Fourth-person sensing for pro-active services

Yumi Iwashita, Kazuto Nakashima, Yoonseok Pyo, Ryo Kurazume
†*Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan*
Email: yumi@ieee.org

***Abstract*—Service robots, which co-exist with humans to provide various services, obtain information from sensors placed in an environment and/or sensors mounted on robots. In this paper we newly propose the concept of *fourth-person sensing* which combines wearable cameras (first-person sensing), sensors mounted on robots (second-person sensing), and distributed sensors in the environment (third-person sensing). The proposed concept takes advantages of all three sensing systems, while removing disadvantage of each of them. The first-person sensing can analyze what a person wearing a camera is doing and details around him/her, so the fourth-person sensing has chance to provide pro-active services, which are triggered by predicted human intention and difficult with the second- and third-person systems, by estimating his intention/behavior. We introduce an example scenario using the fourth-person sensing and show the effectiveness of the proposed concept through experiments.**

I. Introduction

Service robots that can co-exist with humans and provide various services in everyday life are receiving an increasing amount of attention from academics, societies, and industries. These robots must have sufficient ability to sense changes in areas where they are and cope with a variety of situations. However, since everyday environment is complex and unpredictable, it is almost impossible with current methods to sense all the necessary information using only a robot and sensors attached to the robot. One of the promising approaches to develop service robots which co-exist with humans is using ICT technology, such as a distributed sensor network and network robotics. We have developed an informationally structured environment (Fig. 1), which consists of distributed sensors such as laser range finders (LRFs), cameras, RFID sensors [1]. This enables robots to obtain information beyond their abilities, and thus the informationally structured environment can be considered to expand the robot's sensing ability virtually. Here, we can consider that a viewpoint of a distributed sensor with respect to people is a third-person viewpoint and that of a sensor mounted on robots is a second-person viewpoint.

Low-cost high-quality wearable cameras have been available in the market for more than 6 years. Wearable cameras give images/videos obtained from a viewpoint of a person wearing a camera as shown in Fig. 2, and its viewpoint is called the first-person viewpoint. The first-person video analysis has received a lot of attention in the computer vision community. The objective of the first-person computer vision research is to analyze objects around the person, understand activities performed by the person, and predict his/her intention [2] [3].

The first-person viewpoint is suitable for analyzing what a person wearing a camera is doing and details around him/her, but information tends to be limited and local due to its small field of view. The second-person viewpoint is good to get information of both people and an area where the robot is, but the amount of information and the number of sensors are limited due to the processing ability of robots. The third-person viewpoint has the advantage of obtaining information from the entire area including people, robots, and objects in the area, so it is suitable for analyzing things globally, such as motion of robots/people in the area and estimation of a person who needs help and what he needs. However, since the third-person vision gets information from a distance with respect to people, the accuracy of estimated results is not high enough due to the following reasons: (i) resolution is low because of the distance from the sensors to people and (ii) there might be occluded areas between people and sensors because of objects/other people.

We propose a concept of the *fourth-person sensing* which takes advantages of all three viewpoints. In the proposed concept, we combine wearable cameras (first-person sensing), sensors mounted on robots (second-person sensing), and distributed sensors (third-person sensing). Compared with the existing informationally structured environment with distributed sensors, the fourth-person sensing has potential to understand demands of people with high accuracy, and his/her intention/behavior can be a trigger to start pro-active services from the robots. In this paper we introduce an example scenario of pro-active services and carry out experiments to show its feasibility.

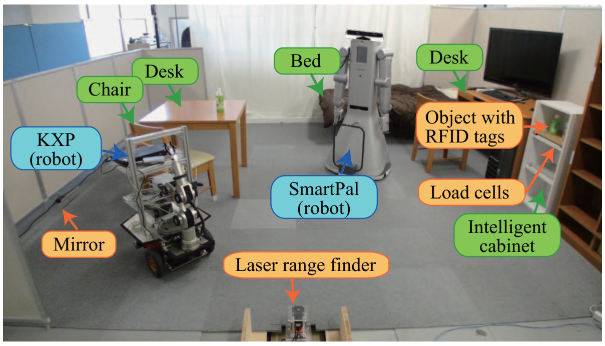


Figure 1. An informationally structured room which we developed [1].

978-1-4799-7007-0/14 $31.00 © 2014 IEEE
DOI 10.1109/EST.2014.17

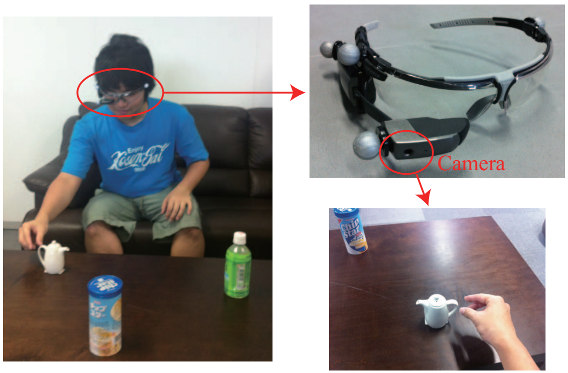


Figure 2. First-person vision and an example of captured images.

A. Related works

The informationally structured environment has received lots of attention in robotics researchers. We developed this environment in ”Robot Town Project” [1]. The aim of this project is to develop a distributed sensor network system covering an area of a block in a town in which there are many houses, buildings, and roads, and manage robot services by monitoring events that occur in the town. The sensed events are notified to the “Town Management System, TMS”, and each robot receives appropriate information about the surroundings and instructions for proper services [1]. We also used TMS in a room with distributed sensors to develop an informationally structured room. There are other researches for the embedded sensor systems in daily human life environment [4],[5],[6]. The ASTROMOBILE system, which is composed of the ASTRO robot and a wireless sensor network, was developed [7]. The above researches focused on either (i) the third-person viewpoint, or (ii) the combination of the third-person viewpoint and the second-person viewpoint. Since the information is obtained from a distance with respect to people, the accuracy of estimated results is not high enough due to low resolution and occluded areas.

Salamin et al. proposed a system which combines the first-person and the third-person viewpoints to relax the occlusion problem of the third-person viewpoint. This system was developed for virtual reality applications and they showed that the first-pereson viewpoint solved the occlusion problem [8]. Martinez et al. introduced an attention recognition system which combined the first-person and the second-person viewpoints [9]. In the original paper [9] they explained that the system combines first-person and the third-person viewpoints. Since the third-person viewpoint was set at head of a person so that he/she captures the other person, in the definition of our paper this viewpoint is considered as the second-person viewpoint.

As we explained above, there is no research which combines all three viewpoints in any research categories. To our knowledge, our paper is the first paper to propose a concept of *the fourth-person sensing* which takes advantages of all three viewpoints.

II. The fourth-person sensing

In this section, at first we explain the concept of the fourth-person sensing and details of each of them. Next, we show an example of possible pro-active services in an informationally structured room.

A. The fourth-person sensing

The fourth-person sensing is the combination of the first-person, the second-person, and the third-person sensing. Regarding the second-person and the third-person sensing, we utilize informationally structured room and service robots developed in the Robot Town Project.

1) First-person vision: Figure 2 shows an example setting of the first-person vision with a wearable glass with a camera (Vuzix, M100) and an example of captured images. The captured images contain information around the person wearing the glass, but areas which images cover tend to be limited due to the small field of view. Using captured images we can recognize his/her activities, and the results of estimated activities is useful to predict what he/she wants to do next. The predicted activities allow robots to provide pro-active services.

2) Second-person sensing: Service robots have own their sensors, such as cameras and LRFs, and information obtained from these sensors on robots is categorized as the information from the second-person viewpoint. Figure 3 shows a service robot ”SmartPal IV” (Yaskawa Electric Coop.). Since robots can move, they can get information around people whom the robots provide services and information of wide area compared with the first-person vision. However, as we mentioned above, the amount of information and the number of sensors are limited due to the processing ability of robots.

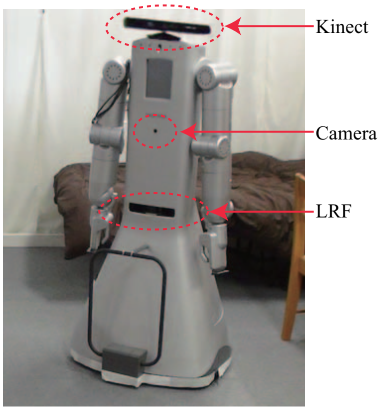


Figure 3. Service robot, ”SmartPal IV” (Yaskawa Electric Coop.).

3) Third-person sensing: Distributed sensors embedded in the environment, such as cameras, LRFs, motion capture sensors, and RFID tags, are used as the third-person sensing. We use the “Town Management System, TMS” to provide appropriate information about the surroundings and instructions for proper services. Figure 4 shows the overview of information flow between sensors and TMS. Some of the functions of TMS are as follows; (i) store all data obtained from the sensors, (ii) analyze obtained data for estimation of status in a room, and (iii) provide services after TMS recognized a trigger, such as abnormal situation. Figure 5 shows an example system for status estimation: an intelligent cabinet. The intelligent cabinet is equipped with RFID readers and load cells to detect the type and position of objects inside. There are other systems for position estimation of robots and objects which have markers of motion capture sensor, and that of people using LRFs. The motion capture system (Vicon MX, Vicon Motion Systems Ltd.) offers millimeter resolution of 3D spatial displacements. After TMS receives requests for information about objects/people in the room, it sends information to robots so that they can provide proper services to human.

The third-person viewpoint has the advantage of obtaining information from the entire area including people, robots, and objects in the area, so it is suitable for analyzing things globally, such as motion of robots/people in the area and estimation of a person who needs help and what he needs. However, as mentioned above, since the third-person vision gets information from a distance with respect to people, the accuracy of estimated results is not high enough.

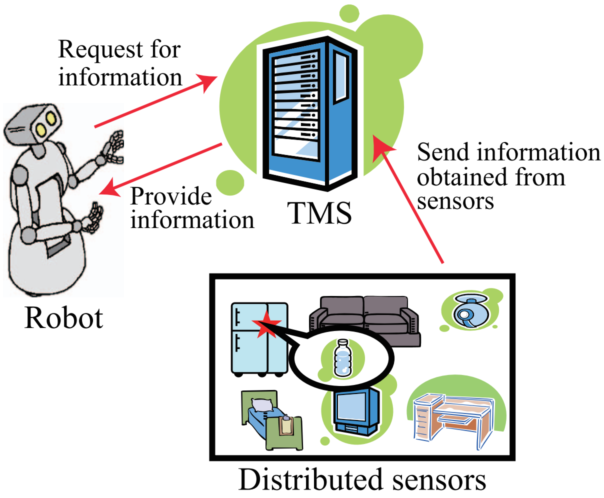


Figure 4. Town Management System, TMS.

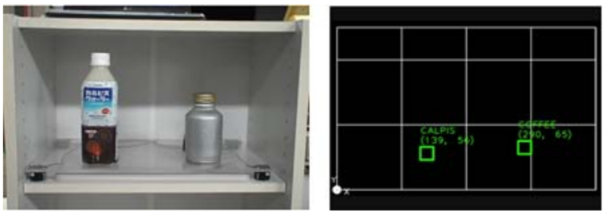


Figure 5. Intelligent cabinet.

B. Example of pro-active services

The proposed fourth-person sensing, which takes advantages of all three viewpoints, has potential to understand demands of people with high accuracy, and his/her intention/behavior can be a trigger to start pro-active services from the robots. In this section we explain an example scenario of possible pro-active services in the informationally structured room.

Suppose that a person asks to a service robot to bring water for him/her. What he wants is a bottle of water, since he is eating chips as shown in Fig. 6. TMS has information of water, which can be used for different purposes, such as water for plants and water for drinking. Details of the information TMS has are the number of bottles/water pots and their positions. The robot requests to TMS to provide the information of water. However, since TMS cannot analyze his behavior to know the information what he wants, the robot may not bring the right one to the person due to several options regarding water, unless the robot estimates the person's intention by its own sensors. The robot may bring a water pot for plants.

The proposed fourth-person sensing system allows the robot to bring the correct one, thanks to the first-person vision which can estimate the person's wish/intention. More specifically, the first-person vision can analyze his activity, so the fourth-person sensing system can understand that he is eating. Thus water he wants should be for drinking, not for plants. This allows the robot to bring the right one to the person.

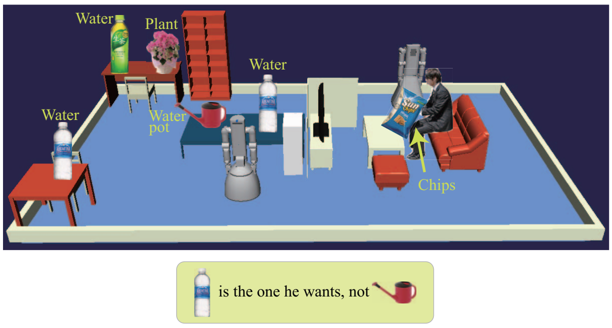


Figure 6. An example scenario of pro-active services.

III. Experiments

In this section, we implement a prototype system of the proposed fourth-person sensing and show the results of the scenario explained in section II.B.

A. Prototype system

Figure 7 (a) shows the setting of a prototype system of the proposed fourth-person sensing. We use distributed sensors including the motion capture sensor and RFID tags as the third-person sensing, to obtain position information of robots, bottles of water, containers of sauce, and chips. We also estimate a view direction of a person wearing glasses with markers using the motion capture system, and estimated view direction is used as a replacement of the first-person vision in the current setting [1]. His voice command, such as "water" and "chips" are used to trigger the service to choose the correct one he wants the robot to bring for him. More specifically, we suppose

[1] We will replace this view direction system with a system using first-person images.

that information of all objects in this room is stored in TMS and it is ready to provide information to the robot whenever TMS receives request for information. After the person gives commands, such as "sauce", TMS is triggered to provide the information of the one he wants. In case that there are multiple containers of sauce in the room, his view direction is used to select the right one as shown in Fig. 7 (b). In this experiments, we recognized voice command using an open software for voice recognition "Julius" [10], and the robot does not provide any service, which will be our future work.

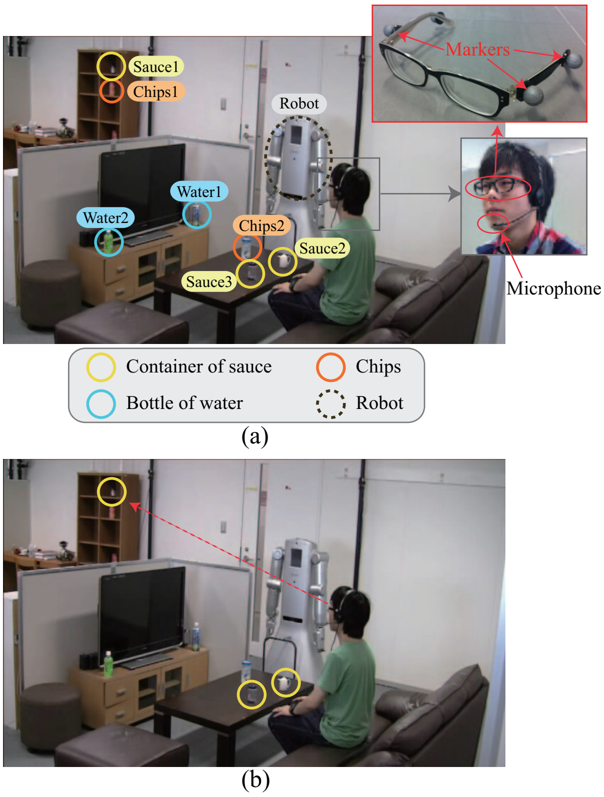
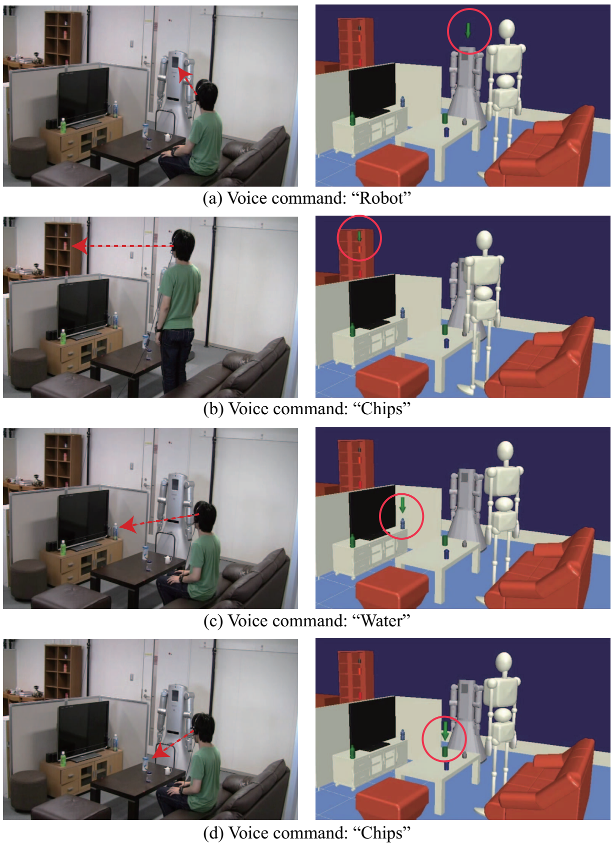


Figure 7. (a) The setting of experiments: there are a robot, two bottles of waters, two chips, and three containers of sauce, (b) after he gives voice command "sauce", the view direction is used to pick up the right one.

B. Experimental results

In experiments, a person gave a voice command while he watched an object corresponding to the voice command. Figures on left column in Fig. 8 show actual scene of experiments while he said four commands: (i) "robot", (ii) "chips" which is "chips1" in Fig. 7 (a), (iii) "water" which is "water1" in Fig. 7 (a), and (iv) "chips" which is "chips2" in Fig. 7 (a). Again, the current positions of the objects are sensed by the distributed sensors on-line (the third-person sensing). Figures on right column in Fig. 8 show selected objects with green arrows in TMS system. On these simulated figures, a person kept standing since we did not analyze his motion but we analyzed only his view direction. From these results, the proposed system could choose the right one even though there are multiple candidates in the scene, while the existing systems using the second-person and third-person sensing could not give the correct answer due to multiple candidates.


(a) Voice command: "Robot"

(b) Voice command: "Chips"

(c) Voice command: "Water"

(d) Voice command: "Chips"

Figure 8. Figures on left column show actual scene of experiments while he said four commands and red dotted lines show his viewing direction. Figures on right column show selected objects with green arrows.

IV. Conclusion

In this paper, we proposed the concept of fourth-person sensing, and we explained an example scenario of the proposed concept. Experimental results of the scenario showed the effectiveness of the proposed concept. The future work includes analyzing images captured by wearable cameras to estimate an intention/behavior of a person wearing the camera.

Acknowledgment

The present study was supported by a Grant-in-Aid for Exploratory Research (26630099).

References

[1] R. Kurazume, Y. Iwashita, K. Murakami & T. Hasegawa, "Introduction to the Robot Town Project and 3-D Co-operative Geometrical Modeling Using Multiple Robots", 15th International Symposium on Robotics Research (ISRR 2011), 2011.

[2] M. S. Ryoo & L. Matthies, "First-Person Activity Recognition: What Are They Doing to Me?", In CVPR, 2013.

[3] Y. Iwashita, A. Takamine, R. Kurazume & M. S. Ryoo, ”First-Person Animal Activity Recognition from Egocntric Videos”, International Conference on Pattern Recognition (ICPR) 2014.

[4] J-H. Lee, K. Morioka, A. Ando & H. Hashimoto, ”Cooperation of Distributed Intelligent Sensors in Intelligent Environment”, IEEE/ASME Trans. Mechatronics, Vol. 9, no. 3, pp. 535-543, 2004.

[5] T. Sato, T. Harada & T. Mori, ”Environment-Type Robot System “Robotic Room” Featured by Behavior Media, Behavior Contents, and Behavior Adaptation”, IEEE/ASME Trans. Mechatronics, Vol. 9, no. 3, pp. 529-534, 2004.

[6] Y. Nakauchi, T. Fukuda, K. Noguchi & T. Matsubara, ”Intelligent Kitchen: Cooking Support by LCD and Mobile Robot with IC-Labeled Objects”, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2464-2469, 2005.

[7] F. Cavallo, M. Aquilano, M. Bonaccorsi, R. Limosani, A. Manzi, M. Carrozza & P. Dario, ”On the design, development and experimentation of the ASTRO assistive robot integrated in smart environments”, IEEE ICRA, 2013.

[8] P. Salamin, D. Thalmann & F. Vexo, ”Improved Third-Person Perspective: a solution reducing occlusion of the 3PP?”, The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, 2008.

[9] F. Martinez, A. Carbone & E. Pissaloux, ”Combining first-person and third-person gaze for attention recognition”, IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013

[10] ”http://julius.sourceforge.jp/”