Cooperative Manipulation Works by a Human and a Humanoid Robot based on Human Imitation and Human Aural Instructions

Hideaki Ito, Masaki Murooka, Iori Yanokura, Shunichi Nozawa, Kei Okada and Masayuki Inaba

***Abstract*— Cooperative works between a human and a humanoid robot are important considering household chores. We have constructed a system to realize various cooperative tasks by a human and a humanoid robot. The cooperative works we worked on are those which have to be done by two workers because of the size or the weight of the objects that are too big or heavy to be manipulated by one person (e.g. folding a large tablecloth or carrying a large board). In this paper, we propose a interactive human-robot cooperative task executing system. In the system, cooperative works are decomposed into two layers; local action generation and global transitions between those actions. In the former, the robot acquires its motions based on human imitation and feedback modification applying force and voice. In the latter, the robot switches its motion according to the human instructions. Superiority of this system can be described with the ability in general use due to the sustainability and the flexibility. The effectiveness of the proposed collaborative system is shown in the conducted experiments, executing cooperation tasks that don't have fixed orders between each contained process.**

I. Introduction

Daily life assistance is one of the important missions for service robots. Especially, when the human tends to execute some tasks that are troublesome to perform alone due to the limitation in reachability or load, the humanoid robot that has a human-like body structure can be a great partner for the human (Fig.1). In this research, as the typical tasks where one needs other's help, we work on the manipulation tasks of both large rigid and flexible objects, and aim at the acquisition of the ability for humanoid robots to accomplish various manipulation tasks in cooperation with humans.

In the daily life environment, there are a wide variety of works needed to be done. One of the difficulties in such cooperative works between a human and a robot is that the robot is expected to take actions autonomously according to what the human wants the robot to do. In this research, we deal with this difficulty by decomposing cooperative works into local action generation and global transitions between those actions, and effectively utilizing multimodal information obtained from the cooperating human.

In the local action generation, we categorize the actions required in a series of cooperative tasks into three styles according to the attention target in the action, and realize each action by the imitation of human hand movement, walking velocity or arm posture taking advantage of the similarity in body structure and the feedback modification based on force and voice.

In the global transitions, the robot switches its motion at appropriate timing in proper order by detecting multimodal simple instructions from the human.

Using this framework, leaving the high order decisions in human hands and following the multimodal signs by the human, the robot can take an action that is autonomously acquired with the local purpose based on perceptual processing, and finally perform various cooperative tasks with the human co-worker.

Fig. 1: Several tasks dealt with in this paper; folding a large cloth, carrying a large board with other objects on it and passing large objects from a human to a robot.

A. Related Works

1) Robot-robot Cooperative Works: There are some researches on cooperative works by two robots [1][2][3], not by a human and a robot. These researches are efficient if we consider the situation where the objective task has a fixed order between its included processes beforehand like a task in factories. However, under the household situations, the tasks don't necessarily come in the same order. Since it is difficult for a standalone robot to perform an advanced judgment like deciding the order of processes in order to execute the whole task or when to go on to the next process, programmers have to prepare the whole robot motions for each task when they want the robots to execute different tasks. It is not suitable for manipulation tasks in the daily life environment including several processes in several orders. In this research, assuming

H. Ito, M. Murooka, I. Yanokura, S. Nozawa, K. Okada and M. Inaba are with Department of Mechano-Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan h-ito at jsk.imi.i.u-tokyo.ac.jp

a household situation where there is a human in front of the robot, we solve such problems by relying on the human making those high-quality decisions. With the robot obeying the human requests, the objective tasks are executed.

2) Human-robot Cooperative Works: Human-robot cooperation can be seen in several types. Passing an relatively small item to a human who needs it will be one collaborative work [4][5]. Wearable robots assisting humans are also in this field [6][7]. Manipulating a single object with a human and a robot, transportation of a large panel or a table [8][9] is also studied. However, although a certain task can be executed following these studies, these works are specialized and thus limited to a single certain task; passing an item, providing an extra hand or carrying a table. To achieve a series of tasks in the household situations, it is not enough and have to be integrated.

On the other hand, when manipulating the same objects with two operators that are too large to be manipulated by one, communication and interaction between the two become more important than manipulating different things. The methodologies for communicating and interacting include voice recognition [10], visual gesture understanding [11][12][13], and also haptic (force/torque) information [8][9][14][15] can be joined.

3) Motion Acquirements by Imitation Learning: There are a lot of attempts for robots to acquire deliberate actions by first imitating human motions and then learning the meaning of them [16]. The object of these studies is that the robot learns and gets a skill to accomplish a certain task on their own. However, not considering cooperativity, these learned actions are hard to be applied to cooperative works where the communication with each other is important. In this paper, we adopt a pure imitation of the human motion when deciding the robot motion, which is the simplest reaction in the field of imitation learning, but is the most basic and the most generic approach.

B. Contributions and Overview of this Paper

Here, we would like to discuss the requirements in the robots among the human-robot collaborative tasks manipulating one object at the same time. The robots

1) need to acquire their motions adequate to achieve the purpose with helping humans,
2) must realize 1) in a safety manner and meet the objects' constraints,
3) should have a framework to obey the human commands sustainably.

We take an approach to satisfy these needs by combining visual recognition, haptic information and speech interaction as shown in Fig.2. First, the robot gets its motions observing human motions and imitating them. Here, we prepare a vision system to obtain human motion information. Then, referring force sensors data, those robot motions are modified to satisfy the objects' constraints. Finally, by introducing a speech recognition system, we constructed an integrated cooperative task executing system where a human can instruct the robot how to and when to take actions by aural communication.

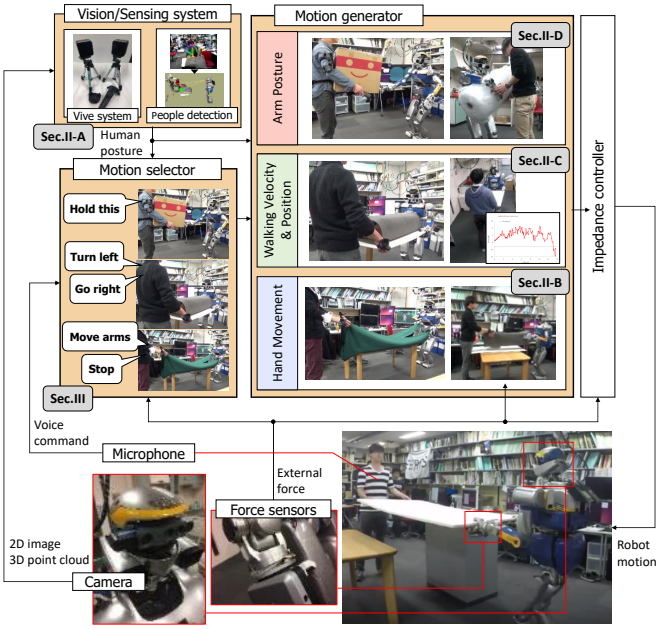

Fig. 2: Proposed system in this paper.

In the following, we explain the methods how the robot gets its motion referring human motion in Sec.II, and the integrated system for collaborative works based on those motions in Sec.III. Then, we show the experiments with the real robot applying the whole system in Sec.IV.

II. Human Motion Based Robot Motion Generation

In this research, we take human motions into account to decide robot motions with imitating them. According to the motion feature in each task, we sort the tasks in 3 classes – manipulation i) by moving hands, ii) by walking, and iii) by holding with whole upper body – and prepare the robot motion based on those human motion feature in each class. First, we describe the vision system used to acquire human motion information in Sec.II-A, and then the methods to decide the robot motions in each class in Sec.II-B, Sec.II-C and Sec.II-D.

A. Visual Recognition to Acquire Human Motion Information

In order for robot to acquire human motion information, several approach can be considered. Using PointCloud measurements to get human skeleton, Microsoft Kinect or Xtion PRO Live can be applied. However, those techniques are limited to the detection of people whose bodies are not largely occluded by other objects. Since the robot needs to observe the human posture while manipulating large objects, these approaches won't be successful. To overcome that problem, we apply a visual recognition system using OpenPose [17], which is the human detection technique using 2D images, to get the 3D positions of human limbs (Fig.3). Since OpenPose has the characteristic in using a bottom-up method to detect human pose, it doesn't require the whole body to be seen. In addition to it, HTC Vive system [18] (Fig.4) is introduced to get the orientations of human limbs. HTC Vive system can measure the position and the orientation of each controller or

tracker using infrared laser measurement. By attaching those devices on the human body, we can get the pose of each human limb (Fig.5).

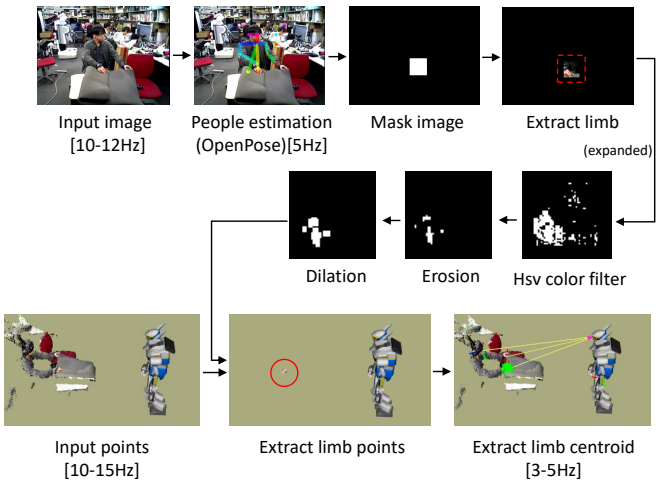

Fig. 3: Getting the three dimensional positions of human hands and face during object manipulation.

Fig. 4: HTC Vive system. A controller (right front) and a tracker (left front) recieve the IR laser from the two base stations called "lighthouse" and calculate the position and orientation.

Fig. 5: The example of attaching the Vive trackers and the controllers on the human body.

B. Object Operation based on Hand Movement

In situations of cooperative works where a human and a robot manipulate the same object at the same time with facing each other, we propose that the robot end effectors' coordinates can be decided by observing human hand motions and imitating them symmetry. Fig.6 illustrates the reference frames and naming convention used in the rest of this paper. We focused on the hand coordinates $\{rh\}$ and $\{lh\}$ represented in the face coordinate $\{f\}$. Using the homogeneous transformation matrixes $\boldsymbol{T}_h^{trgt}$ and $\boldsymbol{T}_f^{trgt}$, two coordinates of the hands $\{rh\}$ and $\{lh\}$ relative to the head coordinates $\{f\}$ can be represented as homogeneous transformation matrixes ${}^f\boldsymbol{T}_h^{trgt}$ as below.

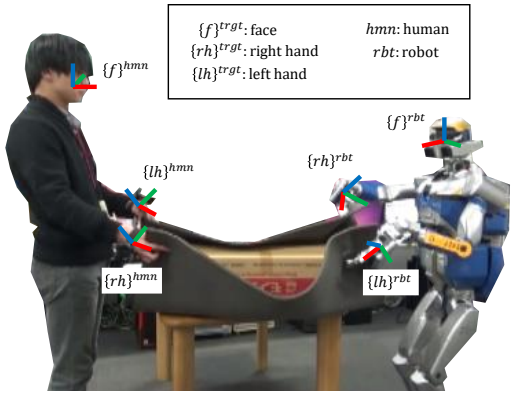

Fig. 6: Coordinate reference used in this paper.

$$
\begin{aligned}
{}^f\boldsymbol{T}_h^{trgt} &= (\boldsymbol{T}_f^{trgt})^{-1}\boldsymbol{T}_h^{trgt} \\
\text{where} \quad trgt &\in \{rbt, hmn\}, \\
h &\in \{rh, lh\}
\end{aligned} \tag{1}
$$

A homogeneous transformation matrix $\boldsymbol{T}$ can be converted into 6-dimensional vector $\boldsymbol{t} = (\boldsymbol{p}^{\mathrm{T}}\ \boldsymbol{\phi}^{\mathrm{T}})^{\mathrm{T}} = (x\ y\ z\ \alpha\ \beta\ \gamma)^{\mathrm{T}}$. $\boldsymbol{p} = (x\ y\ z)^{\mathrm{T}}$ represents the positions and $\boldsymbol{\phi} = (\alpha\ \beta\ \gamma)^{\mathrm{T}}$ represents the Euler angles in the ZYX convention. Considering two coordinates of the human hands, the desired robot end effectors' positions and orientations are determined as below.

$$
\begin{pmatrix} {}^fx_h^{rbt} \\ {}^fy_h^{rbt} \\ {}^fz_h^{rbt} \end{pmatrix} = \begin{pmatrix} {}^fx_{h'}^{hmn} \\ -{}^fy_{h'}^{hmn} \\ {}^fz_{h'}^{hmn} \end{pmatrix}, \tag{2}
$$

$$
\begin{pmatrix} {}^f\alpha_h^{rbt} \\ {}^f\beta_h^{rbt} \\ {}^f\gamma_h^{rbt} \end{pmatrix} = \begin{pmatrix} -{}^f\alpha_{h'}^{hmn} \\ {}^f\beta_{h'}^{hmn} \\ -{}^f\gamma_{h'}^{hmn} \end{pmatrix} \tag{3}
$$

$$
\text{where} \quad h \cup h' = \{rh, lh\}
$$

Here, to decide robot right hand target coordinates, those of human left hand are used. The same can be said for the reversed.

We solve Inverse Kinematics (IK) with these end effectors' constraints and get the robot angle vector. When time series angle vectors are applied to the real robot, there are three major constraints or difficulties, described in [19].

1) joint angle limits
2) joint angular velocity limits
3) motion around the gimbal lock

In the previous researches, human motions of actors or dancers are applied to those of robot [19][20] with satisfying these restrictions with various optimizations. However, since these approaches are executed offline, these methods are not so useful here where online movement is required and not so much about the exact timing. Therefore, in this research, we take temporal solutions; limiting the area where the robot can move its hands in order not to violate 1) and 3), and setting the interpolating time on the real robot joint angles as 4 sec so that 2) can be eased.

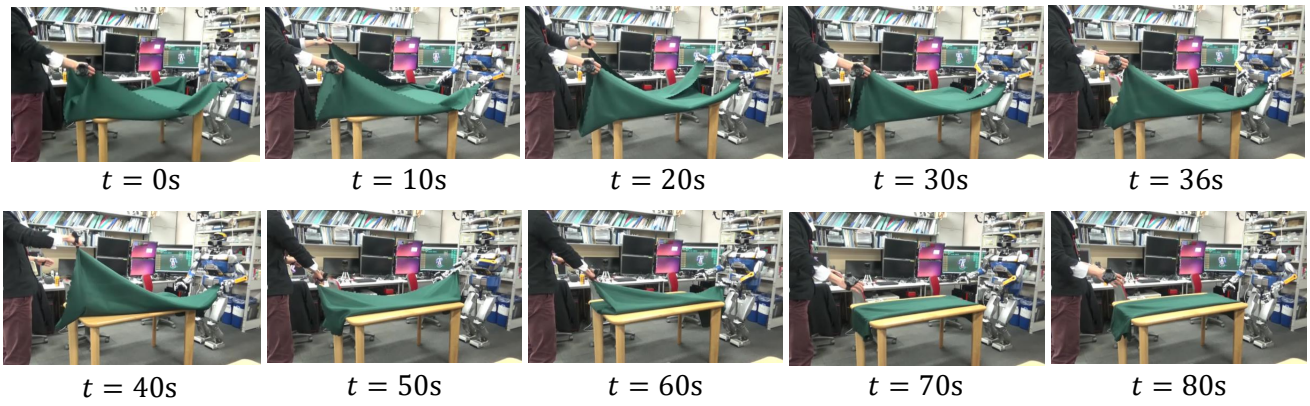

Fig. 7: The experiment of folding a large cloth by a human and a robot. At $t = 36\,\mathrm{s}$ and $t = 60\,\mathrm{s}$, the human make the robot release its hand by using voice instruction.

Fig.7 shows a short experiment proving that these robot motions work suitably in the task of folding a table cloth by a human and a robot grasping each side of it. At $t = 36\,\mathrm{s}$ and $t = 60\,\mathrm{s}$, using also the aural instructions, the human order the robot to release its right or left hand.

C. Cooperative Carrying Following Human Walking

In case that transportation of the object are needed for executing the task, we also introduced a robot walking motion by observing human velocity and position. In the previous researches, the walking direction or the speed of the robot is controlled by the interactive force or torque [8][9]. Although rigid objects can be carried using those methods, they cannot be applied to carry flexible objects like cloth. To overcome that problem, we introduced a method to generate robot velocity considering human velocity and position as below.

$$v^{rbt}_{vel,t+1} = v^{rbt}_{vel,t} + f_{vel}(v^{hmn}_t) \tag{4}$$
$$v^{rbt}_{pos,t} = f_{pos}(x^{hmn}_t - x^{hmn}_d) \tag{5}$$
$$v^{rbt}_{real,t} = v^{rbt}_{vel,t} + v^{rbt}_{pos,t} \tag{6}$$
$$\text{where} \quad v^{hmn}_t = \frac{x^{hum}_t - x^{hum}_{t-\Delta t}}{\Delta t}$$

$v^{rbt}_{vel,t}$ represents the robot velocity considering the human velocity v^{hmn}_t, and $v^{rbt}_{pos,t}$ represents the robot velocity considering the human position x^{hmn}_t compared with the desire position x^{hmn}_d. Note that v^{hmn}_t and x^{hmn}_t are the velocity and the position relative to that of the robot, and not the absolute value. Finally, $v^{rbt}_{real,t}$ represents the robot velocity applied to the real robot. The function f is described as below and the graph of it is shown in the left in Fig.8.

$$f(x) = \begin{cases} -v_{max} & (x < -thre - \frac{v_{max}}{gain}) \\ gain \cdot (x + thre) & (-thre - \frac{v_{max}}{gain} \leq x < -thre) \\ 0 & (-thre \leq x < thre) \\ gain \cdot (x - thre) & (thre \leq x < thre + \frac{v_{max}}{gain}) \\ v_{max} & (otherwise) \end{cases} \tag{7}$$

Eq.4-Eq.6 can be simplified to the equation below.

$$\Delta v^{rbt} = v^{rbt}_{real,t+1} - v^{rbt}_{real,t} \tag{8}$$
$$= K_p(v^{hmn}_{abs} - v^{rbt}) + K_i \sum (v^{hmn}_{abs} - v^{rbt})\Delta t \tag{9}$$

This can be recognized as a PI controller to let the robot velocity match the human velocity giving proper parameters K_p and K_i, which correspond to the parameter $gain$ in the function f.

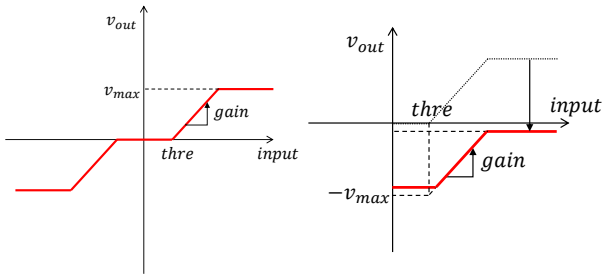

Fig. 8: The left represents a graph of a filter to generate robot velocity from human velocity or human position. Constant $thre$, $gain$ and v_{max} are the given parameters according to the walking direction. The right represents a graph of a filter to generate robot velocity from human position when the human and the robot are moving in the opposite direction back.

On the other hand, there will be a situation that the two operators move in the opposite direction close or away when manipulating flexible objects; folding clothes with approaching each other or stretching them with taking a distance between each other. In such situation, velocity is not so much important than moving in the same direction with carrying the same thing. We consider only position when the two moving against each other.

$$v^{rbt}_t = v^{rbt}_{pos,t} = f_{pos}(x^{hmn}_t) \tag{10}$$

The same function f can be applied when the two move in the opposite direction close. The right in Fig.8 shows a new graph applied when the two move in the opposite direction away.

To secure the use with rigid objects, we take haptic information into account in the feedback layer. Using the

methods above, the robot can recognize when to stop by observing human velocity and position. However, in order to carry rigid large objects by the two, this may not sometimes be enough. Because of the relatively low frequency of the visual recognition, the robot may not be able to respond to the sudden stop of the counterpart. This causes an unexpected falling down of the robot. Preventing this accident, we introduce the feedback layer that the robot stops walking when feeling unexpected force against the walking direction. This can be used not only to prevent the accident, but also to suggest that it's an enough distance to take, when, for example, the two are grabbing each side of the cloth and stretching it.

Finally, the aural instruction of the direction by the human is combined. Which direction to move can be decided by the human and announcing it, robot can start moving with the velocity generation explained above.

D. Passing Large Objects based on Arm Posture

In order to carry large objects, we need to use whole upper body including arms. When passing those objects from a human to a robot, the robot can watch how the human holds them and regard those postures as suitable ones to hold the objects whose weight is unknown. We get the positions and orientations of human 7 points (shoulders, elbows, hands and torso) to reconstruct the human posture carrying the object, and apply them to the robot to make the similar posture to the human. However, passing objects cannot be succeeded only by it. We take a mean of fixing robot posture by human teaching in order for the robot to hold the objects stably.

To acquire the robot angle vector, two methods can be considered; IK based approach and joint-space based approach. The former approach ensures the end effector be fixed to the target coordinate. It is useful when executing tasks where the position and orientation are important; reaching and picking a certain object. However, to hold large objects, the whole posture of arms appears to be more essential than exact position or orientation. Although those arm postures might be able to be generated with adding other constraints when calculating IK, it causes more failure in IK solution. Therefore, we take the joint-space approach considering each direction of human limb where the above weaknesses in IK don't appear.

Using the Vive system, position and orientation information of human shoulders, elbows, hands and the torso are gotten. We apply those information to the robot to consider the robot arm posture. The robot has 7 DoF in its each arm; 3 DoF in shoulder, 1 DoF in elbow, 3 DoF in wrist. The angles of these joints are calculated as below.

1) The roll and pitch joint angle of shoulder are calculated by desired shoulder direction ${}^{sh}\boldsymbol{v}_{el}^{hmn}$.
2) The pitch joint angle of elbow are calculated by the angle between ${}^{sh}\boldsymbol{v}_{el}^{hmn}$ and ${}^{el}\boldsymbol{v}_{hnd}^{hmn}$.
3) The yaw joint angle of shoulder is calculated by the angle between $\boldsymbol{n}^{hmn}$ and $\boldsymbol{n}^{rbt}$ that are the normal vectors of the planes that contain ${}^{sh}\boldsymbol{v}_{el}^{trgt}$ and ${}^{el}\boldsymbol{v}_{hnd}^{trgt}$.
4) The roll, pitch and yaw joint angle of wrist are calculated by solving IK with only 3 DoF in the wrist.

Vectors used above are illustrated in Fig.9.

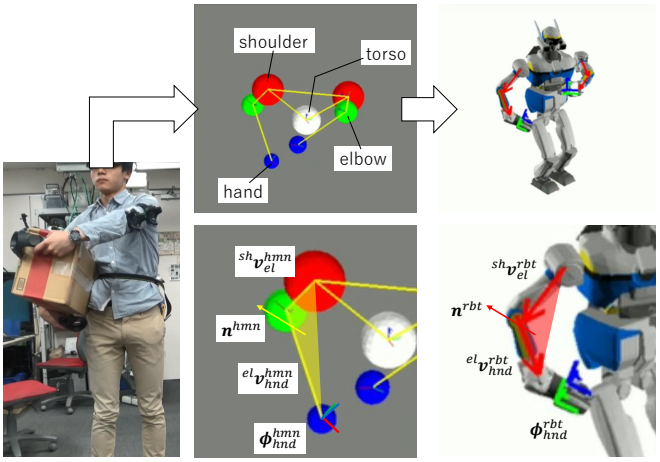

Fig. 9: Arm posture generation in robot.

Taking the average of the time series robot upper-body postures calculated with the above calculation, we decide one pre-grasp robot posture and apply to the real robot, supposing the human postures when holding an large object are almost unchanged.

Holding large objects cannot be realized only by the above process. Although the robot links are configured similar to those of human, each link length of the robot is not exactly the same as human. Moreover, to hold objects, pre-grasp poses leading the final grasping poses are needed [21]. To satisfy those requirements, we settle this problem by introducing physical interaction. As noted above, the posture of the upper body is important to hold large objects. To preserve the posture made through the above process, a system is introduced where a human can move each joint by choosing which joint to move by voice and the extent of the change in the joint angle is determined according to the extent of force.

Fig.10 shows the snapshots of the experiments passing large boxes from a human to a robot. First, the robot takes an pose acquired by considering human holding pose. Here, it has to be noted that the posture the robot takes are right/left inverted as that of human. Then, with human fixing the robot posture by force and voice, the human-robot object passing is executed, finally the robot stepping in the position with holding the boxes.

III. Motion Selection based on Human Signs

We construct a rule-based reaction system for a human user to activate the robot action. Preparing some sets of keyword and associated robot motion, the human can make the robot move with considering what action he/she wants the robot to take and how it can be realized combining the prepared rule-based actions. Although which order the robot moves completely relies on which order the human commands, this system sufficiently works in a household

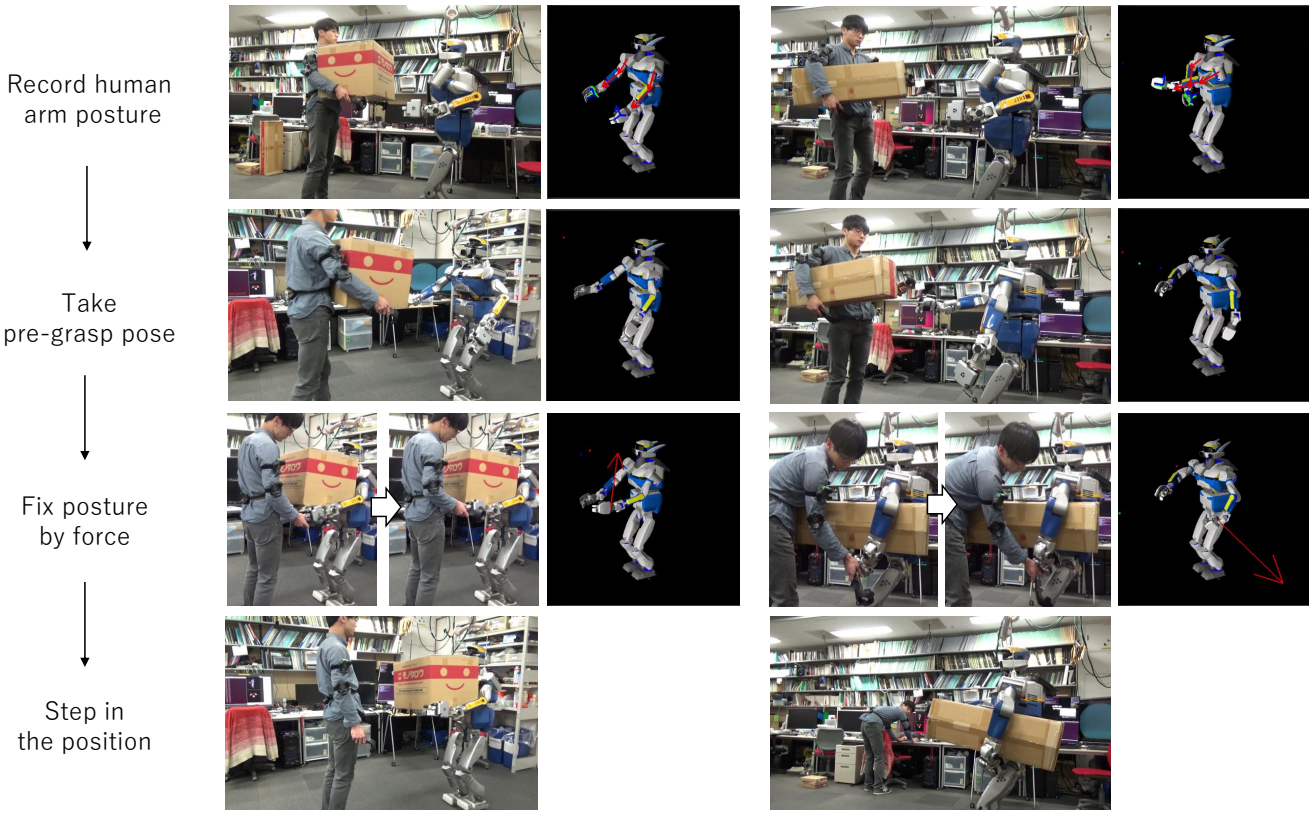

Fig. 10: The experiments of passing objects from the human to the robot.

environment where the human wants the robot to be an obedient assistant. Table I below shows the list of the prepared keywords and associated robot motions. For speech recognition, we use the system produced in our laboratory that is now published as an application for an android phone [22].

TABLE I: Speech commands.

usage	command	detail
Selection of mode	Walk	walking mode
	Move arms	manipulating mode
Direction of walking	Go forward	translate forward
	Go back	translate backward
	Go right	translate right
	Go left	translate left
	Turn right	rotate clockwise
	Turn left	rotate counterclockwise
	Come here	move in front of human
	Go away	move opposite back
Hand motion	Grasp (right/left)	grasp (right/left) hand
	Release (right/left)	release (right/left) hand
Begin/End	Start	start motion
	Stop	stop motion
	OK	completion of motion
Selection of limb	(right/left/both) shoulder	move (right/left/both) shoulder
	(right/left) elbow	move (right/left) elbow
	(right/left) wrist	move (right/left) wrist

Using these rule-based commands, we combine the robot manipulating motions described in Sec.II-B and the walking motions described in Sec.II-C in parallel and enable them to be activated repeatably as shown in Fig.11.

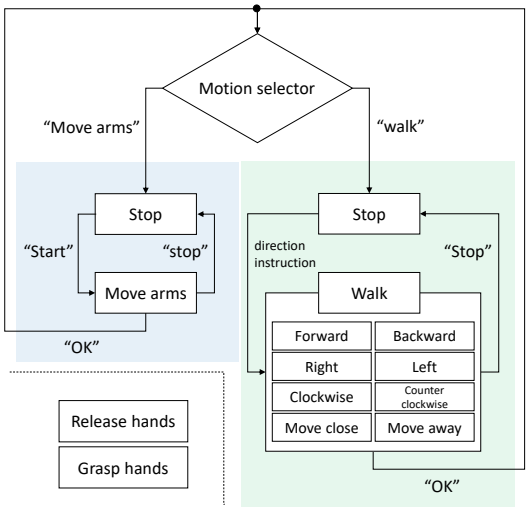

Fig. 11: The state flow to execute several cooperative works. The states colored blue are the motions explained in Sec.II-B, and the the states colored green are the motions described in Sec.II-C. Lining up these motions parallelly with using aural interactions, continuous task execution can be realized.

Added to these aural instruction commands, we attempt to induce robot movements not only by voice but also by other factors; force information and human standing position. Although the voice commands ensure the absolutely correct motion with the robot, considering human-human interaction, there must be other communication means during the cooperative works. For that kind of communication between

a human and a robot, we introduce two characteristics to the robot in cooperative carrying. One is that when the robot feels strong force while the robot is not walking, the robot regards it as the suggestion to move in that direction where the force works. Another is that when the human moves sideward exceeding the threshold, the robot regards it as the sign to move sideward in that direction and starts to move. Note that using the motion described in Sec.II-C, there don't have to be excess internal force during carrying. These characteristics are only for the triggers of the walking motions.

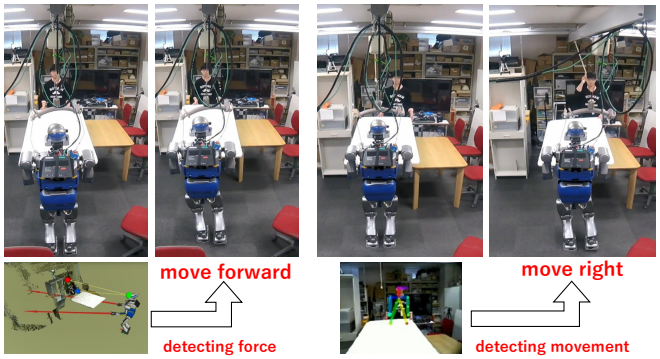

Fig. 12: The examples of multimodal signs inducing the robot to start moving. The left represents the robot starting walking forward due to the force felt through the object. The right represents the robot starting walking to the right due to the human position.

Fig. 13: The first half of the experiments applying the integrated cooperative system. Stretching the cloth (1)-(3), transporting it (4) and wrapping with it (5)-(12) are executed sequentially. At $t = 180$ s and $t = 230$ s, the human make the robot stop its motion, and it realizes the role to keep the cloth wrapping the box, not being released.

IV. Sequential Task Execution Experiment

We conduct a sequential experiment supposing the situations in a household environment.

Fig.13 shows the first half of the experiment. At $t = 55$ s, the robot stops walking back with feeling the force caused by the cloth, and stretching the cloth by the human and the robot is realized. Then, conducted by the human instructions and the movements, the transportation of flexible objects is executed by the two at $t = 110$ s. From $t = 115$ s, wrapping task of the box is accomplished. Switching the robot motion type from walking mode to manipulating mode by a human voice command and applying the robot motion generation based on the human hand movement, the task is done completely. Ordering the robot to stop moving its arms by voice at $t = 180$ s and $t = 230$ s, the role is realized to hold the cloth not to be released until the human put tape on it.

Fig.14 shows the latter half of the experiment. It shows that our collaborative carrying methods also works on rigid objects; the box and the board. Applying human velocity and position to generate robot velocity instead of haptic information, sideward transportations that seem to be difficult and aren't executed in the passed works [8][14] are realized.

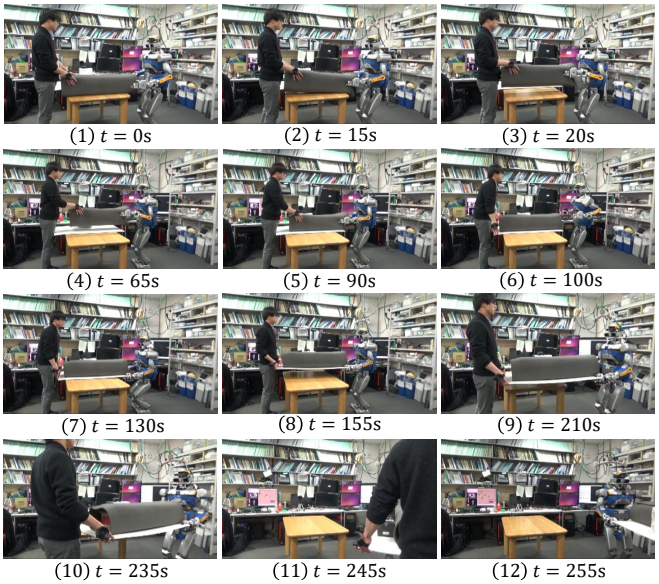

Fig. 14: The latter half of the experiments applying the integrated cooperative system. Carrying not only flexible objects but also rigid objects is executed through the whole experiment; the box (1)-(5) and the board (6)-(12).

V. Conclusions

In this study, we worked on the human-humanoid collaborative works manipulating large objects that will be hard to be executed by one. To realize various tasks in a series of

household chores, the robot has to acquire flexible motions that are not programmed for a certain task. We took an approach to meet that expectation by taking human motions into account for the robot motions. Moreover, modifying those generated motions with the use of force information and aural communication, both rigid and flexible objects can be manipulated in cooperation with a human. Eventually, enabling those motions be activated selectively and continuously based on the multimodal human instructions, tasks can be finely executed in which the order between the included processes is random. Our contributions are mainly based on the decomposition of the cooperative works into the local action generation and the global transition framework. The sequential task experiment in this research confirms that this system can perfectly work in human-robot collaborative works in the household environment where the human has the initiative and wants the robot to be an obedient assistant.

REFERENCES

[1] Y. Hirata et al. Leader-follower type motion control algorithm of multiple mobile robots with dual manipulators for handling a single object in coordination. In *International Conference on Intelligent Mechatronics and Automation*, pp. pp.362–367, 2004.

[2] M. H. Wu, A. Konno, S. Ogawa, and S. Komizunai. Symmetry cooperative object transportation by multiple humanoid robots. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3446–3451, 2014.

[3] M. Dogar, A. Spielberg, S. Baker, and D. Rus. Multi-robot grasp planning for sequential assembly operations. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 193–200, 2015.

[4] G. Maeda, M. Ewerton, R. Lioutikov, H. Ben Amor, J. Peters, and G. Neumann. Learning interaction for collaborative tasks with probabilistic movement primitives. In *IEEE-RAS International Conference on Humanoid Robots*, pp. 527–534, 2014.

[5] W. P. Chan, Y. Kakiuchi, K. Okada, and M. Inaba. Determining proper grasp configurations for handovers through observation of object movement patterns and inter-object interactions during usage. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1355–1360, 2014.

[6] B. Llorens et al. A robot on the shoulder: Coordinated human-wearable robot control using coloured petri nets and partial least squares predictions. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 119–125, 2014.

[7] A. Dollar and H. Herr. Lower extremity exoskeletons and activd orthoses: challenges and state-of-the-art. In *IEEE Transactions on Robotics*, Vol. 24, pp. 144–158, 2008.

[8] K. Yokoyama, H. Hanade, T. Isozumi, Y.Fukase, K. Kaneko, F. Kanehiro, Y. Kawai, F. Tomita, and H. Hirukawa. Cooperative works by a human and a humanoid robot. In *IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 2985–2991, 2003.

[9] J. Stückler and S. Behnke. Following human guidance to cooperatively carry a large object. In *IEEE-RAS International Conference on Humanoid Robots*, pp. 218–223, 2011.

[10] J. Lee, J. Choi, and M. Park. Design of the robotic system for human-robot interaction using sound source localization, mapping data and voice recognition. In *ICCAS-SICE*, pp. 1143–1147, 2009.

[11] C. Hu, M. Q. Meng, P. X. Liu, and X. Wang. Visual gesture recognition for human-machine interface of robot teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 2, pp. 1560–1565, 2003.

[12] F. A. Bertsch and V. V. Hafner. Real-time dynamic visual gesture recognition in human-robot interaction. In *IEEE-RAS International Conference on Humanoid Robots*, pp. 447–453, 2009.

[13] Y. Sato, K. Bernardin, H. Kimura, and K. Ikeuchi. Task analysis based on observing hands and objects by vision. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2, pp. 1208–1213, 2002.

[14] D. J. Agravante, A. Cherubini, A. Bussy, P. Gergondet, and A. Kheddar. Collaborative human-humanoid carrying using vision and haptic sensing. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 607–612, 2014.

[15] E. Berger, D. Vogt, N. Haji-Ghassemi, B. Jung, and H. B. Amor. Inferring guidance information in cooperative human-robot tasks. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 124–129, 2013.

[16] P. Yang, K. Sasaki, K. Suzuki, K. Kase, S. Sugano, and T. Ogata. Repeatable folding task by humanoid robot worker using deep learning. *IEEE Robotics And Automation Letters*, Vol. 2, No. 2, pp. 397–403, 2017.

[17] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition*, 2017.

[18] HTC Vive. `https://www.vive.com/jp/`.

[19] N. S. Pollard, J. K. Hodgins, M. J. Riley, and C. G. Atkeson. Adapting human motion for the control of a humanoid robot. In *IEEE International Conference on Robotics and Automation*, Vol. 2, pp. 1390–1397, 2002.

[20] S. Nakaoka, A. Nakazawa, K. Yokoi, H. Hirukawa, and K. Ikeuchi. Generating whole body motions for a biped humanoid robot from captured human dances. In *IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 3905–3910, 2003.

[21] A. D. Joven, C. Andrea, S. Alexander, W. Pierre-Brice, and K. Abderrahmane. Human-Humanoid Collaborative Carrying. 2016.

[22] ROS Voice Message. `https://play.google.com/store/apps/details?id=org.ros.android.android_voice_message`.