

Original Research Paper

Robot Control Using Natural Instructions Via Visual and Tactile Sensations

Takuya Ikai, Shota Kamiya and Masahiro Ohka

Department of Complex Systems Science, Graduate School of Information Science, Nagoya University, Nagoya, Japan

Article history

Received: 30-08-2014

Revised: 10-08-2015

Accepted: 26-05-2016

Corresponding Author:

Masahiro Ohka

Department of Complex
Systems Science, Graduate
School of Information Science,
Nagoya University, Nagoya,
Japan

Email: ohka@is.nagoya-u.ac.jp

Abstract: Stress-free interaction between humans and robots is necessary to support humans in daily life. In order to achieve this, we anticipate the development of new robots equipped with tactile and vision sensors for receiving human instructions. In this article, we focus on spontaneous movements that do not require training, such as pointing and force adjustment and that are suitable for daily care. These movements, which we call natural instructions, involve the transmission of human instructions to robots. In this experiment, we examine a robot equipped with vision and tactile sensors capable of receiving natural instructions. Our new robot accomplishes a retrieving and passing task using the natural instructions of finger pointing and tapping with the palm.

Keywords: Machine Vision, Tactile Sensation, Natural, Instruction, Robot

Introduction

Background and Purpose

The demand for robots has been increasing not only in industrial fields but also in daily life. It has been suggested that human orders to robots vary according to the human's needs in life. However, it is impossible for robotics companies to provide robot programs that address every single need and circumstance that may exist. Furthermore, requiring users to learn situation specific skills, gestures and utterances in order to command a robot increases user stress.

Although some studies approached human-robot interaction using gestures, as described in the next section, it is often difficult for people to make a gesture during a cooperative task because the hands are already being used to perform the task. Furthermore, utterance recognition is dependent on the current environment and the speaker because noisy conditions and speech difficulties can prevent the successful transmission of commands. Consequently, we believe the use of both gestures and utterance creates stress. Thus, a new interaction that does not involve specific utterances and gestures will most effectively reduce daily human stress.

In this study, we introduce a new interaction between humans and robots. Our method uses spontaneous or unconscious movements and signs that do not require training and that are concomitant with normal movement. For example, the iris position changes with the line of sight. Similarly, force changes

are accompanied by hand actions. In such cases, a human unconsciously moves the iris and exerts force. In other words, there is no stress because these actions naturally accompany the main movement. Therefore, if robots can sense *natural movements*, such as eye direction and force direction changes, the robot can provide stress-free support to the human. We aim to utilize these movements as *natural instructions*. In this article, we mounted vision and tactile sensors capable of receiving such instructions on the robot. With our method, the robot mastered a retrieving and passing task without requiring special training or specific gestures and utterances.

Literature Review

Researchers have strived to develop robots capable of recognizing not only utterances but also nonverbal instructions. Since utterance recognition is being studied by many researchers, we focused our attention on nonverbal instruction. Several studies on nonverbal instructions have been presented, particularly for gestures (Kurata *et al.*, 2002; Ong and Ranganath, 2005; Shotton *et al.*, 2011). Kurata *et al.* (2002) achieved high-speed image tracking using hand gestures, Ong and Ranganath (2005) surveyed recent automatic sign language analysis and Shotton *et al.* (2011) developed real-time 3D pose recognition.

Furthermore, other studies looked into nonverbal instructions other than gestures. Although a method using a mounted tactile pad to accept human commands

(Ito and Tsuji, 2010) is related to our study, this system requires the user to learn special touching patterns in order to make the robot move according to his or her intention, which is burdensome. Recently, researchers have used the human body as a pointing device along with Kinect to make a robot comprehend commands (Quintero *et al.*, 2013). Furthermore, for tutoring and coaching, nonverbal social cues like eye gaze and gesture are effective for socially assistive robots (Admoni and Scassellati, 2014).

As shown by these related works, researchers have presented several methods for nonverbal instructions. It is our goal to transmit the desired intention in a natural, nonstressful manner using nonverbal instructions.

Theory

Overview of Robotic Instructions

We first assume that the robot is situated relatively far from the operator. In order to perform a cooperative task with the robot, we should issue a command via gesture and utterance. Then the cooperative task will be performed over a short distance.

Over a short distance, since we are close enough to touch the robot, communication via touch, rather than utterance, becomes possible. In this study, natural gesture and contact force are used for long and short ranges, respectively.

In this study, the object-retrieving and -passing task is treated as a typical daily task. Accordingly, we explain our scenario using this task as an example.

Human instructions utilize pointing and touch communication for long and short ranges, respectively. For each range, vision and tactile sensors are applied. The task of retrieving and passing an object is accomplished by the steps shown in Fig. 1:

- The operator first points at the object that he or she wants and the robot then recognizes the object
- The robot extends its hands to grasp the object, using tactile information
- The robot recognizes how the grasped object should be treated through tactile information applied as contact communication. For example, the robot recognizes the release time of the object via drawing force applied to it and the conveyance course via force in cooperative tasks

Pointing Method

In order to achieve the scenario explained in the last section, we use visual and tactile sensor systems to recognize human instructions. The human tries to instruct the robot using natural mannerisms, such as pointing out the object, holding out a hand and pushing.

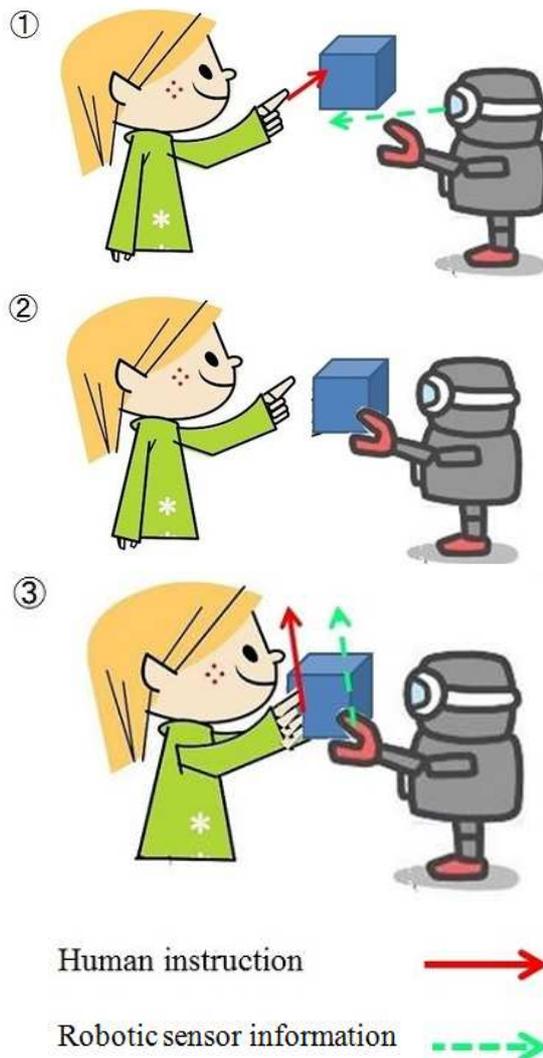


Fig. 1. Scenario of object retrieving and passing task

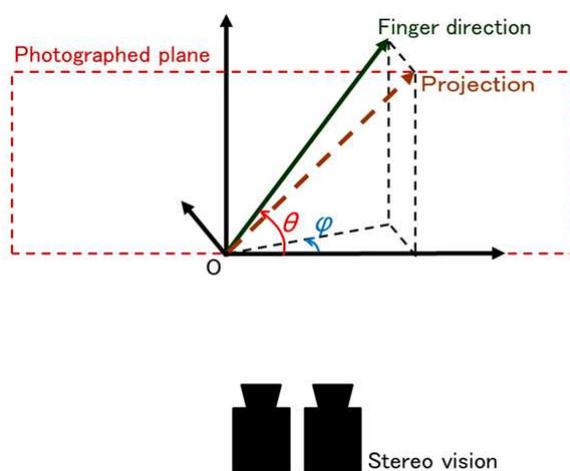


Fig. 2. Definition of θ and ϕ in Finger Direction Recognition (FDR) system

If we have something we wish brought to us, we point out the specific object with our index finger. Although this is a gesture, we understand it even without any prior instruction. We therefore consider this gesticulation to constitute a *natural instruction*.

In our previous work (Ikai *et al.*, 2013), we used stereo vision to create a Finger Direction Recognition (FDR) system that estimates the 3D direction indicated by human pointing (Fig. 2). For a detailed explanation, please refer to the paper, but here we discuss the key issues.

First, in this system, the acute portions including the fingertip are searched for along the contour of the finger image according to the following formulas for three points on the contour as shown in Fig. 3:

$$\cos \Theta_{\alpha} = \max \frac{a_x b_x + a_y b_y}{|\vec{A}| |\vec{B}|}, \vec{A} = (a_x, a_y) \text{ and } \vec{B} = (b_x, b_y) \quad (1)$$

$$a_x b_y - a_y b_x > 0 \quad (2)$$

where, a_x , b_x , a_y , b_y and Θ_{α} are values positioned as shown in Fig. 3. Through Equation 1 and 2, the FDR system finds a fingertip.

Next, using stereo matching for the fingertip and centroid of the hand, the FDR system estimates two finger directions (θ and φ), which are projections on two planes, instead of a 3D finger direction as shown in Fig. 2. Elevation θ is the sum of the vectors along the lines of the finger's sides and it can be calculated by applying a finger straight-line detector from a fingertip image. Azimuth φ is calculated from a vector defined by the center of an image of the hand region and the fingertip point of which coordinates are obtained from stereo matching. Using this method, we get the 3D finger direction of pointing and the approximated position of the object that is being pointed to.

In this study, for simplification, we define the pointing finger and pointed object as existing in the same plane and equidistant from the camera.

Tactile Data for Contact Information

Let us consider a situation in which an object is handed to us. If we instruct the robot to release the grasped object when the object bottom is lightly tapped, the robot recognizes the intentional tapping of the bottom of the object, or a hand holding the bottom of the object, as the release time.

A very sensitive tactile sensor is required to measure subtle tactile sensation such as light tapping. The three axis tactile sensor developed in our previous work is useful for this objective (Abdullah *et al.*, 2011).

Figure 4 demonstrates this tactile sensor, which can measure three axis forces (one vertical and two horizontal) simultaneously. The three force components are measured based on the variations in rotational momentum occurring in the sensor's tactile feelers.

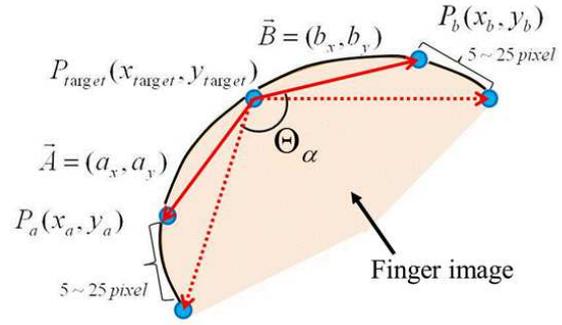


Fig. 3. Θ_{α} on finger contour; in order to search for $P_{target} (x_{target}, y_{target})$ of a fingertip, we chose $P_a(x_a, y_a)$, $P_b(x_b, y_b)$ and $P_{target} (x_{target}, y_{target})$ inside a certain range and checked whether $\cos \Theta_{\alpha}$ exceeds a certain threshold

The tactile sensor contains 41 sensing elements that have local coordinates as shown in Fig. 5. We display element #00 to element #08's locations and coordinates in Fig. 5 because they will be used in the experimental results described in section 4.

In the sensor, three components of applied force (F_x^{sensor} , F_y^{sensor} and F_z^{sensor}) are calculated according to the following formulas:

$$F_x^{sensor} = K_x (u_x^{sensor,t} - u_x^{sensor,t-1}) \quad (3)$$

$$F_y^{sensor} = K_y (u_y^{sensor,t} - u_y^{sensor,t-1}) \quad (4)$$

$$F_z^{sensor} = K_z G \quad (5)$$

where, u_x^{sensor} and u_y^{sensor} are components of centroid displacement in the sensor coordinate. Superscript t and $t-1$ show the current step and preceding step, respectively. G is the obtained grayscale value observed by the fiberscope in Fig. 4. K_x , K_y and K_z are constants determined by calibration tests.

It should be noted that $F_i^{sensor} (i = x, y, z)$ is the force component in the embedded coordinate of the sensor element. When force applied to the robotic finger is obtained, F_i^{sensor} should be transformed through the robotic kinematics to obtain the force component in the world coordinate of the robot ($O_G-x_G y_G z_G$) shown in Fig. 7 to obtain contact information between a grasped object and another object.

Algorithm

Using the pointing method and tactile data processing, the robot performs the task according to the flowchart in Fig. 6, which shows the main flow of the program for robotic instruction.

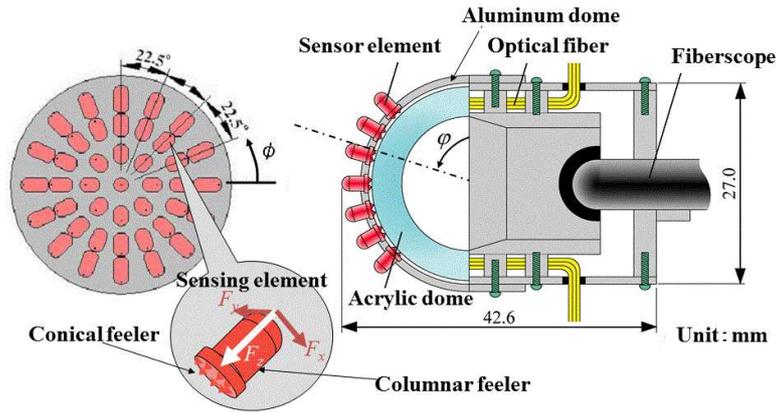


Fig. 4. Three-axis tactile sensor

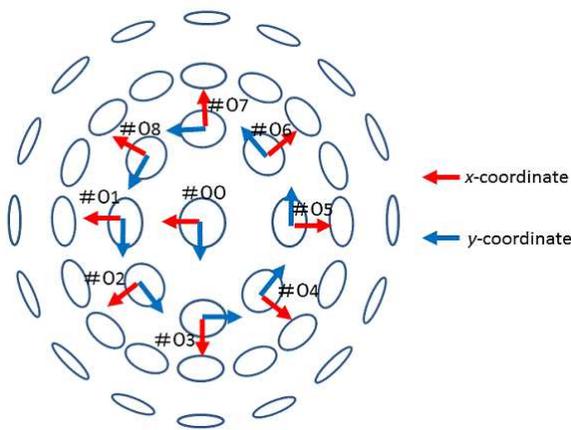


Fig. 5. Position and local coordinates' tactile sensor elements

In the FDR block, first, a pointing finger is identified and then the pointing direction is estimated according to the FDR system as described in the preceding section. Second, the distances between objects and the camera are obtained using a graph-cut algorithm of OpenCV to estimate the specific object pointed at by FDR. Then, an opened palm is determined as the largest skin-colored area and its centroid is calculated.

Next, by solving inversed kinematics of the manipulator, motor control variables are calculated to control the motors so that the robot extends its arm-hand to approach the object. After the intermediate point between two fingertips reaches the object centroid, the hand grasps the object. During the grasping operation, F_z^{sensor} is measured by the tactile sensor to prevent the robot hand from exceeding the limit force.

After the robot brings the object over the palm, it tries to put the object on the hand by lowering the hand. If the object bottom is touched and $|dF_{x \text{ or } y}^{sensor} / dt|$ reaches a specific value, it stops its lowering motion to complete its task.

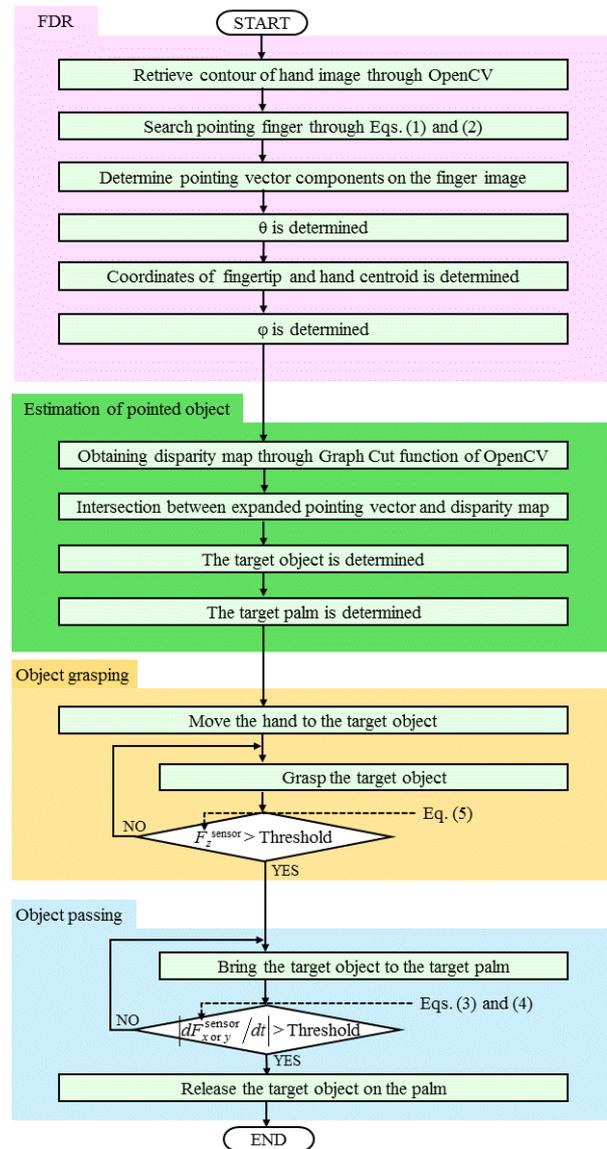


Fig. 6. Flowchart of robotic instructions

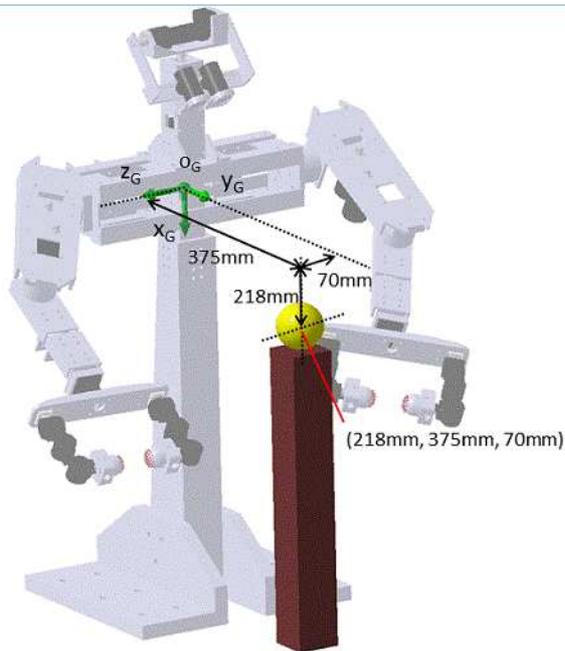


Fig. 7. Robot equipped with two hands and two eyes

Experimental Procedure

Robotic System

The robotic system used in this study is shown in Fig. 7. It has two hands and two eyes. We built this robot by adding a two-eyed robotic head to the robot produced in our previous paper (Abdullah *et al.*, 2011). The robotic head has two Degrees of Freedom (DOF) (pan and tilt) and each arm has six DOF. We mounted two tactile sensors on each hand so that the robot's two fingertips would face each other. Consequently, the robot can recognize not only a human hand but also the tapping force applied to the grasped object, which is the instruction from the human.

Object Recognition

Using the FDR system described in section 2.2, the robot identifies the specific object that the human wants. In this algorithm, we assume that both the specific object and the pointing finger exist on the same photographed plane (in Fig. 2, $\varphi = 0^\circ$ is assumed). The distance is obtained from a disparity map attained from right and left camera images using a stereo matching technique. We adopted OpenCV in the basic program modules used for image processing. The process flow of the object recognition is shown in Fig. 8:

- A human points at a specific object using a pointing gesture with their index finger. The FDR system identifies the direction in which the finger is pointing

- Using a graph-cut algorithm, the disparity map is obtained from the captured images of the right and left cameras
- Regions at the same distance as the pointing fingertip are extracted
- The width of the pointing vector is expanded
- The specific object is identified as the overlapped area of the expanded pointing vector and the object is obtained from the preceding procedure
- After ignoring as noise any small areas with a circumference of length less than a specific threshold, the left area is identified as the specified object. If multiple objects are identified, the nearest object to the hand's centroid is identified as the specified object

Object Retrieving and Passing

Using our method, we performed a series of experiments on passing an object from the robot to a human. The robot grabs the object that the human indicates and places it on the human's palm. First, the human performs a natural pointing gesture to indicate the object. The human then merely extends his hand palm up. The robot recognizes the human signs that are concomitant with the main movement; the human stretches his arm and applies force to the object where the human intends to receive the object. Thus, this system does not need information of a human hand position or detachment timing.

This experiment's method proceeds according to the following steps:

- A human indicates an object by pointing. The robot estimates the position of the indicated object using the procedure explained in *Object Recognition*
- The robot moves its head to set the object image at the center of its eyesight
- The human shows his palm to indicate the destination of the object. The palm is recognized as the largest skin-colored area other than the face area. In addition, the palm's centroid is obtained
- The robot grasps the object. At that time, if the vertical force in the fingertip exceeds a threshold, the robot finishes its grasping motion
- The robot moves its hand over the human's hand recognized in the preceding procedure
- The robot puts its hand down and places the object on the palm of the human's hand. If the shearing force on the fingertip exceeds the threshold, the robot completes this placing motion
- The robot opens its hand and releases the object when the time derivative of the shearing force, which represents slippage, exceeds the threshold due to the tapping of the cube's bottom by the palm

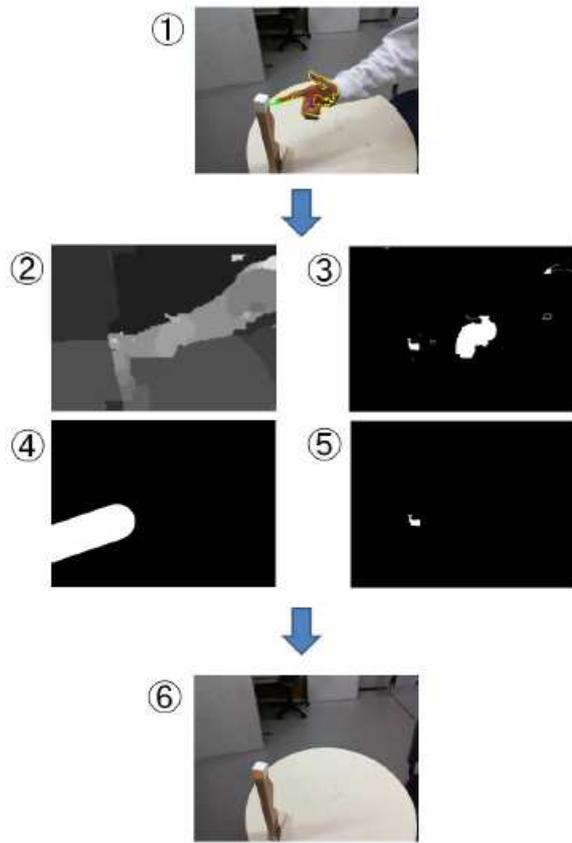


Fig. 8. Object recognition algorithm

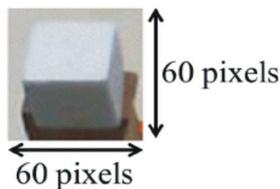


Fig. 9. Cube specimen in image data

In the above task, we used the object in Fig. 9, which is a paper cube that is 40 mm on each side. Our tactile sensor is designed to recognize the grasping force so as to not crush the cube.

Experimental Results and Discussion

Scene of Experiment

In this experiment, the robot's progress in passing the object is shown by the photographs in Fig. 10. Photos (1) to (3) show the object and palm positions using the theory of robot instruction explained in section 2. The robot then retrieves the object and passes it to the human by placing it on the human's palm, as shown in Photos (4) to (7).

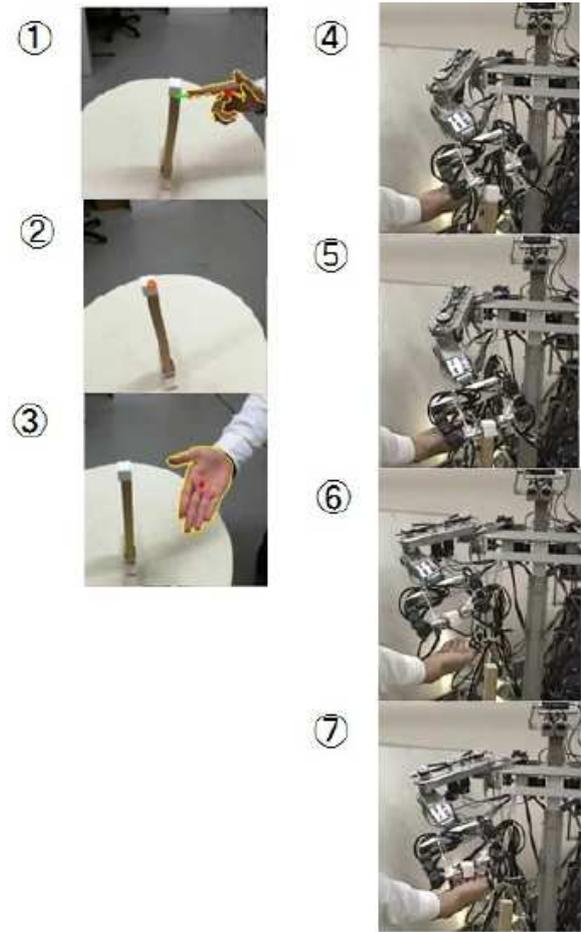


Fig. 10. Sequential photographs showing the robot passing the object (http://www.ohka.cs.is.nagoya-u.ac.jp/~ohka/ohka_lab2/images/JCS.wmv)

After Photo (7), the bottom of the object is tapped by the palm to generate upward slippage force on the robotic fingers. The slippage force generated by the tapping acts as a force sign for the robot to release the object.

Position and Force Data

The experimental results are shown in Fig. 11-14. Figure 11 shows the time variation of the fingertip's position. Figure 12 and 13 show changes in the normal force distribution of fingers #1 and #2, respectively and Fig. 14 shows the time derivative of the tangential force of the specific elements, which indicates slippage (Ohka *et al.*, 2012). The element numbers in Fig. 12 and 14 are the same as those in Fig. 5.

As shown in Fig. 11, after approximately three seconds, the robot arm moves to the position of the object at $(X_G, Y_G, Z_G) = (218, 375, 70)$ [mm], which is obtained by the FDR system. The robot then begins the grasping motion and completes it by exceeding a threshold at around 20 sec, as shown in Fig. 12 and 13,

due to the normal forces of element #04 of finger #1 and element #06 of finger #2 reaching the maximum. As shown in Fig. 5, since elements #04 and #06 are not at the center of the tactile elements, the hand does not grasp the object with just the center of the fingers.

At approximately 48 sec, the robot starts the release motion as the normal forces suddenly diminish. This release motion is induced by slippage force, which is demonstrated as the second peak of the tangential force derivative of elements #04 (finger #1) and #06 (finger #2) in Fig. 14 and is caused by the object's bottom touching the human's palm. The first peak at around 18 sec shows the slippage force when the robot picks up the object and strengthens its grasp to prevent slippage.

As the results demonstrate, the robot can receive instructions and successfully complete the retrieving and passing task without special gestures or utterances.

Multiple Object Status

Even among multiple objects, the robot should recognize the specific requested object. In order to test for this, we checked whether this system could select the specific object by adopting the object recognition program (Fig. 15). We used three objects for this test: A wood cube, a ping-pong ball and the paper cube and they have almost the same size as shown in Fig. 12.

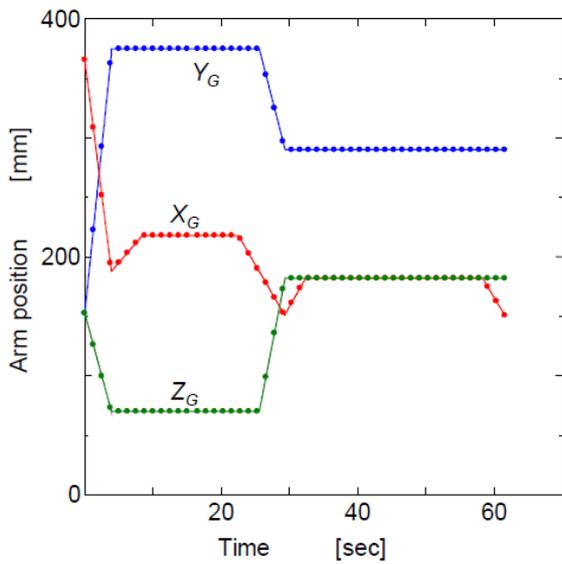


Fig. 11. Time variation in arm position

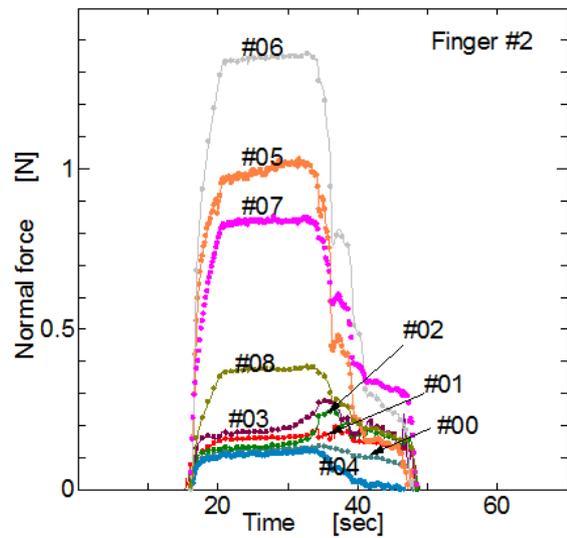


Fig. 13. Time variation in normal force of finger #2

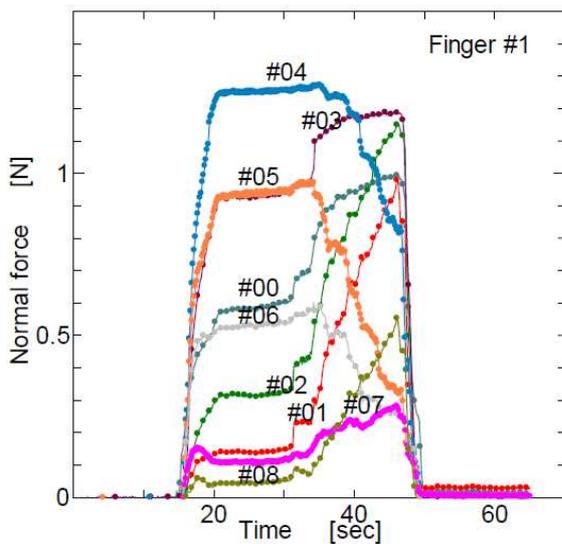


Fig. 12. Time variation in normal force of finger #1

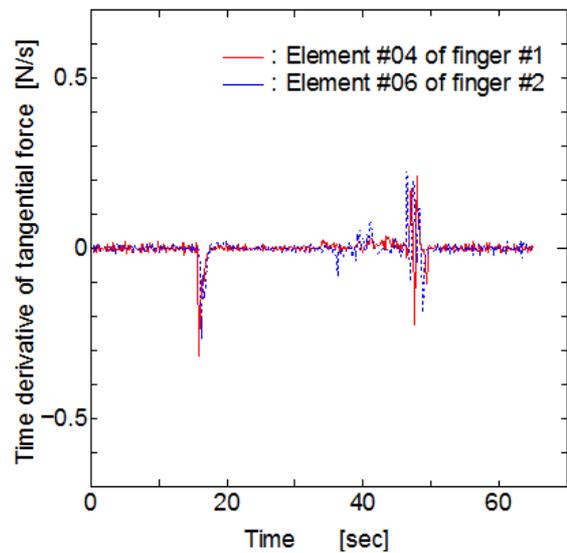


Fig. 14. Time variation in tangential force

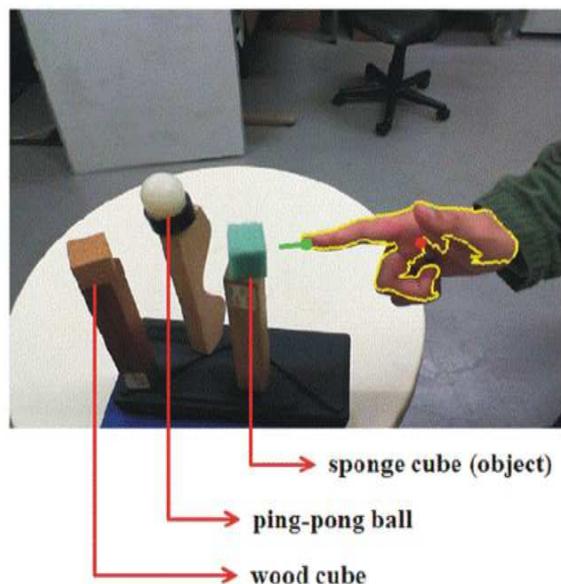


Fig. 15. Time variation in normal force of finger #1

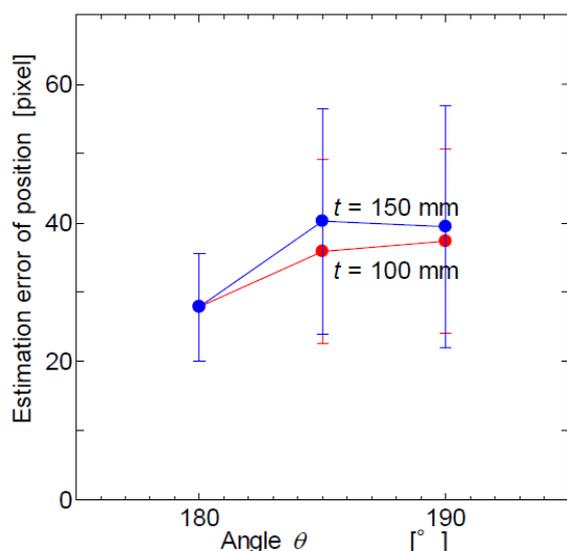


Fig. 16. Estimation error of object position

Table 1. Detection rate of object position

Angle θ [°]	$d = 100$ mm (%)	$d = 150$ mm (%)
180	70	70
185	45	35
190	40	30

Since the robot has tactile sensors, a slight position error does not pose a problem for grasping an object. Since the size of the image data is around 60 pixels, as shown in Fig. 15, the robot can grasp the object even if there is a 30-pixel error. Therefore, we determined that object recognition was successful when the position estimation of the object was within 30 pixels. In this

experiment, we adopted two parameters: The distance between the object and the centroid of hand d and finger inclination θ . For each condition, we performed 20 trials.

Figure 16 shows the estimation error of object position for the $d = 100$ and 150 mm cases. As shown in this figure, if the finger direction deviates from horizontal ($\theta = 180^\circ$), the position error becomes greater than 30 pixels and, when distance d increases, the estimation error becomes larger.

The detection rates are shown in Table 1. If the finger direction maintains a horizontal direction, the percentage of success is 70%, even if there is another object along the finger's direction. However, if the angle deviates from the horizontal plane, the detection rate becomes less than 50%. We will make improvements to handle this issue in the future.

Conclusion

In this study, we proposed a new robot system equipped with tactile and vision sensors for receiving human instructions. With this system, the robot retrieves an object requested by a human and places it on the human's palm. Although the pointing direction is limited to the horizontal plane, there is the possibility of applying this system to housekeeping.

The system does not require a human hand position or detachment timing because it obtains the information through visual and tactile sensations. Since pointing and holding out one's palm to receive an object is natural for humans, the instructions for completing this task with the robot are stress-free.

In the future, we will further develop this system to apply it to cooperative tasks between humans and the robot. Accordingly, we will improve the detection rate at angles outside the horizontal direction.

Acknowledgement

The authors would like to extend their appreciation to Prof. Eisuke Kita and Prof. Takashi Watanabe for several related research discussions.

Funding Information

This research is funded by Hori Science and Arts Foundation.

Author's Contributions

Takuya Ikai: Proposed the main idea of this study and drafted the manuscript.

Shota Kamiya: Performed experiments and analyzed the results.

Masahiro Ohka: Designed the research plan and organized the study.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other author have read and approved the manuscript and no ethical issues involved.

References

- Abdullah, S.C., J. Wada, M. Ohka and H. Yussof, 2011. Object exploration using a three-axis tactile sensing information. *J. Comput. Sci.*, 7: 499-504. DOI: 10.3844/jcssp.2011.499.504
- Admoni, H. and B. Scassellati, 2014. Data-driven model of nonverbal behavior for socially assistive human-robot interactions. *Proceedings of the 16th International Conference on Multimodal Interaction*, Nov. 12-16, Istanbul, Turkey, pp: 196-199.
- Ikai, T., M. Ohka, S. Kamiya, H. Yussof and S.C. Abdullah, 2013. Evaluation of finger direction recognition method for behavior control of robot. *Int. J. Smart Sens. Intelli. Syst.*, 6: 2308-2333.
- Ito, T. and T. Tsuji, 2010. Command recognition of robot with low dimension whole-body haptic sensor. *IEEJ Trans. Industry Applic.*, 130: 293-299. DOI: 10.1541/ieejias.130.293
- Kurata, T., T. Kato, M. Kouroggi, J. Keechul and K. Endo, 2002. A functionally-distributed hand tracking method for wearable visual interfaces and its applications. *Proceedings of the IAPR Workshop on Machine Vision Applications, (MVA' 02)*, Nara, Japan, pp: 84-89.
- Ohka, M., S.C. Abdullah, J. Wada and H. Yussof, 2012. Two-hand-arm manipulation based on tri-axial tactile data. *Int. J. Social Robot.*, 4: 97-105. DOI: 10.1007/s12369-011-0131-x
- Ong, S.C.W. and S. Ranganath, 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Patt. Anal. Machine Intelli.*, 27: 873-891. DOI: 10.1109/TPAMI.2005.112
- Quintero, C.P., R.T. Fomena, A. Shademan, N. Wolleb and T. Dick *et al.*, 2013. SEPO: selecting by pointing as an intuitive human-robot command interface. *Proceedings of the IEEE International Conference on Robotics and Automation*, May 6-10, IEEE Xplore Press, Karlsruhe, pp: 1158-1163. DOI: 10.1109/ICRA.2013.6630719
- Shotton, J., A. Fitzgibbon, M. Cook, T. Sharp and M. Finocchio *et al.*, 2011. Real-time human pose recognition in parts from single depth images. *Commun. ACM*, 56: 1297-1304. DOI: 10.1145/2398356.2398381