Intelligent Autonomous Systems 10
W. Burgard et al. (Eds.)
IOS Press, 2008

doi:10.3233/978-1-58603-887-8-398

Scenario controller for daily assistive humanoid using visual verification, task planning and situation reasoning

Kei OKADA [a,1], Satoru TOKUTSU [a], Takashi OGURA [a], Mitsuharu KOJIMA [a], Yuto MORI [a], Toshiaki MAKI [a], and Masayuki INABA [a]
[a] *7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan*

Abstract.
This paper describes scenario description and control method for a daily assistive humanoid robot. High level representation of a task scenario help user to develop various daily assistive tasks of a humanoid robot. Four key techniques are presented. 1) Three layered architecture integrating a vision based behavior control layer, a task planning layer, a situation based scenario generation layer. 2) Visual verification system for adapting symbol level description into real-world situation. 3) Task level planner providing high-level action primitives for scenario description, 4) Vision and auditory based situation interpretation for controlling the scenario. Finally, we demonstrated a kitchen assistive task by a humanoid robot.

Keywords. Humanoid, Task plannnig, Situation reasoning, Vision based behavior ontrol

1. Introduction

Development of robotic behaviors in human daily environments is one of the most desperately-needed application[1,2,3,4]. Most of humanoid research is focus on realizing individual action or sequence of actions, such as walking, carrying tray, pouring tea, washing dishes and so on, few research deals with how to select a goal of humanoid based on environmental situation and generatie sequence of actions from the goal. In this paper, we present humanoid robot system capable of providing scenario level description. High level task description as scenario help user to develop various daily assistive tasks of a humanoid robot. Previous researches on robot architecture with high level behavior description mainly focus on communication task of a robot[5,6]. On the other hand, we deal with daily assistive tasks of a humanoid robot.

Four key techniques are presented. 1) Three layered architecture integrating a vision based behavior control layer, a task planning layer, a situation based

[1]Corresponding Author: Graduate School of Information Science and Technology, The University of Tokyo, E-mail: k-okada@jsk.t.u-tokyo.ac.jp

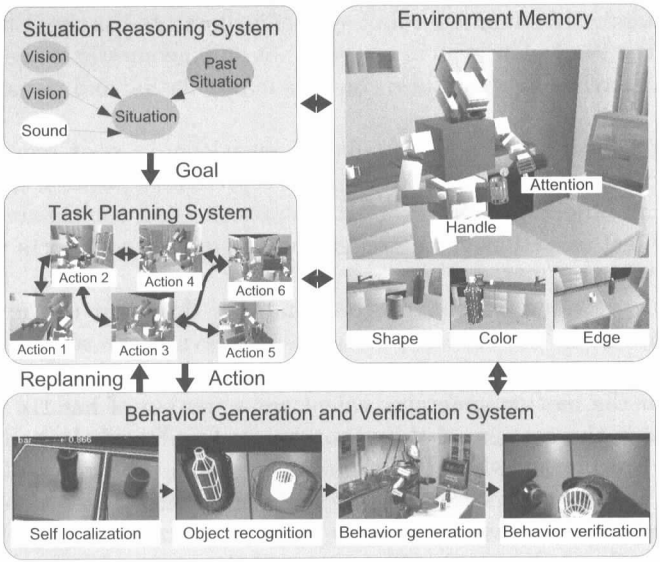

Figure 1. Action-recognition integrated humanoid system with task planning and situation reasoning for scenario level description and control

scenario generation layer. 2) Visual verification system for adapting symbol level description into real-world situation. 3) Task level planner providing high-level action primitives for scenario description, 4) Vision and auditory based situation interpretation for controlling the scenario.

2. Situation based Task Interpretation in Action-Recognition Integrated Humanoid System

Fig.1 presents action-recognition integrated humanoid system. It is composed of three layers: 1) Vision based behavior control layer, 2) Task planning layer and 3) Situation based scenario control layer. These layer required a environment memory which contains knowledges related to a manipulation and a recognition.

2.1. Vision based behavior control layer

This layer receives an action primitive from the task planning system and generate a humanoid behavior with a visual recognition and verification. For example, an action produced by the planning system is "Grasp Cup". This layer decompose this action into vision based cup detection, whole body reaching motion, grasping control and verification of these steps, which we call behavior.

In order to achieve vision based behavior control of humanoid manipulation action, we rely on knowledge based vision-guided robot system, which is achieved through the development of three components: 1) Manipulation knowledge based whole body motion generation system[7], 2) Visual feature knowledge based 3D object recognition system[8], 3) Vision based environment and behavior verification system by using both manipulation and visual feature knowledge[9].

The knowledge of humanoid robot system is illustrated as the "Environmental Memory" in the Fig.1. The system contains not only geometric shape information of objects and environment but also contains manipulation and visual recognition knowledge.

We have defined following manipulation knowledge: a `spot` represents a base coordinate when performing a task, a `handle` represents a position and constrains for grasping an object and an `attention` represents tool coordinates for manipulating an object as well as geometric shape information of objects and environment in the scene.

The sequence of an `attention` coordinates is the input of the whole body motion generator. In the case of pouring tea behavior, the sequence represents a rotating motion of the top of the bottle using an `attention` coordinate of a bottle.Then the motion generator calculates a motion of `handle` coordinates, which indicates the motion of the robot hand. Finally whole body motion is generated by calculating whole body joint angles from the motion of `handle`.

In order to recognize objects, we employ the Particle Filter algorithm[10,11] which is widely used because of its robust characteristics. Each particle represents the hypothesis that indicates the 3D position of the target object and is weighted by likelihood using multi visual cue integration method[8].

Visual feature knowledge for objects and events recognition is also described in the system: `Shape` for calculating 3D distance between this shape and visual 3D feature points, `Color` for calculating similarity between this histogram and the histogram taken from the view images and `Edge` on an object surface for calculating 2D edge distance on the image plane.

Three visual behavior controls are required to perform each motion: 1) Visual self localization, 2) Visual object localization, 3) Visual behavior verification[12]. Before performing each motion, the robot is assumed to be located on the `spot` position and positions of task relevant objects are known in advance. Thus the visual self localization and the visual object localization are required. After the motion, the robot verifies the behavior. We classified the verification process into two groups. One is an indirect verification and another is a direct verification.

Visual self localization, object recognition and behavior verification results are presented in the bottom of the Fig.1.

Thanks to the visual verification process, this layer is able to provides high level autonomous behavior command as "Move to Bar" "Hold Cup", "Hold Bottle" and "Pour Tea". User or other layers do not have to care about visual recognition and verification when executing these commands.

2.2. Task planning layer

Since vision based behavior control layer accepts high level action commands. It is easy to connect a behavior control layer and a high level task planner. We adopt the STRIP type operator for each action. For example the HOLD operator can be described as following:

```
HOLD:
 preconditions : (ON ?OBJECT ?SPOT) (AT ?SPOT)
 action        : (HOLD ?OBJECT ?ARM)
 effects       : (HOLD ?OBJECT ?ARM) ~(ON ?OBJECT ?SPOT)
```

Actions	Visual controls
Actions with self localization	
Move to counter	Recog. counter
Move to kitchen	Recog. sink
Actions with object localization	
Hold a cup	Recog. cup
Hold a bottle	Recog. bottle
Place a cup	Recog. cup
Place a bottle	Recog. bottle
Actionss with visual verification	
Pour tea	Recog. tea
Open tap	Recog. water
Close tap	Recog. water
Wash cup	—

Object	Knowledge
Cup	Shape
Bottle	Histogram, Shape
Counter	Edge
Sink	Edge
Search area	Ttarget
On counter	Cup, Bottle
Counter foot	Counter
Sink foot	Sink
Under tap	water flow
Event	Knowledge
Recog. tea	Color histogram
Recog. water	Water flow model

Table 1. Knowledge description in the kitchen experiment.

An `action` in the operator can be interpolated as a command to be issued to the behavior control layer. A `preconditions` in the operator consists of conjunctive logical expressions which must be true in order to apply the operator. Each expression evaluated not only the symbol level description of the system by but also using the real sensors. Conditions as `(ON ?X ?SPOT)`, `(AT ?SPOT)` `(WATER-FLOW)` are recognized using a vision sensor and a `(HOLD ?OBJECT ?ARM)` condition is recognized using torque sensors at a robot hand.

2.3. Vision and auditory integrated situation reasoning layer[13]

Estimating current situation gives adequate goal for task planning layer. We have developed situation reasoning method using the Dynamics Bayesian Network. Network node corresponds to the result of life sound recognition such as "tap water sound" or "table ware sound" and human location detected by 3D cylindrical model and 3D position of visual feature point.

Situation node have several states as "cooking", "washing" and "others". The brief node of the network is update continuously using vision and sound sources.

Scenario description can be written as the connection between the recognized situation and goal of the task planner. For example, when the situation reasoning layer interpret as "Washing" situation, then the goal status of the task planner is set to generate behavior sequence to wash a cup.

3. Tea Serving Task Experiment

3.1. Knowledge description for vision based behaviors

This section describes knowledges required to perform the experiment. Ten set of actions required for achieving this experiment are listed in the left table in the Table 1. First two actions require a vision based localization, next four requires an object detection and the last four requires visual verification.

```
HOLD:
 preconditions : (ON ?OBJECT ?SPOT)
                 (AT ?SPOT)
 action        : (HOLD ?OBJECT ?ARM)
 effects       : (HOLD ?OBJECT ?ARM)
                 ~(ON ?OBJECT ?SPOT)

PLACE:
 preconditions : (HOLD ?OBJECT ?SPOT)
                 (AT ?SPOT)
 action        : (PLACE ?OBJECT ?ARM)
 effects       : ~(HOLD ?OBJECT)
                 (ON ?OBJECT ?SPOT)

MOVE-TO:
 preconditions : (AT ?FROM)
 action        : (MOVE-TO ?FROM ?TO)
 effects       : (AT ?TO) ~(AT ?FROM)
```

```
OPEN-TAP:
 preconditions : (AT ?SPOT) ~(WATER-FLOW)
 action        : (OPEN-TAP)
 effects       : (WATER-FLOW)

CLOSE-TAP:
 preconditions : (AT ?SPOT) (WATER-FLOW)
 action        : (CLOSE-TAP)
 effects       : ~(WATER-FLOW)

WASH-CUP:
 preconditions : (HOLD CUP LARM) (AT SINK)
                 (WATER-FLOW)
 action        : (WASH-CUP)
 effects       : (WASHED CUP)

POUR-TEA:
 preconditions : (HOLD CUP LARM)
                 (HOLD BOTTLE RARM)
 action        : (POUR-TEA)
 effects       : (POURED CUP)
```

Table 2. Action operators in the kitchen experiment.

We defined four objects in the demo scene. For each object, we described associated visual recognition knowledges as shown in the right top table.

Three search areas are defined for detecting objects for grasping(cup and bottle). In this case, we used 2D search space definition since these objects are spinning objects, which has freedoms along with x and y axis and z position of the object is assumed to be the table height.

The right bottom table shows task relevant visual behavior verification knowledge. Recognizing tea is utilized for pouring tea behavior verifications. Recognizing water is applied to verify the open and close tap actions.

3.2. Action operators for task planning

We have defined seven operators for this experiment as presented in Table 2. Variable domain is defined as followings: `BAR/SINK` for `?SPOT`, `CUP/BOTTLE` for `?OBJECT` and `LARM/RARM` for `?ARM`.

Assuming that starting situation of this experiment can be described as `(AT BAR) (ON CUP BAR) (ON BOTTLE BAR)`, "Making tea" can be describe as setting goal status to be `(POURED CUP) (ON CUP BAR) (ON BOTTLE BAR)`. Then we obtain following behavior sequence by applying the task planner.

1. `(HOLD BOTTLE RARM)`
2. `(HOLD CUP LARM)`
3. `(POUR-TEA)`
4. `(PLACE BOTTLE RARM)`
5. `(PLACE CUP LARM)`

When we set goal status as `(WASHED CUP) (ON CUP SINK)`, then we obtain another task scenario "Cleaning up cup" automatically as `(HOLD CUP LARM) (MOVE-TO BAR SINK) (OPEN-TAP) (WASH-CUP) (PLACE CUP LARM) (CLOSE-TAP)`.

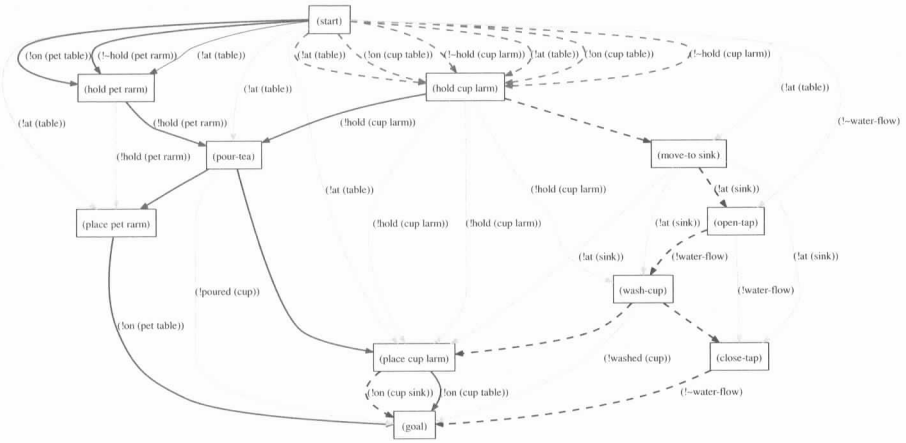

Figure 2. Task plan graph for tea serving experiment

We used POP planner (Partial Ordered Planner)[14] to solve this problem. Task graph to present these actions is illustrated in the Fig.2.

3.3. Tea serving task demonstration

We demonstrated the tea serving task experiment as shown in the Fig.3. The sequence of this experiment is as followings. The number on each line corresponds to the number in the figure.

1. The robot recognizes the cup (1) and holds (2).
2. The robot recognizes the bottle (3) and holds (4).
3. The robot pours tea into the cup from the bottle (5).
4. The robot places the cup (6) and the plastic bottle (7).
5. The human drink tea in the cup and place it (8-9).
6. The robot recognizes the cup (10) and holds.
7. The robot walks to the kitchen (11-12).
8. The robot localizes self position (13).
9. The robot opens the tap (14) and conform it (15).
10. The robot washes the cup (16).
11. The robot closes the tap and place the cup.

Sequence from 1 to 4 is generated by setting a goal situation as `(POURED CUP) (ON CUP BAR) (ON BOTTLE BAR)` and the rest of robot actions are generated by `(WASHED CUP) (ON CUP SINK)`.

3.4. Robustness

This demonstration is repeated very often at the lab on demand. Thanks to the robust object recognition using attention control, visual feature prediction, multi-cue integration and visual behavior verification, the task is rarely failed thus we believe the targeted robustness was reached. The rare case of the failure is when there are droplet on the cup or the bottle. It slips when the robot holds them.

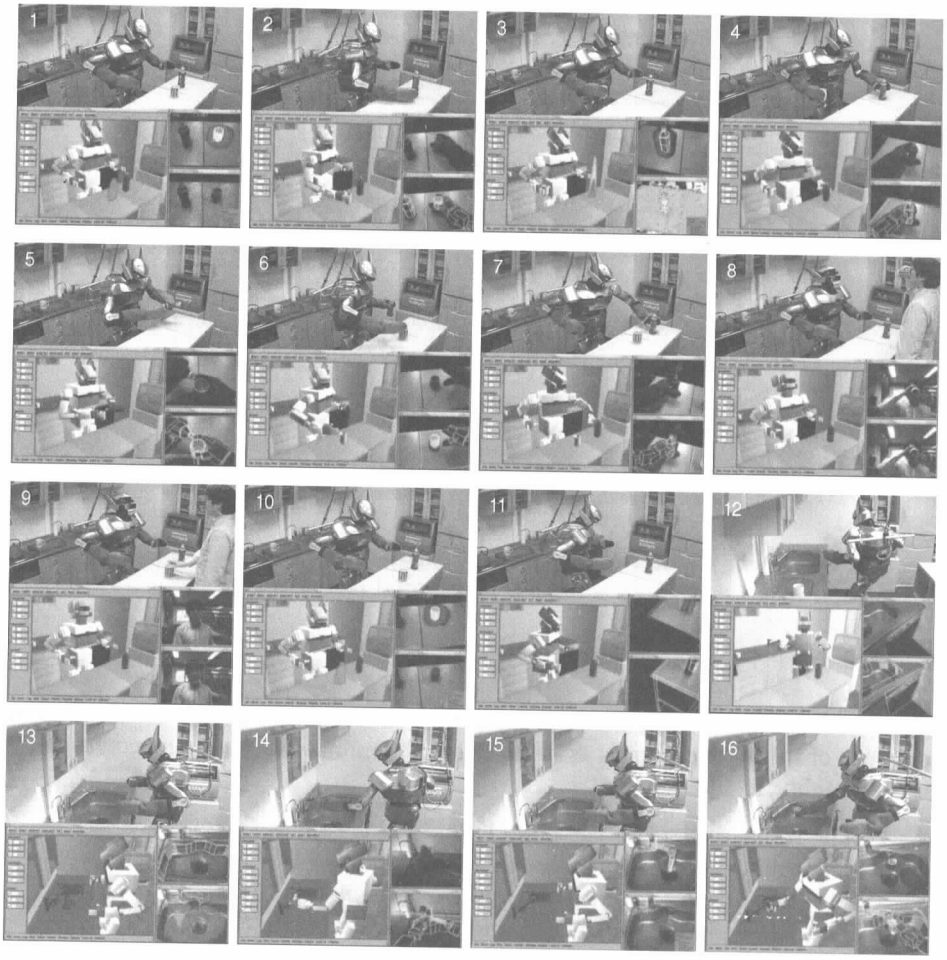

Figure 3. daily life support experiments using knowledge based on recognition system.

Changes of the lighting condition usually affects the recognition result, however our object recognition system is robust enough not to restrict using flash when taking photos. 3D feature points and 2D edges are robust to the illumination change and we use the HSV color space based histogram matching.

4. Conclusion

This paper describes scenario description and control method for a daily assistive humanoid robot. High level representation of a task scenario help user to develop various daily assistive tasks of a humanoid robot. Four key techniques are presented. 1) Three layered architecture integrating a vision based behavior control layer, a task planning layer, a situation based scenario control layer. 2) Visual verification system for adapting symbol level description into a real-world situation. 3) Task level planner which provides high-level action primitives for scenario

description, 4) Vision and auditory based situation interpretation for controlling the scenario.

Finally we demonstrated a tea service task by a humanoid robot. Very robust vision based behavior control layer is able to connect a real humanoid robot to the high level task planner. Then task planner is able to provide high level goal description to generate a action sequence. Situation reasoning used to select goals by interpreting vision and auditory information.

References

[1] H. Inoue, S. Tachi, K. Tanie, K. Yokoi, S. Hirai, H. Hirukawa, K. Hirai, S. Nakayama, K. Sawada, T. Nishiyama, O. Miki, T. Itoko, H. Inaba, and M. Sudo. HRP: Humaonid Robotics Project of MITI. In *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots (Humanoids 2000)*, 2000.

[2] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura. The intelligent ASIMO: System overview and integration. In *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'02)*, pages 2478–2483, 2002.

[3] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann. ARMAR-III: An Integrated Humanoid Plattfrom for Sensory-Motor Control. In *2006 6th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006)*, pages 169–75, 2006.

[4] C.C. Kemp, A. Edsinger, and E. Torres-Jara. Challenges for robot manipulation in human environments. *IEEE Robotics & Automation Magazine*, 14(1):20–29, 2007.

[5] H. Ishiguro, T. Kanda, K. Kimoto, and T. Ishida. A robot architecture based on situated modules. In *Proceedings of the 1999 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS'99)*, pages 1617–1624, 1999.

[6] Y. Hoshino, T. Takagi, U. Di Profio, and M. Fujita. Behavior description and control using behavior module for personal robot. In *Proceedings of The 2006 IEEE International Conference on Robotics and Automation*, pages 4165–4171, 2004.

[7] K. Okada, T. Ogura, A. Haneda, J. Fujimoto, F. Gravot, and M. Inaba. Humanoid Motion Generation System on HRP2-JSK for Daily Life Environment. In *2005 IEEE International Conference on Mechatronics and Automation (ICMA05)*, pages 1772–1777, 2005.

[8] K. Okada, M. Kojima, S. Tokutsu, T. Maki, Y. Mori, and M. Inaba. Multi-cue 3D Object Recognition in Knowledge-based Vision-guided Humanoid Robot System. In *Proceedings of the 2007 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS'07)*, pages 3217–3222, 2007.

[9] K. Okada, M. Kojima, Y. Sagawa, T. Ichino, K. Sato, and M. Inaba. Vision based behavior verification system of humanoid robot for daily environment tasks. In *2006 6th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006)*, pages 7–12, 2006.

[10] Genshiro Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, March 1996.

[11] Michael Isard and Andrew Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[12] K. Okada, M. Kojima, S. Tokutsu, Y. Mori, T. Maki, and M. Inaba. Task Guided Attention Control and Visual Verification in Tea Serving by the Daily Assistive Humanoid HRP2JSK. In *Proceedings of the 2008 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS'08)*, page (Submitting), 2008.

[13] S. Tokutsu, K. Okada, and M. Inaba. Daily Life Sound Recognition based on Cepstrum for Environment Situation Reasoning in Humanoid Robot. In *Proceedings of the 2008 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS'08)*, page (Submitting), 2008.

[14] S. Russell and P. Norvig. Planning and Acting. In *Artificial Intelligence: A Modern Approach, 2nd edition*, pages 445–448, 2002.