Photograph-Based Interaction for Teaching Object Delivery Tasks to Robots

Sunao Hashimoto[1], Andrei Ostanin[2], Masahiko Inami[1,3], Takeo Igarashi[1,4]

[1]JST, ERATO, IGARASHI
Design UI Project,
Tokyo, Japan

[2]School of Computing
University of Utah,
Salt Lake City, USA

[3]Graduate School of Media
Design, Keio University,
Yokohama, Japan

[4]Graduate School of
Information Science and Tech,
The University of Tokyo,
Tokyo, Japan

hashimoto@designinterface.jp, ostanin@cs.utah.edu, inami@inami.info, takeo@acm.org

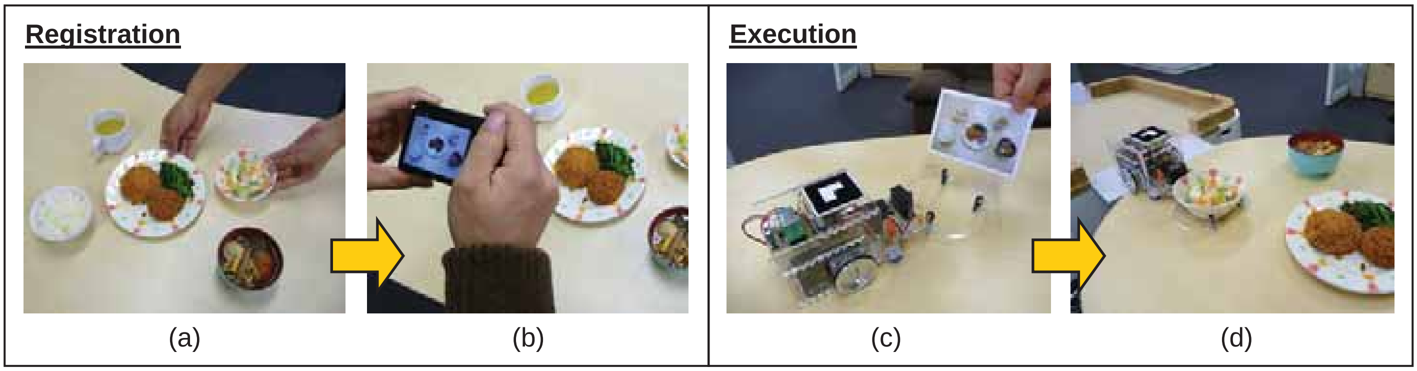

Figure 1. The user manually arranges dishes (a) and takes a photograph (b) at registration. The uer shows the photograph to a robot (c), and the robot arranges the dishes accordingly (d).

***Abstract*— Personal photographs are important media for communication in our daily lives. People take photos to remember things about themselves and show them to others to share the experience. We expect that a photograph can be useful tool for teaching a task to a robot. We propose a novel human-robot interaction using photographs. The user takes a photo to remember the target in a real-world situation involving a task and shows it to the system to make it physically execute the task. We developed a prototype system in which the user took a photo of a dish arrangement on a table and showed it to the system later to then have a small robot deliver and arrange the dishes in the same way.**

Keywords-Delivery robots, Object arrangement, Photograph-based interaction

I. Introduction

Computers with eyes (sensors) and hands (actuators), i.e., robots, are being introduced into homes. Unlike industrial robots, these home robots need to interact with non-expert users to receive orders and provide feedback to them. One strategy for designing an effective human robot interface is to mimic human-human communication such as that conveyed through speech and gestures. However, speech and gestures suffer from limitations such as being inherently ambiguous, making it difficult to represent visual information. Our goal in this research was to investigate alternatives to the modality of human-robot interaction to address these limitations.

This paper focuses on object-delivery tasks and proposes a user interface for remembering and specifying specific tasks. Object delivery is one of the fundamental tasks we hope home robots will be able to undertake to minimize human labor. People might want a robot to set cups and plates on a table, carry them to the kitchen, or prepare their tools before they start to carry out do-it-yourself construction at home.

The question we want to find an answer to is how to teach specific delivery tasks to robots, more specifically, what object is to be delivered where. One can specify a task by using speech and gestures such as saying “deliver A and B to here” pointing to the target objects and their destination [1-3]. However, such seemingly natural methods present various difficulties in practice. For example, speech recognition requires the user to use a specific name for individual objects and it is difficult for a pointing gesture to specify the position and orientation of numerous objects. Furthermore, these methods lack good representations for reference that can be reused later.

We propose using a photo as a reference for an object-delivery task. A user registers a new task by taking a photograph of the required result for an object-delivery task, and he/she can ask the robot to conduct the task later by selecting the corresponding photograph. For example, the user places cups and plates on a table and takes a photograph. Later, he/she can have the robot carry out the task by showing the robot the printed photograph or selecting it on a GUI. The robot delivers and arranges the cups and plates based on the photograph (Figure 1 shows this concept). This method has various advantages. First, taking a photograph is a natural one-button-action people are already familiar with. Second, a photograph contains all the necessary information for specifying a delivery task (what to deliver and where). Third, it serves as a good tangible reference for specifying the task for later reuse.

II. PHOTOGRAPH-BASED INTERACTION

People immediately understand various pieces of information such as "what is the target object" and "where is the target position" by looking at a photograph, even without verbal descriptions. Our goal is to leverage these features of photographs to enrich human-robot interactions.

There are two phases of user interaction in our method: registration and execution. We will explain each of these in more detail in the following.

Registration phase

The user first creates the wanted setting by using real objects, and takes a photograph of the setting by using a special camera device. The device has a small display like a digital still camera, and it is connected to the central server via a wireless network. The system recognizes "which objects appear in the camera frame" based on either tracking the position and direction of the camera device (the range of its viewing field is known), or the camera device recognizing the image. The recognition results, such as the object's names, are presented on the device's display and when the user presses the button for the shutter, the system records how the objects are arranged in the camera frame. The arrangement data contains the object's ID and its absolute position and direction, and these data are associated with the photograph that has been taken. The camera device has a simple GUI for previewing and editing. The user can exclude unwanted objects in the photograph on the display and it can be printed later with a printer connected to the central server if necessary.

Execution phase

We considered three ways of requesting the robot to perform a task: (a) showing a printed photograph to a reader placed on the table or a wall, (b) showing the robot's "eyes" a printed photograph, and (c) selecting a corresponding photograph (image) on a touch display. In (a) and (b), a 2D barcode (such as QR-code) that contains information on the arrangement is printed on the back of the photograph. The robot or the reader reads this barcode, and acquires the information for the arrangement. In addition, when there are multiple robots as in (b), the robot that reads the photograph notifies the others of the user's order and works in cooperation with them.

We mainly used physical photos for the following reasons. First, the user could simultaneously see multiple photos when searching and start execution simply by showing a photo to the system without starting an application on a computer. Second, he/she could freely write notes on the photograph with a pen, e.g., "Everyday Breakfast Setting". Finally, he/she could store the printed photograph in any convenient location. For example, the printed photograph could be placed between the pages of a recipe book, on the door of a refrigerator, or on a box of cereal. The physical locations work as a spatial memory and help users to identify the target layout. This is especially helpful when the time interval between registration and execution is very long.

III. PROTOTYPE SYSTEM

We built a prototype system for arranging dishes using a mobile-camera device and a small tabletop robot (Figure 2). Visual markers were attached to the dishes and the robot whose IDs and positions were recognized by ceiling-mounted cameras. When the user took a photograph of the target layout for the dishes, the IDs and position of each object in the camera frame were recognized and memorized. The photograph was printed with a QR-code with a link to necessary information printed on the backside. The user later showed the barcode to the system to ask it to start arranging the dishes. The system then remotely controlled the small armless robot to deliver the dishes by pushing them according to the arrangement in the photograph.

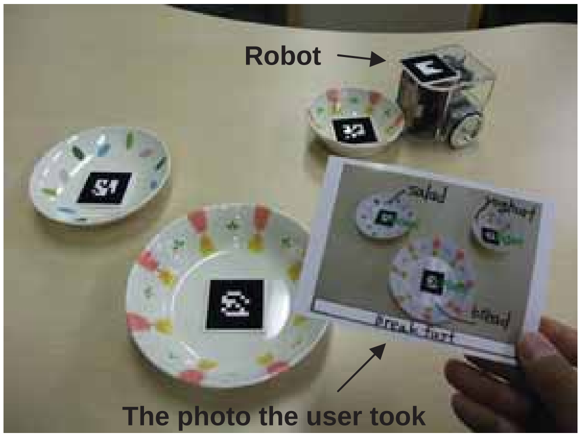

Figure 2. Developed prototype system.

IV. CONCLUSION

The proposed method is applicable for various robots such as humanoids and small tabletop robots. Our method allows the user to register the specific states of the real-world, and recall them later. We believe that it will be useful for various household tasks. For example, one can take a photo of a bed to teach a robot how a bed is made, a photo of an entire room to teach it how to clean it, and a photo of storage facilities to teach it how to store objects. Our method provides cameras and photographs with new uses as human-robot communication tools. We are going to run a user study using the developed prototype system.

REFERENCES

[1] Bolt, R. A. "Put-That-There": Voice and Gesture at the Graphics Interface. In Proc. SIGGRAPH'80, ACM Press (1980), 262–270.

[2] Kemp, C. C., Anderson, C. D., Nguyen, H., Trevor, A. J., and Xu, Z. A point-and-click interface for the real world : Laser designation of objects for mobile manipulation. In Proc. HRI 2008, ACM Press (2008), 241–248.

[3] Ishii, K., Zhao, S., Inami, M., Igarashi, T., and Imai, M. Designing Laser Gesture Interface for Robot Control. In Proc. INTERACT2009, 479–492.