ダイレクトポインティングや PTZ カメラを
用いた見えないオブジェクトを選択する手法

金 石煥† 高橋 伸† 田中 二郎†

来る未来では、実世界にある多くのインテリジェントなオブジェクトとインタラクションが出来ると考えられ、それらオブジェクトの選択を容易にする手法の開発は重要である。現在までに提案された手法の中では、ダイレクトポインティングが直感的かつ容易な選択手法であると言えるが、ユーザの視界に入らないオブジェクトは選択出来ないという問題がある。我々はこの問題を解決するための手法を用いたプロトタイプを開発した。本システムはユーザのポインティングジェスチャを距離カメラにより認識する。天井に取り付けた PTZ カメラは認識したポインティングの方向を向き、取得した画像をユーザのモバイルデバイスに送る。ユーザはモバイルデバイスに表示された画像からオブジェクトの位置をクリックすることで、直接見えないオブジェクトを選択することが可能である。本稿ではこの選択手法と実装したプロトタイプについて詳しく述べる。

A Novel Selection Technique for Hidden Object
Using Direct Pointing and PTZ Camera

SEOKHWAN KIM† SHIN TAKAHASHI† JIRO TANAKA†

In the future, it is expected that we can have interaction with various kinds of object in real environment. To interact with them, intuitive and convenient selection technique is necessary. Among developed techniques, direct pointing is promising due to its intuitiveness from self-revealing feature. However, a problem of direct pointing is that user is very difficult to select an object which is hidden by another object. To resolve this problem, we developed a novel selection technique. Our prototype system recognizes user's pointing gesture using depth camera. After recognition of pointing, system changes the direction of PTZ camera, which is installed on ceiling of space, to designated point. Then, the image is transferred to user's mobile device. User can select an object by clicking that image on mobile device. In this paper, we describe our approach and implementation in more detailed manner.

1. Introduction

There are several attempts to build futuristic interactive spaces that allow user to interact with most of objects in that space [1][2][3]. Intuitive and convenient selection technique is very necessary for those environments. With this purpose, several selection and control methods are developed using touch screen or universal controllers [4][5][6].

Among existing methods, direct pointing technique is promising as a selection technique. Direct pointing is excellent for the casual interaction and it is very intuitive due to its self-revealing feature. Direct pointing is used for several application domains such as virtual reality or large display [7][8][9][10][11].

However, direct pointing has significant problem on selecting hidden object. User cannot designate an object when that object is hidden by other objects. In virtual reality, solutions for this problem are studied since decades ago and various approaches are proposed. However, most of them are based on virtually generated graphical environment and employee 3D interaction, thus it is difficult to be used by real environment and 3D interaction is basically not so easy for novice users [12].

To solve this problem, we developed a novel method for selecting an object using direct pointing gesture and an image from PTZ camera. Our prototype recognizes user's pointing gesture at initial step and it makes PTZ camera change its direction to designated position. After then, the image in PTZ camera is transferred to user's mobile device and shown to user. Then, user can make a selection by selecting an object shown in his mobile device.

This approach contributes to resolve two problems. First is it enables user to select hidden object because camera at

---

† 筑波大学 コンピューターサイエンス専攻
Department of Computer Science,
University of Tsukuba

ceiling has broader sight. Second is it helps user to select very fine object. In real environment, user is difficult to select very small sized object using only direct pointing because of hand jitter and technological limitation of tracking device. However, using our approach user can select even very fine object by clicking the screen. In following section, we will describe details of our scenario and interaction method.

Figure 1 System highlights the region containing selectable object
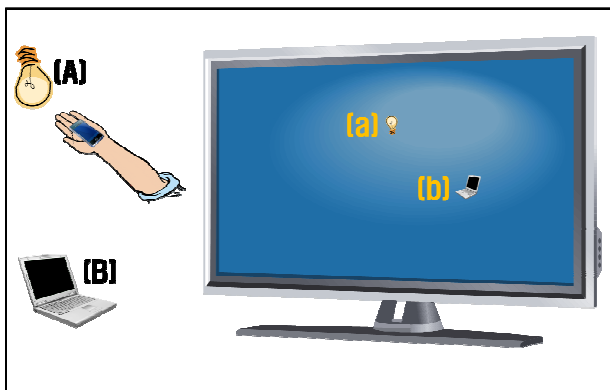

Figure 2 Selected objects can be dropped on desktop, and they become links from metaphor of real objects. Links are connected control applications

2. Prototype System and Application

In this presentation, we show two applications. First is selection example application and second is application which stores selected objects in common desktop as a link. In this section, we describe details of two applications.

2.1 Selection application

Selection application basically allows user to select an object that is not directly shown to user. User is asked to make a pointing gesture and stops movement for 1 second. Then system recognizes pointing gesture and sends the information of designated point to PTZ camera. PTZ camera changes its direction to that point and the image is transferred to user's mobile phone automatically. Figure 1 shows the mobile phone that shows the image from PTZ camera. When there are selectable objects mobile phone shows small transparently colored region is shown on that region. (A) – (C) in figure 1 shows the example. After user selects it, screen shows its coordination values and information of selected device.

This application helps user to select an object even if it is not shown to him and helps to select even fine object. As we described in introduction section, user is difficult select not seen object and small sized object using direct pointing. In this interaction, PTZ camera works as a complementary tool for direct pointing, and we expect that user can confirm its effectiveness.

2.2 Storing metaphor links from real world objects

Even if user excellently selects hidden or very fine objects using proposed methods, it is true that selecting object repeatedly is cumbersome. To resolve this problem, we provide a method to store links (E.g. shortcut) of objects in common desktop environment. After user selects an object in real environment, user can drop it on currently accessible display. Figure 2 shows it. When user selects light (Capital A in figure 2) or laptop (Capital B in figure 2) and drop them on desktop and they become links (small a and b in figure 2). When user selects those links, control applications are executed. Thus, user can select objects using more conveniently. We expect touching link is more convenient than making pointing gesture when available display is near to user.

This application allows user to store links for real objects in graphical environment, and those links are from real objects not from virtual objects. It helps user to recognize meanings of links in desktop intuitively. This application is also promising for future environment. Most of gesture based researches focus on only gesture itself because gesture and GUI are at different interaction spaces. However, in the future more displays and mobile devices will around us; it is possible to design more powerful

interaction when combining gestures and available devices. We expect this application shows a convincing possibility on combination of gesture and GUI.

3. Recognizing Pointing Gesture

3.1 Hardware environment setup

We use a depth camera for recognizing pointing gesture. As shown in (a) of figure 3, we use Microsoft Kinect as a depth camera, and it is installed at corner for cover more range of area in space [13]. This depth camera has VGA (E.g. 640 X 480) resolution and depth information is coming with the rage from 1 to 2047. PTZ camera, shown in (b) of figure 3, is released by AXIS communication. It is installed at the center of ceiling for cover whole range of the space. PTZ camera has resolution 704 X 480 (E.g. NTSC 16:9 video format) and it is up to 30/25 frames per second [14].
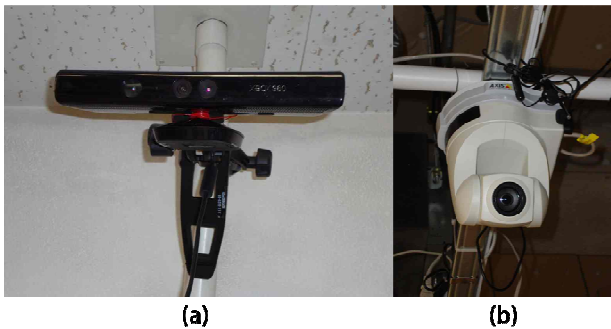

Figure 3 Microsoft Kinect depth camera(a)
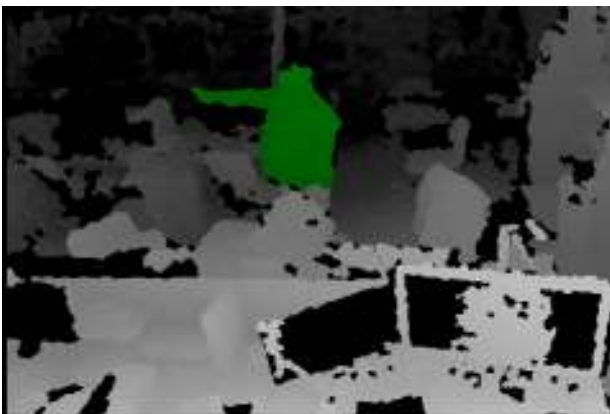and networked PTZ camera(b)

Figure 4 Tracked human (green colored) from OpenNI framework

3.2 Tracking user using depth camera

We used OpenNI framework for getting depth data and tracked human information from Kinect camera [15]. As shown in figure 4, OpenNI framework provides tracked humans' depth data and it also sends converted 3D point information. Green colored area in figure 4 shows tracked user. Our prototype recognizes user's forearm part. More specifically, it tracks two points; those are hand and elbow. Using these two 3D points, we can generate a 3D vector and we can recognize which point is designated by extending this vector.

3.3 Selecting an object in image from PTZ camera

When user's pointing gesture is recognized, system sends a command to PTZ camera and asks to change its direction to designated point. Then, the image from PTZ camera is shown to user's mobile device and the location of interactive object shows a blob. Our system has the information of camera's location and its properties such as zoom level and focal length. Using these values, it is possible to expect which portion of space is shown in camera. Thus, by projecting 3D position of interactive object to 2D screen, relatively scaled, of camera, it is possible to expect which part of image is mapped to a real object. This process is exactly same to 3D vertex projection to 2D plane in 3D graphics. After this process, system is possible to recognize which object is selected in user's mobile device.

4. Related Work

4.1 Ray casting

Ray casting technique selects an object using ray from specific means. Our prototype is basically use ray casting technique. There are various researches that investigate ray casting techniques. Forsberg et al. pointed out that this technique erodes because of hand jitter [16]. Bowman et al. pointed out that it is very powerful except when high precision is necessary [12]. Both researches point out that direct pointing is powerful but erroneous when high precision is required. Our approach is to use only its benefit and eliminate disadvantage by using mobile device.

4.2 Tracking body and hand

The advantage of body tracking is it can result in higher accuracy because it can use more information from several parts of body. Nickel and Stiefelhagen developed the system tracks face and hand together [9]. Their experiment result shows that user's hand pointing gesture is not very precise. Thus, when they use face tracking information together, less error occurs. This research supports our

approach that complementary tools, which are mobile device and PTZ camera in our case, is required. A disadvantage of their work is heavy burden of wearing device. Their prototype requires user to wear head tracking equipment.

4.3 Pointing in real space

Wilson and Pham developed WorldCursor and XWand device [17]. XWand device is equipped with sensors and those sensors enable to track absolute orientation of device. Using tracked data, WorldCursor, installed at ceiling in space, shoot the laser beam to the point user designating. They also simulate the real world space in 3D graphics world. However, their work do not contains the consideration of the objects which are out of sight.

5. Conclusion and Future Work

Direct pointing using depth camera has advantage from device free feature and self-revealing intuitiveness. However, it is difficult to be sufficient to cover hidden or fine object because of feature from its nature. In this paper, we described our featured approach for overcoming those problems. Our effort would contribute to application area trying to control house or office appliances using futuristic interaction. In the future, we will have user test using developed prototype. Through the test we will test two main points. First is its performance. Most of existing research work has experiment of their implantation on experimental environment not in real world. We will have experiment in real environment. Second is its effectiveness. We will have experiment with common mobile device or pointing itself only environment and we will compare results with each other and prove the effectiveness of our approach.

参 考 文 献

1) Microsoft EasyLiving Project
   http://research.microsoft.com/easyliving/
2) Johanson, B., Fox, A. and Winograd, T. The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms," IEEE Pervasive Computing, pp. 67-74, April-June, (2002).
3) Abowd, G. D. Classroom 2000: an experiment with the instrumentation of a living educational environment, IBM Systems Journal, number 4, volume 38, pp.508-530, Dec. (1999).
4) Seifried, T., Haller, M., Scott, S., Perteneder, F., Rendl, C., Sakamoto, D. and Inami, M. CRISTAL: Design and Implementation of a Remote Control System Based on a Multi-touch Display, The ACM International Conference on Interactive Tabletops and Surfaces, pp. 37-44, (2009).
5) Nakajima, T. and Hasegawa, A. Universal Interaction with Networked Home Appliances, 22nd IEEE International Conference on Distributed Computing Systems, pp. 451, (2002).
6) Logitech Harmony Remote Controller
   http://www.logitech.com/en-us/remotes/universal-remotes
7) Poupyrev, I., Weghorst, S., Billinghurst, M. and Ichikawa, T. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. In Proceedings of EUROGRAPHICS, Computer Graphics Forum, volume 17, pp. 41–52. IEEE.
8) Vogel, D. and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In Proceedings of the 17th annual ACM symposium on User interface software and technology, pp. 137–146, (2004).
9) Nickel, K. and Stiefelhagen, R. Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In Proceedings of the 5th international conference on Multimodal interfaces, pp. 140–146, (2003).
10) Kumar, M., Paepcke, A. and Winograd, T. Eyepoint: practical pointing and selection using gaze and keyboard. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 421–430, (2007).
11) Vogel, D. and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology, pp. 33–42, (2005)
12) Bowman, D. A., Johnson, D. B. and Hodges, L. F. Testbed evaluation of virtual environment interaction techniques. In VRST '99: Proceedings of the ACM symposium on Virtual reality software and technology, pp. 26–33, (1999).
13) http://www.xbox.com/en-US/kinect
14) AXIS 214 PTZ Network Camera Datasheet
    http://www.axis.com/files/datasheet/ds_214ptz_33006_en_0809_lo.pdf
15) http://www.openni.org
16) Forsberg, A., Herndon, K. and Zeleznik, R. Aperture based selection for immersive virtual environments. Proceedings of the ACM Symposium on User Interface Software and Technology, pp. 95-96, (1996).
17) Wilson, A. D. and Pham, H. Pointing in intelligent environments with the WorldCursor. In Proceedings of the Ninth IFIP International Conference on Human-Computer Interaction, pp. 495-502, (2003).