A study toward cognitive action with environment recognition by a learning space robot

Proceedings of the 2003 IEEE
International Conference on Robotics & Automation
Taipei, Taiwan, September 14-19, 2003

A Study toward Cognitive Action with Environment Recognition by A Learning Space Robot

Kei SENDA*
Department of Mechanical Systems Engineering
Faculty of Engineering, Kanazawa University
senda.k@t.kanazawa-u.ac.jp

Tsutomu MATSUMOTO and Yuzo OKANO
Former Students
Graduate School of Engineering
Osaka Prefecture University

Abstract — **This paper addresses an experimental system simulating a free-flying space robot, which has been constructed to study autonomous space robots. The experimental system consists of a space robot model, a frictionless table system, a computer system, and a vision sensor system. The robot model is composed of two manipulators and a satellite vehicle, and can move freely on a two-dimensional planar table, without friction, using air-bearings. The robot model has successfully performed the automatic truss structure assembly, including many jobs, e.g., manipulator berthing, component manipulation, arm trajectory control collision avoidance, assembly using force control, etc. Moreover, even if the robot fails in a task planned in advance, the robot re-plans the task by using reinforcement learning, and obtains the task goal for basically kinematic problems. But, for a class of complicated dynamic problems, the computational periods and efforts are infeasible for on-line learning. Some approaches are proposed to accelerate the learning speed, which also give models of cognitive actions and approaches to so-called a frame problem. The experiment demonstrates the possibility of the autonomous construction and the usefulness of space robots.**

I. INTRODUCTION

Space robots are necessary for future space projects to construct, repair and maintain satellites and space structures in orbits. Hence, it is an important subject to develop a free-flying space robot consisting of manipulators and a satellite vehicle, which can fly freely in an orbit (this paper calls it just a space robot). Lots of new complicated dynamic problems have been raised, e.g., an interaction between the manipulators and satellite, a structural flexibility caused by lightweight requirements, etc. There exist many papers focused on the dynamic problems [1]–[5], whereas the references cited here are not extensive. Some studies using hardware equipments on the ground have been reported to examine the control and identification methods [5]–[7].

Moreover, studies of autonomous systems, e.g., recognition using force and vision information, planning and reasoning, etc., are necessary to realize the autonomous space robots that can achieve their mission commanded by human operators [8]. The Stanford University has developed an experimantal space robot with low level autonomy that achieves collision avoidance[9]. In addition, the following projects emphasize the present point: the Telerobotics Research Program [7], the space robot technology experiment (ROTEX) [10], the Ranger Telerobotic Flight Experiment [11], and the Engineering Test Satellite-VII (ETS-VII) [12]. As of year, those projects have been almost finished, but there remain many subjects for autonomous space robots. There are many tasks autonomous space robots can accomplish, thus replacing human astronauts. For such autonomous robots, adaptation and learning in real work environment are key issues. Therefore, testbeds are necessary for the research and development.

For that purpose, this study has developed a ground experimental system simulating a free-flying space robot under micro-gravity condition in orbit (Fig. 1) and started researching in the autonomy. Using the system, lots of control techniques make the space robot model assemble a truss structure automatically. In the assembly demonstration, the robot model performs several tasks, e.g., the manipulator berthing, the component manipulation, the arm trajectory control collision avoidance, the assembly using force control, etc. Repeating the sequence would enable construction of large structures.

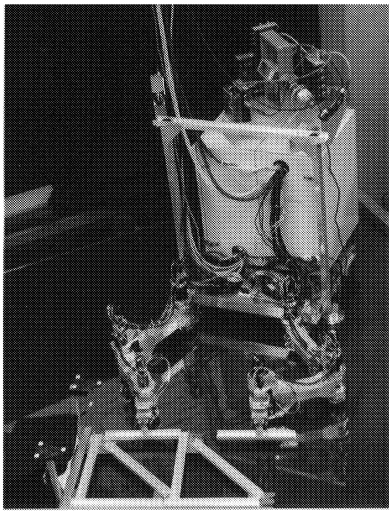

Figure 1: Photograph of space robot model and truss

*Acknowledgement: A part of this work is financially supported by a Grant-in-Aid for Scientific Research from Ministry of Education, Science, Culture and Sports of Japan.

But, the space robot may fail in a task planned in advance because of uncertainties and variations of the worksite. To obtain the task goal, the robot must modify the task suitably for the real work environment. For this purpose, the robot re-plans by using reinforcement learning with trial-and-error processes. The robot experimentally achieves the goal by the re-planned task.

The reinforcement learning is applicable for the basically kinematic problems. For a class of dynamic problems, the computational periods and efforts are infeasible for online learning. To accelerate the learning speed, this paper proposes some approaches. They also give models of cognitive actions and approaches to so-called frame problem obstructing efficient learning and action. The experiment demonstrates the possibility of the autonomous construction and the usefulness of space robots.

The rest of this paper is organized as follows. The experimental system is introduced in the next section. In the third section, the autonomous truss structure assembly is experimentally demonstrated by synthesizing the techniques. The fourth section illustrates the method using reinforcement learning to plan the task-sequence appropriately for the real work environment when the robot fails in the task planned in advance. The fifth section gives some methods to accelerate the reinforcement learning, which is considered as a model of cognitive actions. Some concluding remarks are given in the final section.

II. EXPERIMENTAL SYSTEM

Figure 1 is a photograph of the space robot model and a truss structure under assembly. Figure 2 shows a schematic diagram of an experimental system constructed in this study. The robot model is supported on the horizontal table without friction by using air-pads. The experimental system simulates a free-flying space robot in orbit while motion of the robot model is restricted in a two-dimensional plane.

Information from the robot model is put into the computer system placed beside the table. In the vision sensor system, the stereo images are taken by the CCD cameras and sent to an image-processing unit. After appropriate process in the image-processing unit, the visual information is sent to the computer system. The computer system processes the sensing data and computes control commands to the robot model.

The robot model consists of a satellite vehicle and dual 3 degree-of-freedom (DOF) selective compliance assembly robot arm (SCARA) type manipulators. A pair of charge-coupled device (CCD) cameras for a stereo-vision and a position/attitude control system are installed on the satellite vehicle. The position/attitude control system consists of four thrusters and a control momentum gyro. The total length from the right hand to the left is approximately 1.7 m and the total mass is about 70 kg.

See [13] for details of the experimental system.

III. TRUSS ASSEMBLY

The fundamental control techniques for the space robot have been developed, e.g., the visual servoing, the position and attitude control of the satellite vehicle, the positioning control of the free-floating space robot, path planning of arms for avoiding collision with the local work environment, force controls considering contact with the work environment, etc. After that, truss assembly experiments

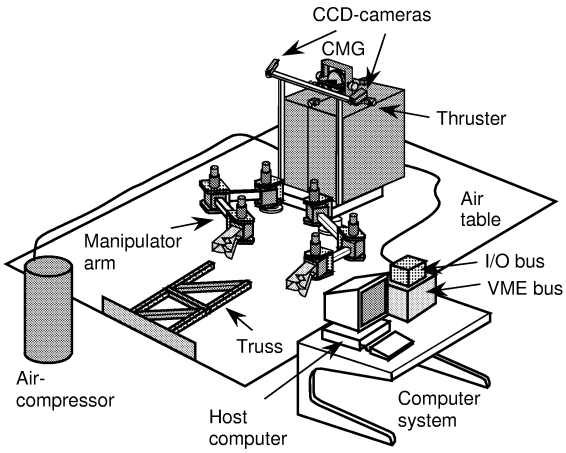

Figure 2: Schematic diagram of experimental system

are conducted. This section represents the experimental results. See [14] for details of the control techniques.

Manipulating a truss component and connecting it to a node precede the assembly. The component is installed in the planned position and direction because the connector at the node has a notch to insert the component. Corners at the notch are planed off to insert the component easily. The installed component would not be detached since the connector has a latch mechanism. The truss is designed as robot-friendly and can be assembled by using one arm.

Figure 3 is a series of photographs of the experimental assembly. An experimental manipulator berthing is shown in Scene (i) of Fig. 3, where the visual servoing with the sensory feedback control for space robots is used. The right manipulator hand is controlled well and the manipulator berthing is successful, whereas the satellite vehicle is moved by the reaction of the arm motion and the disturbance of cables suspended from above. The robot holds on to the worksite by the right arm to compensate reaction force through the assembly. The arm path is planned and the manipulator is controlled to track the obtained path. The robot installs the first component, member 1, during scenes (ii) and (iii). The component installation is performed well by the position-force hybrid control called saturated-proportional and differential feedback (SP-DF) control [15]. The robot installs other members successively and assembles one truss unit from scene (iv) through (vi). Repeating the sequence enables construction of a large truss structures.

The robot-friendly truss is one of the main reasons why the robot has succeeded assembling whereas the vision system has a 2 mm mean measurement error after a hand-eye calibration. However, success is not ensured because of the measurement error.

IV. AUTONOMY WITH LEARNING

In section III, the robot has successfully achieved the truss structure assembly of the task-sequence planned in advance. However, the space robot may sometimes fail in the task because of uncertainties and variations of the work site. To recover from the error and obtain the task goal, the robot must re-plan the task suitable for the real work environment.

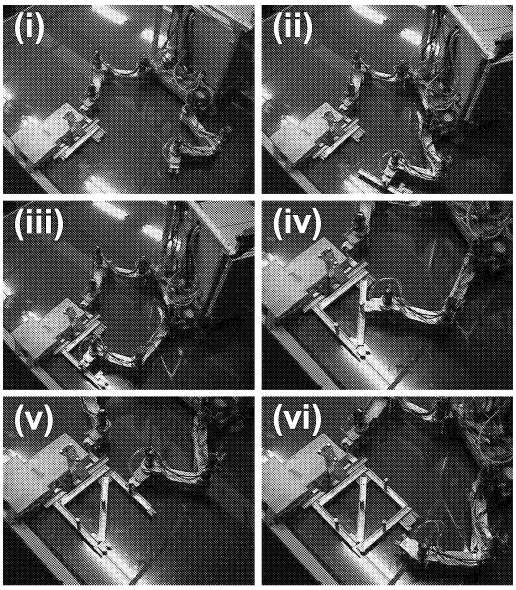

Figure 3: Photograph of truss structure assembly

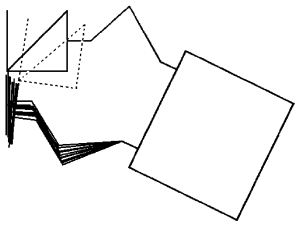

Figure 4: Experimental result of case 1 (dashed line represents the measured position of truss by the vision sensor with measurement error)

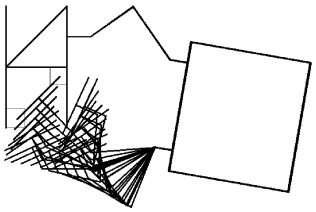

Figure 5: Experimental result of case 2

A. Application of Reinforcement Learning

For the re-planning, one of typical reinforcement learning algorithm, Q-learning [16], is used. The reinforcement learning is used because the robot learns how to do suitably for the real environment so as to maximize a numerical reward that is given by the designer to describe what to do, where the environment cannot be modeled exactly.

Time t, state s, and action a are discretized following a general Q-learning formulation. The Q-learning algorithm estimates the optimal action-value function $Q(s,a)$ through interactions between the robot and the environment with trial-and-error processes. The Q evaluate a at s. During learning, the robot chooses a from s using policy derived from Q. The robot takes action a, observes new state s' and reward r, and updates Q as

$$Q(s,a) \quad \leftarrow \quad (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a')]$$

where α $(0 < \alpha \leq 1)$ and γ $(0 < \gamma \leq 1)$ are a learning rate and a discount rate, respectively. It has been shown that the estimated Q converges to the optimal if the system is modeled as a finite Markovian decision process and all actions are chosen enough times. To choose the action appropriately through learning, this study uses the ϵ-greedy policy [16] where any action is selected randomly with probability ϵ, otherwise the optimal action is chosen by using the current estimated $Q(s,a)$.

B. Autonomous Actions with Learning

1) Case 1: Compensation for Measurement Error:: The bad lighting condition in the space environment often yields measurement error in visual sensing. Consider a situation that the space robot fails in scene (iii) of Fig. 3 because of the measurement error. Let the robot acquire suitable actions for the situation by using the reinforcement learning with trial-and-error process. Task achievement is examined by sensor information of joint angles and applied forces because the component is not moved when it is installed in the right node and latched correctly.

A discrete state space with $9^3 = 729$ states is made for the reinforcement learning, where each coordinate of hand position (x, y) and orientation θ is discretized in 9 states. The x and y coordinates are quantized every 0.02 m and θ every 2.0°. Each state has $2 \times 3 = 6$ actions that are one-step movements of discrete coordinates to the neighbor. Parameters in (1) for updating Q are $\alpha = 0.1$, $\gamma = 0.6$, and $r = 10$. For the ϵ-greedy policy, $\epsilon = 0.1$ is initially used and reduced gradually to be the policy deterministic. The manipulator is controlled by the force control for contact situation.

Figure 4 is the graphic of the experimental robot motion. After an adequate trial-and-error process, the robot obtains adaptive action suitable for the measurement error. The learning method enables it to accomplish the task of compensating the measurement error. This action with learning can be an approach to the sensor-fusion problem.

2) Case 2: Adaptive Action to New Environment:: Consider a situation that the diagonal element of scene (iv) is lost after scene (v) during the truss structure assembly sequence of Fig. 3. In this situation, the robot cannot assemble the diagonal element into the truss structure by the sequence planned in advance because the element attached in step 3 becomes an obstacle. In the previous section, manipulator path is planned by the artificial potential method. But, the artificial potential method is not suitable for the on-line planning, because it needs more computation cost as the work environment gets more complicated.

A discrete state space with $15^3 = 3375$ states is made for the reinforcement learning, where each coordinate of hand position (x, y) and orientation θ is discretized in 15 states. The x and y coordinates are quantized every 0.05m and θ every 10°. Other conditions are the same as case 1.

Figure 5 is the experimental robot motion planned by the reinforcement learning. The generated task-sequence in the state space is illustrated in Fig. 6. The learning method enables to accomplish the task avoiding collision against the environment.

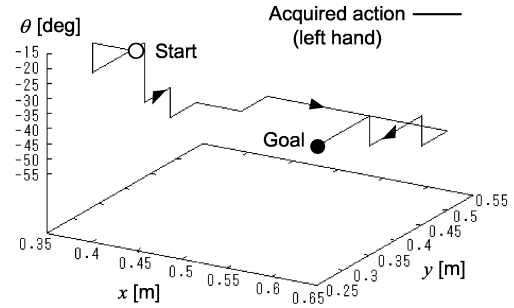

Figure 6: Acquired action of position and orientation of left hand

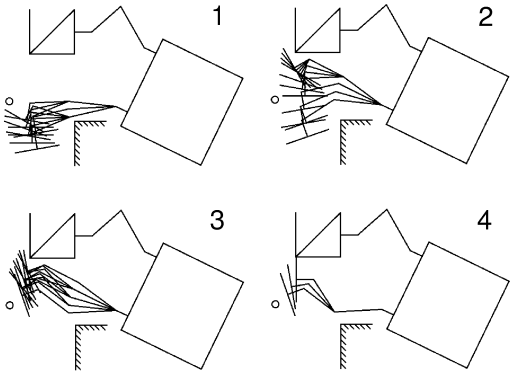

Figure 7: Experimental action acquired to avoid collision

Most path planning methods generate a path from the initial state to the desired state. On the other hand, the reinforcement learning estimates the optimal action-value function Q for all states that derives a policy to choose the best action. Therefore, the robot can take the best action at any state in the environment after the optimal Q is estimated.

3) Case 3: Complicated Obstacle Avoidance:: Consider a situation that the robot cannot assemble a component into the truss structure by the sequence planned in advance due to unexpected obstacles that interfere with the manipulator motion. Figure 7 is the graphic of this experimental robot motion, where the suitable action for the environment is obtained by the reinforcement learning using the same conditions of case 2. The learning method enables the robot to acquire such a complicated action to avoid collision with the obstacles in the environment.

C. Evaluations and Discussions

For the above three cases, Table 1 shows the step numbers of the trial-and-error process, the episode numbers, and the periods for learning convergence, where an episode is a process from start to goal. The computation periods are measured by Pentium II 266 MHz CPU for numerical simulations using modeled environments. From case 1 to case 3, the learning method needs the longer period as the environment becomes more complicated. The computation periods are within a few tens of seconds and the learning method is feasible for the class of problems.

Here, the learning method acquires actions for the basically kinematic problems. For a dynamic problem, it needs a larger number of states and actions to treat the state space with a higher dimension and to model the interaction between the robot and the environment. For this class of dynamic problems, the computational periods and efforts are infeasible for on-line learning. An approach to this class of problems is still an open problem.

Table 1: Computation numbers and periods for convergence of learning

	Case 1	Case 2	Case 3
State no.	729	3,375	3,375
Trial & error no.	33,666	238,092	154,043
Episode no.	1041	854	886
Period [s]	3	15	25

V. COGNITIVE ACTION

A. Models of Cognitive Actions

Investigations of skilled human operators point out a change of "observation". At the beginning, the operators must recognize, plan, choose from actions, etc. and difficult to work quickly. As the persons repeat working, they skip the internal processes relating the environment recognition with much effort, and their environmental observation change to indicate efficient and right action. This can be considered that a knowledge-based behavior changes to a rule-based or skill-based behavior in Rasmussen's model and amount of the information process reduces[17]. The efficient observation is similar to feature-based action[18]. It is called co-provision with a dual-loop feedback structure that the environmental observation provides and organizes behavior and the resultant behavior provides observation again[19]. In the following sections, the change of observation and the co-provision are modeled as the selection of state variables, the categorization of state, and the use of categorized state space. They are also approaches to the frame problem[20] using recognition.

There are some studies to identify the environment[21]. They relate to this example, but direction is different.

B. Formulation by Reinforcement Learning

As shown in Fig. 8, considered here is a task where the 3-link SCARA type manipulator places the component and presses it against the corner of walls in desired direction and force to assemble.

This is simplified from the Peg-in-Hole task and no friction is contained for simplicity. Visual information is not used on the grounds that the robot uses only the forces at hand and joint angles in the final assembling because vision measurement error is not ignored.

As a solution for the reinforcement-learning problem, Q-learning[16] is employed Its formulation is base on a finite discrete space, where time, state, and action are discretized as well as general Q-learning. The system state is defined as follows. Convergence of learning is guaranteed only if the system's state space is constructed so as to determine its future state relating the task form current state and action. Hence it is reasonable to use the state variables of the equations of motion of the robot manipulator with the geometric endpoint constraint as:

$$[x, y, \alpha, \dot{x}, \dot{y}, \dot{\alpha}, f_x, f_y, n_z] \quad (1)$$

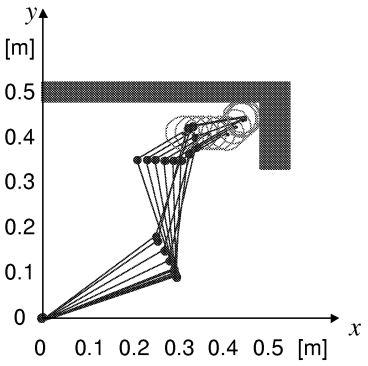

Figure 8: Experimental result following an acquired motion

where they are x, y positions, $\alpha = \theta_1 + \theta_2 + \theta_3$ direction of hand and component, their velocities, and applied forces in x, y, α directions from the environment, respectively. They all are measured form the sensors.

The following controller is used to generate actions:

$$\boldsymbol{\tau} = -\boldsymbol{J}^T \boldsymbol{K}_P (\boldsymbol{y} - \boldsymbol{y}_r) - \boldsymbol{K}_D \dot{\boldsymbol{q}} \tag{2}$$

where $\boldsymbol{\tau}$ is control input to the manipulator, $\boldsymbol{J} = \partial \boldsymbol{y} / \partial \boldsymbol{q}^T$ Jacobian matrix, $\boldsymbol{y} = [x, y, \alpha]^T$ manipulation variable vector, $\boldsymbol{y}_r$ reference of $\boldsymbol{y}$, $\boldsymbol{q}$ joint variable vector, $\boldsymbol{K}_P$ and $\boldsymbol{K}_D$ feedback gain matrices, respectively. For the reference manipulation variable $\boldsymbol{y}_r^{(i)}$ at time i, $\boldsymbol{y}_r^{(i+1)}$ is given by $\boldsymbol{y}_r^{(i+1)} = \boldsymbol{y}_r^{(i)} + \delta \boldsymbol{y}_r^{(i)}$. Action at time i is considered as the $\delta \boldsymbol{y}_r^{(i)}$. The robot regards as the task having been achieved at a target state, where reword is given. In the target state, all velocities and n_z are zeros, and α, f_x, and f_y are specified values.

In this example, number of state is a few millions because of many degree of the state space. The learning has not been converged in 50 hours by using Pentium III 500 MHz/Matlab since much time is consumed for numerical simulations as well as the may states. One must reduce the states from a point of view of the recognition.

C. Change of Observation and State Space

One may wonder if all state variable in (1) are really needed. The learning has been converged using

$$[x, y, \alpha, f_x, f_y] \tag{3}$$

as state variables. One of the obtained optimal behavior is illustrated in Fig. 8. The optimal behavior is achieved from any initial state after the learning is converged. The state variables are reduced because some of the state variables in (1) are not necessary for the task achievement and the sampling time for learning is longer than the settled time of the control (2). The learning has been converged in 4 hours because the number of states has been reduced by 1/1000. An algorithm with a decision tree is used to find the state variables in (3). It takes 5 hours for the convergence of learning including this state variables finding algorithm.

This is an approach to find the minimum sufficient state space for the learning convergence as well as to ease the frame problem. This is also a model of the change of observation because the notable information in state variables is becoming clear as one is learning.

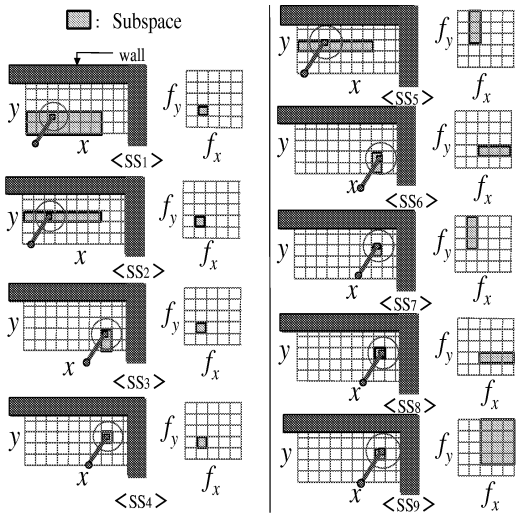

Figure 9: Categorized states

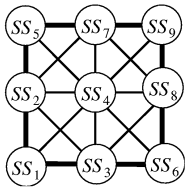

Figure 10: Connection of sub-spaces

D. Categorization and Learning

One can categorize the states by differences of interaction between the manipulator and the environment. There are many states, e.g., the robot move freely without constraint force, it moves freely except x direction because of constraint force f_x, etc. The states are categorized by the interaction as illustrated in Fig. 9

A state is described by a pair of two graphs of the x–y and the f_x–f_y at hand where 4 variables are used in (3). In the figure, each of SS_1 to SS_9 is a set of states belonging to each category. They are subspaces of the entire state space. For instance, the manipulator does not contact to any walls in a state of SS_1, it contacts to the upper wall in SS_5, and SS_2 is a subspace of their border. In each subspace, state transitions from any states by an action are same. Note that the categorization is dependent on the actions that one can take.

Connecting relation among the subspace is illustrated in Fig. 10. The lines between subspace shows existence of actions and the thickness indicates number of actions. The connecting relation enables to decide the action rule through a reinforcement-learning problem where each subspace is treated as a state. In the example in Fig. 8, the manipulator moves from SS_1 to SS_9. In order to achieve the determined action, the manipulator decides its action rule in each subspace through a reinforcement-learning problem whose subtask is the tangent between subspaces. The number of states can be reduced again for learning in each subspace and the learning becomes more efficient. In this example, the number of states is reduced by 1/10 form

that in (3). As a result, the learning has been converged in 30 minutes.

The categorization dependent on the selectable actions can be regarded as the change of observation dependent on selectable skills. The action decision based on the subspace can be considered as the behavior organization followed by the change of observation. Moreover, if the organized behavior with the subtasks of subspace transitions becomes a skill, one can consider the rule-based behavior changes to skill-based behavior. The change from the rule-based to the skill-based may change the observation. The co-provision of observation and action can be modeled in the above. This is also an approach to ease the frame problem.

VI. CONCLUDING REMARKS

This study has demonstrated the autonomous truss structure assembly by the experimental autonomous space robot system. The fundamental techniques have been developed and synthesized for the assembly task, i.e., the stereo image measurement, the visual servoing, the positioning control of free-floating space robot, the arm path planning, and the force control considering contact with the work environment. The robot successfully achieved the autonomous truss structure assembly. Furthermore, the robot re-planned the task-sequence by using reinforcement learning and obtained the goal even when the robot failed in the task-sequence planned in advance. The reinforcement learning was applicable for the basically kinematic problems, whereas it often requires a large number of computation for a dynamic problem. To accelerate the learning speed, some approaches have been proposed. They also give models of cognitive actions and approaches to so-called frame problem obstructing efficient learning and action. As a result, this study has shown a possibility of the autonomous truss structure construction and the usefulness of space robots.

There remain some subjects for autonomous space robots. The approach to the autonomy and/or intelligence is the biggest subject to realize useful space robots. This study has approached this issue by the reinforcement learning algorithm where its application has been still limited.

VII. REFERENCES

[1] Umetani, Y. and Yoshida, K., "Continuous Path Control of Space Manipulators Mounted on OMV", *Acta Astronautica*, vol. 15, no. 12, 1987, pp. 981–986.

[2] Masutani, Y., Miyazaki, F., and Arimoto, S., "Sensory Feedback Control for Space Manipulators," *Proceedings of International Conference on Robotics and Automation*, IEEE, 1989, pp. 1346–1351.

[3] Yamada, K. and Tsuchiya, K., "Efficient Computation Algorithms for Manipulator Control of a Space Robot," *Trans. of Society of Instrument and Control Engineers*, vol. 26, no. 7, 1990, pp. 765–772. (in Japanese)

[4] Murotsu, Y., Tsujio, S., Senda, K., and Hayashi, M., "Trajectory Control of Flexible Manipulators on a Free-Flying Space Robot," *IEEE Control Systems*, vol. 12, no. 3, 1992, pp. 51–57.

[5] Murotsu, Y., Senda, K., Ozaki, M., and Tsujio, S., "Parameter Identification of Unknown Object Handled by Free-Flying Space Robot," *AIAA J. of Guidance, Control, and Dynamics*, vol. 17, no. 3, 1994, pp. 488–494.

[6] Toda, Y., et al., "Development of Free Flying Space Telerobot: Ground Experiments on Two-Dimensional Flat Testbed," *Proc. of Guidance, Navigation and Control Conference*, AIAA, Washington, DC, 1992, pp. 33–39.

[7] Bejczy, A. K., "Toward Advanced Teleoperation in Space," *Teleoperation and Robotics in Space*, AIAA, Washington, DC, 1995, pp. 107–138.

[8] Skaar, S. B. and Ruoff, C. F. (eds.), Teleoperation and Robotics in Space, AIAA, Washington, DC, 1995.

[9] Ullman, M. A., "Experiments in Autonomous Navigation and Control of Multi-Manipulator Free-Flying Space Robots," Ph. D. Thesis, Stanford University, Stanford, CA, 1993.

[10] Brunner, B. et al., "Multisensory Shared Autonomy and Tele-Sensor-Programming ," *Proc. of International Conference on Intelligent Robots and Systems*, Institute of Electrical and Electronics Engineers, New York, 1993, pp. 2123–2139.

[11] David, L., "Robots for All Reasons," *Aerospace America*, AIAA, September 1995, pp. 30–35.

[12] Special Issue, "Robot Experiments on ETS-VII", *Journal of Robotics Society of Japan*, vol. 17, no. 8, 1999, pp. 1055–1104. (in Japanese)

[13] Senda, K., Murotsu, Y., Mitsuya, A., Adachi, H., Ito, S., and Shitakubo, J., "Hardware Experiments of Autonomous Space Robot," *J. of Robotics and Mechatronics*, vol. 12, no. 4, August 20, 2000, pp. 343–350.

[14] Senda, K., Murotsu, Y., Mitsuya, A., Adachi, H., Ito, S., Shitakubo, J., and Matsumoto, T., "Hardware Experiments of A Truss Assembly by An Autonomous Space Learning Robot," *AIAA Journal of Spacecraft and Rockets*, vol. 39, no. 2, 2002, pp. 267–273.

[15] Arimoto, S., "Quasi-Natural Potential and Saturated-Position Feedback," Control Theory of Non-linear Mechanical Systems, Oxford University Press, Oxford, 1996, pp. 93–121.

[16] Sutton, R. S. and Barto, A., Reinforcement Learning, MIT Press, Cambridge, 1998.

[17] Rasmussen, J., "Skills, Rules, and Knowledge," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-13, no. 3, pp. 257–266, 1983.

[18] Bertsekas, D. P. and Tsitsiklis, J. N., Neuro-Dynamic Programming, Athena Scientific, Belmont, 1996.

[19] Sawaragi, T., "Proficient Skills Embedded with in Human, Machine and Environment," *J. Society of Instrument and Control Engineers*, vol. 37, no. 7, pp. 471-476, 1998. (in Japanese)

[20] Brown, F. M., The Frame Problem in Artificial Intelligence, Kaufman, 1987.

[21] Naghdy, F., Lidbury, J. and Billingsley, J., "Robot Force Sensing Using Stochastic Monitoring of the Actuator Torque," *Robots and Automated Manufacture*, pp. 139–156, 1985.