Taxonomy of Atomic Actions for Home-Service Robots

Hyunseok Kim, Yuchul Jung and Yong K. Hwang
Digital Media Laboratory
Information and Communications University
Dogok-dong, Gangnam-gu, Seoul 135-854, Korea
email:{hertzkim,enthusia77,ykhwang}@icu.ac.kr

Abstract—**In household environments, robots are expected to perform many kinds of tasks. It is difficult, however, to write programs for all the tasks in advance because of the diversity of the tasks and changing environmental conditions. One of the essential processes of autonomous multi-functional robots is to define a set of basic robot actions that can be executed unambiguously and also checked for completion. A task planner then can use these actions to accomplish complex tasks that home-service robots are expected to do. This paper first proposes a set of tasks for the first-generation home-service robots, and then systematically decomposes them into sequences of smaller but meaningful actions called molecular actions. The molecular actions are then decomposed into yet more primitive actions called atomic actions. Because vision, sound, range sensors, and force sensors are the primary means of monitoring task progress and completion, the atomic actions are classified according to the complexities and frequencies of the sensing algorithms used. The resulting taxonomy of atomic actions serves as basic building blocks for a knowledge-based task planner, and its advantages are justified and demonstrated through experiments.**

Index Terms—**Task classification, planning, monitoring.**

I. INTRODUCTION

Robots are expected to perform a variety of tasks in household environments. It is a difficult problem, however, to write programs to perform all robot tasks in advance because of the sheer diversity of tasks and changing environments. One of the essential processes of developing multi-functional robots is to define a set of basic robot actions that can be executed unambiguously and also checked for completion. A task planner then can use these actions to accomplish more complex tasks expected of home-service robots. This paper takes a set of prototypical tasks for the first-generation home-service robots proposed in the literature, analyzes them, and then proposes taxonomy of basic actions called atomic actions. This allows task planners to efficiently plan robot actions to perform a variety of tasks that humans command robots to do.

Taxonomy of atomic actions is equivalent to standardization of robot actions into primitive units so that they are reusable in as many different tasks as possible. For reasons justified in this paper, the basic actions are classified into a two-level hierarchy: molecular actions and atomic actions. There is no unique way to define the number of levels in the hierarchy or the complexity of the basic actions, but a well-designed method would facilitate more effective task planners. Because of the diversity of tasks, we have used several principles to define basic actions and then empirically verified their usefulness with experiments. One measure we have used to define basic actions is that their completions have to be verifiable via available sensors.

There has been a steady stream of definition of primitive robot action sets since the invention of robots. Common industrial robots usually come with *joint_move, Cartesian_move, reset* and *halt* as their basic motion set. For mobile robots with various sensors, vision sensing or range finding naturally define a set of basic sensing actions [1]. For more complex, multi-functional robots such as humanoid robots, a rich set of basic actions is defined that includes walking, object grasping and manipulation, human interaction algorithms such as face recognition, gesture recognition, voice recognition, speech synthesis, emotion expression, and so on [2]. In this paper, however, we concentrate on a set of tasks for the first-generation, commercial, home-service robots, and they usually involve navigation and object manipulation and not so much dealing with humans emotionally.

The balance of this paper first reviews previous work in Section II, and then discusses household tasks, molecular and atomic actions in Section III. We then propose taxonomy of these atomic actions in Section IV, and Section V describes how we evaluate the taxonomy empirically by implementing one of the household tasks. Section VI concludes this paper and describes future work.

II. PREVIOUS WORK

There has been a variety of research in the taxonomy of primitive robot operations. Cutkosky [3] constructed taxonomy of grasp and developed an expert system that chooses a grasp type according to the task requirement and object geometry. His taxonomy is based on the facts that a human hand consists

This work has been supported by the Ministry of Information and Communication of the Korean Government

of 25 joints and can execute 58 distinct motions [4], and humans tend to use one of two of grasps: the power grip and the precision grip [5]. Since there are many kinds of objects to be grasped in household environments, it is impossible for robots to be programmed with every grasping rule for each object. Stansfield [6, 7] used vision information to get object geometry and used a rule-based system to determine grasping configurations. Andrew [8] developed an automatic grasp planning system which classified the object to grasp as one of the simplified models: spheres, cylinders, cones and boxes. The final grasp configuration was selected whose hand preshape was similar to the object model. Besides grasping, there are research results that define a set of basic actions to be used for general task planning and learning. Chen and Hwang [9] developed a natural-language like user interface for a robotic system. It initially has a basic action set that the robot knows how to perform, and learns to perform complex tasks in terms of these basic actions through demonstration, memorization and generalization.

There has been relatively fewer research on taxonomy of actions for general-purpose robots because there have not been robot applications with enough varieties of tasks so that systematic taxonomy of robot actions is necessary. Recently, however, home-service robots in ubiquitous computing environments provide a rich set of scenarios for many robot tasks. The role of robots in ubiquitous computing environments has been described by Hwang [10], which is divided as a servant and a companion for humans. Tanie [11] on the other hand has suggested five applications of humanoid robots. Two of them, human care and home-security services, are applicable for household robots, while the others are for robots in industry. Other papers suggest scenarios for guide robots and public-service robots performing tasks such as picking up trash, going up and down in an elevator, and so forth [12]-[16].

The next section analyzes a set of relatively complex tasks to be performed by home-service robots. Each task is decomposed into a set of atomic actions that can be executed by the robot without future interpretation. They are meaningful chunks of robot algorithms that use sensor information to monitor the progress of actions, and the taxonomy of such will make robots function effectively in a household environment.

III. DEFINITION OF MOLECULAR AND ATOMIC ACTIONS

We develop taxonomy of atomic actions based on a set of prototypical tasks for home-service robots. These tasks have been proposed by governments [18], academia and industry as the tasks for the first-generation home-service robots. Our aim is to define a set of atomic actions, sequences of which can be used to perform the prototypical tasks defined below. The atomic actions have to be simple enough to be executed by the robot unambiguously and without further interpretation, but have to do a meaningful chunk of work. To further enhance the effectiveness of task planners, robot tasks are first decomposed into what we call *molecular* actions, which span a short sequence of atomic actions. This will enable task planners to describe actions needed to perform a task with a much fewer number of steps than when without the intermediate molecular actions. We now describe the common household tasks, molecular actions and atomic actions.

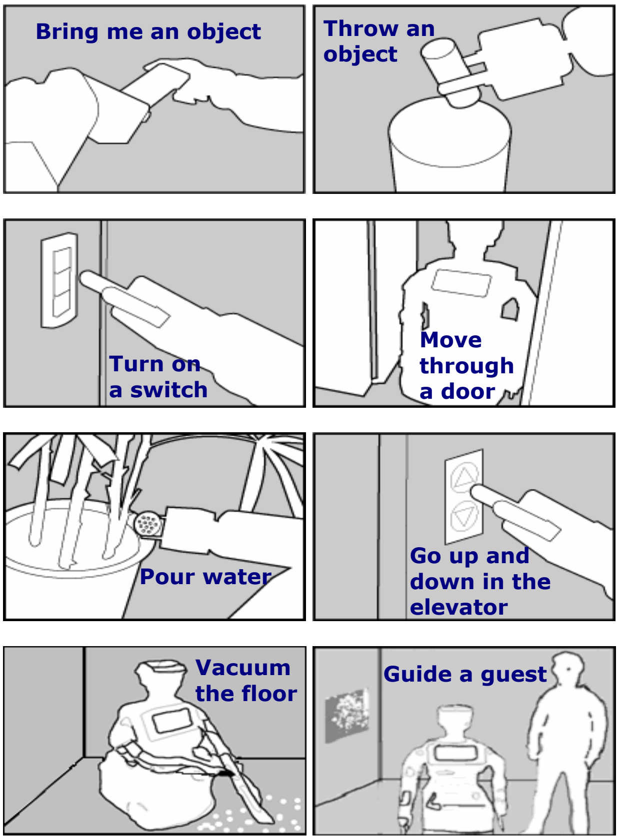

Fig. 1. Eight prototypical tasks for home-service robots.

A. Tasks

A task is defined as a specific piece of work to be done, whether commanded by a human or generated by the robot itself. The number of tasks that robots can do will increase as more technologies are developed, making it a moving target to define prototypical tasks. In this paper, we have selected eight tasks proposed in the recent literature [10]-[18] as follows:

1. Bring me an object
2. Throw an object into a wastebasket
3. Turn on/off a switch
4. Open a door and move through it
5. Pour water into an object
6. Go up and down in the elevator
7. Vacuum the floor
8. Guide a guest

Figure 1 depicts the scenes for the eight tasks mentioned above. For the rest of this paper, the term *task* refers to only these eight robot operations. We analyze these tasks in the next section to derive a minimal set of molecular and atomic actions that are required to perform these tasks. All of these are multi-step tasks and the required sequences of steps are stored in a task planning system we have developed. Our ultimate goal is that with a sufficient number of molecular and atomic actions the task planner would automatically generate efficient plans to perform most household tasks.

B. Molecular Actions

All eight tasks seem just simple actions for humans, but from the robotics point of view, they are fairly complex multi-step operations. We define molecular actions to be of such granularity as *find_object* or *move_to_location*. They usually involve a single object or location parameter, and their execution requires coordination of sensing, planning and action modules of the robotic system. The actual execution of molecular actions requires a sequence of more primitive atomic actions. Table 1 shows the number of molecular actions required for each of the eight household tasks.

TABLE 1. The number of molecular actions for the tasks

Tasks	Molecular Actions
Bring me an object	6
Throw an object into a wastebasket	6
Turn on/off a switch	5
Open a door and move through it	6
Pouring water into an object	3
Going up and down in the elevator	7
Vacuum the floor	5
Guide a guest	3

Figure 2 shows the flow chart for *bring-me-an-object* task in detail. It has 14 stages. The solid arrows denote the normal flow of operations while the dashed arrows show routes for possible error handling and recovery. This flow chart in fact represents a piece of knowledge in our task planner, and it can be hard programmed initially in the robot or may be taught by human on the fly. Table 2 lists all the molecular actions required to perform the eight household tasks.

C. Atomic Actions

An atomic action should be a primitive operation with two important properties: recoverability and non-interference [19]. It is a notion used in the areas of databases and distributed system, and is a compound word with atomicity and action. An atomic action should either complete or leave the state of the system unchanged and should not interfere with any other processes/algorithms in the robot system. It is difficult, however, to implement the atomic actions satisfying these properties because robots operate in the real world where undoing actions involves another action or often impossible. Therefore we have replaced the two properties with the property of completion verifiability. Sensors have to be used to test whether an atomic action completes or not, whether vision, sound or other sensors. It is important to note that we therefore define and classify atomic actions based on the kinds of sensors and algorithms used during action execution.

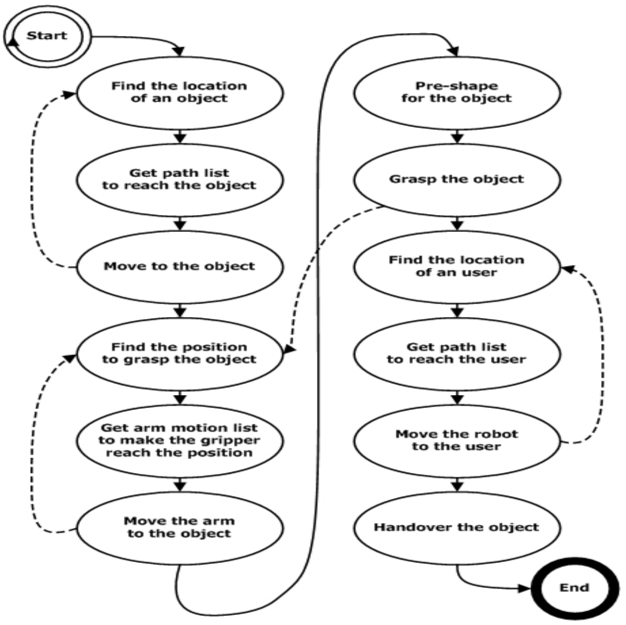

Fig. 2. A flow chart of "bring me an object." The dashed arrows indicate possible routes for error-recovery.

TABLE 2. List of molecular actions

Molecular Actions	Description
get_command	Get a command such as map, path and motion
find_object	Find an object
find_number	Find the number using recognition
find_human	Find a human using recognition
move_to_location	Navigate to the location
move_arm	Move arm to the position
preshape	Preshape gripper
grasp	Grasp an object
throw_object	Release an object while moving arm
flip_object	Move object to the other position
open_door	Open a door
push_object	Push object
pull_object	Pull object
pour_water	Pour water
vacuum	Vacuum the floor
wait	Wait for an event
guide_human	Navigate while checking the human following

■ Vision verification

Atomic actions can utilize vision information to maintain the distance between a robot and a landmark, to monitor a collision with obstacles, to calculate the possibility of grasping an object, and so on. It is a difficult problem, however, to decide what types of vision information is used in a task. We have examined, therefore, human-vision behaviors while executing a task. There were seven subjects in the field tests: six males (five right-handed and one left-handed) and one female (right-handed). Each subject used one video camera to see instead of their eyes, and there was another camera to monitor their behaviors. All scenes were video recorded for further analysis.

We have analyzed the video to find out where humans watch

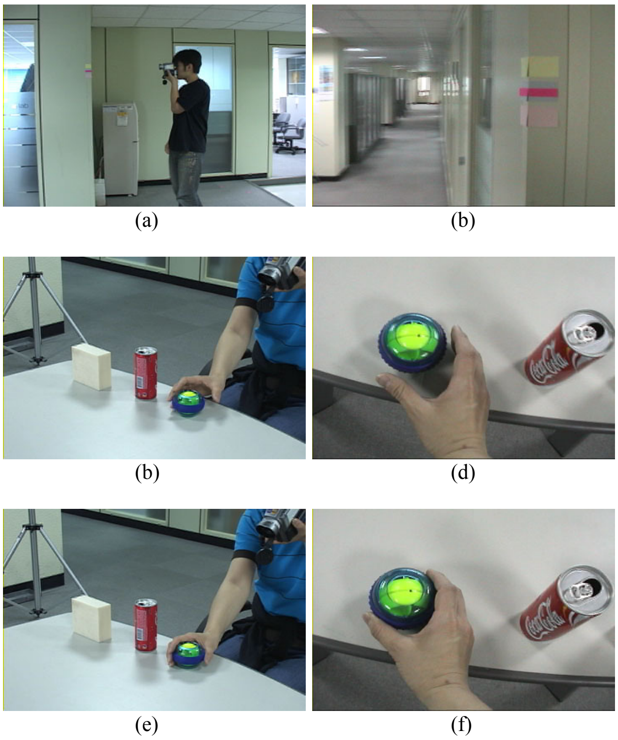
(a) (b)
(b) (d)
(e) (f)

Fig. 3. (a) and (b) show the behavior pattern and utilization of vision of one male doing the atomic action, *GoAround*, (c) and (d) show preshape for spherical object in executing *PreshapeGripper*, (e) and (f) show grasp for the object in executing *GraspSphere*.

when performing various actions, and then used the results to define atomic actions. For example, as shown on Figure 3(a) and (b), *GoAround* is an atomic action of navigating around a specific landmark. It uses vision information such as the position of a known landmark by object recognition, and does visual servoing to maintain a reasonable distance between the landmark and the robot. All subjects watched the landmark in a starting position as a precondition, maintained a distance between them and the landmark as a progress condition, and let the landmark disappear from their vision as an end condition. These results are used to define atomic actions. In defining atomic actions for grasping operations, we use the methods developed by Andrew [6]. As shown on Figure 3(c)-(f), humans make a preshape of their hands similar to the object before grasping it, and check the orientation and the width between their fingers and the object. If these all conditions check out ok, the grasping action is initiated.

- **Sound verification**

The sound is another important sensing modality to check the completion of an action. The completion of turning on/off an electrical switch, for example, can be readily checked with a flip sound. By complementing the vision sensor with a sound sensor, many kinds of actions can be checked for completion with less computational costs.

- **Force and tactile verification**

Humans use an abundance of contact and force sensors distributed throughout their skin and joints to their advantage. For many kinds of robot actions, their completion can be checked with force/contact sensors. Turning the door knob before opening the door is a good example. The knob has to be turned until it cannot be turned any more before the door can be pulled open. Slip detection via tactile sensors can be used to maintain stable grasps during object transfers. Our taxonomy of atomic actions includes these types of sensors.

IV. TAXONOMY OF MOLECULAR AND ATOMIC ACTIONS

We have defined the molecular and atomic actions corresponding to the eight prototypical household tasks, and it remains to classify them to accomplish two purposes. First, good taxonomy facilitates task planners to generate efficient plans for other kinds of tasks. Second, sensor resources can be judiciously allocated to various processes running in the robot system so as to maximize robot's overall performance. The classification of the molecular actions is relatively simple compared to the atomic actions. They are classified by the functionalities they serve: navigation, object manipulation, etc. Atomic action classification on the other hand is very complex due to the fact that they have to maintain completion verifiability via sensors.

A. *Taxonomy of Molecular Actions*

Each molecular action represents a sub-step for complex tasks, and is decomposed into atomic actions. The taxonomy of molecular actions, therefore, is done according to the purpose of action such as human interface, vision sensing, navigation, and object manipulation. Table 3 summarizes the molecular-action taxonomy.

TABLE 3. Taxonomy of Molecular Actions

Cateogry	Molecular Actions
Human Interface	*get_command*
Vision Sensing	*find_object, find_number, find_human, wait*
Navigation	*move_to_location, vacuum, guide_human*
Object Manipulation	*move_arm, preshape, grasp, throw_object, flip_object, open_door, push_object, pull_object, pour_water*

B. *Taxonomy of Atomic Actions*

As mentioned in the previous section, our atomic actions are derived based on the completion verifiability via sensors. It is therefore natural to classify the atomic actions according to the types of sensors and sensor related algorithms used during the execution of the actions. The types of sensors depend on the sensor hardware used and there is no ambiguity about it. Our first classification dimension thus spans the space of sensors. The types of processing done on the sensed data, however, are all different even if the data is obtained by the same sensor. For example, a visual image can be used to extract a target location or the color of an object. Face recognition can be done occasionally off-line, while target tracking requires a stream of images.

The computational complexity and the frequency of usage have important implications in the robot's intelligence and performance. We thus span our second and third classification dimensions with processing complexity and processing frequency. Our classification space is then $(R\text{x}R)^{s}$ where s is the number of sensing modalities of the robot.

■ Vision-based taxonomy

The complexity of vision algorithms is used for classifying atomic actions on the vertical axis, and it can be expressed in terms of the difficulties of the algorithms. For instance, 1D vision such as edge or line detection has a low value, 2D vision such as extraction of a convex hull or a surface normal vector has a higher value than 1D vision, and 3D vision or video processing has a much higher value than the former two. Second, the frequency of vision algorithms used can be expressed in terms of how often they are used during the action. For example, if vision information is used only one time, then the action has a low value on the horizontal axis. If an action is used all the time such as video processing, then the action has a higher value. Table 4 shows the numerical designations of the algorithm complexity and frequency.

TABLE 4. Criteria for the vision-based taxonomy

Complexity	Cost
3D motion estimation	9
3D position	8
Object recognition	7
Face recognition	7
Number recognition	6
Optical flow estimation	6
Convex hull	5
Normal vector estimation	4
Shape	3
Color	2
Edge	1

Frequency	Cost
All times	3
Some times	2
One times	1
Not used	0

■ Sound-based taxonomy

Some atomic actions use sound information. The *WaitingDoorOpened* action waits for the elevator door to open. A modern elevator makes a sound or says the floor number when it touches down at the selected floor. The atomic action, therefore, checks whether the door is open using vision as well as sound. The actions using sound information are shown in blue circles in Figure 4.

■ Force/tactile-based taxonomy

Force or tactile sensors are playing an important role when robots are grasping or touching. Some of the atomic actions use force-feedback and tactile sensors, and these actions are shown in yellow circles in Figure 4. In addition, atomic actions using both the sound and force/tactile sensors are marked with green color.

It turns out that there are a total of 30 atomic actions to cover all 8 prototypical household tasks according to our definition of atomic actions. Since it is hard to visualize the atomic actions in $(R\text{x}R)^{s}$ space, we project them onto the vision-sensor subspace of the classification space (Figure 4). Three of the atomic actions, *InvokeArmMotionPlanner*, *InvokePathPlanner* and *VoiceRecognition*, do not use vision information, so these actions have a low frequency value. *Preshape* is the action to make the gripper preshaped like the object to be grasped, and uses vision information to compute the object shape and its convex hull. The action has the value of 8 (shape 3 + convex hull 5) on the vertical axis and 1 (uses vision one time) on the horizontal axis. *Rotate* and *Stop* actions use vision information all the time, therefore, they are put on the right-side. The other atomic actions are plotted in the classification space in a similar manner.

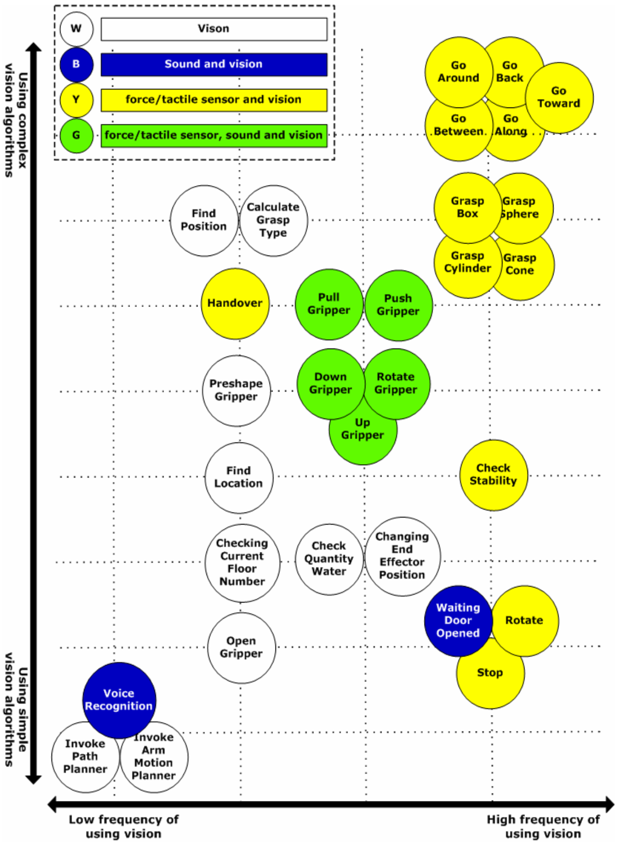

Fig. 4. Taxonomy of atomic actions for household robots. Yellow color denotes atomic actions using force/tactile sensors, blue color denotes atomic actions using the sound, and green color denotes actions using both the sound and force/tactile sensors. All actions use vision except voice recognition, invoke path/arm planner.

It is clear that the taxonomy of atomic actions offer many advantages. For example, when a task planning system has to select only one from a set of candidate tasks, it could use the taxonomy to select the task with the lowest total cost. In addition, the searching time to find an optimal solution would be shorter than other systems built without proper taxonomy.

V. IMPLEMENTATION

The eight prototypical tasks require a total of 66 stages to be implemented, and each of the stages should be able to be executed by choosing an appropriate set of atomic actions we have defined in this paper. Since the eight household tasks require a total of only 30 atomic actions, we have thus achieved an efficiency factor of 2.2 (the number of stages divided by the number of atomic actions) by using a well-defined atomic actions. If the number of tasks increases, the efficiency would

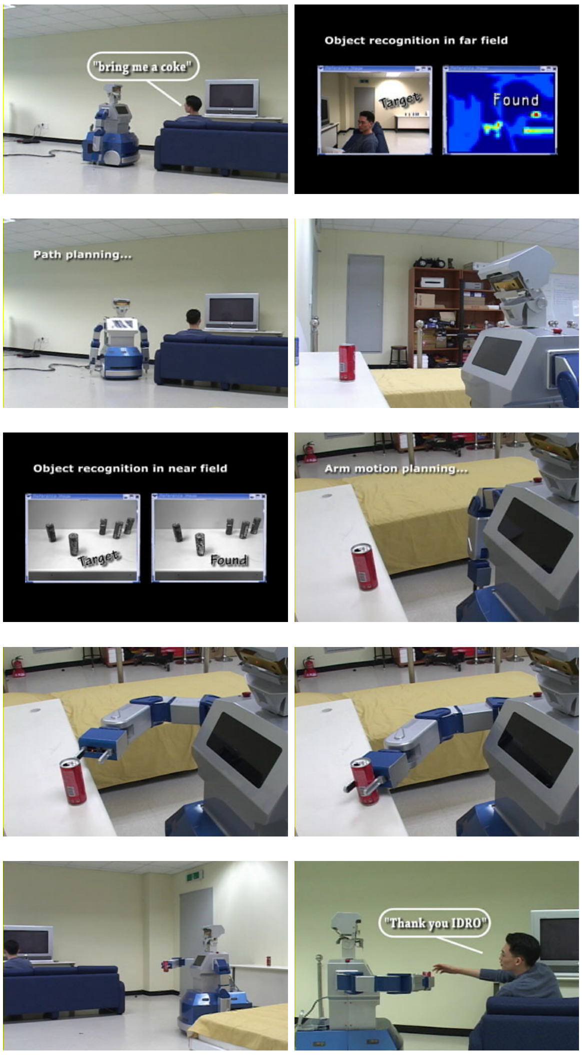

Fig. 5. Atomic actions implementing, "Bring me an object:" *VoiceRecognition, FindLocation, InvokePathPlanner, GoToward, FindPosition & CalculateGraspType, InvokeArmMotionPlanner, PreshapeGripper, GraspCylinder, GoAlong,* and *Handover*. For arm motion planning, we use the algorithm in [20].

further increase. We have implemented the atomic actions in our robot system called IDRO. For the task of "bring me an object," 6 molecular actions and 13 atomic actions are used. Figure 5 shows the snapshots during the execution of the task.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented taxonomy of molecular and atomic actions for home-service robots, which have been classified according to the sensor type, the algorithm complexity and the frequency of algorithm usage. The taxonomy is useful for building a procedural, rule-based task planning system, resource allocation of sensors, CPU and actuators, and design of robot architecture. We are in the process of developing a task planner that is based on the atomic actions presented in this paper. Our goal is to make the task planner work for a home-service robot operating in a real environment. For future work, we plan to develop make the atomic actions and the task planner more powerful in two ways. First, recoverability property will be enforced by augmenting atomic actions with information necessary to plan a recovery action in case errors arise. Second, a re-planning capability will be implemented in the task planner so that robots can deal with unexpected situations while executing a commanded task. We believe our task planner and the taxonomy of atomic actions will be widely applicable to many types of service robots in the near future.

REFERENCES

[1] Ryan Eustice, Oscar Pizarro, and Hanumant Singh, "Visually Augmented Navigation in an Unstructured Environment Using a Delayed State History," *IEEE Conference on Robotics and Automation*, pp. 25-32, New Orleans, LA, April 2004.

[2] Kenji Kaneko, Fumio Kanehiro, Shuuji Kajita, and et al., "Humanoid Robot HRP-2," *IEEE Conference on Robotics and Automation*, pp. 1083-1090, New Orleans, LA, April 2004.

[3] Mark R. Cutkosky, "On Grasp Choice, Grasp Models, and the Design of Hands for Manufacturing Tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no 3, pp. 269-279, June 1989.

[4] B. Campbell*, Humankind emerging, Eighth edition*, Needham Heights, MA: Allyn and Bacon, pp. 105-107, 2000.

[5] J. Napier, "The Prehensile Movements of the Human Hand," *Journal of Bone and Joint Surgery*, vol. 38B, no. 4, pp. 902-913, 1956.

[6] S. A. Stansfield, "Knowledge-Based Robotic Grasping,*" IEEE Conf. on Robotics and Automation*, vol. 2, pp. 1270-1275, May 1990.

[7] S. A. Stansfield, "Representing Generic Objects for Exploration and Recognition," *IEEE Conference on Robotics and Automation*, vol. 2, pp. 1090-1095, April 1988.

[8] Andrew T. Miller, Steffen Knoop, Henrik I. Christensen, and Peter K. Allen, "Automatic Grasp Planning Using Shape Primitives," *IEEE Conference on Robotics and Automation*, pp. 1824-1829, Sept. 2003.

[9] Pang C. Chen and Yong K. Hwang, "Generating Natural-Language like Robot Motion Command Set through Machine Learning," *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 1, pp. 114-191, Oct. 1994.

[10] Yong K. Hwang, Manjai Lee, and Dongman Lee, "Robots' Role in Ubiquitous Computing Household Environments," *International Symposium on Robotics*, Paris, France, March 2003.

[11] Kazuo Tanie, "Humanoid Robot and its Application Possibility," *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 213-214, July 30-Aug. 1 2003.

[12] M. Ehrenmann, R. Zollner, O. Rogalla, and R. Dillmann, "Programming Service Tasks in Household Environments by Human Demonstration," *IEEE International Workshop on Robot and Human Interactive Communication*, pp. 460-467, Sept. 2002.

[13] Lars Petersson, David Austin and Danica Kragic, "High-level Control of a Mobile Manipulator for Door Opening," *IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems*, vol. 3, pp. 2333-2338, 2000.

[14] R. J. Firby, P. N. Prokopwicz, and M. R. Swain, "Plan Representations for Picking Up Trash," *International Conference on Tools with Artificial Intelligence*, pp. 496-497, Nov. 1995.

[15] Keiji Nagatani and Shin' ichi Yuta, "An Experiment on Opening-Door-Behavior by an Autonomous Mobile Robot with a Manipulator,*" IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 45-50, Aug. 1995.

[16] U. D. Hanebeck, C. Fischer, and G. Schmidt, "ROMAN: A Mobile Robotic Assistant for Indoor Service Applications," *IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems*, vol. 2. pp. 518-525, 1997.

[17] R. D. Schraft, E. Degenhart, and M. Hagele, "Service Robots: The Appropriate Level of Automation and the Role of Users/Operators in the Task Execution," *International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 163-169, Oct. 1993.

[18] Institute of Information Technology Assessment, "21st Century Leading Technology Development Initiatives – Intelligent Service Robots," *Call for Proposals by Korean Government*, January 2004.

[19] Gerard Boudol, "Atomic Actions," *EATCS*, vol. 38, pp. 136-144, 1989. available: http://www-sop.inria.fr/mimosa/personnel/Gerard.Boudol.html

[20] P. C. Chen and Y. K. Hwang, "SANDROS: a Dynamic Graph Search Algorithm for Motion Planning," *IEEE Trans. on Rob. and Auto.*, vol. 14, no. 3, pp. 390-403, Feb. 1998.