Identify the Interaction of Multiple Users by Heterogeneous Features Grouping

Pi-Hsien Wang[1], Yu-Hsin Lin[1]

wangpihsien@itri.org.tw, Lawrence.Lin@itri.org.tw
[1] Industrial Technology Research Institute, 195, Sec. 4, Chung Hsing Rd., Chutung, Hsinchu, Taiwan 31040, R.O.C.
Keywords: transparent display, interactive technologies, multiple users.

ABSTRACT

Conventional multi-person interaction methods cannot identify who is touching. We combine touching behavior judgment and eye detection to achieve identification and eye-to-hand pointing. It could apply in transparent display's information fusion to human eyes. We also analyze the accuracy and response time of the 2 users' interaction.

1 INTRODUCTION

In recent years, there has been an increasing interests in the intuitive interaction of the human-machine interface, which can use gestures, touching, and sight. The common approaches do not combine these interactive methods. When the field was used by multiple people, everyone has the same priority and cannot distinguish the users. Under two-person touching and watching, a touching point may have two eye pairing situations. The two pairs of different pointing correspond to different objects behind the transparent display, and incorrect information will be reported (Fig.1). In this article, the heterogeneous features grouping method with touching hand and eyes, and composite directional interaction system is developed. In the case of multiple users, the touching hand corresponds to the user identification, which can provide personalized information and correspond to the object behind the transparent display viewed by the sight direction. (Fig.2).

Fig. 1 Multi-person touching and eyeball matching

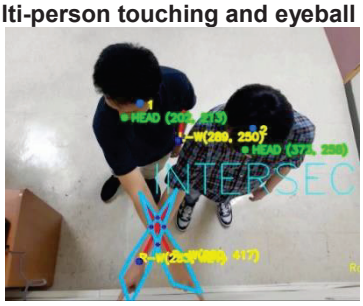


Fig. 2 Intersecting touching behavior

2 THE PROPOSED METHOD

The section 2.1 describes users' identification by heterogeneous features grouping method. The section 2.2 describes the composite directional interaction system with heterogeneous features grouping method and hardware.

2.1 Heterogeneous Features Grouping Method

This method combines recognizing at least a global feature (body) and local feature (eye) by cameras. The wide field of view to the global features as body joint is good for keeping tracking the user, especially when the local feature as eyes is shadowing by hat or by other people. The global feature is also used to judge touching behavior. When a hand is raised for touching, the hand-to-display distance becomes shorter, and the hand's joints-to-joints distance becomes longer in the nearer view to the camera. When two users' arms intersect, the arm's joint-to-joint vectors intersect. Moreover, the more focused view of camera to the local feature could identify the biological characteristics like face or eye. The features' locations of the same user in different cameras could be mapped by the coordinates transformation. In this study, Openpose is used for body recognition [1-2]. All the body joints like hand and head are detected. Then, the head position is used to map the eye position. The hand position is used to map to the touching position.

2.2 System architecture

Fig. 3 illustrates the interaction system. In this study, the transparent display is a high transparent OLED display with Transmittance 70% produced in ITRI laboratory. The computing unit is a laptop compose Intel Core i7- 8850H and NVIDIA GeForce Dual GTX 1080. The first camera is a RGB camera with 1280*720pixels for body detection. The vision covers the top view of the display and the users, then the touching behavior could be detected by the camera. The second camera is a stereo camera with 1280*720pixels for eye detection. The vision covers the front view of the users and the sight direction is detected. The third camera is a stereo camera with 1280*720pixels for object detection. The vision covers the object behind the display. The method is described as follows. Firstly, the body, eye, object image is captured. Secondly, eyes, body and objects in the images is recognized by computing unit. Thirdly,

mapping the touching point on display and the hand on the first camera images, then, mapping the eyes location on the first camera's image and the second camera's image. Fourthly, the eye and hand are paired, and the eye and objects are paired.

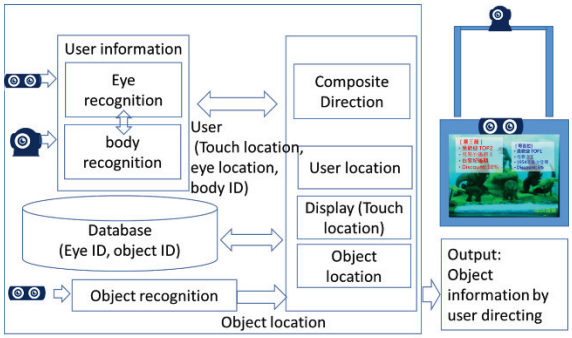


Fig. 3 System architecture

3 EXPERIMENT

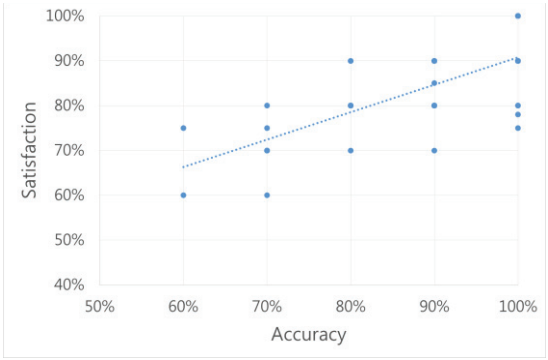
The section 3.1 describe the two users testing experiences. And the section 3.2 shows the users' testing results for the system.

3.1 Two users pointing experiences

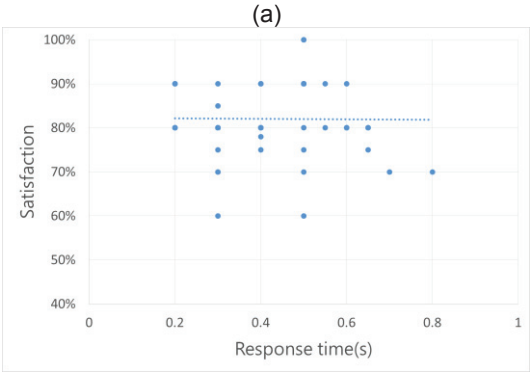
To evaluate the users' satisfaction of accuracy about the composite directional interaction system, we design a two users pointing experiment. There are toy monsters, numbered 1 to 5, standing in the 17" transparent display demo box. When user touched the monster position, the display showed fusion information about the monster's introduction, 20-25words for each monster. The response time is from touching time to display the fusion information time. Two users touched the display following their sights on the monsters, and pointed the monsters in a test round from number 1 to 5 then from number 5 to 1. After a test round, the users feedback their thought about the satisfaction of the accuracy and response time. Then, the 2 users change location and do this test round again.

3.2 Results

Figure 4 illustrates the two users' pointing experiment results, and each point means 1 test round. The accuracy range is 60-100% (Fig.4 (a)). The accuracy lost by fail detection of the eyes, body, monster, pairing for eye and hand, or pairing for eye and monster. These would be discussed in section 4. The satisfaction is positively correlated with the accuracy rate. The image response time is in the range of 0.2 to 0.8 second and no obvious correlation with the satisfaction rate (Fig.4 (b)). The user reply that the response time is good enough that they thought the fusion information was showed right after their touching. However, the users were usually disappointed that no information response or wrong information.



(a)



(b)

Fig. 4 Users satisfaction experience with (a) accuracy, (b) response time

4 DISCUSSION

In section 3, the accuracy was lost by fail detection of the eyes, body, monster, pairing for eye and hand, or pairing for eye and monster. The detection rate is shown in table 1. The detection of eye-monster pairing is lower for the coordinate transformation of stereo camera and the eye position accuracy having large error and needing more calibration. In this study, the eye-hand pairing by heterogeneous feature group method have be proved as detection rate 99%.

Item	Eye	Body	Monster	Eye-hand pairing	Eye-monster pairing
Detection rate	99%	99%	99%	99%	88%

Table 1 Detection rate

The role of the stereo camera here is not only for coordinates transformation, but also to exclude overlapping people. When people's eyes locate closely, the head may group to two too close eyes. (Fig.5). The stereo camera provide distance of users and the user in the using distance to the display has higher priority. This method improved the detection rate of eye-hand pairing in more people condition.

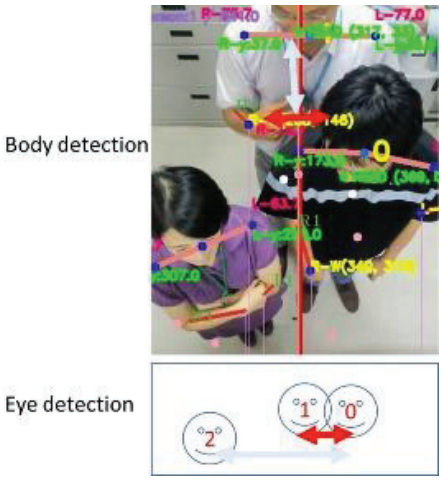


Fig. 5 Close features

5 CONCLUSIONS

This paper presented the heterogeneous feature group method by creating relation of local feature as eye, and global feature as body, and demonstrated two users' pointing identification on the composite directional interaction system. The novel interaction method with transparent display provide users more interesting and intuition interactive experiences.

REFERENCES

[1] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7291-7299).

[2] Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4724-4732)