Technical Report

Multi-user available information providing system for tourism by pointing gesture

Shogo Ohya (Graduate School of Engineering, Kanazawa Institute of Technology, b6700542@planet.kanazawa-it.ac.jp)
Takashi Kawanami (College of Engineering, Kanazawa Institute of Technology, t-kawanami@neptune.kanazawa-it.ac.jp)

Abstract
This research proposes a tourism information provision system that can be used by multiple users. The system allows that visitors intuitively obtain information of an exhibited object by a pointing gesture. To realize multi-user operation, this paper proposes an authentication method using directional speakers and smartphones. The system is highly robust because it does not malfunction when people cross.

Keywords
pointing gesture, sightseeing information, tourism, Kinect, multi-user

1. Introduction

An audio commentary that introduces many exhibits in museums etc. provides information by pressing a button near the exhibits. When the number of exhibits is small, visitors can easily understand the correspondence between the exhibits and the buttons and select the button. However, in the case of a large number of exhibits or in the case of a large diorama, the distance between each button and the exhibits are separated, and the visitor must carefully select the corresponding button.

Therefore, the authors have proposed an information providing system by a pointing gesture using Kinect sensor [Watanabe et al., 2018]. By this system, the time and effort of the visitor who takes the correspondence between buttons and exhibits are reduced, and the visitor can obtain information smoothly by pointing to the exhibit. However, this system could only provide information for one user. This research extends the system to be available to multiple users.

In this research, the authors cooperate with a music facility that holds a large number of record jackets in Kanazawa Institute of Technology, and propose an information providing system using record jackets as an example of applied cases.

2. Related projects

There is research in which robots recognize the target pointed by a person and use it in daily life [Ueno et al., 2014]. A mathematical model of a gesture based pointing interface system has also been proposed to simulate pointing behavior in various situations [Kondo et al., 2018].

As described above, although there are multiple studies using Kinect pointing, there have been no studies that have been used for tourism etc. yet.

3. Overview of base system

3.1 Base system architecture

Figure 1 shows the basic system architecture targeted for single user implementation.

The number of record jackets displayed was twelve: a total of 11 actual jackets and one virtual record jacket. The virtual one is displayed by a projector. The reason for displaying virtual record jackets is that there are many rare and unobtainable record jackets, and even such record jackets can be exhibited. Each record jacket was embedded in a concave frame made of styrene board.

The range of projection mapping was expanded by using two projectors. The authors calibrated positions to be able to project appropriately according to the shape of the jacket. In addition, a client PC for Kinect control and projection mapping, and a

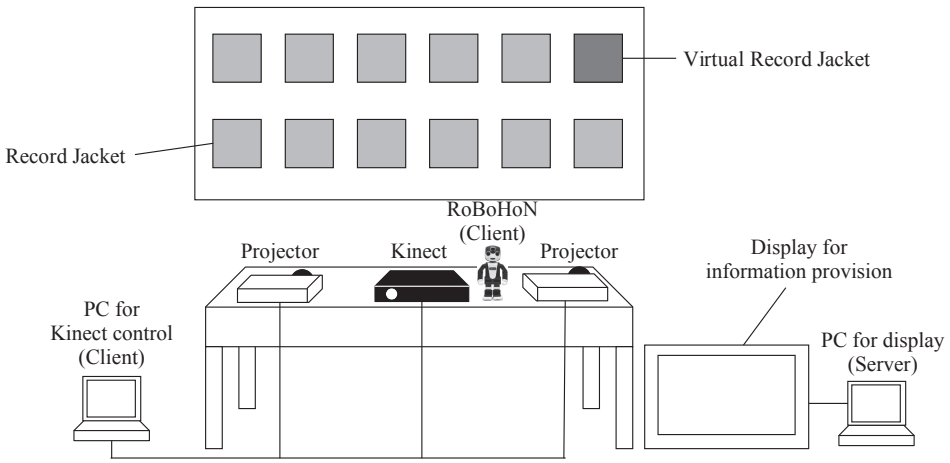


Figure 1: Base system architecture

server PC for information provision display and audio commentary robot control were prepared. Although these PCs were connected by Socket communication, they can be operated with one PC if the performance is satisfied.

Sharp's RoBoHoN was used for audio commentary. RoBoHoN gived an initial explanation of how to use the system to visitors and a detailed explanation of the record jackets. In addition to voice commentary, RoBoHoN is capable of gestures, and can provide more human-friendly commentary.

3.2 Behavior of base system

A processing sequence of the system is shown in Fig.2. When a visitor stands in front of the system, (1) Kinect recognizes a person and transmits data indicating that a person has arrived from the client PC (Kinect side) to the display server PC. (2) The server PC transmits the received data to RoBoHoN. (3) When RoBoHoN receives data indicating that a person has arrived, it explains to the visitor how to use the system. (4) When the visitor points at the record jacket, the system calculates the pointing position by Kinect, and (5) the pointer is projected at the pointed position (like a laser pointer). (6) When the pointer points over the position of the record jacket, the record jacket ID is sent to the server PC, (7) and the server PC immediately displays an overview of the record jacket according to the record jacket ID on the information provision monitor. (8) Continuing to point at the record jacket, a frame is projected gradually by projection mapping around the record jacket. (9) If the visitor continues pointing to the same record jacket for 2 seconds (the frame drawing is completed in 2 seconds), the ID of the record jacket pointed to and the confirmation flag are sent to the server PC. (10) The server PC sends the ID of the record jacket pointed to the RoBoHoN. (11) Finally, RoBoHoN gives the visitor an audio commentary of the record jacket according to the received ID.

3.3 Calculation of pointing direction

There are several ways to calculate the pointing direction, which is based on the direction of the forearm, the direction of the straight line passing from the head and hands, the straight direction passing from the head to the fingertip, and the direction of the face.

According to studies that have evaluated the performance of multiple methods, it is concluded that the method using the direction of the straight line passing from the head and hand, and the straight direction passing from the head to the fingertip has higher accuracy in direction estimation [Le et al., 2010; Nickel et al., 2007].

Therefore, this research uses the method of finding pointing direction based on the straight line passing from the center of head to the fingertip [Watanabe et al., 2018].

3.4 Calculation of pointed position

All record jackets were placed on the same plane. The pointing direction can be represented by a vector. In order to calculate the pointed position, the following equation was used to obtain the intersection coordinates from the plane and the vector.

$$p = \frac{h - (\vec{n} \cdot x0)}{n \cdot \vec{m}} \tag{1}$$

h represents the inner product of the center of the head and the fingertips. Pointing is calculated with the right or left hand. Therefore, the heights of the detected right hand fingertip and left hand fingertip are compared, and the hand at the higher position is regarded as the hand performing pointing. n is the normal to the wall on which the record jacket is displayed, and

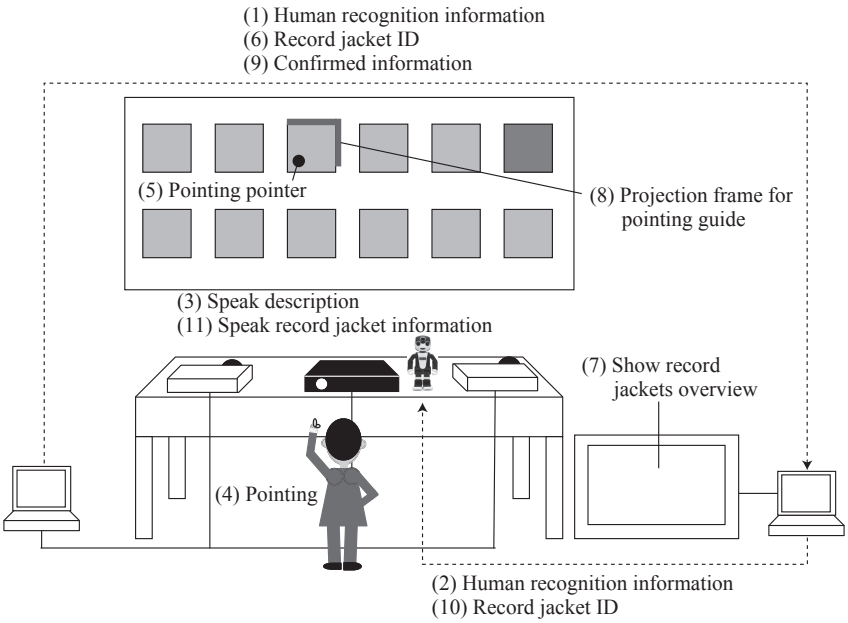


Figure 2: Processing sequence of base system

$x0$ is the position of the fingertip. m represents the pointing direction, and the vector is obtained from the result of subtracting the coordinates of the fingertip from the center coordinates of the head.

3.5 Pointing accuracy improvement

The detection of the center of the head by Kinect may be less accurate if the part of the head is hidden by a hand. According to Ueno et al. [2015], the position of the head is corrected on the extension of two points using the center of the body and both shoulders of the skeleton information.

This correction improves the accuracy of the head x and z axes. However, the authors confirmed that the value of the y axis slightly fluctuates. Therefore, in addition to the above correction, the y-coordinate of the center of the head is stored when the values of the y-coordinates of the right and left fingertips are lower than the waist. Then, when calculating the pointing direction, correction is performed using the stored y-coordinate value. As a result, even when part of the head is hidden by the pointing, the value of the y-coordinate of the head is stabilized.

3.6 Pointer projection to pointed position

The point of intersection between the wall and the pointed position obtained in 3.4 is the coordinate with Kinect as the origin. In this research, the space of projection mapping is developed in Unity and needs to be converted to screen coordinates in Unity. Therefore, coordinate conversion is performed by the following equation.

$$Cx = \frac{(Px - DPx)\, RW}{DW} \tag{2}$$

$$Cy = \frac{(Py - DPy)\, RH}{DH} \tag{3}$$

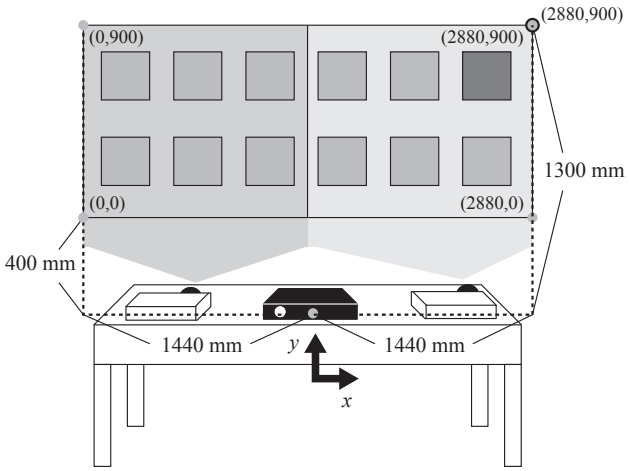
Cx and Cy represent the transformed coordinates of the x and y coordinates. Px and Py represent the x and y coordinates of the intersection of the pointing direction and the wall calculated by equation (1). RW and RH are the screen width and height in pixels, respectively. DW and DH are the width and height of the screen in meters. DPx, DPy is the length in meters from Kinect to the lower left corner of the screen. In this research, the length from the Kinect to the lower left corner of the screen projected on the left projector is used.

The transformation process of coordinates is shown in Figure 3. Since the resolution of the projector used in this research is 14440 px × 900 px, RW and RH are 1440 and 900 respectively. In addition, although the size of the record jacket display panel is 2880 mm × 900 mm, since the width of the screen is equally divided and projected by two projectors, DW = 1.44 and DH = 0.9. Assuming that the lower left coordinates of the screen projected by the left projector as in Figure 3 are 1440 mm in the left direction and 400 m in the upper direction from Kinect, DPx and DPy are –1.44 and 0.4, respectively.

For example, the visitor points at the upper right corner of the screen (1440 mm to the right of Kinect and 1300 mm to the top of Kinect (ie Px = 1.44, Py = 1.3)). In this case, coordinates Cx and Cy are 2880 and 900, respectively. In screen coordinates, the pixel at the lower left corner is (0, 0) and the upper right corner pixel is (2880, 900). Cx and Cy each have the same value as the upper right corner pixel. As a result, it was confirmed that the coordinate conversion was correct.

3.7 User selection method for single-user

Since there is only one user who can operate the base system at the same time, it is necessary to select the operating user. In the base system, since it is assumed that the user stands and operates right in front of Kinect, the operating user is defined as a person standing in the range of 1 m in front of Kinect. However, if the exhibition area and the aisle are the same space,



$$Cx = \frac{(1.44 - (-1.44)) \times 1440}{1.44} = 2880 \quad Cy = \frac{(1.3 - 0.4) \times 900}{0.9} = 900$$

Figure 3: Coordinate configuration

there is a possibility that the passerby may be misidentified as the user. Therefore, the system assumes that the visitor is not the operating user when the visitor is at least 1.5 m away from Kinect. Among the people identified under these conditions, the person closest to Kinect is selected as the operating user.

3.8 Determination of the object to be described

In this system, if the position of the record jacket pointed to continues to overlap for 2 seconds, the record jacket is selected. If this overlap period is too short, there is a high possibility that a record jacket, which the visitor does not want, is selected. On the other hand, if the overlap period is too long, the visitor will be bored. In the experiment, the overlap period was changed between 2 and 4 seconds, and as a result of a questionnaire and interview of about 30 persons, there were many opinions that about 2 seconds was appropriate. Therefore, the overlap period is 2 seconds in this study.

Also, to make the selection clearer, display a frame around the record jacket was displayed while pointing as shown in Figure 4. This frame is drawn gradually in 2 seconds, and there are 40 types of effect patterns. The effect pattern is a combination of the display appearance of the frame, the start position of the frame drawing, and the color. This allows visitors to enjoy visually various expression patterns each time they point.

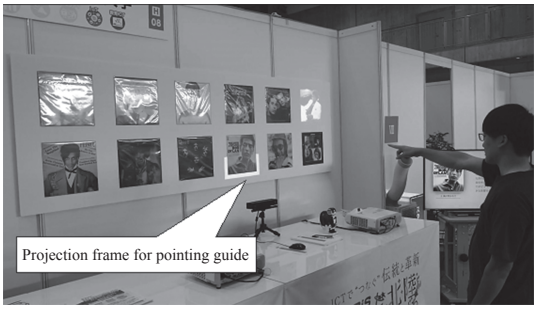


Figure 4: Projection frame for pointing guide

4. Multi-user extension

Depending on the exhibition target, several visitors may request an explanation at the same time. Therefore, this system also supports multi-user use.

4.1 System extension for multi-user usage

In a multi-user environment, providing information about the record jacket changed from the RoBoHoN and information providing monitor to a Web browser in the visitor's smartphone. In this system configuration, it is necessary to pair a person detected by Kinect with a smartphone of a new visitor. Correspondence between the person newly detected by Kinect and the smartphone is performed using a directional speaker. The directional speaker is rotated towards the person detected by Kinect, and sounds a specific frequency. Then, the smartphone receiving the specific frequency is authenticated. Directional speaker is used only for authentication and not for transmission of exhibition information.

4.2 Behavior of multiuser system

A schematic diagram of the system and its processing sequence are shown in Figure 5. (1) A visitor reads a QR code with a smartphone and accesses a dedicated web page. (2) When the visitor is within the detection range of Kinect while accessing the web page, Kinect recognizes the position of the person and sends the position information to the server PC. (3) The server PC calculates the motor angle from the position information of the person and the positional relationship of the directional speaker, and sends the rotational angle of the directional speaker to the microcontroller. (4) The microcontroller changes the angle of the servomotor to which the directional speaker is fixed, and then sounds a specific frequency from the directional speaker. (5) The smartphone that accessed the dedicated page initially has the frequency analysis function activated. When the smartphone detects a specific frequency, the smartphone accesses the server PC. (6) The server PC links

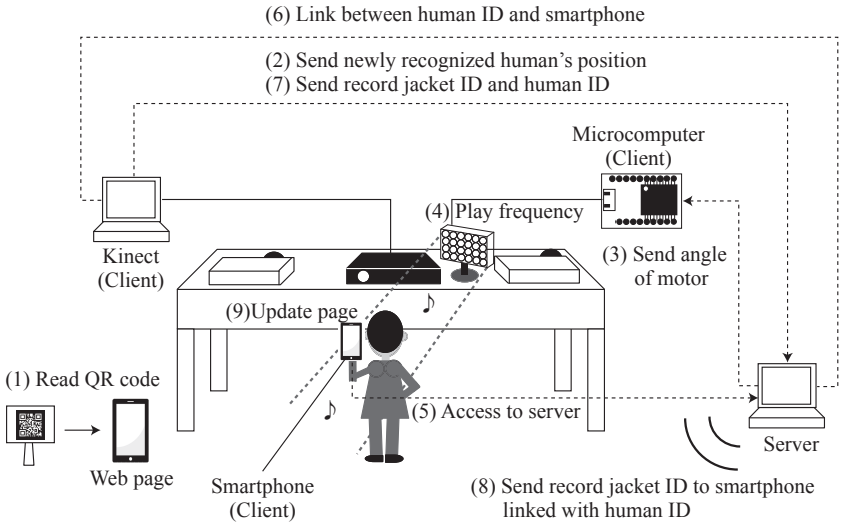


Figure 5: Multi-user system overview and processing sequence

the IP address of the accessed smartphone with the ID of the person recognized by Kinect, and completes the authentication. (7) When the authenticated visitor points to the record jacket, client PC sends the record jacket ID and the ID of the person to the server PC. (8) The server PC transmits the record jacket ID to the smartphone of the corresponding IP address according to the link information of (6). (9) The smartphone switches the web page according to the ID of the record jacket, and presents the information of the record jacket to the smartphone user.

4.3 Elimination of passers in multi-user systems

In a multi-user system, the recognition range is not limited as in the single-user case, and all people within 3 m from Kinect are targeted for recognition. With the expansion of the recognition range, it is necessary to separate passersby from the operators. When the detected person is a passerby, the position data obtained by Kinect is largely different from the data at the previous frame. Therefore, the current position data is compared with the position data of the previous frame, and if the difference is within the threshold for 2 seconds, it is identified as the operator. The authors determined by experimentation that 2 seconds were optimal. As a result, this system can distinguish between passersby and operators.

4.4 Calculation of rotation angle of directional speaker

In the system operation procedure (3) in Section 4.2, the authors show how to calculate the rotation angle of the directional loudspeaker. When Kinect is the origin, the angle between the directional speaker and the position of the person is calculated by the following equation.

$$\theta = \tan^{-1} \frac{Hz - Sz}{Hx - Sx} \tag{4}$$

Hx and *Hz* represent the position *x* and *z* coordinates of the person with Kinect as the origin. *Sx* and *Sz* also indicate the position *x* and *z* coordinates of the directional speaker with Kinect as the origin, and the coordinate of the directional speaker is manually input in advance. Based on these coordinate information, the angle from the directional speaker is calculated in radians. Then, the radian angle is converted to degrees and sent to the microcontroller as the rotation angle of the servomotor.

4.5 Directional speaker and controller

The directional speaker is a parametric speaker kit developed by Tri-State. This directional speaker is composed of 50 ultrasonic transducers. In this research, as shown in Figure 6, the directional speaker, the amplifier and the horizontal axis servomotor were fixed by a frame, which was printed by 3D printer. This directional speaker and servomotor can be connected to a microcontroller to enable angle control and sound reproduction.

4.6 Correspondence method when person crossed

In Kinect, each recognized person is assigned a unique ID.

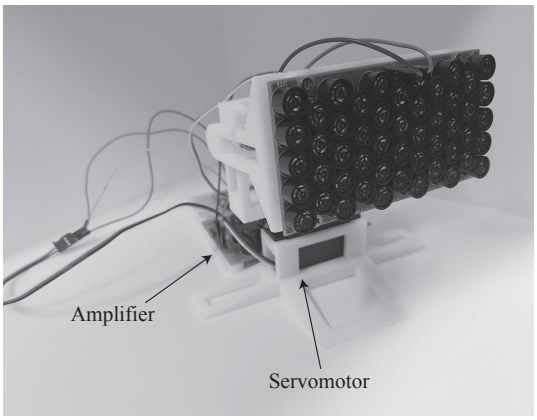


Figure 6: Directional speaker and controller

While Kinect continues to recognize, the ID of the person being identified does not change. However, when another person crosses in front of the recognized person, the recognition state may be released. Also, when the same person is recognized again, even the same person may have a different ID from the previous ID. If this occurs, in the operation procedure (6) in Section 5.2, the system may be misidentified as another person. Therefore, it is necessary to prevent the ID from changing even if another person crosses in front of the identified person.

The method is as follows. The system holds the 3D coordinates of the person whose identification has been released. When the coordinates of the person identified again are close to the coordinates held, the system overwrites the ID of the person whose identification has been canceled with the ID of the person identified again. Note, the retention period of the 3D coordinates of the person whose recognition was canceled was set to 5 seconds in consideration of the case where the visitor left the system (end of use). This is because the time taken to identify again after crossing another person is completed within 5 seconds at the latest.

5. Conclusions

In this study, the authors proposed a tourism information provision system by a pointing gesture. The system also supports multi-user use. This system is applicable not only to record jacket display but also to many applications by changing the display object coordinates. For example, in an art museum, this system can be used to simply explain the actual painting by pointing. Also, in combination with projection mapping, this system can interactively superimpose images or comments on the actual object. In addition, this system can switch and display exhibits that are not owned by the museum at regular intervals.

As future extensions, the authors plan to use a new depth sensor and implement using a human body skeleton detection library with a common Web camera and artificial intelligence.

Acknowledgments

The authors would like to thank K. Takano, I. Suzuki, G.

Sato, and T. Watanabe for useful discussions.

References

Kondo, K., Mizuno, G., and Nakamura, Y. (2018). Feedback control model of a gesture-based pointing interface for a large display. *IEICE Transactions on Information and Systems*, Vol. 101, No. 7, 1894-1905.

Li, Z. and Jarvis, R. (2010). Visual interpretation of natural pointing gestures in 3D space for human-robot interaction. *Proceedings of 11th International Conference on Control Automation Robotics & Vision*, 2513-2518).

Nickel, K. and Stiefelhagen, R. (2007). Visual recognition of pointing gestures for human-robot interaction. *Image and vision computing*, Vol. 25, No. 12, 1875-1884.

Ueno, S., Naito, S., and Chen, T. (2014). An efficient method for human pointing estimation for robot interaction. *IEEE International Conference on Image Processing*, 1545-1549.

Ueno, S., Naito, S., and Chen, T., (2015). A calibration method using one point for instructing robot by pointing gestures. *The Journal of the Institute of Image Information and Television Engineers*, Vol. 69, No. 2, 53-57.

Watanabe, T., Ohya, S., Takano, K., and Kawanami, T. (2018). An indicated area prediction system for exhibitions. *Journal of Global Tourism Research*, Vol. 3, No. 1, 25-30.

(Received April 15, 2019; accepted May 20, 2019)