Remote Assistance Using Visual Prompts for Demented Elderly in Cooking

Sei Ikeda
Nara Institute of Science and Technology, Nara, Japan
sei-i@is.naist.jp

Jaakko Hyry
University of Oulu
Oulu, Finland
jaakko.hyry@oulu.fi

Antti Pitkanen
University of Oulu
Oulu, Finland
antti.pitkanen@oulu.fi

Zeeshan Asghar
University of Oulu
Oulu, Finland
zeeshan.asghar@oulu.fi

Petri Pulli
University of Oulu
Oulu, Finland
petri.pulli@oulu.fi

Hirokazu Kato
Nara Institute of Science and Technology, Nara, Japan
kato@is.naist.jp

ABSTRACT

This paper presents a smart kitchen system for supporting demented elderly cooking in a kitchen. In this system, a remote supporter provides a user with verbal and visual prompts manually. As a prompting method for this system projection should be chosen as a priority because it enables direct pointing of targets in a real scene. Even if projection is used as a priority, this system should be designed to solve three practical problems: 1) there are many areas where sharp images cannot be projected in a kitchen, 2) kitchen is a too complex environment to display visual prompts using only a single method, 3) remote supporters do not know the kitchen environment well enough. In this paper we propose a novel system overcoming the above problems as follows, respectively. [1)] The system first tries indirect projection of circular markers as visual prompts around target objects. [2)] Proper method is automatically selected from using projector, embedded monitor and illumination devices based on environmental information. [3)] The system automatically acquires flat surfaces to project on and their reflectance properties as the environmental information. Based on an observation of two elderly suffering from dementia, we discuss the feasibility and concrete problems in displaying visual prompt and how to solve them technically.

Categories and Subject Descriptors

H.4.3 [**INFORMATION SYSTEMS APPLICATIONS**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*

General Terms

Design

Keywords

smart kitchen, visual prompt projection, elderly with dementia, remote assistance

1. INTRODUCTION

Cooking and shopping are the most important activities for elderly people living solitarily in a sense of both necessities being about life and pleasure. For many people with mild to moderate dementia, these multi-step tasks are difficult without support. Our research group is developing smart services for supporting these activities in order to encourage their independency and communication. This paper focuses on assisting cooking activities in a smart kitchen system using computer techniques.

We assume a typical smart kitchen environment equipped with graphical displays and speakers for prompting, and with cameras and microphones to capture a situation around a supported user [17]. Wherton and Monk [19] clarified that the problems that would occur in such environments can be categorised into nine types, and that systems need to recognise these types and to present verbal and visual prompts properly depending on the types. Even though some research groups [3, 14] have tried to automate the recognition by using various sensors and detection methods, they have not achieved systems which fully automatically support a series of tasks in a kitchen because recognition has essential difficulties on pattern recognition or computer vision. Incomplete automation might cause problems involving human lives. For supporting solitarily living elderly people, these assisting functions should be operated by remote users through network.

There are various networked assistance systems [5,6,10,11, 13]. These systems support simple tasks, but do not cover cooking activities in a kitchen. In addition, they commonly assume supporters are familiar with, or experts of target environments and tasks. This is not always applicable to kitchen environments where elderly people cook because informal and formal caretakers live or work far from the eldery so they do not know the environments well enough. Supporters may be required to find something or to consider a specific procedure together with the supported person. Our system needs a design satisfying such special conditions.

This paper presents a novel system satisfying the conditions for cooking activities of the elderly. The system switches multiple modes for displaying visual prompts even if supporters do not know the environment well enough. Which mode to switch to is decided based on images acquired by cameras for survey. Based on an observation of elderly with dementia problems, we discuss remaining concrete problems in displaying visual prompts and how to solve them technically. The remainder of this paper is organised as follows. In Section 2 we discuss related work and position of our own work. Section 3 presents the proposed system for displaying visual prompts. Section 4 describes the observation and findings from several videos recorded in a senior center. In Section 5, we discuss what functions can be made to improve the current system. Finally, Section 6 concludes this paper.

2. RELATED WORK

2.1 Smart kitchen for demented elderly

There are some proposed systems for supporting a person cooking by displaying computer information. CounterActive [7] by an MIT group is a pioneering implementation of smart kitchens. This system shows users recipes and instructual videos by touching buttons in a projected image on a counter top. A similar system was proposed by Terrenghi et al. [18] in the Kitchen stories project. It enables multiple people to check their favourite recipes and to share them. These systems should be actively operated by user interactions. They are not applicable to demented elderly people because of difficulties in learning how to operate such system.

In contrast to these systems, several systems basically do not assume any active user operations. CounterIntelligence [1, 14] is a system which has multiple discrete functions to gather information from the kitchen and to directly display it around related objects. Chen et al. [2, 3] developed a smart kitchen system for supporting healthier cooking. This system displays amount of fat, vitamin, sodium and mineral in ingredients, which were detected with weight sensors and RFID tags. These methods of detecting current status is applicable to our kitchen system for the elderly. The systems referred to above have not achieved supporting users in actual situations because it is essentially difficult to cover exceptional actions and situations though automation is achieved in simple environments or limited situations [16].

2.2 Remote assistance

Our smart kitchen has another aspect as a remote assistance system. When supporting from remote sites, it is difficult to express positions of needed objects only with verbal communication [13]. There are some methods achieving remote pointing to real objects instead of using demonstrative pronouns.

Kuzuoka et al. [12, 13] first tried to solve this problem with SharedView [12], which is a system that enables the assisted operator to see appearance of a remote supporter pointing at an object of attention in the operator's view image. The operator's view is captured with a head-mounted camera (HMC) and sent to the supporter site. The appearance image of the supporter is acquired with a camera fixed in the supporter site to sent back to a display in the operator site. The advantage of this system is possible to present visual prompts independently of surface property of the object. GestureCam [13] is another system capable of pointing directly at a real object by a finger-like actuator and laser pointer. The disadvantages are that it is not always possible to put actuators in a kitchen due to physical limitation, and that availability of laser pointing depends on surface property of the target object.

Figure 1: An example of projected visual prompt.

Telepointer [15] and WACL [11] are similar wearable pointing systems except using different types of laser pointers. According to the work by Fussell et al. [5], providing remote supporters with a wide-angle, static view of the workspace is more effective than using head- or body-mounted cameras. Attaching such a device to head or body of elderly people [6, 10, 11] gives them large amount of physical discomfort even if it is for short-duration experiment. Hence, in this research, camera, projector and all other heavy devices are assumed to be fixed in the kitchen environment.

2.3 Position of our own work

This research focuses on a method for displaying visual prompts in kitchen environments where there are various objects with glossy or transparent surfaces which are difficult to project images with a projector or laser pointer. If the target object is on a surface such as a counter top, it is highlighted by indirect projection of a circular mark around the object, as shown in Fig 1. Otherwise, the target object is being picked up or put on unsuitable surfaces for projection, visual prompts are displayed on a wall by projectors or embedded monitors, which is similar to SharedView using a fixed camera instead of an HMC. The distinction which mode to be applied is one of the things which a supporter cannot easily know and one of obstacles to smooth communication. The primary contribution of our research is a proposal of a novel method to automatically switch these display modes based on analysing images acquired with camera and projector.

3. VISUAL PROMPTING SYSTEM

3.1 Insertion and Presentation of Prompts

In the proposed system, visual prompts are displayed on flat surfaces such as counter tops and wall of cabinets. In order to project undistorted markers on the surface, it is required to compute how the markers are distorted on the projector image, in other words, to which point on the projector image an arbitrary point on the flat plane corresponds.

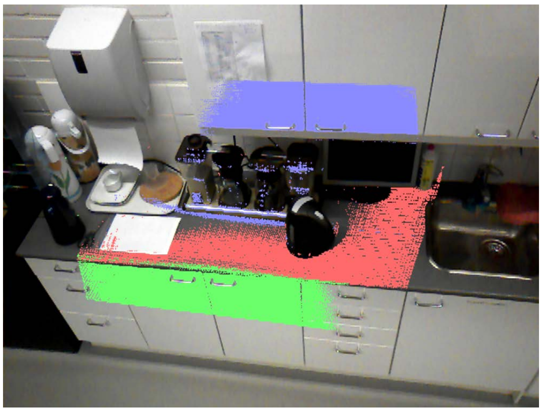

Figure 2: Camera image for a supporter and visualized projectable regions.

On the other hand, the supporter inserts markers through a camera image displayed on some information devices such as desktop monitor and smart phone. When adding visual prompts at an arbitrary position in the camera image by pointing on the screen, it is necessary to calculate to which point on the flat surface the given position corresponds.

For these purposes, we should consider three kinds of coordinate systems fixed on camera image C, projector image P and planar surface S_i $(i = 1, 2, \ldots, n)$. Given a position on each surface represented by three-dimensional homogeneous coordinate $\mathbf{u}_X$, the transformation from one coordinate system X to another Y can be computed using homography matrix H_{XY}. The transformations from camera C to surface S_i and from projector P to surface S_i are computed as H_{CS_i} and H_{PS_i}, respectively. In order to draw markers without perspective distortion as shown in Fig 1, $\mathbf{u}_P$ should be computed from each point $\mathbf{u}_{S_i}$ which the marker consists of, as follows.

$$\mathbf{u}_P \propto \mathsf{H}_{PS_i}^{-1}\ \mathbf{u}_{S_i}. \tag{1}$$

Then, as the following equation, we can directly compute the center position $\mathbf{u}_P$ of the marker corresponding to the position $\mathbf{u}_C$ user pointed at.

$$\mathbf{u}_P \propto \mathsf{H}_{PS_i}^{-1}\mathsf{H}_{CS_i}\ \mathbf{u}_C \equiv \mathsf{H}_{CS_iP}\ \mathbf{u}_C. \tag{2}$$

Here we summarise the required information to realize smooth visual prompting. For inserting visual prompts from a remote site, the following information for each surface S_i must be acquired in advance.

- Homography H_{PS_i} from projector P to surface S_i,
- Homography H_{CS_iP} from camera C to projector P,
- Projectable areas in surface S_i.

3.2 Mode Selection of Visual Prompts

Considerable visual prompt methods for pointing are direct projection [11, 13, 15], indirect projection, embedded monitors [12, 18], projected windows [7, 17] and illumination devices [12]. Since embedded monitors and projected windows are not essentially different, projected windows are not addressed below. There are rarely the items which are directly projectable in typical kitchen environments. Therefore, this paper assumes direct projection is used as an optional method when a supporter gives a special input.

Given a point $\mathbf{u}_C$ on a camera image by a supporter, a proper method is selected by the following steps. The system first makes a corresponding illumination device emit light if the point $\mathbf{u}_C$ in a illumination device region, whose detection method is shown in the next sub-section. The system next draws a circular marker if the following conditions are satisfied.

- The given point $\mathbf{u}_C$ is in a region of a movable object, which is detected by background subtraction.
- There is a projectable region around the object.
- The region is larger than a certain threshold.
- The pixel color at $\mathbf{u}_C$ is the same as the colors of the corresponding pixels on the other camera images [8].

Finally, the system displays, in an embedded monitor, a pointing marker drawn on a captured image.

3.3 Acquisition of Environmental Information

As described in the previous section, there are at least three kinds of information to display visual prompts from a remote site. This information must be acquired at least once after temporarily replacing movable objects when installing of this system. The devices we primarily use are projectors and cameras. Therefore, this setup is proper to measure 3D environments geometrically and photometrically.

If the environment contains non-coplanar surfaces where image can be projected, it is possible to acquire the environmental information by the following steps.

1. **Project structured light pattern** such as gray code to find pixel correspondences between projector and camera images.
2. **Extract projectable areas**. The pixels which have large different intensities between black and white pixels is extracted as projectable region, shown in Fig 2.
3. **Estimate homography H_{CS_iP} of the largest flat area.** Homography of the largest flat area is extracted by RANSAC [4] robust estimation using the correspondences.
4. **Remove the largest flat area** to detect the second largest area in the next iteration.
5. **Iterate from step 3)** until the removed area is smaller than a threshold.
6. **Calculate a fundamental matrix**, which describes epipolar geometry between camera and projector, by Kimura et al.'s calibration method [9] using H_{CS_iP}.

For embedded monitors and illumination devices, corresponding regions of these devices are acquired by toggling their switches.

Fig. 2 shows three projectable regions detected by our prototype system, which consists of a projector (EPSON EMP-73, 1500 ANSI Lumens, 1024x768), a camera (Buffelo Kokuyo Supply Inc. BSW13K05HBK, 640x480 px. 30fps) and a PC (CPU: Intel Core2 2.66GHz, RAM: 2.94GB).

4. OBSERVATION

4.1 Method

We have observed two participants making coffee with an electric drip coffee maker as one of simplest cooking activities in kitchen. The participants in this observation are the following Finnish elderly, who agreed with our observation and research purposes.

Figure 3: Kitchen environment in a senior center.

JO 85-year-old male with Moderate Alzheimer's. MMSE and GDS are 18/30 and 3/15, respectively. Uses a hearing aid.

MI 86-year-old female with vascular dementia due to a stroke. MMSE and GDS are 24/30 and 8/15, respectively. Uses a hearing aid.

The place where the observation was conducted in was a kitchen at a senior center shown in Fig 3. Both participants did not have any experiences in making coffee at that location. In this environment, they made two cups of coffee with support from senior center staff.They were given instructions of item locations and procedures by the staff before making coffee. The procedures they actually took were almost the same as the following one due to these instructions.

1. Take coffee filter from the cabinet,
2. Open and take a dripper,
3. Put the filter in the dripper,
4. Take a coffee can from the cabinet,
5. Take a tablespoon from the can,
6. Put two spoons of coffee powder inside the filter,
7. Insert and close the dripper,
8. Take a pitcher from the counter,
9. Put water into the pitcher,
10. Pour the water into a reservoir,
11. Turn power switch on.

The participants were allowed to ask the staff for support if needed.

4.2 Results

Both participants succeeded in the coffee making task. First, the situations in which each participant was given any support are shown below.

JO completed those steps in 5 minutes. He was supported by the staff 5 times when he was looking for the coffee powder can (step 4), the spoon (step 5), the pitcher (step 8), a fill port of the reservoir (step 10) and the power switch (step 11). All the support for JO were position indications with both visual and verbal prompts except in step 11, in which the staff used only a verbal prompt.

MI completed the same steps in 7 minutes. She was supported 4 times when she was looking for the coffee filter (step 1), a cup instead of the pitcher (step 8), put the dripper and carafe pot on the coffee maker (step 7) and turning the power switch on (step 11). All the support for MI were indications of positions using both visual and verbal prompts. All the supports for both participants involved verbal prompts. Next, we describe the problems which have possibility to occur in the above situations if the proposed system is used.

In the above situations, step 4, 5, 9 and 11 of JO, and step 1 and 11 of MI were situations in which indirect projection of visual markers seems possible, because the target item in each situation touches or is close to a projectable surface. In step 5 of JO, it seems possible to project a marker to the coffee can instead of the spoon JO was searching for. However, the staff did not know whether the spoon was in the can. The staff instructed him that the spoon was inside the can after checking it themselves.

Other than those above are the situations in which indirect projection seems impossible. In step 10 of JO, it was required to instruct him of the position of the fill port by pointing at a part of the coffee maker. It might be difficult to apply indirect projection because the coffee maker's part did not contact any projectable surfaces. For any direct projection, the surface of the part is too black and glossy. In step 7 of MI, she tried to set the dripper and carafe pot on the coffee maker but did not understand the process. Furthermore, she touched those items with both hands, peered down and leaned forward upon them, so that her body covered both hands and the objects. In this situation her hands cannot be observed easily even if seen from any direction.

5. DISCUSSION

In order to confirm the feasibility of the proposed system, especially from the aspect of availability of indirect projection for visual prompts, two demented elderly making coffee were observed. According to the recoded video, they were given supports eleven times. Ten of them were visual prompts for pointing at required items, five were in the situations where indirect projection was available. This result is compatible with the proposition that the indirect projection method is available in various situations. Direct projection was available in two situations, which is less than in the indirect projection method. It suggests that it is possible to expand supportable situations by using the indirect projection method. However, future work should involve construction the proposed system to evaluate effectiveness.

Finally, we discuss considerable solutions for the problems addressed in the previous section. For the problem in step 10 of JO, instead of marker projections, a still or video image showing around a target can be captured and displayed in an embedded monitor or a projectable area away from the target. By doing this, any makers can be drawn in the displayed image. This is almost the same as SharedView system [12] if the operator uses a normal monitor instead of an HMD.

For the other problem in step 7 of MI, on the other hand, this method is not applicable because any good images to show around the target may not be obtained by any fixed cameras. In this case, the only remaining way is to display existing instructual videos. This is equivalent to Counter-Active [7] system.

Given these facts, it is inevitable to switch multiple modes for displaying visual prompts. We reasonably infer that measuring environments with camera and projector provides supporters with important information. In the future, we also speculate on how to switch display modes for intuitive prompting.

6. CONCLUSION

In this paper, we presented a novel smart kitchen system for supporting demented elderly people cooking in a kitchen. In this system, a remote supporter provides the user with verbal and visual prompts, manually. To avoid reflective or transparent surfaces of objects such as cup and pot, circular markers as visual prompts are projected around small objects on a counter top. The system automatically acquires flat surfaces to project and their reflectance properties. Proper methods are automatically selected from using projector, embedded monitor and illumination devices, based on the above information. We found some concrete problems in displaying visual prompts and how to solve them technically.

7. ACKNOWLEDGMENTS

This work is partially supported by Young Researcher Overseas Visits Program for Vitalizing Brain Circulation of Japan Society for the Promotion of Science (JSPS). The authors also would like to thank Academy of Finland for the support of this work under funding for Value Creation for Smart Living Environment for Senior Citizens (VESC) research.

8. REFERENCES

[1] L. Bonanni and C.-H. Lee. The kitchen as a graphical user interface. In *ACM SIGGRAPH 2004 Art gallery*, SIGGRAPH '04, pages 109–111, New York, NY, USA, 2004. ACM.

[2] J. Chen, K. Chang, P. Chi, and H. Chu. A smart kitchen to promote healthy cooking. In *Proceedings of the 8th International Conference on Ubiquitous Computing (ACM UBICOMP 2006)*, 2006.

[3] P. Chi, J. Chen, H. Chu, and J. Lo. Enabling calorie-aware cooking in a smart kitchen. In *Proceedings of the 3rd International Conference on Persuasive Technology*, PERSUASIVE '08, pages 116–127, Berlin, Heidelberg, 2008. Springer-Verlag.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[5] S. R. Fussell, L. D. Setlock, and R. E. Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 513–520, New York, NY, USA, 2003. ACM.

[6] B. Hestnes, S. Heiestad, P. Brooks, and L. Drageset. Real situations of wearable computers used for video conferencing - and implications for terminal and network design. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*, ISWC '01, pages 85–93, Washington, DC, USA, 2001. IEEE Computer Society.

[7] W. Ju, R. Hurwitz, T. Judd, and B. Lee. CounterActive: An interactive cookbook for the kitchen counter. In *in Extended Abstracts CHI 2001*, pages 269–270. ACM Press, 2001.

[8] N. Kasuya, I. Kitahara, Y. Kameda, and Y. Ohta. Robust trajectory estimation of soccer players by using two cameras. In *19th International Conference on Pattern Recognition*, ICPR2008, pages 1 –4, Dec. 2008.

[9] M. Kimura, M. Mochimaru, and T. Kanade. Projector calibration using arbitrary planes and calibrated camera. In *IEEE International Workshop on Projector-Camera System (Procams2007)*, pages 1 – 2, 2007.

[10] R. E. Kraut, M. D. Miller, and J. Siegel. Collaboration in performance of physical tasks: effects on outcomes and communication. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96, pages 57–66, New York, NY, USA, 1996. ACM.

[11] T. Kurata, N. Sakata, M. Kourogi, H. Kuzuoka, and M. Billinghurst. Remote collaboration using a shoulder-worn active camera/laser. In *Proceedings of the Eighth International Symposium on Wearable Computers*, ISWC '04, pages 62–69, Washington, DC, USA, 2004. IEEE Computer Society.

[12] H. Kuzuoka. Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92, pages 533–540, New York, NY, USA, 1992. ACM.

[13] H. Kuzuoka, T. Kosuge, and M. Tanaka. GestureCam: a video communication system for sympathetic remote collaboration. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, pages 35–43, New York, NY, USA, 1994. ACM.

[14] C. J. Lee, L. Bonanni, J. H. Espinosa, H. Lieberman, and T. Selker. Augmenting kitchen appliances with a shared context using knowledge about daily events. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 348–350, New York, NY, USA, 2006. ACM.

[15] S. Mann. Telepointer: Hands-free completely self contained wearable visual augmented reality without headwear and without any infrastructural reliance. In *Proceedings of the 4th IEEE International Symposium on Wearable Computers*, ISWC '00, pages 177–178, Washington, DC, USA, 2000. IEEE Computer Society.

[16] A. Mihailidis, J. Boger, M. Canido, and J. Hoey. The use of an intelligent prompting system for people with dementia. *Interactions*, 14:34–37, July 2007.

[17] P. Olivier, G. Xu, A. Monk, and J. Hoey. Ambient kitchen: designing situated services using a high fidelity prototyping environment. In *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '09, pages 47:1–47:7, New York, NY, USA, 2009. ACM.

[18] L. Terrenghi, O. Hilliges, and A. Butz. Kitchen stories: sharing recipes with the living cookbook. *Personal Ubiquitous Comput.*, 11:409–414, June 2007.

[19] J. P. Wherton and A. F. Monk. Problems people with dementia have with kitchen tasks: The challenge for pervasive computing. *Interacting with Computers*, 22:253–266, July 2010.