.J.Flynn and A.K.Jain, "BONSAI: 3D Object Recognition using Constrained Search", IEEE Trans.on PAMI, Vol.13, No.10, pp.1066-1075, 1991.
- [6] M.D.Wheeler and K.Ikeuchi, "Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition", IEEE Trans.on PAMI, Vol.17, No.3, pp.252-265, 1995.
- [7] K.Sato and S.Inokuchi, "Three-Dimensional Surface Measurement by Space Encoding Range Imaging", Journal of Robotic Systems, Vol.2, No.1, pp.27-39, 1985.
- [8] H.Totani, H.Goto, M.Ikeda and T.Yoda, "Miniature Optical Scanning Range Sensor Utilizing a Silicon Micromachined Scanner", Proceedings of TRANSDUCERS, pp.1275-1278, 1997.
- [9] M.J.Magee, B.A.Boyter, C.H.Chien and J.K.Aggarwal, "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects", IEEE Trans.On PAMI, Vol.7, No.6, pp.629-637, 1985.
- [10] K.Umeda, K.Ikushima and T.Arai, "3D shape recognition by distributed sensing of range images and intensity images", Proceedings of ICRA, pp.149-154, 1997.
- [11] M.R.Stevens, "Localized Scene Interpretation from 3D Models, Range, and Optical Data", Computer Vision and Image Understanding, 80, pp.111-129, 2000.
- [12] E.R.van Dop, and P.P.L.Regtien, "Multi-sensor Recognition of Electronic Components", Machine Vision and Applications, Vol.12, No.5, pp.213-222, 2001.
- [13] G.Xu, H.Kondo, and S.Tsuji, "A Region-Based Stereo Algorithm", Proceedings of IJCAI, pp.1661-1666, 1989.
- [14] H.Hirai, K.Tsutsumi, and Y.Yuba, "3-D Object Profilometry using the Checker-Pattern Projection", Memoirs of the Faculty of Eng.And Design Kyoto Institution of Technology, 1996.
- [15] S.B.Kang, J.A.Webb, C.L.Zitnick, and T.Kanade, "A Multibaseline Stereo System with Active Illumination and Real-time Image Acquisition", Proceedings of ICCV, pp.88-93, 1995.
- [16] A.Koschan, V.Rodenhorst and K.Spiller, "Color Stereo Vision using Hierarchical Block matching and Active Color Illumination", Proceedings of ICPR, pp.835-839, 1996.
- [17] M.Hashimoto, K.Sumii, and S.Kawato, "High Speed Template Matching Algorithm using Contour Information", Proceedings of SPIE Symposium on Electronic Imaging & Science and Technology, Vol.1657, pp.374-385, 1992.