Proceedings of 2004 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 28 - October 2, 2004, Sendai, Japan

Flexible Cooperation between Human and Robot by interpreting Human Intention from Gaze Information

Kenji Sakita
The University of Tokyo
Tokyo, Japan
Email: sakita@cvl.iis.u-tokyo.ac.jp

Koichi Ogawara
The University of Tokyo
Tokyo, Japan
Email: ogawara@cvl.iis.u-tokyo.ac.jp

Shinji Murakami
Kyushu Electric Power Co., Inc.
Fukuoka, Japan
Email: Shinji_B_Murakami@kyuden.co.jp

Kentaro Kawamura
Kyushu Electric Power Co., Inc.
Fukuoka, Japan
Email: Kentarou_Kawamura/KYUDEN@kyuden.co.jp

Katsushi Ikeuchi
The University of Tokyo
Tokyo, Japan
Email: ki@cvl.iis.u-tokyo.ac.jp

***Abstract*— This paper describes a method to realize flexible cooperation between human and robot which reflects the intention and state of human by using gaze information. This physiological information expresses the process of thinking directly, so it enables us to read the internal condition such as hesitation or search in decision making process. We propose a method to interpret the intention and condition from the latest history of gaze movement and determine an appropriate cooperative action of a robot based on it so that the task proceeds smoothly. Finally, we show experimental results by using a humanoid-type robot.**

I. INTRODUCTION

In recent years, many laboratories and companies have been studying on humanoid-type robots and have gotten considerable results especially in realization of stable locomotion with two legs. Meanwhile, research on more intelligent tasks, such as skillful manipulation or cooperative tasks between human and robot, has also been investigated[1], [2], [3], [4]. In these frameworks, a demonstration of a task is performed by a human operator and a task model, an abstract representation which describes necessary conditions for the task to proceed, is generated. Then, reproduction of the task[1], [2], [3] or cooperative behavior[4] is realized by a robot system based on the task model. However, the behavior of the robot is determined from a static task model, thus a human must exactly follow the procedure described in the task model even when under cooperative tasks. This is far from a natural cooperative task in which one dynamically determines appropriate cooperative behavior by taking the intention and state of the partner in consideration. To realize natural cooperative tasks, the robot needs to know the information which can be used to estimate the intention and state, i.e. the process of thinking, of a human partner as well as the information about the procedure of the task.

To estimate the intention and state of a human at work, we think gaze movement is very useful whose physiological nature represents the attention or interest of him/her directly. Gaze movement reflects the process of thinking under intellectual activity and contains useful information to infer them. Thus, the use of gaze information is popular in the field of psychology, as well as in the field of engineering[5], [6], [7]. If the behavior planning of the robot can be modified based on the intention and state estimated by analyzing gaze movement, it is possible to realize flexible and natural cooperation between human and robot as we do. Besides, gaze movement is appeared as a by-product of thinking, it does not impose extra burden on a human compared with other methods like oral command or explicit interval before the robot starts to behave. We select LEGO assembly task as an example task, and propose a method to estimate the intention and state of human from gaze information and a strategy to generate an appropriate cooperative action based on them.

In this paper, our framework for cooperative tasks is discussed in section 2. In section 3, a method to estimate the intention and state of human from gaze movement and a strategy to determine an appropriate cooperative action are proposed. In section 4, implementation details about the proposed method on a humanoid robot system is described and some experimental results are shown. Finally, we conclude in section 5.

II. FRAMEWORK FOR COOPERATIVE TASKS

A. Related Research

We extend the cooperative framework[4] proposed by Kimura, et al., and realize a much flexible cooperative task between human and robot based on the estimation of the intention and state of human. Fig.1 shows a task model for assembly tasks used in Kimura's framework. A model is represented as a sequence of Events, each of which indicates a pair of "Pre-condition" and "Result." "Result" is a state after an assembly action is performed and "Pre-condition" is a conditional action required to achieve the corresponding "Result." A task model is constructed at

0-7803-8463-6/04/$20.00 ©2004 IEEE

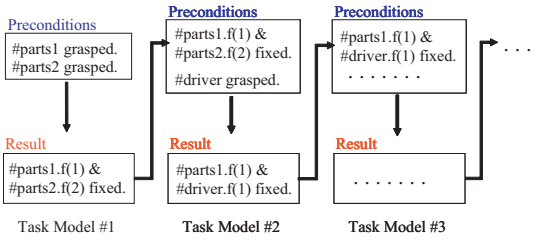

Fig. 1. One-way Task Model

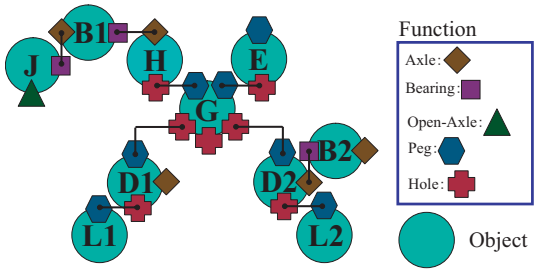

Fig. 2. Branching Task Model

a teaching phase. At a cooperation phase, a sequence of actions performed by a human is observed by a robot system and if the next "Result" is not satisfied for a certain period of time, the robot performs the action described in the corresponding "Pre-condition" instead. Hereby, each "Result" is guaranteed to be satisfied sequentially and the task proceeds.

However, the order of assembly motion is fixed and the robot behavior generated at any time is uniquely determined from the task model. Thus, this framework cannot be applied to much more general cooperative tasks in which one will choose the most appropriate action among many possible candidates according to the process of thinking. Further, the robot starts to take an action after confirming that the next "Result" is not satisfied for a certain period of time. This interval is determined so as to be sufficiently long and has nothing to do with the process of thinking. So, even if the generated cooperative action is appropriate, it make the progress of the task slow compared with the cooperative action performed by human.

To summarize, there are two problems.

1) The order of the action is uniquely determined.
2) A certain amount of delay is required before the robot starts to cooperate.

B. Framework for Flexible Cooperative Tasks

To solve the above problems, we propose the following framework.

1) Task representation with a branching task model.
2) An appropriate cooperative action is determined from the intention and state of a human at that time.

In a branching task model, the final configuration of the assembly task is fixed as in the previous model, however there are many possible paths, i.e. the order of actions, to reach the goal and the choice of the path is deeply affected by the intention and state of a human. Fig.2 is an example of a branching task model.

This model is composed of "Object" and "Function;" where "Function" characterizes the use of "Object," while "Objects" are assembled by connecting "Functions" with each other. In this paper, LEGO assembly task is selected, thus "Object" corresponds to a LEGO part and "Function" corresponds to one of "Axle," "Open-Axle," "Bearing," "Peg" and "Hole." Each "Function" has connectable "Function" as shown in TABLE I. The task proceeds by connecting "Functions" sequentially so that the "Object" pair are assembled as described in the task model. Some combinations of "Functions" needs extra action, screwing by "Driver," after connection.

TABLE I
ASSEMBLY PATTERN

Connectivity	Driver required
Axle ↔ Bearing	Yes
Open-Axle ↔ Bearing	No
Peg ↔ Hole	No

The branching task model represents the final configuration. Suppose the task is partially completed, and L1, D1, G in the figure is assembled so far. Then the next candidate object to be assembled is one of H, E, D2. In this situation, if one of the following conditions is met, the cooperative action by a robot might help; (1) a human is unable to reach the next object because the both hands are occupied or the object is placed far, (2) a human is unable to decide which assemble action is the correct one. Furthermore, in the case of (1), if the cooperative action is delayed, a human tries to break the situation by releasing the held object or by moving from his/her position to bring it. In this case, the delayed cooperative action may conflict with the action performed by a human and may block the task instead.

So, the robot must know in which situation the human is and choose the right assembly action according to the intention of the human if there are many candidates, and start the cooperative action without delay. To estimate those information while the human is in his process of thinking, we employ "gaze movement."

III. ESTIMATION OF THE INTENTION AND DETERMINATION OF COOPERATIVE ACTION

A. Acquisition of Gaze Movement

To know the role of gaze movement in assembly tasks, 5 subjects were asked to perform several LEGO assembly tasks and gaze movement was measured. First, the final plan (Fig.3) was presented to the subjects and they were requested to memorize it within 30 seconds. Then, they were asked to assemble the LEGO object based on the memorized plan. Note that we have assumption that the decision making process for selecting next assembly operation is largely affected by the relationship between "Functions" in the plan, because only the parts which have a connectable "Function" can be connected to the

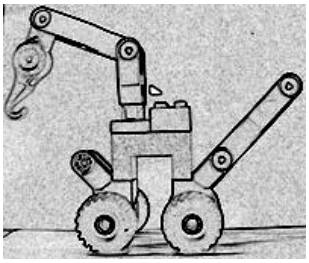

Fig. 3. Final Plan of LEGO Assembly

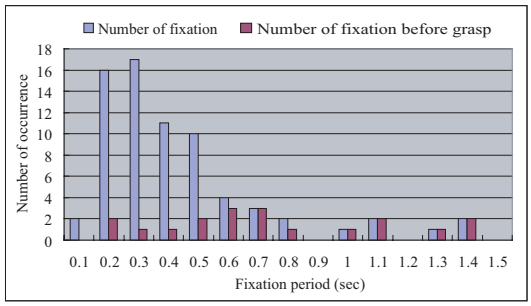

Fig. 4. Distribution of the Fixation Time in Searching and Determining State

incomplete construction at hand. Thus, for ease of analysis, color information is removed from the presented final plan.

Gaze movement is measured by using an gaze tracking system. The history of the gazed objects and its fixation time is obtained.

B. Cooperative Action by a Robot

After analysis of the obtained gaze movement, the following 3 types of cooperative actions are found to be useful.

1) Taking over
2) Settlement of hesitation
3) Simultaneous execution

In the following sections, the detail of the cooperative actions is discussed.

C. Cooperative Action 1: Taking Over

The flow of LEGO assembly can be summarized as follows.

1) search for the next part in the environment
2) determine the next part to be assembled
3) grasp the part and assemble it
4) goto 1)

If the transition from 1) the searching state to 2) determining state can be detected, the cooperative action "Taking over" is possible by passing the selected part to the subject at the time of transition. This is useful under the situations below.

- The both hands of the subject are occupied and he/she is unable to grasp the selected object.
- The selected object is far from the subject and it is efficient for the robot to pass it to him/her.

Here we focus on the fixation time during gaze movement and try to separate the searching state and the determining state.

1) Characteristics of the fixation time during search: We measured the distribution of the gaze fixation time in the searching state and that just before a grasp, i.e. determining state, during LEGO assembly tasks. Fig. 4 shows the distribution. If the fixation time is larger than 0.6 [s], the 70 % of the samples are in determining state. Moreover, a half of the samples where the fixation time is less than 0.6[s] is captured when using "Driver." Because the subjects used "Driver" several times in the measurement, the subjects got to know where and how "Driver" was placed and did not require longer time to check it before grasping.

If we remove the data related to "Driver," the 77 % of the samples where the fixation time is larger than 0.6[s] are in determining state. So, we can say that there are meaningful difference in fixation time in between searching state and determining state, and this can be used to separate determining state from searching state.

2) Proposed Cooperative Action: If fixation time T_i at i-th transition during searching state is greater than the threshold T_{thr} determined from the above distribution, the robot decides that the object of interest is what the human tries to grasp next. So, we propose "Taking Over" cooperative action as follows.

1) $i = i + 1$, *measure* T_i
2) *if* $T_i < T_{thr}$ *then* goto 1)
3) *if* $|P_i - P_{human}| < dist$ *then* goto 1)
4) *If* $isEmpty(hand)$ *then* the robot passes the part *else* the robot assembles the part
5) goto 1)

where P_i is the position of the object of interest and P_{human} is the position of the subject.

Of course, the long fixation time does not always mean the signal of grasping, but at least the robot can advice whether the object of interest is one of the correct candidate objects to be assembled next or not.

D. Cooperative Action 2: Settlement of Hesitation

"Hesitation" is a common state appeared during LEGO assembly tasks. The subject knows a specified function on the incomplete construction at hand must be connected with a part in the environment, but cannot be sure which one is required. If the robot notice this state and also notice which function the subject is looking for as a counterpart, the robot can guess the right part that is what the subject thinks in mind.

In this research, the robot knows the final configuration, i.e. the task model. So, the correct object to be assembled at a certain time is also known. However, because the branching task model is employed, multiple correct answers can exist in some cases. For this reason, when the subject is in "Hesitation" state, the correct answer depends on which "Functions" he/she is trying to find. To estimate the true candidate, we employ the history of gaze movement.

1) Estimation of Intended Object based on Voting from Fixation History: First of all, we try to distinguish the following two situations; (1) the subject is unable to determine the next part possibly caused by partial loss of his/her memory, (2) The subject is just looking for the already determined object in the scene. Only in the former case, the cooperative action serves well, but the robot further needs to know the intended part which will be assembled with the incomplete construction at the subject's hand.

When the subject cannot determine the next part which should be assembled to "Function(A)" on the incomplete construction, we assume that he/she looks for a part which has connectable "Function" to "Function(A)" and then compares the part with the plan in his/her memory to determine whether this is the right part. If we focus on gaze movement, the gazed "Function(B)" of "Object" at a certain fixation period means that the subject is looking for a part which is connectable to the counterpart of "Function(B)" on the incomplete construction. If the counterpart of "Function(B)" is known, the right part can be estimated from the task model and it can be presented to the subject as a cooperative action.

Here, we will explain in detail how to estimate the intended part by utilizing the task model and gaze information. Suppose we have an incomplete construction and other parts in the environment as shown in Fig. 5.

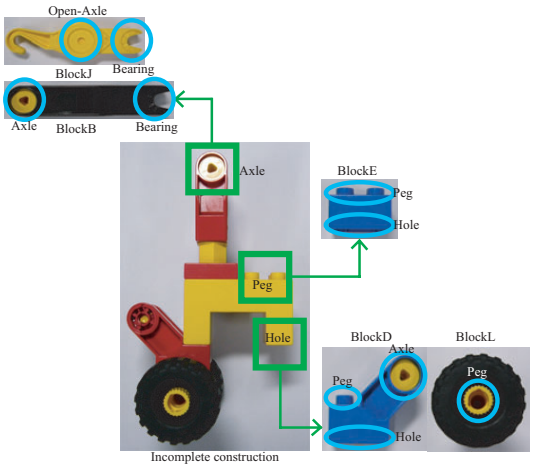

Fig. 5. Incomplete Construction and Parts in Progress

Based on the task model, the candidate parts at this time is one of "BlockB," "BlockD" and "BlockE." We try to estimate the intended part by using voting mechanism based on the "Functions" of the gazed part. A gazed part can be classified into following 2 types.

1) Correct candidate parts which is connectable to the incomplete construction as described in the task model ("BlockB","BlockD","BlockE")
2) Wrong candidate parts ("BlockJ","BlockL")

First, the robot extracts the connectable "Functions" to one of the "Functions" on the incomplete construction among the "Functions" on the gazed part. Then, the robot votes for all the parts in the environment which have one of the extracted "Functions." For example, when "BlockD" is gazed, the not-yet-connected "Functions" on the incomplete construction is Axle, Peg and Hole(marked in squared box in Fig. 5) and the counterpart of those in the gazed part are "Peg" and "Hole." So if "BlockD" is gazed, the subject is expected to look for a part which will be connected to one of "Peg" or "Hole" on the incomplete construction. Then the robot extracts the parts which have the counterpart "Function" from the correct candidate parts list and votes for "Peg" and "Hole" by one for each of the extracted parts, e.g. "BlockD" and "BlockE" in this case.

During searching state, this voting process is repeated while gaze transition continues from one part to the other. If the number of votes of an part is greater than a certain threshold, the part is estimated as the intended part. The threshold value is determined from the measured gaze movement data.

Under the framework of accumulating votes over a constant value, the cooperative action starts only when the searching time is long enough, i.e. sufficient number of gaze transitions is counted. The estimated object is always the right answer on the task model. Fig. 6 shows an example of voting process.

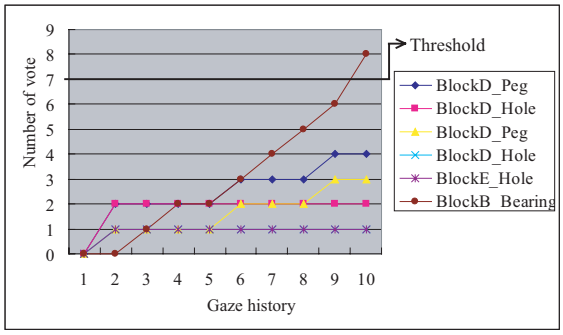

Fig. 6. Analytical Result from Voting on Functions

2) Proposed Cooperative Action: The decision making process of this cooperative action is summarized as follows.

1) *at i-th gaze transition, object O_k is gazed*
2) *vote* $\forall j\{N_j = N_j + 1 |$ *func*$(O_j) \subset$ *c_func*(*func*$(O_{const}) \cap$ *c_func*(*func*$(O_k))\}$
3) *if* $\exists i N_i > N_{thr}$ *then* present O_i to the subject

where *func*(O_k) means the not-yet-connected function set O_k has. O_{const} is the incomplete construction. *c_func*(f) means the function set of the counterpart of function set f.

If the number of votes N_i of object O_i reaches a threshold N_{thr}, the subject is considered to be in "Hesitation" state and the object O_i is presented to the subject by a robot as a cooperative action to settle the "hesitation".

E. Cooperative Action 3: Simultaneous Execution

Consider the situation where the subject is working on an assembly. If the subsequent assembly action is estimated and the robot can execute a part of it simultaneously, the

task will proceed much more efficiently. Further, if an assemble (A) is always accompanied with another assembly (B), "Simultaneous Execution" of assembly (B) is realized when the current assembly is estimated as assembly (A) before it is completed.

1) Utilization of Gaze Movement: Simultaneous execution can be possible based on the task model, however if the same parts in the scene may lead to different assembly patterns as in Fig. 7 and both of them are used in the task model, it is impossible to determine which one is intended from the task model. In this case, gaze movement can help the estimation.

Fig. 7. Different Assembly between the Same Parts Pair

The subject usually gazes both "Functions" to be connected with before assembly as in Fig. 8 and Fig. 9.

Fig. 8. Attention Point ex.1 before Assembly

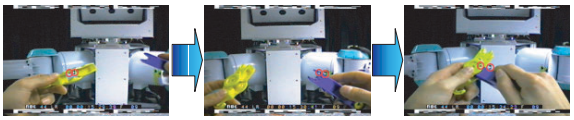

Fig. 9. Attention Point ex.2 before Assembly

So, by investigating the gazed "Functions" just before the assembly starts, the type of assembly(A) can be estimated. If we know that an assembly(B) always follows the assembly(A), the robot can realize simultaneous execution of assembly(B).

2) Proposed Cooperative Action: Simultaneous execution of the subsequent action is performed as follows.

1) The assembly pattern is estimated by investigating the gazed "Functions" before the subject completes the assembly.
2) If the subsequent action is uniquely determined, the robot executes it simultaneously while the subject is working on the current assembly.

IV. Implementation of Cooperative Action and Verification Experiment

A. Experimental Platform

Among other behavior of human, we focus on manipulation tasks and our purpose is to realize the integration of learning process and reproduction process of manipulation tasks in a real platform which employs dexterous hands and

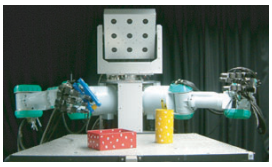

Fig. 10. CVL Robot

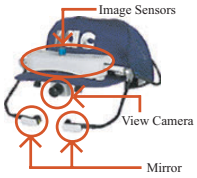

Fig. 11. EMR-8

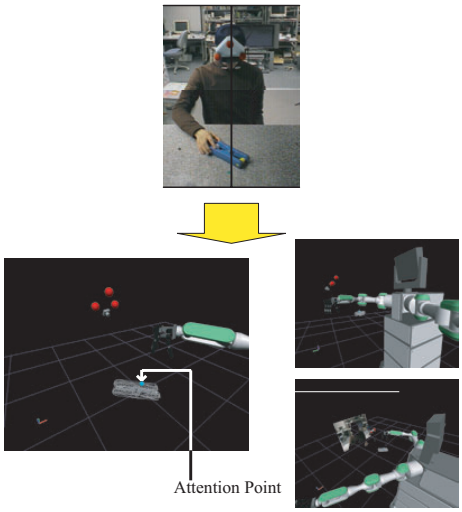

Fig. 12. Visualization of Attention Point in Virtual Space

high level vision system . For that purpose, a humanoid-type robot (Fig. 10) which has the similar capabilities to humans upper body has been developed. In this paper, this platform is used to realize cooperative tasks between human and robot,

To measure gaze movement, a gaze-tracking system, Eye Mark Recorder (EMR-8) Fig. 11, is employed.

We have developed a real-time 3D gaze tracking system by integrating the vision system of the robot and EMR-8, which can visualize the gazed point in the integrated 3D space (Fig.12). With this system, we can measure the 3D position of the gazed location in the same coordinates frame of the object recognition system, so that the gazed object can be easily identified.

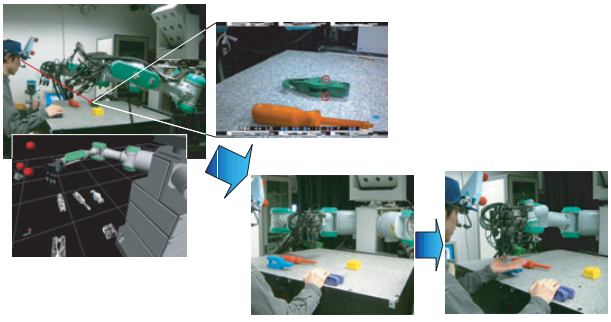

Fig. 13. Experimental Result: Taking Over

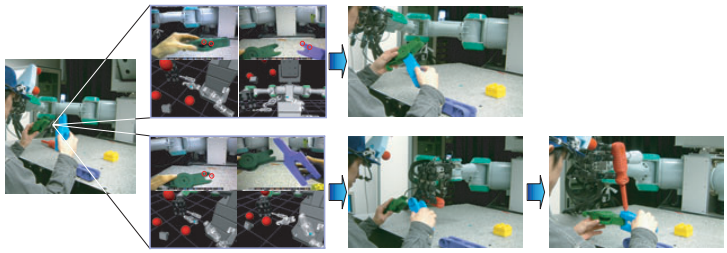

Fig. 15. Experimental Result: Simultaneous Execution

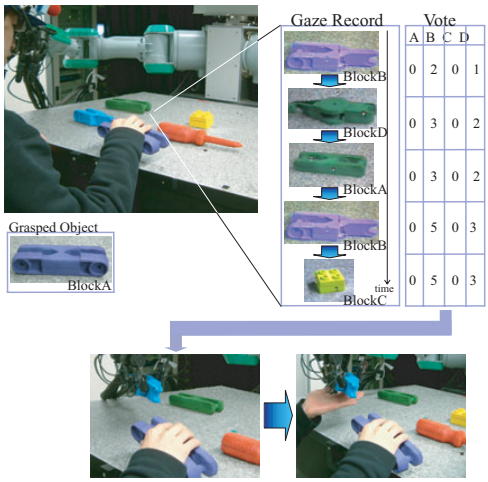

Fig. 14. Experimental Result:Settlement of Hesitation

B. Implementation of Cooperative Action

1) Cooperative Action 1: Taking Over: Fig. 13 shows the experimental results, in which the robot passed the object which had been gazed over a certain period of time during searching state.

2) Cooperative Action 2: Settlement of Hesitation: The experimental result is shown in Fig. 14. The history of gaze movement is obtained. The upper-right of Fig. 14 shows the record of accumulated vote for each object. When either of the number of vote exceeds the threshold, the subject is considered to be in "Hesitation" state and the robot passed "BlockB(Light Blue)" to the subject in this case.

3) Cooperative Action 3: Simultaneous Execution: Assembly of "Shovel" and "BlockB" is selected. There are 2 possible patterns to assemble these 2 parts.

1) Shovel:Bearing $\Longleftrightarrow$ BlockB:Axle (Fig. 7Right)
2) Shovel:Open-Axle $\Longleftrightarrow$ BlockB:Bearing (Fig. 7Left)

Assembly 1) requires screwing using 'Driver" immediately after this action, while assembly 2) does not.

Fig. 15 shows the experimental result. The upper row shows the assembly of "Bearing" and "Open-Axle." This does not require the screwing using "Driver," so the robot does nothing. Meanwhile the lower row shows the assembly of "Bearing" and "Axle." This requires a screwing action, so the robot tries to grasp the "Driver" simultaneously when the subject is doing assembly, and passes the "Driver" to the subject immediately after he/she finishes the assembly.

V. Conclusion

To ensure the flexible cooperative task between human and robot, a branching task model is introduced to represent an assembly task. Under this task model, human worker can freely choose the next assembly action from the possible candidates. In this case, the robot has to determine which action the subject is intended to take next during the process of thinking and has to take an appropriate cooperative action without delay when a situation occurs where the subject is unable to advance the task smoothly. For that purpose, we propose a method to estimate the intention and state of human working on an assembly task from the recent history of gaze movement. We also propose 3 types of cooperative actions, "Taking Over," "Settlement of Hesitation" and "Simultaneous Execution" to deal with 3 typical blocking situations. These methods are implemented on our gaze-tracking system and humanoid robot system, and experimental results are presented.

Acknowledgment

This work is supported in part by the Japan Science and Technology Agency (JST) under the Ikeuchi CREST project, and in part by the Grant-in-Aid for Scientific Research on Priority Areas (C) 15017222 of the Ministry of Education, Culture, Sports, Science and Technology.

References

[1] Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching. *IEEE Trans. Robotics and Automation*, 10(6):799–822, 1994.
[2] K. Ikeuchi and T. Suehiro. Toward an assembly plan from observation part i: Task recognition with polyhedral objects. *IEEE Trans. Robotics and Automation*, 10(3):368–384, 1994.
[3] K. Ogawara, J. Takamatsu, H. Kimura, and K. Ikeuchi. Extraction of essential interactinos through multiple observations of human demonstrations. *IEEE Transactions on Industrial Electronics*, 50(4):667–675, 2003.
[4] H. Kimura, T. Horiuchi, and K. Ikeuchi. Task-model based human robot cooperation using vision. In *Int. conf. on Intelligent Robots and Systems*, volume 2, pages 701–706, 1999.
[5] N. Mukawa, A. Fukayama, T. Ohno, M. Sawaki, and N. Hagita. Gaze communication between human and anthroporphic agent -its concept and examples. In *10th IEEE Int. Workshop on Robot and Human Communication (ROMAN) 2001*, pages 336–370, 2001.
[6] K. Talmi and J. Liu. Eye and gaze tracking for visually controlled interactive stereoscopic displays. *Signal Processing: Image Communication*, 14:799–810, 1999.
[7] Y. Matsumoto and T. Ogasawara T. Ino. Development of intelligent wheelchair system with face and gaze based interface. In *10th IEEE Int. Workshop on Robot and Human Communication (ROMAN) 2001*, pages 262–267, 2001.