Vision-Based Motion Recognition of the Hexapod for Autonomous Assistance

Hiroshi Kimura and Hideaki Katano

Univ. of Electro-Communications
5-1 Chofu-ga-oka 1-chome, Chofu City, Tokyo 182-8585, JAPAN

Abstract

In a cooperation style in which a human operates one robot and the other autonomous robots assist that robot, the autonomous robots must be able to recognize the motion of the human operating robot in real time. Vision is the most useful sensor for this purpose. By using vision, the assistant robot recognizes the current target object and actions of the human operating robot to the target. For the working robots, we use two hexapod robots which are able to utilize two forelegs as arms in order to manipulate objects. Joint motion of each leg and occlusion by another leg make it difficult to recognize and track the motion of the hexapod robot. In this study, we propose a method for the recognition and tracking of the 3D position and posture of the hexapod. Through the use of 2D image matching, the body of the hexapod, several legs, and toes of the legs can be recognized in turn. The positions of the body and toes are measured by binocular-stereo, and the angles of the legs can then be obtained, which are used for motion recognition. In this study, we propose a motion recognition mechanism using eigenspace method which can reduce the dimensions necessary for matching. The effectiveness of this method is tested through an experiment using two hexapod robots.

1 Introduction

On construction sites and in other hazardous environments robots are often expected to manipulate objects of a size comparable to their own. A single robot usually encounters difficulty performing such a task. Cooperation of multiple robots is the solution to this problem. Because the simultaneous operations of multiple robots are difficult for a human operator, we propose a cooperation style in which the human operates one robot and another autonomous robots assists (Fig.1).

In order to assist the human operating robot, the autonomous robots must recognize its motions in real time[1]. Vision is the most useful sensor for this purpose. By using vision, the assistant robot recognizes the current target object and actions of the human operating robot to the target. For the cooperation experiment we constructed two hexapod working robots which were able to walk on irregular terrain and to utilize two forelegs as arms in order to manipulate an object. The motions of the hexapod are not easy to recognize because of joint motion of each leg and occlusion by another leg. In the field of research surrounding the recognition of objects with multi-links and multi-joints by vision, much attention has been paid to studies of the motions of humans or human hands[2]-[7]. In these studies, the following two problems have been identified:

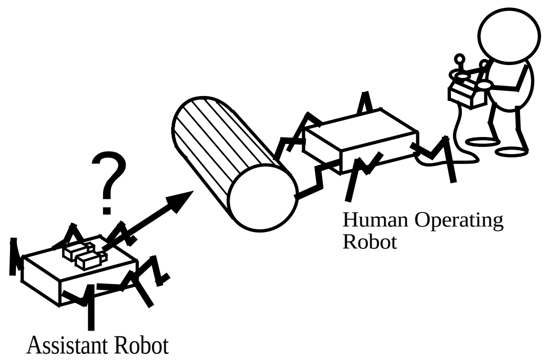


Figure 1: Motion recognition-based assistance.

(1) dealing with the 3D motion of multi-joints,

(2) dealing with occlusion between multi-links.

For problem (1), many methods have been proposed in which the 3D model is projected onto the 2D image and 3D information is inferred from 2D matching on the image[2, 3]. For problem (2), a method has been proposed in which occlusion is avoided by observing from multi-viewpoints[4, 5]. Another method has also been proposed in which occlusion is predicted from a 3D model[6].

In a hexapod robot, problem (1) becomes easier to manage than in situations with a non-rigid object such as a human hand. However, since the hexapod consists of many complicated parts, it is still difficult to recognize the motion of the joints. Therefore, we don't use a 3D geometric model but instead use an appearance model because it does not require the extraction of geometric features. Regarding problem (2), it is impossible for a robot to simultaneously observe the other robot from multiple viewpoints. Therefore, it should be possible to recognize the hexapod based on partial information observed from a single viewpoint. For this purpose, we use 2D matching based on minute templates. This method is robust for occlusion. By using a hierarchical recognition method with these methods, the body of the hexapod, several legs, and toes of the legs are recognized in turn. That is, the 3D position and posture of the hexapod are recognized and tracked stably in spite of the joint motion and occlusion.

Recognition of human motions by vision has been studied in the field of robot teaching. "Teaching by Showing" by Kuniyoshi[8] and APO (Assembly Planning from Observation) by Ikeuchi[9] have been proposed. Mori[10] has described the development of holding tools which can assist humans with soldering work through behavior understanding. In the field of multiple robot cooperation, Kuniyoshi[11] has proposed "Cooperation by Observation" and has realized such cooperation skills as posing, unblocking, and passing.

In this research, the target of motion recognition is the hexapod, which has a body, 6 legs and 18 joints. Because the motion pattern of this hexapod is more complicated than those worked with in previous studies, it has been necessary to develop a new, effective method for motion recognition. In this study, we propose a motion recognition mechanism using eigenspace method, which can reduce the dimensions for matching.

In the case of cooperation between robots, a robot can ascertain the state of the other robot by computer communication as described in our previous work[1]. The wireless communication necessary for this method, however, has problems of its own. In addition, when we consider cooperation of a robot with the human or competition among robots where cooperative computer communication is not expected, motion recognition by vision becomes essential.

2 3D Recognition and Tracking

In order to recognize the motion of the hexapod working robot, it is necessary to detect not only the position and orientation of the body but also the joint angles of each leg. For this purpose, we introduce the 3D vision system. It can recognize and track the 3D position and posture of the body and legs in spite of joint motion and occlusion. The features are as follows:

- 2D image matching using the appearance model and minute templates,
- 3D position measurement of the feature points by binocular stereo cameras,
- magnification of the scene image based on depth,
- hierarchical recognition of the body, legs, and toes,
- the use of the color tracking vision hardware.

2.1 2D Recognition and 3D Measurement

2.1.1 Data Images for Appearance Model

Recently, a method which uses the image itself as a model for the recognition of a 3D object has received attention in conjunction with advances in computation power because it does not require the extraction of geometric features[12]. The data image for recognition is called the "appearance model". When we locate the hexapod on the rotary table with its body yaw axis coinciding with the rotation axis, we can obtain 36 stereo images of the hexapod while rotating the table by 10-degree steps (Fig.2). These stereo images are saved on the disk. One of the stereo image is a color image. These color images obtained from 360-degree whole views are used as the data image for the appearance model.

Because the rotary table is calibrated by the stereo cameras, we can get the 3D position and orientation of the body on each data image. In addition, the angles of the legs are known. Therefore, we can project a 3D geometric model of the hexapod on each data image and obtain the areas of the body, legs, and toes on this image. These areas on each data image are registered and used for the 2D matching of each part as described in **2.2.**. Information from the geometric model itself (faces, edges, etc), however, is not used for recognition.

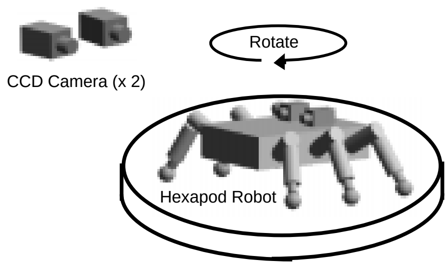


Figure 2: Data images from 360-degree whole views.

2.1.2 2D Matching by Minute Templates

Among the small areas (for example, 5×5) on the data images and the scene image, we select the ones which have a high trackability measure[13] and call them "minute templates". Ohba[14] has proposed a method in which the minute templates on both images are compared, and matched templates are clustered by using the geometric constraints. By this method, we can determine the best data image matched to the scene image and obtain the area of the object recognized on the scene image. This method has the following advantages:

(a) Segmentation on the scene image is not necessary.

(b) The method is robust for occlusion.

In this study, based on the above method, we use a 3D object recognition method on a 2D image.

2.1.3 3D Measurement by Stereo Vision

The 3D positions of the body and toes of the hexapod are obtained by subpixel stereo measurements of the minute templates. Because the corresponding points of the stereo measurements are searched only for the minute templates which have a high trackability measure, we can easily obtain the correct results.

2.1.4 Image Magnification Based on Depth

The 2D matching method described in **2.1.2.** cannot deal with large changes in appearance that occur in response to movement along the depth axis. Therefore, the color marks on the hexapod are searched and found before 2D matching, and the rough depth from the camera to the hexapod is measured by the stereo cameras. Because the depth to the rotary table origin in the data images is known, the scene image is magnified based on two depths. This magnified scene image is used for 2D matching with the data images (Fig.3).

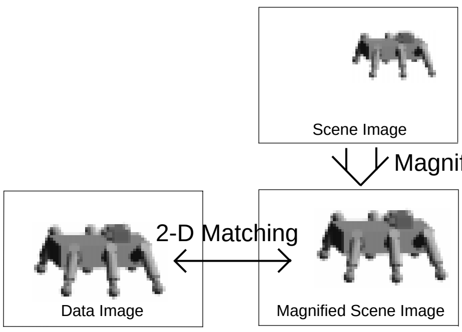


Figure 3: Image magnification.

2.2 Hierarchical Recognition of the Body, Legs and Toes

The hexapod consists of a body and legs that are attached to the body. The body translates in the 3D space and rotates mainly around the yaw axis while walking and working. Each leg has three joints and moves in 3D space. In addition, the body moves more slowly than the legs. Therefore, although the occluded areas of the body change in response to movement of the legs, changes in the appearance of the body are smaller than those of the legs. For this reason, we search and find the body, legs and toes in turn on the scene image as follows (Fig.4):

(1) The best data image matched to the scene image is determined, and the area of the body on the scene image is obtained by 2D matching described in **2.1.2.**. Because the 3D orientation of the body on the data image is known as described **2.1.1.**, we can obtain the orientation of the body on the scene image.

(2) With the stereo measuring of the matched minute templates described in **2.1.3.**, we can obtain the translation vector of the body between the data and scene images. Because the 3D position of the body on the data image is known, we can calculate the position of the body on the scene image from the translation vector.

(3) By (1) and (2), we can make the frame of the body relative to the camera frame of the scene image and project the 3D geometric model of the hexapod onto it.

(4) The search area for each leg is determined based on the geometric constraints of the joint motion in relation to the body. The image area of each leg registered in the data image is searched within this search area on the scene image by using the 2D matching described in **2.1.2.**.

(5) When the image area of one leg is found on the scene image, the toe of the leg is searched by 2D matching, similar to (4).

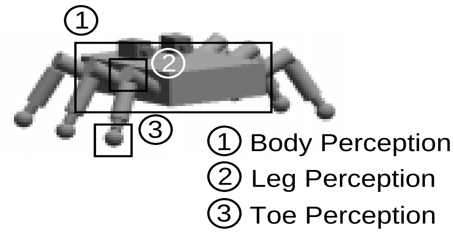
(6) When the image area of the toe is found on the scene image, the 3D position of it is measured by the stereo cameras.

(7) The 3D position of the toe is transformed to the body frame, and the joint angles of the leg are calculated by using inverse kinematics.

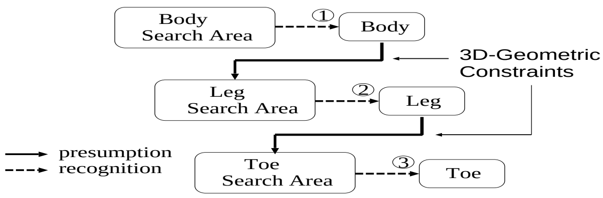
2.3 Experiments of Recognition

The results of the experiments related to recognition of the body, legs, and toes of the hexapod on the scene image are shown in Fig.5-8. In Fig.8, the toes of the left three legs are recognized, but those of the right three legs are not due to occlusion. The angles of the left foreleg obtained by vision are shown in Table.1 while being compared to those measured with photo encoders attached to joints. The maximum error is less than 5 degrees.

Recognition of the body takes about 30 (sec) on UltraSparc-I because we must compare many data images to the scene image. Recognition of legs and toes takes several seconds.



(a) Hierarchical recognition of the hexapod.



(b) Presumption of search area.

Figure 4: Recognition process of the hexapod.

Table 1: Measurement results for the angles of the left foreleg.

joint	vision (degrees)	encoder (degrees)
1st	-39.6	-41.0
2nd	50.1	48.0
3rd	-143.1	-142.0

2.4 Tracking of the Hexapod

Although recognition of the body takes much time, it can be tracked relatively quickly once it is recognized. The tracking system is shown in Fig.9. The best data image and minute templates of the body, legs and toes on it obtained by the 3D recognition system (upper part of Fig.9) are given to the color tracking vision system (CTRV: lower part of Fig.9). These minutes templates are tracked using the functions of the CTRV. In addition, the 3D positions of the tracked minute templates are measured by the stereo cameras using the functions of the CTRV, and the 3D positions of the body and toes are determined. This 3D tracking is repeated within a period of about 350 (msec), during which the 3D position of the body and the joint angles of observed legs are obtained.

The results of tracking are used in the next 3D recognition step. Because the number of data images and the search area can be restricted from the results of tracking, 3D recognition is repeated within a shorter period of about 10 (sec). The recognition process is

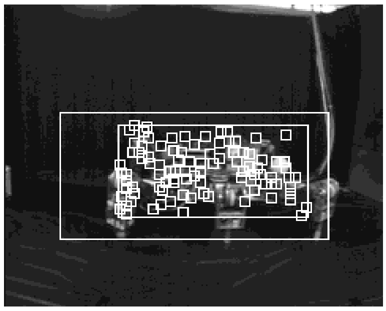


Figure 5: Recognition result for the body on a scene image. The outer box is a search area for the body. Matched minute templates are shown.

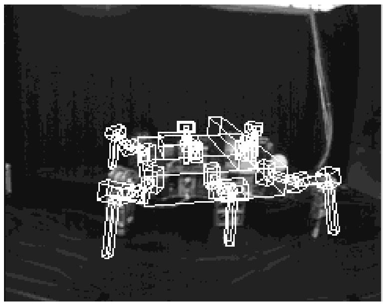


Figure 6: Geometric model projected on the scene image.

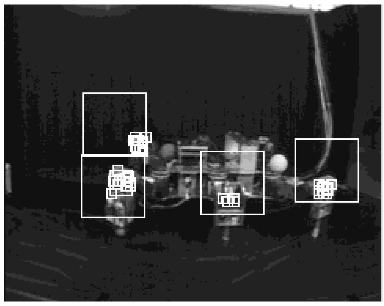


Figure 7: Recognition result for the legs. The outer boxes are the search areas for each leg.

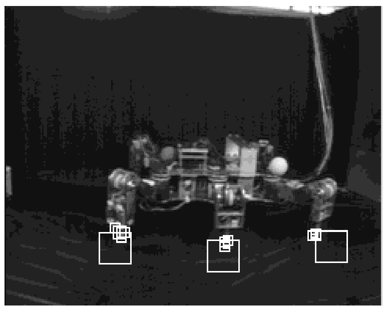


Figure 8: Recognition result for the toes. The outer boxes are the search areas for each toe.

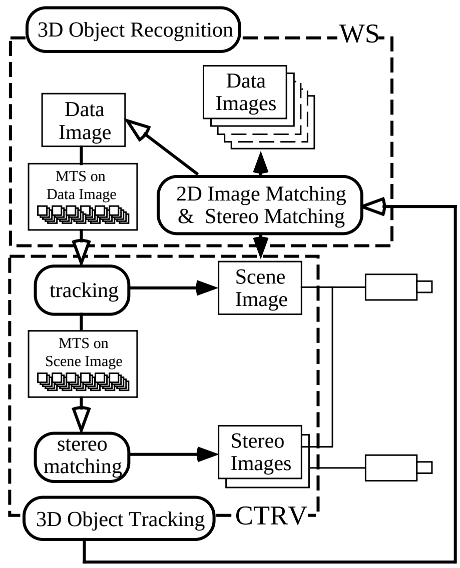


Figure 9: Process of recognition and tracking. MTS indicates the minute templates.

concurrent with the tracking process, and the results of recognition process are used in the next tracking process.

In conclusion, the 3D translation of the hexapod is tracked using CTRV in every 350 milliseconds. The rotation of the hexapod around the yaw axis is recognized by 2D matching using the appearance model on the computer in every 10 seconds.

3 Motion Recognition of the Hexapod

The authors have implemented the motion symbolization and motion recognition methods in previous work[1]. The motion feature templates and patterns, however, have been generated manually. It is now necessary to develop a mechanism by which these templates and patterns can be generated automatically. In this section, we describe a motion classification and recognition method using eigenspace method which can solve this problem.

The eigenspace method is well-known as the "K-L procedure" in the area of pattern matching. It projects the data vector to a space of which the unit vectors are the normalized eigenvectors of the covariance matrix of the pattern vectors. By using eigenvectors corresponding to the large eigenvalues, an effective reduction of the dimension or classification is possible. We apply this method to the classification and recognition of ambiguous information such as the motion of the hexapod.

3.1 Motion Feature Space

As the angles of the legs express the motion pattern of the hexapod, it is natural to use the angles of the legs as state vectors for motion recognition. When we use vision to observe motion, however, we cannot get the joint angles of all legs due to occlusion. In such cases, it is necessary to obtain the motion features from partial information. Here it is assumed that we can get the joint angles from enough legs to recognize the motion by controlling the viewpoint.

When we get the joint angles of k legs $(i_1, \cdots, i_k)$, we consider the m $(= 3 \times k)$ dimension vector to be as follows:

$$\boldsymbol{x} = [\boldsymbol{\theta}_{i_1}{}^T, \cdots, \boldsymbol{\theta}_{i_k}{}^T]^T \in \boldsymbol{R}^{m\times 1}, \tag{1}$$

where $\boldsymbol{\theta}_j = [\theta_{j1}, \theta_{j2}, \theta_{j3}]^T$ is a vector of the joint angles of the jth leg.
Let the standing state of the hexapod be the neutral point $\boldsymbol{x}_a$ and $\boldsymbol{z}$ $(= \boldsymbol{x} - \boldsymbol{x}_a)$ be the state vector. By observing the motion of the human operating robot for several tasks in the training mode and by sampling N state vectors, we obtain $\boldsymbol{z}_1 \sim \boldsymbol{z}_N$ and a sampling matrix as follows:

$$\boldsymbol{Z} = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N] \in \boldsymbol{R}^{m\times N}. \tag{2}$$

$\boldsymbol{Z}_c$ is the covariance matrix of $\boldsymbol{Z}$.

$$\boldsymbol{Z}_c = \boldsymbol{Z}\boldsymbol{Z}^T \in \boldsymbol{R}^{m\times m} \tag{3}$$

This covariance matrix provides a series of eigenvalue λ_i $(i = 1 \sim m)$ in descending order. The eigenvector corresponding to λ_i is $\boldsymbol{e}_i$. We make the following matrix by the upper n eigenvectors:

$$\boldsymbol{E} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n] \in \boldsymbol{R}^{m\times n}, \tag{4}$$

where n is determined so that the contribution factor W_n becomes larger than the given threshold.

$$W_n = \sum_{i=1}^{n} \lambda_i \ / \sum_{i=1}^{m} \lambda_i \tag{5}$$

The state vector $\boldsymbol{z}_i$ is projected to the n dimension eigenspace by $\boldsymbol{E}$ and $\boldsymbol{g}_i$ is obtained.

$$\boldsymbol{g}_i = \boldsymbol{E}^T \boldsymbol{z}_i \in \boldsymbol{R}^{n\times 1} \tag{6}$$

Because the eigenspace can reduce the dimension of the state vector, m, to the eigenspace dimension, n, while keeping the important features, we call this eigenspace a "motion feature space".

In the eigenspace, $\boldsymbol{g}_1 \sim \boldsymbol{g}_N$ make some tracks for each task (Fig.10). For the sake of memory efficiency, we pick up points $\hat{\boldsymbol{g}}_1 \sim \hat{\boldsymbol{g}}_r$ so that the distance between each point becomes larger than the given threshold on these tracks. These points are the motion features obtained in training.

In this study, we prepare the eigenspaces for every combination of observed legs by simulation as training.

For example, the eigenspace constructed by the left three legs in the representative tasks such as hexapod-walking, quadruped-walking, and manipulating a box with the forelegs is shown in Fig.10, where m is 9 and n is 3.

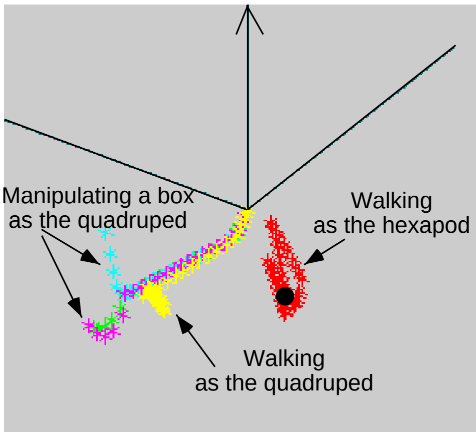


Figure 10: Eigenspace in several tasks.

3.2 Motion Recognition

In the recognition mode, we generate the state vector $\boldsymbol{z}_p$ from the joint angles of the legs observed by vision. By using (6), we calculate $\boldsymbol{g}_p$ and obtain $\hat{\boldsymbol{g}}_q$ as the point closest to $\boldsymbol{g}_p$ among the motion features $\hat{\boldsymbol{g}}_1, \cdots, \hat{\boldsymbol{g}}_r$. This $\hat{\boldsymbol{g}}_q$ is the motion feature observed at this moment.

The motion feature for the joint angles obtained in Fig.8 is shown in Fig.10 as a point $\bullet$ on the track of the hexapod-walking. By repeating such motion feature observation for a while, the present motion of the hexapod is recognized as hexapod-walking.

4 Conclusion

In order to cooperate with a robot controlled by a human operator, autonomous assistant robots must recognize its motion in real time. Joint motion of each leg and occlusion by another leg make it difficult to recognize and track the motion of the hexapod robot.

In this study, we proposed a method by which we could recognize and track the 3D position and posture of the robot by vision. This method used the appearance model and minute templates. These technique were effective in 2D image matching for a complicated object such as the hexapod under self-occlusion. By 2D image matching, the body of the hexapod, several legs and toes of the legs were recognized in turn. Once recognized, the body and toes were tracked by using the functions of color tracking vision hardware. The 3D positions of the body and toes were measured by binocular-stereo cameras, and the angles of the legs were obtained. These angles were then used for motion recognition.

The motion pattern of the hexapod is so complicated that we need an effective method for motion recognition. In this study, we proposed a motion recognition mechanism using eigenspace method which could reduce the dimensions for matching.

The effectiveness of these methods was tested through experiments using two hexapod robots.

References

[1] H.Kimura and G.Kajiura, Motion Recognition Based Cooperation between Human Operating Robot and Autonomous Assistant Robot, Proc. of ICRA97, pp-297-302, 1997.

[2] Y.Kameda, M.Minoh and K.Ikeda, Three Dimensional Pose Estimation of and Articulated Object from its Silhouette Image, Proc. of ACCV93, pp.612-615, 1993.

[3] J.Kuch and T.Huang, Vision Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration, Proc. of ICCV95, pp.666-671, 1995.

[4] D.M.Gavrila and L.S.Davis, 3D model-based tracking of Humans in action: a multiview approach, Proc. of CVPR96, pp.73-80, 1996.

[5] I.A.Kakadiaris and D.Metaxas, 3D Human Body Model Acquisition from Multiple Views, Proc. of ICCV95, pp.618-623, 1995.

[6] J.M.Rehg and T.Kanade, Model-Based Tracking of Self-Occluding Articulated Objects, Proc. of ICCV95, pp.612-617, 1995.

[7] D.Metaxas and D.Terzopoulos, Shape and Non-rigid Motion Estimation through Physics-Based Synthesis, PAMI, Vol.15, No.6, 1993.

[8] Y.Kuniyoshi, M.Inaba and H.Inoue, Learning by Watching: Extracting Reusable Task Knowledge from Visual Observation of Human Performance, IEEE Trans. on Robotics and Automation, Vol.10, No.6, pp.799-822, 1994.

[9] S.B.Kang and K.Ikeuchi, Grasp Recognition and Manipulative Motion Characterization from Human Hand Motion Sequences, Proc. of ICRA94, pp.1759-1764, 1994.

[10] T.Mori, T.Sato and H.Mizoguchi, Situation Reactive Handiwork Support through Behavior Understanding, Proc. of ICRA97, pp.587-593, 1997.

[11] Y.Kuniyoshi, Behavior Matching by Observation for Multi-Robot Cooperation, Proc. of ISRR'95, 1995.

[12] H.Murase and S.K.Nayar, Visual Learning and Recognition of 3-D Objects from Appearance, Int. Journal of Computer Vision, Vol.14, No.1, pp.5-24, 1995.

[13] C.Tomasi and T.Kanade, Shape and Motion from Image Stream: a Factorization Method-2. Point Features in 3D Motion, CMU-CS-91-105, 1991.

[14] K.Ohba and K.Ikeuchi, Recognition of the Multi Specularity Objects using the Eigen-Window, CMU-CS-96-105, 1996.