Task-model based human robot cooperation using vision

Conference Paper · February 1999

DOI: 10.1109/IROS.1999.812762 · Source: IEEE Xplore

Katsushi Ikeuchi
The University of Tokyo

Task-Model Based Human Robot Cooperation Using Vision *

Hiroshi Kimura and Tomoyuki Horiuchi
Graduate School of Information Systems, Univ. of Electro-Communications
1-5-1 Chofu-ga-oka, Chofu, Tokyo 182

Katsushi Ikeuchi
Institute of Industrial Science, Univ. of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo 106

Abstract

In order to assist the human, the robot must recognize the human motion in real time by vision, and must plan and execute the needed assistance motion based on the task purpose and the context. In this research, we tried to solve such problems. We defined the abstract task model, analyzed the human demonstration by using events and an event stack, and automatically generated the task models needed in the assistance by the robot. The robot planned and executed the appropriate assistance motions based on the task models according to the human motions in the cooperation with the human. We implemented the 3D object recognition system and the human grasp recognition system by using the trinocular stereo color cameras and the real time range finder. The effectiveness of these methods was tested through an experiment in which the human and the robotic hand assembled toy parts in cooperation.

1 Introduction

For the purpose of the child care and nursing care, we are developing the robot which can assist the human in a cooperative task. In order to assist the human, a robot must autonomously recognize the human motion in real time. The vision is the most useful sensor for this purpose.

As an example of a cooperative task, we take up the assembly of toy parts by the human and a robot hand. The robot recognizes the current target objects and the human grasp by vision, and must plan and execute the needed assistance motion based on the task purpose and the context. In this study, we are constructing the human robot cooperation system in which a robot can recognize the motion of the human without any special facilities such as a 'data glove'. For this purpose, we implemented the 3D object recognition system and the human grasp recognition system by using the trinocular stereo color cameras and the real time range finder.

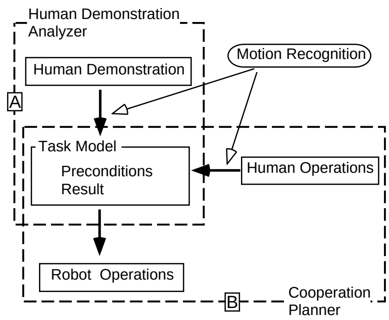


Fig.1 Overview of task model based cooperation

Even though we can prepare programs for several particular patterns of human robot cooperation, we need some frameworks for autonomous assistance by robot in general. In addition, in order to cope with so much variety of human robot cooperation, we need some teaching or learning mechanisms within the framework. In this study, the assistance operations by a robot are generated based on a task model. Task models are created by observation of the human demonstration (Fig.1-A) with vision. For this purpose, we defined the abstract task model and implemented events and an event stack mechanism. We also implemented the mechanism for making the robot select the appropriate assistance motions according to the human motions based on task models (Fig.1-B). The experiments showed the effectiveness of our task-model based human robot cooperation system.

*Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems Kyongju, pp.701-706, 1999.10.

Vision-based cooperation between the human and a robot[1, 2] and between robots[3, 4] had been studied. The automatic generation of the robot motions by observation of the human motions[5, 6] had been studied. But how to generate cooperative motions of a robot automatically has not been considered.

2 Robot System

The robot system is shown in Fig.2. The robot hand is mounted on the 7-dof arm. The robot hand with three fingers manufactured is controlled by VxWorks. Each finger has three joints and 6-axes force/torque sensor. The two sets of trinocular stereo color cameras are connected to the color tracking vision boards and the real time range finder are used for the vision system. The real time range finder can generate 24x24 depth image in video rate.

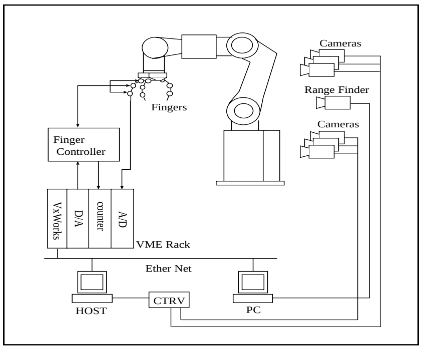
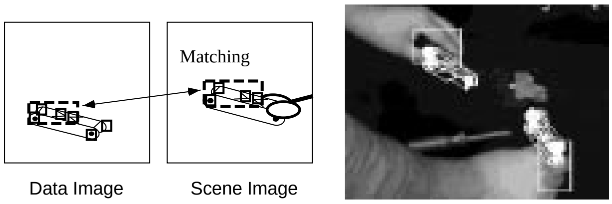


Fig.2 Robot system

3 Recognition of Target Object and Human Grasp

In this study, we implemented 3D object recognition method on a 2D image referring to [7] and 3D position measurement of minute templates by subpixel stereo[8] (Fig.3).

Although the recognition of the human hand gesture has been actively studied on in the field of the human interface, the online 3D recognition of the human hand grasp is still difficult. But as the human uses various grasps according to a task[9], the grasp information is important in recognition of assembly motions. Therefore, we implemented the human hand configuration recognition system, which cannot detect the angles of finger joints but can classify the grasps into several types such as precision grasp, power grasp and so on[6]. In addition, this system can recognize



(a) (b)

Fig.3 (a) Templates cluster matched between data image(left) and current scene(right) by considering correlation and geometric constraints. (b) The geometric models of recognized objects could be overlapped on the scene image.

"request by hand" motion of the human, which is used in **5.**.

In our vision system, the trinocular stereo cameras is not suitable for the real time recognition of the human hand configuration and the real time range finder cannot cover the wide area. Therefore, we combine these two sensors for the online recognition of the human hand configuration. The process of recognizing the human hand configuration and detecting the finger tips are shown in Fig.4[8].

4 Analysis of Human Demonstration

4.1 Definition of Task Model

In this paper, we assume the following attributes about target objects are given:

- frames of the functions such as a hole, a shaft, etc. in the frame of a target object.

We define the following notations about the frames of the functions of a object and operations by the human hand or a robot hand.

- 'obj(a).f(i)' means the i-th **function** of the object 'a'. In the case that the function is not specified, '.f(i)' is eliminated.
- 'operate obj(a).f(i) {to obj(b).f(j)}' means an **operation** to a function or functions. 'obj(a).f(i) {& obj(b).f(j)} operated' means a **state** of a function or functions caused by the operation. The part '{}' means that the operation affects two functions.

For examples, 'obj(a) grasped' is the state 'grasped' of the unspecified function of the object 'a'. 'obj(a).f(i) & obj(b).f(j) fixed' is the sate 'fixed' of paired functions such as the shaft-hole pair, the driver-screw pair and

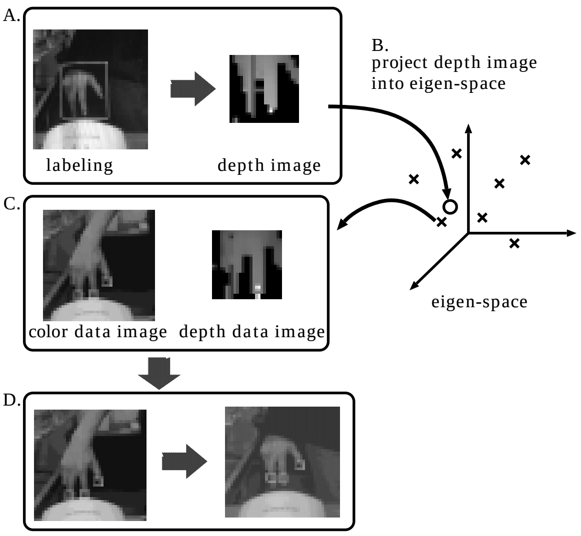


Fig.4 Recognition process of the human hand configuration and positions of finger tips. When the human hand area found by using color labeling gets into the area covered by the range finder, a 16x16 depth image is taken:(A). The human hand configuration is recognized by projecting this depth image into the data images eigen-space:(B). As a result, the depth data image matched with the current hand configuration are obtained:(C). The areas of finger tips registered on the color data image:(C) are searched on the current scene image:(D). The 3D positions of the found finger tips are measured:(D).

two faces attached. The state 'fixed' is translated into the following states considering the coincided points and the aligned axes of the frames of functions.

obj(a).f(i) & obj(b).f(j) fixed
    obj(a).f(i).point   & obj(b).f(j).point   coincided
    obj(a).f(i).axis(k) & obj(b).f(j).axis(l) aligned

Under such definitions, the **abstract task model** consists of the following two elements (Fig.5-(a)):

- the **result** as the state of a function or the state of fixed functions of two objects,
- the **preconditions** for the result.

The preconditions are also states of functions. For an example, the fixation of a hole and a shaft is the precondition of the screwing of the shaft by a driver.

4.2 Generation of Task Model

For the each scene of the human demonstration, the 3D recognition of objects (toy parts and tools) and the human hand configuration is executed by the vision system described in **3.** and the relations between functions of objects and the human hand grasping are analyzed. The results of the 3D recognition and the analysis are shown in Fig.6. The areas of the hands grasping a object are shown on each image.

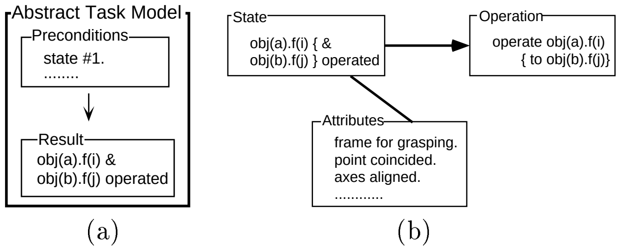


Fig.5 (a):Abstract task model consists of the preconditions and the result. (b):Operation is generated from state using its attributes.

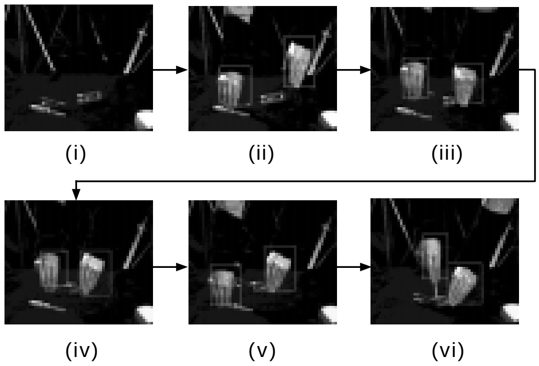


Fig. 6 Example of analysis of human demonstration: parts axes fixation (iv) and screw hole-tip fixation (vi) were detected.

The **human demonstration analyzer** (Fig.1) is able to execute the following procedures about the states of the toy parts and tools.

(a) The analyzer generates an **event** from the newly recognized state. An event has the same notation as a state.

(b) The analyzer recognizes the states such as 'obj(a) grasped' and 'obj(a).f(i) & obj(b).f(j) fixed'.

(c) The analyzer calculates parameters in order to generate the robot operation from the recognized state, and registers the parameters as attributes into the state (Fig.5-(b)). For an example, the axes which should be aligned in the operation 'fix obj(a).f(i) & obj(b).f(j)' are extracted from the state 'obj(a).f(i) & obj(b).f(j) fixed'.

By using the above described procedures, the human demonstration analyzer generates task models. For examples, the events and task models generated by the analyzer for demonstration in Fig.6 through the following process are shown in Fig.7.

1) to push events detected at the present scene onto the stack called an **event stack** (Fig.7-(a)),
2) to generate the task model when the state 'fixed' appears at the top of the event stack,
   The state 'fixed' becomes the result in the task model. Other states below the state 'fixed' on the event stack become the preconditions in the task model (Fig.7-(b)-1, (c)).
3) to return to 1).

As results, the task model #1 and #2 are generated in Fig.7. The whole tasks demonstrated by the human can be expressed by a series of task models shown in Fig.8, where the result of task_model(i) (i-th task model) becomes one of preconditions of task_model(i+1).

5 Cooperative Assistance by Robot

5.1 Planning of Cooperative Assistance

The **cooperation planner** (Fig.1) analyzes the human operations by using events and the event stack in the same way as the human demonstration analyzer. The cooperation planner is able to execute the following procedures in addition to the procedures (a)-(c) described in **4.2.**.

(d) The planner generates an operation from a state and its parameters registered as attributes (Fig.5-(b)). That is, the planner generates the trajectories of the robot arm and fingers. The generated operations are put into a first-in-first-out buffer called a **OP-FIFO**. The robot is able to remove an operation from OP-FIFO and execute it.

(e) The planner confirms the event which should occur by the robot operation by using vision system and pushes the confirmed event on the event stack.

The cooperation planner generates assistance operations of robot comparing events on the event stack and preconditions of task models. One of problems in

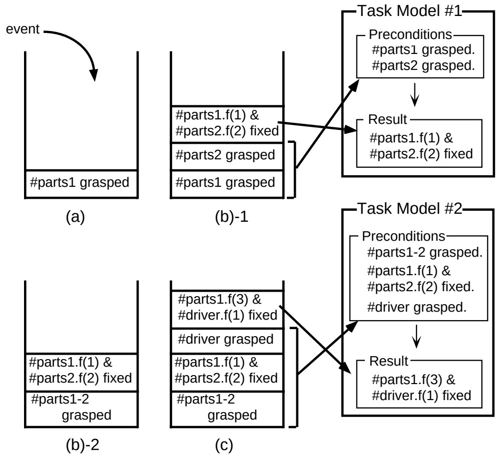


Fig.7 Events, event stack and task models generated by the analyzer for demonstration in Fig.6: (a), (b)-1,2 and (c) are correspondent with (ii), (iv) and (vi) of Fig.6, respectively.

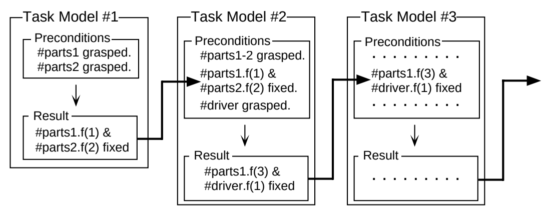


Fig.8 Series of task model generated by the analyzer for demonstration in Fig.6.

this planning is how to divide preconditions and result in a task model into the states caused by the human operations and the states which should be caused by the robot operations (Fig.1-B). Therefore, the mechanism by which the robot can select the appropriate assistance motion according to the human motion at each stage is necessary.

In this study, we consider following three types of assistant motions and make the robot select appropriate assistance motion by using series of task models and the event stack.

(A) the assistance in order to realize the purpose of the task
   When the human executed a part of preconditions of task_model(i) and a new event has not been detected for constant period, the planner infers that the assistance by the robot is needed in order to realize the result of task_model(i) and puts preconditions and re-

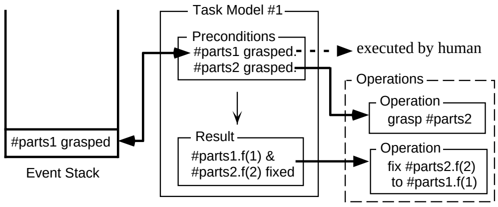


Fig.9 The assistance operations in order to realize the purpose of the task were generated.

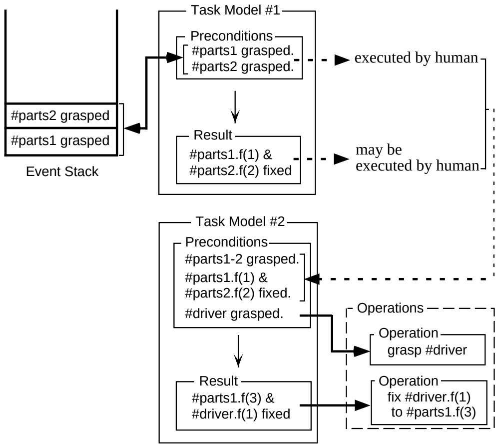


Fig.10 The assistance operations in order to execute a series of tasks efficiently were generated.

sult of task_model(i) into OP-FIFO except for states on the event stack which was executed by the human (Fig.9).

(B) the assistance in order to execute a series of tasks efficiently

When the human executed all preconditions of task_model(i), the planner infers that the result of task_model(i) will be executed by the human and puts preconditions of task_model(i+1) into OP-FIFO except for preconditions and the result of task_model(i) (Fig.10) as preparation for the next task in order to improve efficiency of cooperation. In addition, the planner puts the result of task_model(i+1) into OP-FIFO in order to realize the purpose of this task.

(C) the assistance by passing the grasping object to the human hand

When the human requested the robot hand grasping object by hand motion, the planner infers that the fixation using the object will be executed by the human and makes the robot hand pass the object to the human hand. In addition, the operation for the fixation using the object is removed from OP-FIFO.

5.2 Experiments

The scenes of cooperation experiments are shown in Fig.11. A series of task models shown in Fig.8 was used in this experiments.

In this experiments, six patterns of cooperation were executed according to the human motions, even though only one series of two task models was used. The meanings of typical patterns are as follows:

(a)→(b)→(c)→(d)

Since the human had grasped both #parts1 in (a) and #parts2 in (b), the robot grasped #driver in order to prepare for the next task model #2 in (c) as assistance-(B) described in **5.1.**. Since a new event had not been detected for constant period after #parts1.f(1) and #parts2.f(2) fixation event in (c), the robot fixed #driver to #parts1.f(3) in (d) as assistance-(A).

(a)→(b)→(c)→(e)→(f)

Since a “request by hand” event had been detected in (e), the robot passed #driver to the human hand in (f) as assistance-(C).

(a)→(g)→(h)→(i)→(j)

Since a new event had not been detected after human’s grasping #parts1, the robot grasped #parts2 in (g) and fixed #parts2.f(2) to #parts1.f(1) in (h) as assistance-(A). Since a new event had not been detected after (h), the robot grasped #driver in (i) as assistance-(A). Since a “request by hand” event had not been detected after (i), the robot hand fixed #driver to #parts1.f(3) as assistance-(A).

As a result, appropriate assistance operations were executed by a robot hand in response to the human motions based on the series of task models. Video footage of these experiments can be seen on WWW (*http://www.kimura.is.uec.ac.jp*).

6 Conclusion

In this study, we constructed the framework for the vision-based cooperation of the human and a robot hand taking toy parts assembly as an example. The task models for cooperation were generated by observation of the human demonstration. The operations of the robot assistance were generated based on the task models. We proposed the mechanism by which the robot was able to select the appropriate assistance

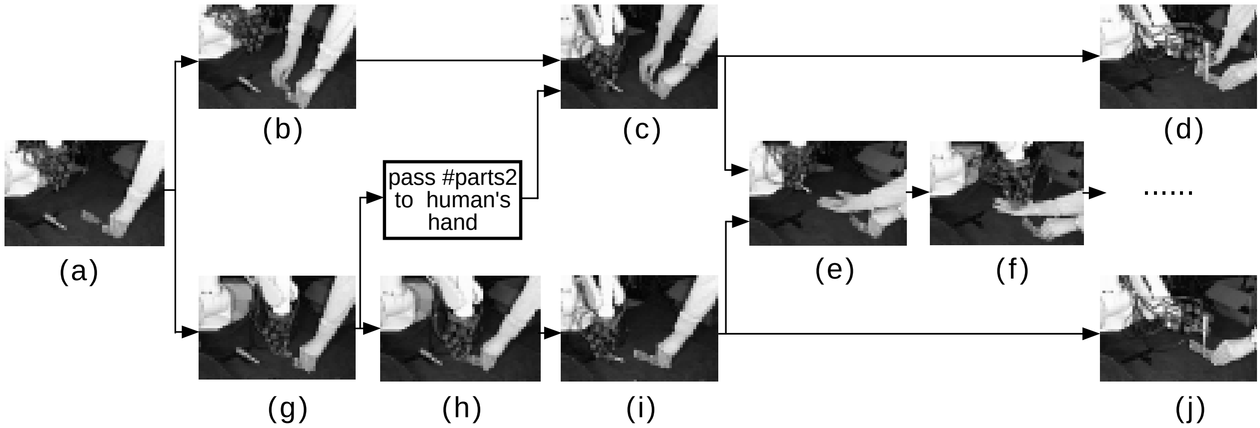


Fig.11 Photos of cooperation experiments based on a series of task models. The meaning of each scene is as follows: (a):The human hand grasped #parts1. (b):The human hand grasped #parts2. (c):The robot hand grasped #driver. The human hand fixed #parts2.f(2) to #parts1.f(1). (d):The robot hand fixed #driver to #parts1.f(3). (e):The human hand requested #driver by hand motion. (f):The robot hand passed #driver to the human hand. (g):The robot hand grasped #parts2. (h):The robot hand fixed #parts2.f(2) to #parts1.f(1). (i):The robot hand grasped #driver. (j):The robot hand fixed #driver to #parts1.f(3). Six patterns of cooperation were executed according to the human motions. Those were (a)→(b)→(c)→(d), (a)→(b)→(c)→(e)→(f), (a)→(g)→(c)→(d), (a)→(g)→(c)→(e)→(f), (a)→(g)→(h)→(i)→(e)→(f) and (a)→(g)→(h)→(i)→(j).

motion according to the human motion at each stage. It must be noticed that several patterns of cooperation appear even for only one series of two task models when we consider passing an object from the robot to the human.

As the future works, when several series of task models are applicable to the current scene and the human motions, the mechanism for selecting an appropriate one is needed. For more flexible cooperation, estimation of human states[10] and error recovery based on it are significant. For the general mechanical assembly, more general definition[11] for functions and states is needed. In addition, the operation control mechanisms using force sensors and skills[12, 13] are needed.

This work was supported by the Research for the Future Program of the Japan Society for the Promotion of Science, 'Cooperative Distributed Vision for Dynamic Three Dimensional Scene Understanding' Project (JSPS-RFTF96P00501).

References

[1] T.Mori, T.Sato and H.Mizoguchi, Situation reactive handiwork support through behavior understanding, Proc. of ICRA97, pp.587-593, 1997

[2] Y.Hayakawa, I.Kitagishi and S.Sugano, Human intention based physical support robot system in assembling work, Proc. of IROS98, pp.930-935, 1998

[3] Y.Kuniyoshi, Behavior matching by observation for multi-robot cooperation, Proc. of ISRR'95, 1995

[4] H.Kimura and H.Katano, Vision-based motion recognition of the hexapod for autonomous assistance, Proc. of IROS98, pp.1-6, 1998

[5] Y.Kuniyoshi, et al., Learning by watching, IEEE Trans. on RA, 10-6, pp.799-822, 1994

[6] S.B.Kang, K.Ikeuchi, Grasp recognition and Manipulative motion characterization from human hand motion sequences, Proc. of ICRA94, pp.1759-1764, 1994

[7] K.Ohba and K.Ikeuchi, Recognition of the multi specularity objects using the Eigen-Window, CMU-CS-96-105, 1996

[8] H.Kimura, et al., Human Robot Cooperation for mechanical assembly using cooperative vision system, Proc. of 2nd Int. Workshop on CDV, pp.213-228, 1998

[9] M.R.Cutkosky and R.D.How, Human grasp choice and robotic grasp analysis, *Dextrous Robot Hands*, Springer-Verlag, pp.5-31, 1990

[10] C.R.Wren and A.P.Pentland, Dynamic models of human motion, Proc. fo IEEE Int. Conf. on automatic face and gesture recognition, 1998.

[11] G.V.Paul and K.Ikeuchi, Modeling planar assembly task: representation and recognition, Proc. of IROS95, pp.17-22, 1995

[12] T.Suehiro and K.Takase, Skill based manipulation system, J. of RSJ, 8-5, pp.551-562, 1990

[13] T.Saito, et al., Assembly task plan generation from observation and constraint states control using force sensor, Proc. of 16th Conf. of RSJ, pp.61-62, 1998