Recognition and Manipulation Integration for a Daily Assistive Robot Working on Kitchen Environments

Kimitoshi YAMAZAKI, Yoshiaki WATANABE, Kotaro NAGAHAMA, Kei OKADA and Masayuki INABA

Abstract— **This paper describes a system integration of a daily assistive robot. Several tasks on cooking are focused on, recognition and manipulation functions are developed and integrated. It is often the case that kitchen tools and foods have less distinctive texture on its surface, and kitchen environments which are made of reflective materials are susceptible to the effect of illumination. From these fact, recognition functions are implemented with a basic policy which composes simple image features. On the other hand, tasks on kitchen often include relatively complicate dual arm manipulation. In such case it is effective in generating a robot pose by considering several manipulators at the same time. Experiments doing several cooking tasks with handling daily tools showed the effectiveness of our system.**

I. Introduction

One of the abilities needed for daily assistive robots is to handle various tools existing in real world. It means that the robot will be able to do chores, for instance, cleaning floors, tidying up devices for eating and cooking.

The purpose of this research is to develop recognition and manipulation functions to perform relatively complex routine by a daily assistive robot. We especially focus on cooking, which needs multiple functions as foods state recognition, handling tools and so on. In this paper, an integrated robot system and its application will be introduced. For handling various objects, essential functions for recognition and manipulation are developed, and robot's behavior for achieving tasks are defined by connecting these functions. 3D geometrical simulator existing in the middle of these functions creates the combinations.

Recognition functions are built up with a policy that basic image features like edges, color and geometrical shape are used. This is because images captured from kitchen environments can have much illumination influence, and foods in the images have almost no distinguishable texture on it surface. From the same reason, elimination of noises is also important.

Manipulation functions are the tools for designing various robot poses for handling kitchen tools and foods. One of the difficulties on our challenges is that a manipulation pose should satisfy a constraint of the body configuration with considering several end-effectors; left arm, right arm and perhaps, a head. Our approach is to make a robot pose with considering above effectors simultaneously.

Department of Mechano–Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan {yamazaki, watanabe, nagahama, k-okada, inaba}@jsk.t.u-tokyo.ac.jp

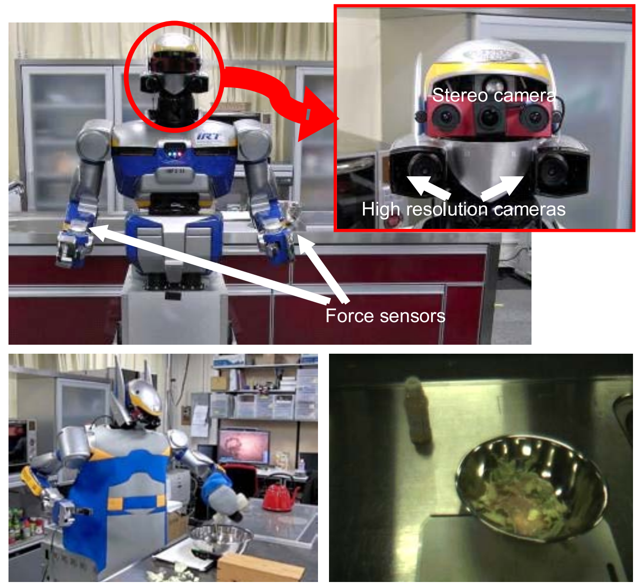

Fig. 1. A lifesized robot which is making a salad

With the software system combining these two types of functions, we actually implemented several behaviors related to cooking. Moreover, a challenging task "making a salad" was tried by a dualarm robot. In this demonstration, the robot recognized and handled several objects; a cutting board, a knife, a bowl and vegetables.

This paper is organized as follows: Section II describes related work and our approach. Section III and VI introduce our integrated system and explain each functions. Section V explains about system integration and preparation of the implementation of cooking tasks. Secion VI describes experimental results, and section VII concludes this paper.

II. Related work and our Approach

Daily assistive robots have been developed over several decades. Researchers have evaluated their control system, intelligent system or teaching system with applying their method to a single daily task in real environment [3], [4], [5], [10]. In the viewpoint of system integration, Petersson [9] et al. developed a mobile manipulator system which could pick an instructed object up, convey, and hand it to a person.

In recent years, daily assistance by using humanoid robots becomes an active area of robotics research [1], [7]. Sugano et al. presented assistance behavior by using a human symbiotic robot which has object manipulation skills [14]. We also have developed daily assistive robots provided perception, learning and motion planning skills. Several daily tasks or cooperative working etc. were implemented [8], [12].

However these results achieved object manipulation in real environments, these technological elements are insufficient to apply to cooking tasks. The reasons are that almost of conventional object manipulation by a robot have been assumed solid and distinguishable object. On the other hand, foods such as vegetables can have various shape and appearance, and tools such as knife and bowl will not have rich texture. Moreover, the robot must recognize and manipulate several tools and foods at the same time. These facts make it difficult to achieve cooking task by a autonomous robot. In this paper, we utilize several simple image processing and combine them for overcoming above conditions. Dual arm manipulation is also focused on.

Fig.1 shows a lifesized robot to be used in this research. The upper body is constructed from two 7-DOF arms and a head. A stereo camera and 2 high resolution cameras are mounted on the head. An arm, a force sensor is embedded on the wrist, and its end-effector is gripper type hand.

III. Visual functions for perceiving kitchen works

A. Basic policy

In general, almost parts of kitchen environment are formed by plane plates. It facilitates the prior knowledge making by using geometrical shapes. On the other hand, top board which is made of stainless in many cases can have high level reflection, it prevents the reliability of image-based recognition. Moreover, vegetables have various shapes, and less and unstable texture can be observed on its surface. From these facts, we use basic image features for overcoming noises and illumination influences.

Firstly, we prepare several simple image processing functions as follows:

- Edge Extraction,
- Area segmentation based on color or intensity,
- Background subtraction,
- 2D geometrical shape detection,
- Edge segments based matching.

After that, a visual function is designed by combining above functions properly. Some of these visual functions incorporate probabilistic approach or multiple hypothesis approach for robust processing.

Following subsection explains about some of the functions.

B. Model-based cutting board recognition

In order to recognize a cutting board which is a very basic equipment for cooking, model based pose estimation is applied first. As shown in upper figure in Fig.2, a simple 3D geometrical model is used for the recognition. Basically this process can be achieved by matching model edges with image edges, but image noises and light reflection interfere to make unique correspondence. From this reason, particle filter based pose recognition [8] is applied.

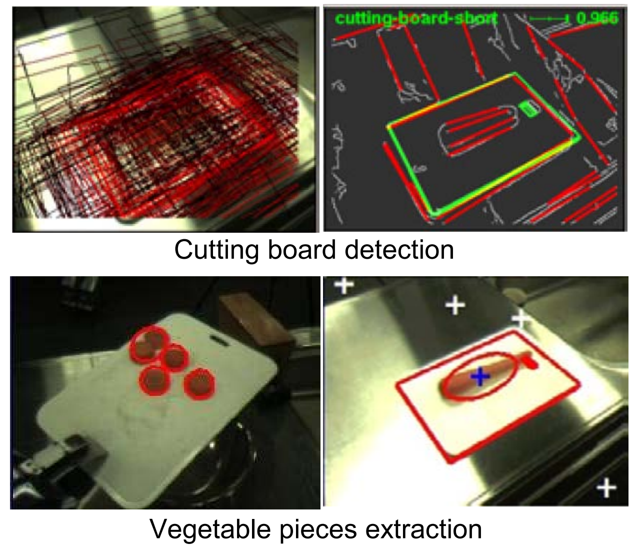

Fig. 2. Recognition based on elemental visual functions

In the recognition procedure, a pose of a target object $\mathbf{x}_t$ is estimated from measurements by external sensors $\mathbf{z}_t$ according to following two equations:

$$p(\mathbf{x}_t|Z_{t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|Z_{t-1})dx_{t-1}, \quad (1)$$

$$p(\mathbf{x}_t|Z_t) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|Z_{t-1}), \quad (2)$$

where $\mathbf{z}_t$ indicates a sensor measurement, that is, image features in our case. The Z_t is a group of $\mathbf{z}_t^i$ $(i = 1, ..., n)$ at time t. The former equation is a prior probability which is calculated before image processing at time t, and the latter is a posterior probability which includes estimation result.

In our approach, a likelihood $p(\mathbf{z}_t|\mathbf{x}_t)$ is calculated by comparing edges extracted from an input image with projected cutting board model into the image. Evaluation equation is as follows:

$$p(\mathbf{z}_t|\mathbf{x}_t) = exp\frac{(D_{edge}(E^{2D}, E^{2D}_{ref}))^2}{2\sigma^2_{edge}}, \quad (3)$$

where E^{2D} and E^{2D}_{ref} are sets of edges extracted from image processing and projected results of the 3D model, respectively. The D_{edges} is a function which investigates the nearest pair between E^{2D} and E^{2D}_{ref}. If the distance between them is short, the likelihood will have high value.

Fig.2 shows an example of recognition results. Green lines shows the contour of a cutting board found. After this process, the robot registers the color of board region. It will be used for background subtraction to extract cut vegetables.

C. Vegetable modeling and detection

Because vegetables can have various shape and appearance, such information is difficult to give in advance. In our assumption, a target vegetable is placed on the cutting board first, and it is modeled just before manipulation for cooking. By using the result of cutting board recognition,

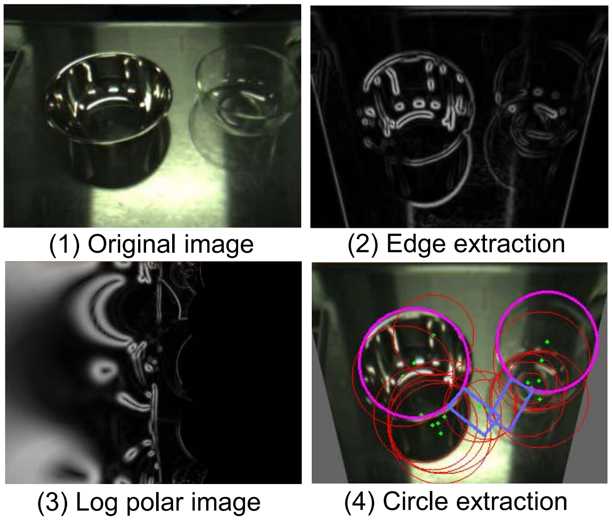

Fig. 3. Recognition based on elemental visual functions

the existing region of the target vegetable is specified. Background subtraction is used in this modeling phase, its contour and color are registered.

The contour is used as initial shape model of the vegetables. More detail information is obtained just before cutting or peeling, that is, the robot grasping a knife or other tools verifies the created model by thrusting the tool to the vegetable from side.

In the middle of cutting routine, the vegetable is divided into several pieces. Color information is used to recognize the state. For achieving the robustness against illumination influence, we apply Hue and Saturation because they are insusceptible to illumination. It is also able to be used to judge whether or not the task is finished when the robot transfers the cut vegetables to a bowl. These results are shown in section IV.

In this research we also tried to implement a peeling task. This means that the robot removes coat of a vegetable by using a peeler. Because peeled region will have different appearance from original coat, it enables the robot to recognize the degree of attainment of the peeling task.

D. Circular container detection

In general, many parts of kitchen environment is made of reflective fabric. The same is equally true of containers such as bowl and dish. It means that adequate filtering process for canceling reflection influence is very important.

In order to achieve a stable container detection, we proceed the detection in two phases; candidates extraction and their selection. In the first half process, perspective transformation is applied to an original image under the condition that target containers are placed on a horizontal plane. As this procedure enables to observe a circular object as a 2D circle in image, and next, circular hough transform is applied. Because the transformed image can have quantization error, hough transform is performed with low threshold. On the other hand, gabor filter is applied to the transformed image, and the image is expanded by means of log-polar transformation. An origin coordinates is a center of the one of extracted circles. From this result, plausibility of each candidate is judged by scanning the log-polar image with a vertical line detector.

Gabor filtering is performed based on following equation:

$$F(\mathbf{u}) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} e^{a} \sin(2\pi f_u(u-u_{cen}) + 2\pi f_v(u-v_{cen}) + p), \tag{4}$$

where f_u and f_v indicate frequency domain. $\mathbf{u} = (u, v)$ is pixel coordinates, while u_{cen} and v_{cen} are the center of gaussian. The variable a is calculated as follows:

$$a = -\frac{1}{2}\left(\frac{(u-u_{cen})^2}{\sigma_u^2} + \frac{(v-v_{cen})^2}{\sigma_v^2}\right). \tag{5}$$

This filter is designed to have strong reaction at relatively low wavelength for ignoring noises. As shown in image (2) in Fig.3, string edges are able to be found from an outline of a bowl. Log-polar image generated from this result is suitable to detect lines robustly. In the case of a situation shown in Fig.3, two bowls were found. A bowl made of glass was extracted even though intermissive edges were only observed.

IV. Manipulation functions to handle kitchen tools and foods

A. Motion control based on force information

Because of occlusion caused of viewpoint constraint, visual information do not always provide the robot with perfect results to understand its manipulation target. Moreover, it is difficult to perform image processing with high frame rates. From these reasons, force sensors equipped on wrists are used. We mainly prepare two functions; (i) to check a manipulation success by pressing a handled tool against food and (ii) to control an end-effector trajectory to conform to irregular surfaces.

The function (i) is used in conditions to cut or peel a vegetable, or to get the size of a knife and so on. Simple failure detection can be achieved by this function. The function (ii) is mainly used in peeling motion. Conventional impedance control was implemented, a peeler is controlled to follow the irregular surface on a vegetable. The equation is as follows:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{D}(\dot{\mathbf{x}} - \dot{\mathbf{x}}_d) + \mathbf{K}(\mathbf{x} - \mathbf{x}_d) = \mathbf{P}(\mathbf{f}^{act} - \mathbf{f}^{ref}), \tag{6}$$

where $\mathbf{M}, \mathbf{D}$ and $\mathbf{K}$ are virtual inertia matrix, virtual stickness matrix and virtual stiffness matrix, respectively. The matrix $\mathbf{P}$ cotrols the influence from force and moment. $\mathbf{x}$ is a 6D vector representing end-effector pose, $\mathbf{f}^{act}$ and $\mathbf{f}^{ref}$ are an input value and a reference value of the wrist force and moment.

In this control procedure, the $\mathbf{f}^{ref}$ is calculated from

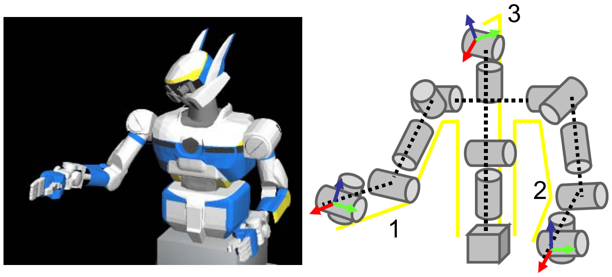

Fig. 4. Kinematic links model of a humanoid robot

following equation:

$$
\begin{aligned}
&\mathbf{x}(t) = \\
&dt\mathbf{D}\mathbf{x}(t-dt)] \\
&+[\mathbf{M}+dt\mathbf{D}+dt^2\mathbf{K}]^{-1}[\mathbf{M}[2\mathbf{x}(t-dt)-\mathbf{x}(t-2dt)] \\
&+[\mathbf{M}+dt\mathbf{D}+dt^2\mathbf{K}]^{-1}[dt^2\mathbf{K}\mathbf{x}_d(t)+dt^2\mathbf{P}(\mathbf{f}^{act}-\mathbf{f}^{ref})]
\end{aligned} \tag{7}
$$

B. Inverse kinematics for dual arm manipulation

One of the important things to consider a cooking behavior is that several objects such as tools and foods are needed to be manipulated simultaneously. This means that an unified manipulation by both arms is essential, and how to decide a viewpoint of cameras should also be considered.

We take an approach to generate a robot motion by using a single jacobean matrix. As shown in Fig.4 our robot consists of three parts of cascaded links which starts from waist to left arm, right arm and head. From this fact, we can get an integrated jacobean matrix $\mathbf{J}$, that is,

$$
\mathbf{J} = \begin{bmatrix} \mathbf{J}_{waist\_to\_Ls} & \mathbf{J}_{Ls\_to\_Lend} & \mathbf{0} \\ \mathbf{J}_{waist\_to\_Rs} & \mathbf{0} & \mathbf{J}_{Rs\_to\_Rend} \end{bmatrix}. \tag{8}
$$

The captions 'Ls' and 'Rs' mean 'left shoulder' and 'right shoulder', respectively. In this equation, a jacobean matrix related to a serial manipulator is divided into two parts which are dependent and independent on other manipulators. In the case of eq.(8), matrices of leftmost column include only elements of waist joints. Of course it is easy to add a jacobean matrix $\mathbf{J}_{head}$ related to a head by dividing the matrix into $\mathbf{J}_{waist\_to\_neck}$ and $\mathbf{J}_{neck\_to\_viewpoint}$.

By using this jacobean matrix, pose calculation is performed based on following equation:

$$
\dot{\mathbf{q}} = \mathbf{J}_w^{\#}\dot{\mathbf{x}} + \lambda(\mathbf{I} - \mathbf{J}_w^{\#}\mathbf{J}^{\#t})\mathbf{y}, \tag{9}
$$

where $\dot{\mathbf{q}}$ reveals velocity and angle velocity of end-effectors. $\mathbf{J}_w^{\#}$ indicates the multiplication result between $\mathbf{J}^{\#}$ and a weight matrix $\mathbf{W}$. The $\mathbf{J}^{\#}$ is a SR-inverse[6] of $\mathbf{J}$, which is calculated through following equation:

$$
\mathbf{J}^{\#} = \mathbf{J}^t(k\mathbf{I} + \mathbf{J}\mathbf{J}^t)^{-1}. \tag{10}
$$

$\mathbf{y}$ is an evaluation function. This should include some kind of constraints such as joint limit avoidance, collision checking and so on. In our implementation, we refer the method proposed in [2].

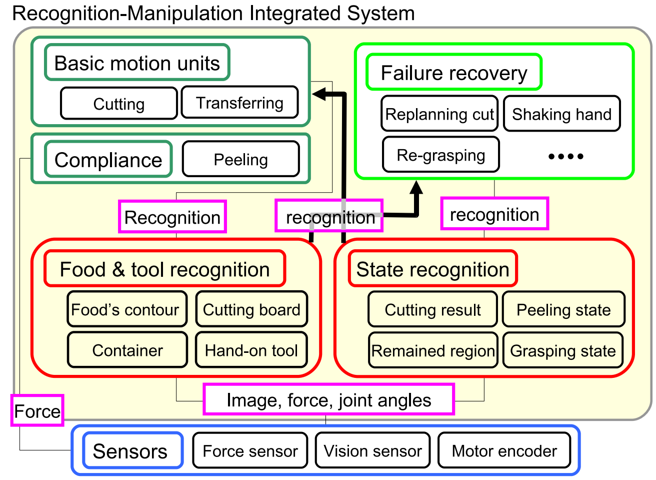

Fig. 5. A software system integrating recognition with behavior

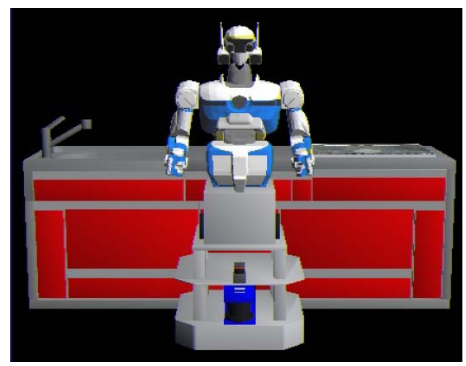

Fig. 6. 3D geometrical simulator

For instance, vegetable cutting is performed on a cutting board. It can be said that target end-effector poses of both arms are decided from the position of the cutting board. Our approach easily copes with the changing of the position because arm poses are calculated simultaneously.

V. System Integration

A. Software system

Fig.5 shows existing functions and its connection in our software system. Red frames indicate recognition functions, and dark green frames indicate motion generation functions. Basically, the motion generation functions are called after accepting recognition results. The results as a pose of a target object are reflected into 3D geometrical simulator shown in Fig.6, robot poses are calculated to handle the objects and to avoid collisions with itself and environments.

Several functions for failure recovery are also implemented to cope with simple mistakes on manipulation. Because recognition function detects an irregular condition in some cases, the result is able to be used to call a motion for recovery. In our cases, several simple motions were prepared, and failure recovery was achieved while manipulating objects.

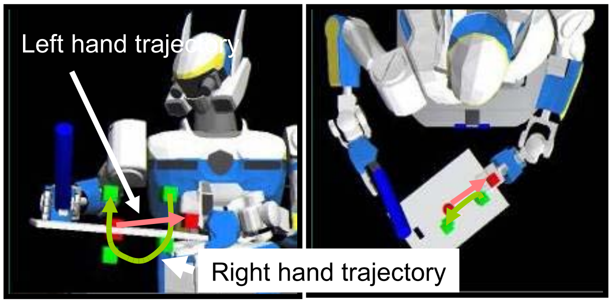

Fig. 7. Recognition based on elemental visual functions

B. Motion design and description

As a robot used in our experiment has similar joint configuration with human, our own cooking behavior is a good reference to design the robot motion. For instance, when we cut a vegetable placed on a cutting board, a shoulder of a dominant arm which is having a knife is separated from a cutting board comparing with the other shoulder. This pose makes sense because high manipulability is kept to handle the knife. In other cases, a person who transfers vegetable pieces on a cutting board into another container will move both a knife and the cutting board as shown in Fig.7. If we only consider a dominant arm for pose calculation, it is difficult to ensure sufficient working volume. From these facts, we take an approach to set initial manipulation pose with referring our own behavior, and to modify the pose by using inverse kinematics described in the former section.

All of inputs to make a robot pose are coordinates. If the robot attempts to grasp a vegetable placed on a cutting board, a coordinates is defined on the vegetable, and another coordinates fixed on the robot hand is moved until corresponding with the vegetable coordinates. On the other hand, if the robot attempts to grasp a knife and cut a vegetable, the coordinates of an end-effectors is translated on the edges of the knife. This enables to calculate a knife trajectory considering with cut position directly.

VI. EXPERIMENTS

A. Cooking tasks

To confirm the effectiveness of our approach, several cooking behaviors (i) cutting, (ii) peeling and (iii) transferring were arranged to robotic tasks. Some vegetables such as cabbage, carrot and eggplant were chosen to be cooked.

Cutting was a task that a vegetable was divided into some pieces. Under the condition that the vegetable was placed on a cutting board, the cutting was performed while the robot grasped a knife in one hand, and bore down the vegetable by other hand. **Peeling** was a task to remove coat of a vegetable. By using special tools such as a peeler, a part of the vegetable's surface was peeled in incremental steps. **Transferring** was a task that some pieces of vegetables placed on a cutting board were moved to another container as a bowl. The robot lifted up the cutting board, and pushed away the vegetable pieces by using a knife.

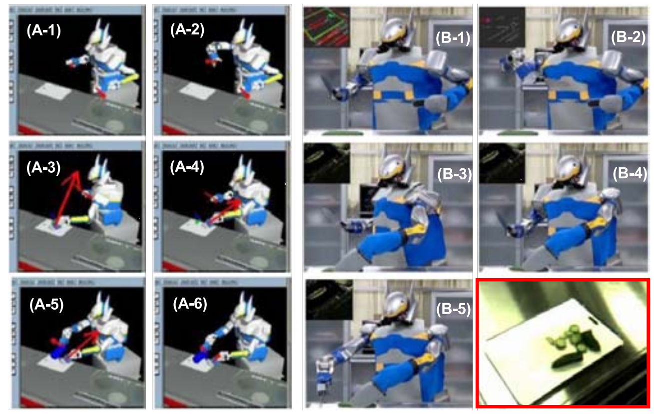

Fig. 8. An experiment of vegetable cutting using a knife

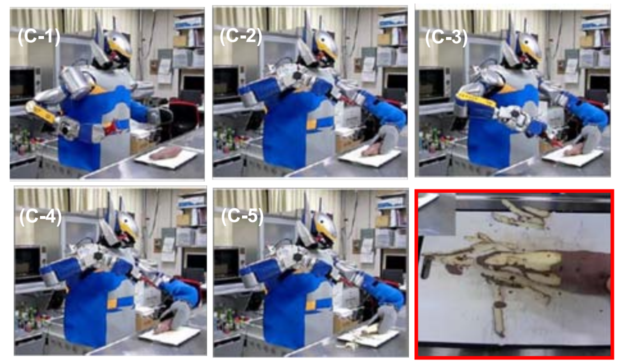

Fig. 9. An experiment of vegetable peeling

B. Experimental results

Fig.8 shows an experiment of a cutting task. Left figures shows the robot state in our simulator, it shows not only 3D geometrical models but also force orientation and strength against wrists. Right figures show motion sequence and processed vegetable. In this sequence, the robot first confirmed the tip of a knife by using vision, and then found a cutting board and a target vegetable. Next, it picked up and moved the vegetable to the center of the cutting board. Before cutting, the robot measured the length of the vegetable by pressing the knife against the vegetable.

In this case, a cucumber was targeted, the robot cut it into 8 pieces in 9 seconds. Whole cutting sequence was performed smoothly, and then all of pieces had about 5 [mm] in width. We also tried to cut more hard vegetable such as carrot. In such case, however the robot was sometimes unable to cut it, failure detection functions by means of force sensor worked well. The cutting was retried with changing the knife trajectory and the task was accomplished.

Fig.9 shows an experiment of peeling task. A sweet potato was targeted, the robot peeled the coat. If one peeling motion was achieved, the robot rotated the potato by using left hand, and continued the peeling. Finally, almost all of the coat on target region was removed as shown in lower right figure in Fig.9.

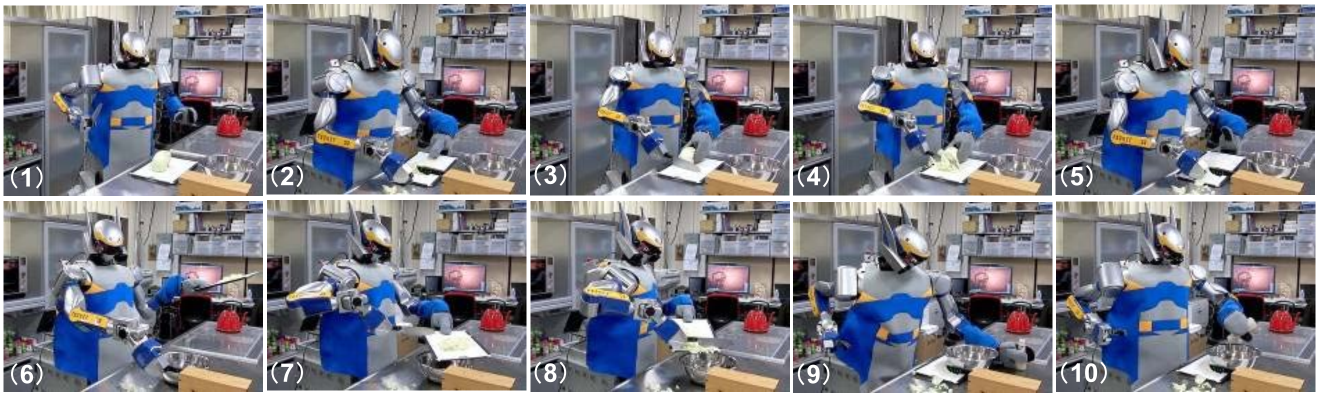

Fig. 10. Salad making sequence

C. Making a salad

Finally, we tried to implement sequential cooking task; making a salad. The procedure is as follows:

- move a cabbage onto a cutting board,
- cut the cabbage by using a knife,
- transfer the cut pieces into a bowl,
- dress the pieces.

Fig.10 shows one of the experimental results of making a salad. In this experiment, though a person intervened on several occasions because of shifting tools grasped on the robot hand, other behaviors including recognition and motion generation were performed automatically.

VII. Conclusion

In this paper we described a system integration of a daily assistive robot. Recognition and manipulation functions needed for cooking tasks were developed, Several complicated tasks were achieved by using a real robot. In the recognition functions, we take an approach to use simple image features to overcome difficult visual conditions, such as illumination influence and poor texture objects. On the other hand, manipulation functions to handle various cooking tools were constructed by considering with both arms simultaneously. Our own behavior of cooking was also useful to design the poses. Through experiments doing several cooking tasks with handling daily tools, we could show the effectiveness of our approach.

References

[1] T. Asfour et al. “ARMAR-III: An Integrated Humanoid Plattfrom for Sensory-Motor Control,” IEEE-RAS Int’l Conf. on Humanoid Robots, 2006.

[2] T. F. Chang and R.-V. Dubey: “A weighted least-norm solution based scheme for avoiding joint limits for redundant manipulators,” in IEEE Trans. On Robotics and Automation, 11((2):286-292, April 1995.

[3] N.Y. Chong and K. Tanie: “Object Directive Manipulation Through RFID,” Proc. Int’l Conf. on Control Automation and Systems, pp.22-25, 2003.

[4] H Jang et al.: “Spatial Reasoning for Real-time Robotic Manipulation,” Proc. of IEEE/RSJ Int’l Conf. on Intelligent Robotics and Systems, pp.2632–2637 , 2006.

[5] R. Katsuki et al.: “Handling of Objects with Marks by a Robot,” Proc of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, pp.130–135, 2003.

[6] Y. Nakamura and H. Hanafusa, ”Inverse Kinematic Solutions with Singularity Robustness for Robot Manipulator Control,” Journal of Dynamic Systems, Measurement, and Control, Vol.108, pp.163-171, 1986.

[7] E. S. Neo et al.: “Operating Humanoid Robots in Human Environments”, Proc. Workshop on Manipulation for Human Environments, Robotics: Science and Systems, 2006.

[8] K. Okada et al.: “Vision based behavior verification system of humanoid robot for daily environment tasks”, 6th IEEE-RAS Int’l Conf on Humanoid Robots, pp 7–12, 2006.

[9] L. Petersson, et al.: “Systems Integration for Real-World Manipulation Tasks,” Proc. of Int’l Conf. on Robotics and Automation, Vol.3, pp.2500–2505, 2002.

[10] T. Takahama, K. Nagatani, Y. Tanaka: “Motion Planning for Dual-arm Mobile Manipulator –Realization of “Tidying a Room Motion” –,” Proc. of Int’l Conf. on Robotics and Automation, pp.4338-4343, 2004.

[11] K. Yamazaki and M. Inaba: “A Cloth Detection Method Based on Image Wrinkle Feature for a Daily Assistive Robots,” IAPR Conf. on Machine Vision Applications, 2009. (to appear)

[12] K. Yamazaki, R. Ueda, S. Nozawa, Y. Mori, T. Maki, N.Hatao K. Okada and M. Inaba: “System Integration of a Daily Assistive Robot and its Application to Tidying and Cleaning Rooms,” Proc. of IEEE Int’l Conf. on Intelligent Robots and Systems, 2010 (to appear).

[13] K.Yokoi et al.: “Experimental Study of Humanoid Robot HRP-1S,” The International Journal of Robotics Research, 2004.

[14] http://twendyone.com/index.html