Ob ect Recognition using Smart Tag and Stereo Vision System on Pan-Tilt Mechanism.

Jin- oung Kim, Chang-Jun Im, Sang-Won Lee and Ho-Gil Lee
Sensing & Perception Research Group
Korea Institute of Industrial Technology
Ipjang, Cheonan, ChungNam, South Korea
{kjy, catrinne, last879, leehg}@kitech.re.kr

Abstract: We propose a novel method for object recognition using the smart tag system with a stereo vision on a pan-tilt mechanism. We developed a smart tag which included IRED device. The smart tag is attached onto the object. We also developed a stereo vision system which pans and tilts for the object image to be the centered on each whole image view. A Stereo vision system on the pan-tilt mechanism can map the position of IRED to the robot coordinate system by using pan-tilt angles. And then, to map the size and pose of the object for the robot to coordinate the system, we used a simple model-based vision algorithm. To increase the possibility of tag-based object recognition, we implemented our approach by using as easy and simple techniques as possible.

Keywords: object recognition, tag-based vision, smart tag, pan-tilt vision

1. INTRODUCTION

Object recognition is very important for a robot to carry out it s task. For example, a robot should avoid obstacles, get object sizes, handle objects and so on, in order to execute it s task. Many approaches and methods have been proposed and implemented of which, vision based approaches have been most popular among them. Vision based object recognition however, is still very difficult to implement and is not yet reliable only by using pure image processing. There are three pieces of information which robots should know. These are the geometrical, physical and semantic information of an object. It is difficult for robots to obtain information by using pure image processing and it is also not realistic. For example, a robot has to store all of the object template data dealt with in it s data storage device to compare with the captured object image. Moreover, in order to find that one object corresponds with the captured object image, the robot would spend a lot of time searching. This may be impossible or a waste of resources.

Recently, to solve the above problems, a novel approach has been proposed by using ubiquitous computing technology. Namely, to be getting reliable information, and to be information distributed, tag based object recognition has been hot issued. Asama et.al. proposed it has distributed guidance knowledge management by intelligent data carriers [1][2]. They developed the IC tag which is called IDC. The IDC provides knowledge of guidance for autonomous mobile robots. Ando et.al. proposed that tag-based vision [3][4], which is applied to use the IC tag to localize objects with 3D CAD models, because the RFID technology gives us information about the object. Chong et.el. proposed a knowledge distributed tag-based vision system [5][6] in which every object is posted with an IC tag containing the manufacturer s network address and the knowledge information required for a robot to handle the object concerned.

These new approaches are efficient to distribute information and more reliable to get information directly by the robot communicating with IC tags. But there are some problems to solve for realization of tag-based object recognition. For example, tag-based vision can not give us the positional information of an object. The IDC has been used only to intermediate between the robot and environmental information. knowledge distributed tag-based vision system can give us positional information, but it is necessary to have additional devices besides hand-eye vision system on a robot, such as an ubiquitous vision system and a table equipped with a tag reader.

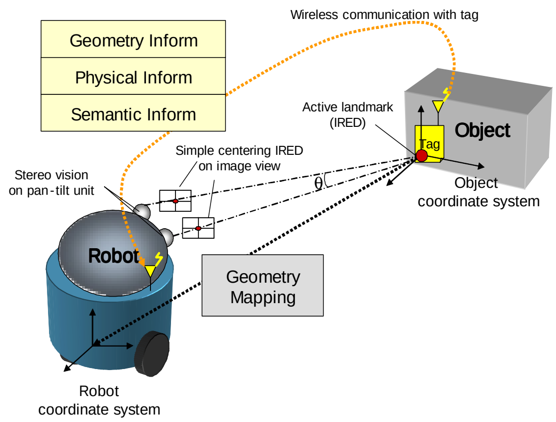

Fig. 1 Concept of our approach

To solve some problems of tag-based object recognition and increase the possibility of the above, we propose another method for realization of tag-based object recognition which is both easier and simpler to implement by using a smart tag with an active landmark and stereo vision on a pan-tilt mechanism.

Fig. 1. illustrates our concept. To recognize an object and then perform a task, the robot needs to know three types of information; the geometrical, physical, and the semantic information about an object. More specifically, the geometrical information of an object is for example, the shape, size, position of reference point and pose. The physical information of an object describes the weight, material, color and strength, and the semantic information of object is whether it is a cup , or a chair . The other kind of semantic information pertains to whether it is movable or fixed and so on. We adopt an IC tag for the robot to get the information, therefore the robot can obtain the information of the object by simply communicating with the IC tag. Generally, the robot initially needs to know the geometrical information for object recognition. The robot gets the geometrical information of an object from the tag. but to use the information, the robot should map the geometrical information of an object to the robot coordinate system. For that purpose, we adopt an active landmark, which is IRED, and a stereo vision on a pan-tilt mechanism. That is, if the robot communicates with the tag and gets the geometrical information, the robot then should find the reference point of the object for mapping for the robot to coordinate. As a reference point of object, we use IRED as an active landmark defined because it can be operate actively by the robot when it needs to. Also we adopt a stereo vision system on a pan-tilt mechanism to map the position of IRED to the robot coordinate system. The operation is very simple and includes a left and right camera pan and tilt for the IRED image to come to be in the center of each whole image view. This is not necessarily a difficult camera calibration. After the reference point of object is mapped to the robot coordinate system, a traditional model-based vision[7] has to be applied to map the shape of an object to the robot coordinate system. Only it remains for the robot to map the pose of the object to the robot coordinate system. But it is simple and easy because the robot knows the reference point of object and distance to the object from the robot.

For the realization of our approach, a setup system is constructed with a 6 DOF robot arm system(Samsung AS-1), smart tag(our development) and a stereo vision on pan-tilt mechanism(our development). The IC tag which we developed is called smart-tag because it has a smart function, such as operating an active landmark besides containing information and communicating with the robot. The next sections will introduce more details about our system and method of object recognition.

2. FRAMEWORK

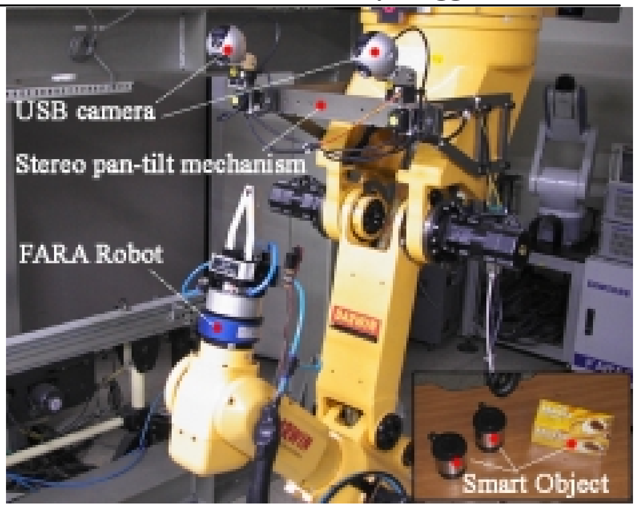

Fig. 2 System overview

We applied a stereo pan-tilt Robotis Dynamixel and a manipulator SAMSUNG FARA-AS1. The stereo pan-tilt moves 3.6 rad/sec and the robot manipulator moves 50mm/s. We made the smart-tag by using a blue-tooth module SAMSUNG BTKZ1702 A. Fig. 3 shows an overview of the object and pose recognition method.

Step 1) Attaches the smart-tag contained IRED internally to an object and flickers the IRED.

Step 2) when the searching of IRED is completed, locates the IRED image to image center by adjusting the stereo pan-tilt angle.

Step 3) Sets the pan to (θ_1, θ_2)and sets the tilt to (ϕ_1, ϕ_2) and calculates the distance between the IRED and the stereo pan-tilt.

Step 4) extracts the edge line of the each image by canny method.

Step 5) Creates a template image by the distance d and the model data which receives from the smart-tag.

Step 6) converts the edge extracted image to a contour line image by the distance transform algorithm.

Step 7) compares the contour line image and template image which is eliminates the hidden line.

Step 8) Calculate the cross correlation function and extracts maximum value.

Step 9) the object recognition and the pose recognition are completed.

It is very easy to detect the image difference between IRED which is turned on and IRED which is turned off because the IRED flickers for a split second. So it is able to search the IRED in the complex environment through a simple subtraction calculation and also get a strong result about a light. Also it has an important advantage; the robot does not need to store information because the smart-tag is capable of this.

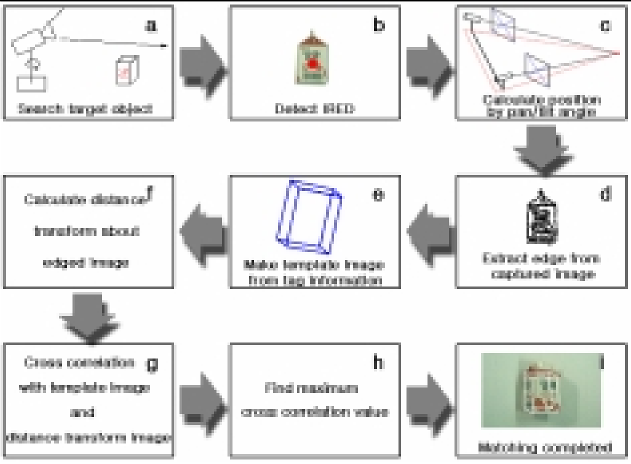

Fig. 3 Overview of pose recognition

2.1 Smart-Tag

The smart-tag should offer information of the object to the robot and it can attach to the object. Also, it should be detected to the camera vision easily and be able to communicate with other devices. So it should have the memory and communicating module. Fig. 4 shows the smart-tag prototype. The smart-tag prototype satisfies all the above requirements. The smart-tag consists of CPU, memory, blue-tooth module and IRED.

The features of the smart-tag are as follows. It has an IRED and the IRED is operated by wireless commands. Also it is possible for networking with other devices for the blue-tooth and it can store information on the objects. So we can do work without gathering the information of the objects through the wireless network. The structure of the information of the smart-tag is defined in Table 1. Object ID is identical information of the smart-tag and the robot communicates with the smart-tag by the Object ID . Position of active landmark on object and Object bounding box item means the geometrical information of an object. Namely, Position of active landmark on object is the position of IRED in object coordinate system. Object bounding box item is the size of object s bounding box. We modeled an object as a simple cubic model using this object bounding box. The physical information of an object describes the weight, material, color and strength, and the semantic information of object is whether it is a cup , or a chair . The other kind of semantic information pertains to whether it is movable or fixed and so on. In this paper, Object ID , Position of active landmark on object and Object bounding box items are used. Other item will be used for more intelligent task in the future.

2.2 Stereo Pan-Tilt Camera

The stereo pan-tilt camera is used for the object recognition and calculating a coordinate. We locate the IRED image to the center of an image by adjusting the stereo pan-tilt angle. Then, it sets the pan to (θ_1, θ_2) and the tilt to (ϕ_1, ϕ_2). Next, it calculates the distance between the IRED and the stereo pan-tilt (fig. 5).

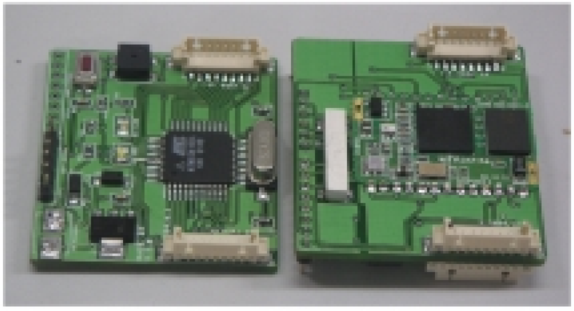

Fig. 4 Smart-tag prototype

Table 1 Data structure of smart-tag

Object ID		
Position of active landmark on object		
X	Y	Z
Object bounding box		
Width	Height	Length
Physical information		
Weight	Strength	Material
Semantic information		
Semantic Name	Mobility	

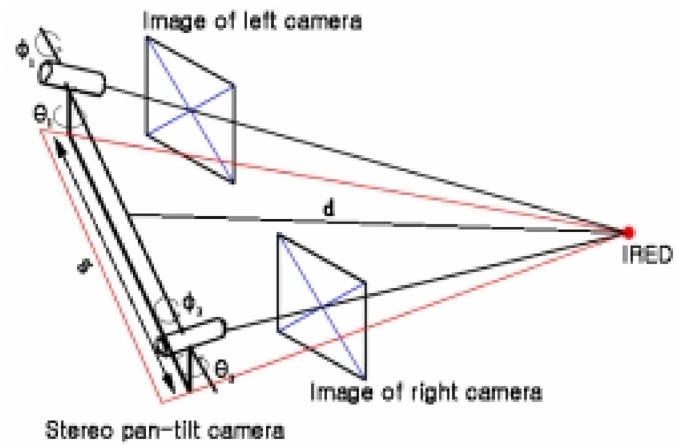

Fig. 5. Concept of distance measuring

$$d = \frac{s}{2} \cdot \frac{1}{\sin\frac{\phi_1 + \phi_2}{2}} \cdot \frac{\sin\theta_1}{\sin\frac{\pi - \theta_1 - \theta_2}{2}} \qquad (1)$$

So we are able to calculate the position of the object. A method of searching IRED is as follows.

Step 1) turn on the IRED which captures the left camera; A.
Step 2) turn off the IRED which captures the left camera; B.
Step 3) find the point of difference between A and B.
Step 4) same step is applied in the right camera.
Step 5) if there is no difference, rotate pan or tilt the camera and apply steps 1 to 4.

In an ideal case, the difference exists at the IRED point exactly but it is not possible to obtain the perfect image difference because the image has a noise. To solve this problem we remarked the pattern that a large difference is asserted continuously in the subtracted image. So now, we can extract the IRED position by the projection histogram analysis about - direction.

This stereo pan-tilt structure has advantages that a camera calibration does not. Namely, it is able to extract the exact position of the smart-tag by the stereo pan-tilt camera because the center of image has no distortion and the coordinate calculation is very simple and fast

2.3 Noise Filtering for Specify Fre uency

General USB camera contains noise of many quantities. We must calculate the difference of a two images to detect an image difference of IRED.

In case we use a general USB camera which contains noise of many quantities, we solve this problem by eliminating the noise by applying the blur image process. The left image (a) of fig. 6 is the FFT image that is not applied. The right (b) is the applied FFT image. We know the right image (b) is better.

When we capture an image difference of IRED, there will be another flickering object. Fig. 7 shows the situation. When a is the point of object which we want to find, there is another flickering object b . So we obtain results such as depicted at c . To solve this problem in this paper we use the method which flickers IRED in certain frequency. If we use this method we can obtain the certain frequency signal. (Refer to fig. 8.) So now we can obtain a certain IRED position.

2.4 Ob ect Pose Recognition

We have already extracted the exact position of the smart-tag (smart object) by the stereo pan-tilt camera. In this section we estimate the object pose by using the object position.

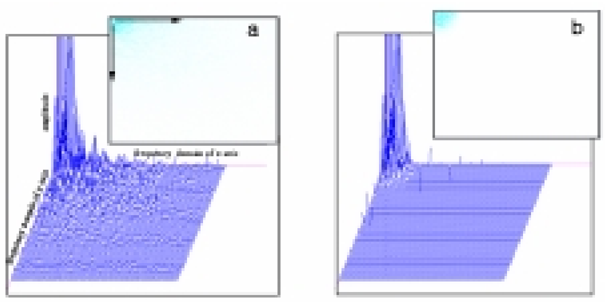

Fig. 6 Appling blur method

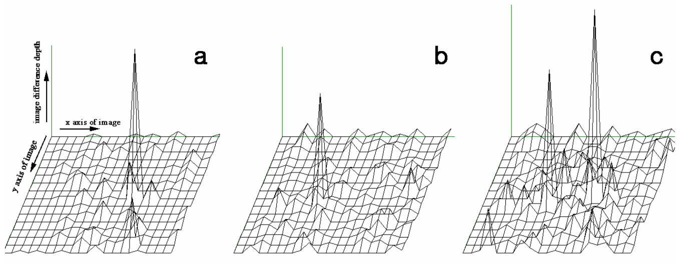

Fig. 7 Flickering noise

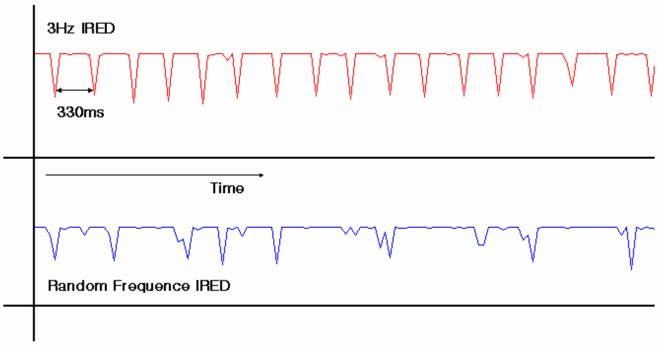

Fig. 8 Obtaining certain frequency signal

In this paper, we estimate the object pose by making a simple object model. The object model information includes only the object type and size that is needed to estimate the pose and distinguish the clearness of the object. The model information is obtained from the smart-tag and then it is regenerated to the template image. Next, it is compared with the object image of the camera. Finally we can find the maximum matching value between the template image and the object image. In this paper we defined a very simple cubic model but for the enhanced task we must classify the models in more detail

Fig. 9 shows the concept of the geometric method. The τ, θ, ϕ parameter is a known parameter by stereo pan-tilt. So we can calculate scale factor in regards to width, height, and length by using r . Therefore we can calculate the exact template image size.

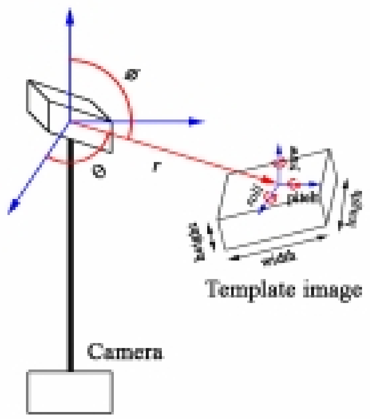

Fig. 9 Matching concept of geometry method

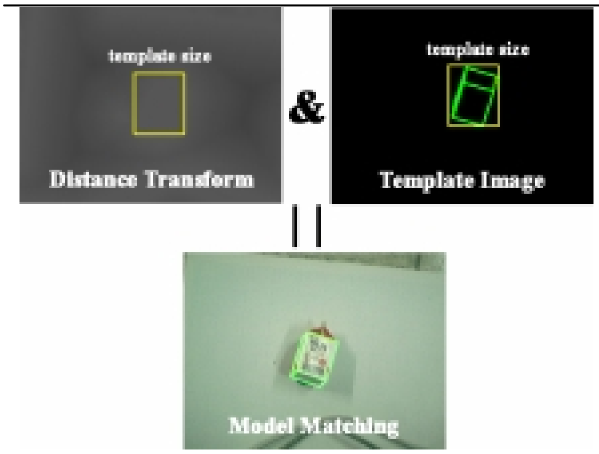

Fig. 10 Model matching

When the exact template image size is calculated we should apply the distance transform algorithm to the edge extracted image for the contour line image because the edge extracted image is not clear. There are many image transform algorithms. But the distance transformation is a very simple and easy algorithm. So we apply it in this paper. Finally we can compare the contour line image and template image which is eliminated the hidden line. The pose estimating process is shown in fig. 10.

3. EXPERIMENT AND RESULT

The first experiment is a simple experiment that finds a small box and a cup and estimates pose. The Initial setting of the experiment is shown in fig. 11(a) and the condition is shown in table 2. The stereo pan-tilt camera is attached to the robot. And in this proposed method the object should be not seen at the camera because the stereo pan-tilt camera can be rotated for searching. The task flow is as follow.

Step 1) search the ID: BO by using the smart-tag.
Step 2) flicker the IRED and search the IRED. Refer to fig. 11(a).
Step 3) when the searching of IRED is complete, it locates the IRED image to image center by adjusting the stereo pan-tilt angle. Refer to fig. 11(b).
Steps 4) complete a job. Refer to fig. 11(c, d, and e).

Fig. 11(f, g and h) shows the finding of a cup. . And fig. 11(j, j and k) shows the experiment when occlusion is occurred. Fig. 11(e, f and i) shows edge extracted images and Fig. 11 (d, g and j) shows distance transform images.

In the first experiment, we often got inaccurate results because the object image included a shadow. To solve this problem, we eliminated the hidden line of the template image. The effect was considerably better as there were subsequently no wrong results. The robot recognized the object with speed and ease. But the IRED of the smart-tag (the object or the smart object) must always face each other with the camera because if the IRED is not seen, the camera will not find the difference between the turned on image and the turned off image.

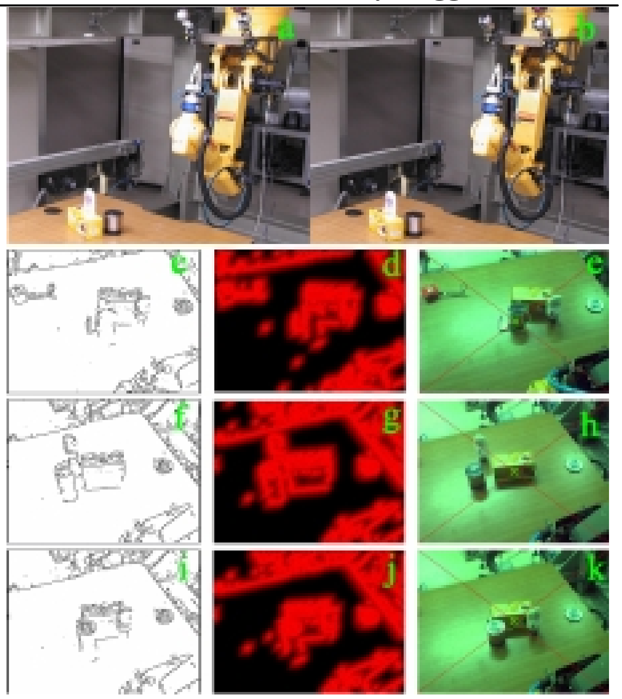

Fig. 11 Result of experiment that finds objects

Table 2. Experiment condition

Pan - tilt resolution	0.35
Image size	320 X 240
Object number	3
Software	Intel OpenCV
PC	P4 - 1.5G, 512 Mbyte memory
OS	Win2000

The second experiment is an experiment that grasps a cup. The result of the experiment is shown fig. 12 and the process of the second experiment is equal to the first experiment. When the robot knows the position and pose it attempt the grasping task.

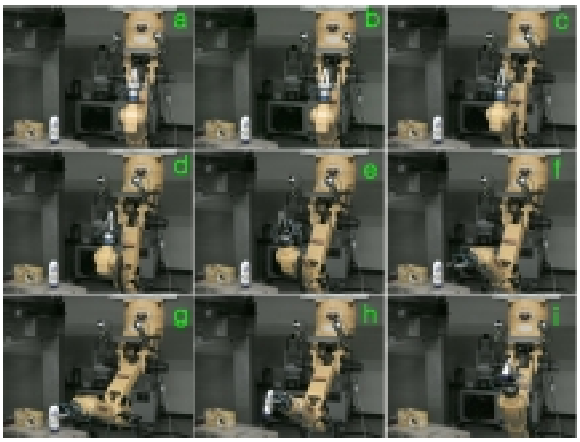

Fig. 12 Result of an experiment that grasps a cup

4. CONCLUSION

We propose a novel method for object recognition using a smart tag and stereo vision on a pan-tilt mechanism. For that purpose, we developed a smart tag with an active landmark (IRED) and a stereo vision system on a pan-tilt mechanism. Therefore, using them we can acquire the position of IRED from the smart tag attached onto an object and map the size and pose of the object to the robot coordinate system by using a simple model-based vision algorithm. The experimental results show the possibility of our tag-based object recognition using easy and simple techniques. But we know that some problems still exist as follows. Problems in the traditional model based vision still exists, such as determining a proper threshold for good extraction of the interested object from whole image view and occlusion is a problem. The bounding box template of an object used for matching it with an object may not cope with various shapes of objects. To be more realistic, a smart tag has to be smaller than it is now, the battery of a smart tag has to withstand long use, and the cost of a smart tag has to be very cheap.

Although problems such as the ones named above still exist, the experimental results show that our approach has sound benefits. Namely, the recognition time is fast and recognition results are reliable compared with traditional methods of object recognition. To search where an interesting object is, traditional methods have to spend very long time or it will be impossible searching itself.

REFERENCES

[1] D. Kurabayashi, K. Konishi, H. Asama, Distributed guidance knowledge management by Intelligent Data Carriers , International journal of Robotics and Automation, vol. 16, no. 4, pp. 207-216, 2001.

[2] D. Kurabayashi, K. Konishi, H. Asama, Autonomous knowledge acquisition and revision by intelligent data carriers in a dynamic envirionment , Journal of Robotics and Mechatronics, vol. 13, no. 2, pp. 154-159, 2001.

[3] M. Boukraa, S.Ando, Tag-based vision: assisting 3D scene analysis with radio-frequency tags , ICIP02, pp. 269-272, 2002.

[4] M. Boukraa, S.Ando, A computer vision system for knowledge-based 3D scene analysis using radio-frequency tags , C2002 IEEE, pp. 412-418, 2002..

[5] N. . Chong, H. Hongu, K. Ohba, S. Hirai, and K. Tanie, Knowledge distributed robot control framework , Proc. Int. Conf. on Control, Automation, and Systems, pp. 1071-1076, 2003

[6] K. Takemura, K. Ohara, K. Ohba, N. . Chong, S. Hirai, and K. Tanie, Knowledge distributed robot system tag-based vision system , SI2003, pp. 317-318, 2003

[7] oshiaki Shirai, Model based vision , vol. 2, no. 6, pp. 593-598, RSJ, 1984