

# A Touch Perception Algorithm based on Depth Information\*

YU Haitao<sup>1</sup>, GE Shuiying<sup>2</sup>, LIU Shuman<sup>3</sup>

1. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

E-mail: [haitao.yu@ia.ac.cn](mailto:haitao.yu@ia.ac.cn)

2. Research Center for Scientific Communication, Chinese Academy of Sciences, Beijing 100190, China

E-mail: [shuiying.ge@rcsc.cas.cn](mailto:shuiying.ge@rcsc.cas.cn)

3. National Science Library, Chinese Academy of Sciences, Beijing 100190, China

E-mail: [liushuman@mail.las.ac.cn](mailto:liushuman@mail.las.ac.cn)

**Abstract:** The paper presents an algorithm of touch perception based on depth information. Firstly, the area of perception is determined according to the depth information, which corresponds to the two types of action: movement and click. Secondly, the two types of perception are defined based on the movement of hand, which provide the basic interactive datum. Thirdly, the relationship between the perception and display is built on the basis of mapping from depth to pixel, which is helpful to keep the position consistency between touch and feedback. The result shows that new algorithm is effective and accurate for touch perception in arbitrary flat area.

**Key Words:** Touch Perception, Depth Information , Calibration

## 1 Introduction

Because of popularity of Smartphone, multi-touch has become an important way of human-computer interaction, which is a more simple and natural in interactive process compared with contact-type methods, i.e. mouse, keyboard and joy stick.

Multi-touch perception technology can be divided into three types: "FTIR" (Frustrated Total Internal Reflection), "ToughLight" and "Optical Touch".

For "FTIR", it used light reflection to implement reaction through adding led light into screen[1-3]. Although the precise touch position could be obtained by the method, the area of perception was restricted by the screen size which was hard to be extended.

For "ToughLight", hand tracking was done based on the interruption of infrared which was projected to the plane or screen[4-5]. Although the restriction of perception area had been alleviated, the accuracy of perception is affected on the condition of infrared shielding.

For "Optical Touch", hand position is obtained by the two cameras[6]. Although the light shielding can be avoid, like the FTIR, the area of perception was hardly to be extended.

To solve the problem, the paper presents an algorithm of touch perception based on depth information. It can get rid of the light shielding through putting the depth sensor parallel to the area of perception. Besides, the different data types perception are defined based on the movement of hand. Finally, the area of perception can be extended by fusing the perception datum from neighboring depth sensors.

## 2 Depth Perception

In order to simulate the actions, such as movement and click in a large area, the zones of movement and click are defined separately based on depth information first. In

terms of spatial relationship, movement zone contains click zone. In other words, the testing of movement is a prerequisite for judging click. The two lines are defined to distinguish the movement zone and click zone, as in Fig.1.

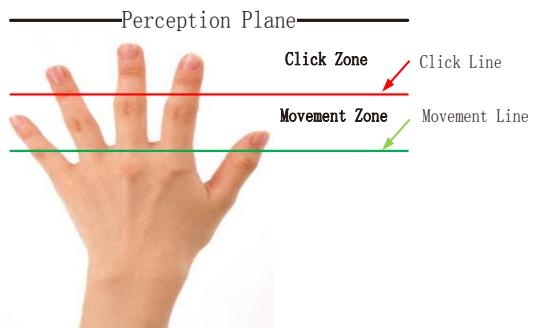


Fig.1: Perception Area

The two types of action: movement and click are detected through the distance between fingertips and perception plane. The movement perception is detected when the fingertips enter into the movement zone. Accordingly, the click position is obtained while fingertips are included in click zone.

In the depth sensor, the depth region for movement and click is recorded separately, and the centre of minimal circumscribed circle which include the depth region for click is regarded as the exact position of perception, as in Fig.2.

Besides, due to the diversity of hand gesture, if more than one depth regions are detected in the depth image for click, the centre of minimal circumscribed circle for each click depth zone is computed respectively. And the average of these centre is computed as the exact position of perception, as in Fig.3.

\*This work is supported by National Science and Technology Support Project of China under Grant 2012BAH42F02.

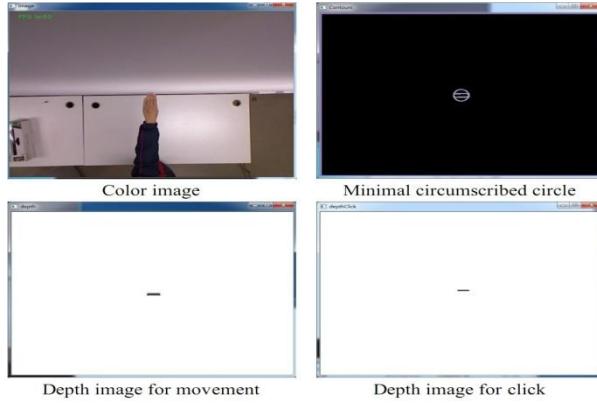


Fig.2: Depth for Movement and Click



Fig.3: Perception for Different Gesture

### 3 Perception Datum

On the basis of position of perception, the various of perception datum are defined for different multi-touch application.

The perception datum can be divided into two types: position and action.

#### 3.1 Position Datum

For position datum, we can divided it into single-perception position and multi-perception position.

##### (1) single-perception

According to the classification of movement and click, perception position can be further decomposed as movement position and click position.

##### (2) multi- perception

Multi-perception is built based on the single-perception. Besides, in order to distinguish the different multi-perception position, we set the unique identifier(UUID) for each single perception on the basis of the consistency of its movement. The consistency of single perception can be judged by the comparison of position and orientation with those at previous frame, shown as in Fig.4.

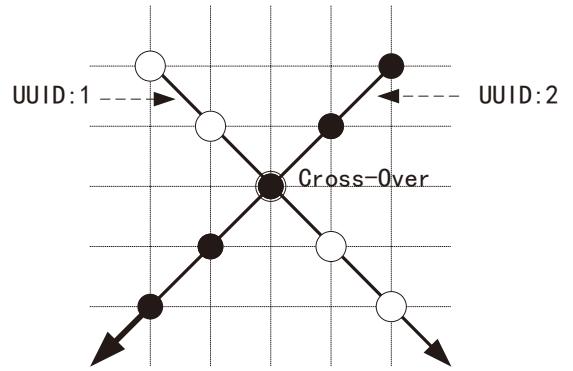


Fig.4: Consistency of single perception

#### 3.2 Action Datum

Action Datum are built on the basis of position datum. Compared with position datum, action datum can be used to provide more natural human-computer interaction. Action can be divided into two types: one -hand operation and double-hand operation, compared with single -perception position and multi-perception position.

##### (1) One-hand operation

One-hand operation is the basic interactive method. Based on the process of the operation, it can be divided into simple one-hand interaction and complex one-hand interaction.

###### ● Simple one-hand operation

Simple one-hand operation include click and movement in the same direction. Click can be detected by judging the position relationship between fingertips and click zone mentioned above.

For the movement in the same direction, we need take several factors into consideration to improve the robustness of recognition.

The factors include time (T), distance(D) and angle(A), offset(O). They represent the time of recognition, distance of movement, deviation range of angle along the direction during the operation and the offset of misoperation during the operation .

The concrete process of recognition is shown as in Fig.5.

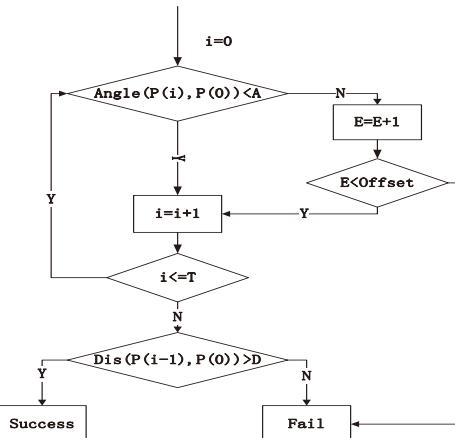


Fig.5: Recognition of movement and click

In the figure,  $\{P(i), i=0,..n\}$  represent the consecutive perception points;  $P(0)$  is the starting point;  $\text{Angle}(P(i), P(0))$  represents the angle between  $P(i)$  and  $P(0)$ ;  $\text{Dis}(P(i), P(0))$  is

the distance between the  $P(i)$  and  $P(0)$ ;  $E$  represents the number of misoperation.

Take the recognition of wave to the right for example, the results of different operations are shown as in Fig.6.

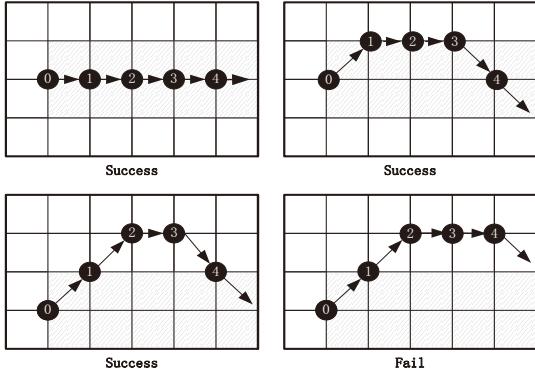


Fig.6 The recognition of wave to the right  
(Time=4, Distance=3, Angle=45°, Offset=2)

### ● Complex one-hand operation

Complex one-hand operation is built based on the simple one-hand operation and can be used to provide more natural interactions.

Complex one-hand operation is composed of several simple one-hand operation in order. For example , L-shape interaction can be divided into two operation: wave down and wave to the right.

### (2) Double-hand operation

The recognition of double-hand operation is not only built on the recognition of one-hand operation but also depends on the cooperation of two hand.

In order to express the position datum and action datum simultaneously , the data structure of perception is designed as follows:

```
{
    dataType: {Position, Action};
    actionType: {wave left, wave right,
                zoom in, zoom out,.....};
    PosX: int
    PoxY: int
}
```

## 4 Calibration between Perception and Display

In order to project the position or action of hand into correct area of display, the correspondence between perception and display should be calibrated first.

We use  $Dep(X, Y, Z)$  to represent the 3D datum in depth and  $Screen(x, y)$  to represent the 2D datum in display. What we need to do is to build the association between them. For  $Dep(X, Y, Z)$ , the  $Z$  axis is perpendicular to perception plane and  $X, Y$  axes are parallel to perception plane.

According to the classification of depth perception mentioned above, the value in  $Z$  axis is used to judge classification of depth perception. So the correspondence between  $Dep(X, Y, Z)$  and  $Screen(x, y)$  is simplified to the relationship between  $Dep(X, Y)$  and  $Screen(x, y)$ .

Considering the difference between the display area and the perception area, we build the bilinear map from  $Dep(X, Y)$  to  $Screen(x, y)$ , shown as Fig.7.

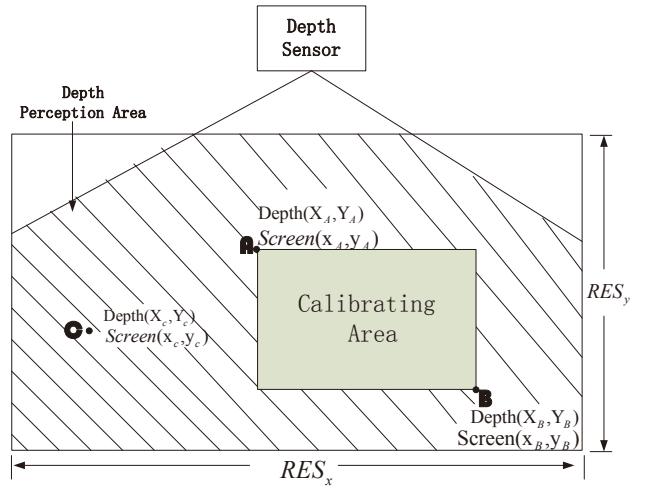


Fig.7 Calibration between Depth and Display

As shown in the figure, we use the Calibrating Area to compute the correspondence between the  $Dep(X, Y)$  and  $Screen(x, y)$ . A and B is the vertices of diagonal in Calibrating Area. C is the any point in Depth Perception Area. We can compute the relationship between the  $Dep(X, Y)$  and  $Screen(x, y)$  as formula (1) and formula(2).

$$\begin{aligned} Screen(x_c) = & \frac{Screen(x_B) - Screen(x_A)}{Depth(X_B) - Depth(X_A)} * Depth(X_c) \\ & + \frac{Depth(X_B) * Screen(x_A) - Depth(X_A) * Screen(x_B)}{Depth(X_B) - Depth(X_A)} \end{aligned} \quad (1)$$

$$\begin{aligned} Screen(y_c) = & \frac{Screen(y_B) - Screen(y_A)}{Depth(Y_B) - Depth(Y_A)} * Depth(Y_c) \\ & + \frac{Depth(Y_B) * Screen(y_A) - Depth(Y_A) * Screen(y_B)}{Depth(Y_B) - Depth(Y_A)} \end{aligned} \quad (2)$$

## 5 Distributed Data Fusion

Due to the limitation of perception area for one depth sensor, the distributed framework for is built for a large perception area, as in Fig.8.

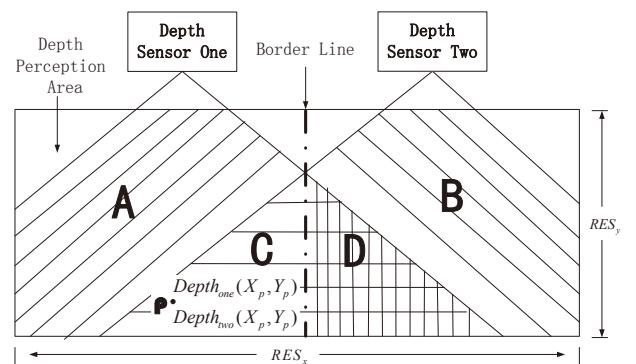


Fig.8 Calibration between Depth and Display

In the Fig.8, there are two Depth Sensors and the Depth Perception Area is divided into four parts :A,B,C and D. P is the any point in Depth Perception Area.

For area A and B, we can compute the correspondence between perception and display as mention above. For area C and D, both of them can be covered by Depth sensor one and Depth sensor two simultaneously. So in order to avoid acquiring the redundant perception datum, the data fusion should be done in advance of display.

Assuming the resolution of display is ( $\text{RES}_X, \text{RES}_Y$ ), the border line is the bisector of Depth Perception Area and is set to distinguish area C and area D. If perception point P in the area C, the depth datum ( $\text{Depth}_{\text{one}}(X_p, Y_p)$ ) is acquired only from Depth Sensor One. Likewise, If perception point P in the area D, the depth datum ( $\text{Depth}_{\text{two}}(X_p, Y_p)$ ) is acquired only from Depth Sensor One.

## 6 Experiment

We use Kinect as the depth sensor and build the experimental environment to prove the effectiveness of algorithm. The width and height of perception area are 3.5m and 2m respectively. And the resolution of display is (1920,1080).

According to the classification of Perception Datum mentioned above, the experiment is divide into two part: perception for position and perception for action.

### ● perception for position

We compute the position of perception with different touching count under the circumstance which includes various count of Kinect. For convenience, we set unique lighting effect for each touch perception, shown as Fig.8. And the result of perception for position is shown as Table 1.

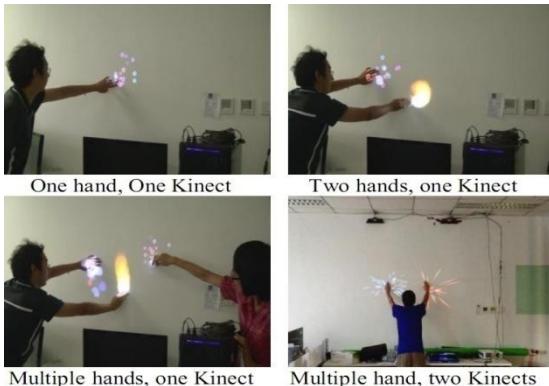


Fig.8 Perception for Position

Table 1: Perception for position

Total	Correct	Accuracy	Mean of distance error (cm)
550	517	93.5%	4.7

Based on this result, random hand gestures contributed to the distance error, shown as Fig.10.

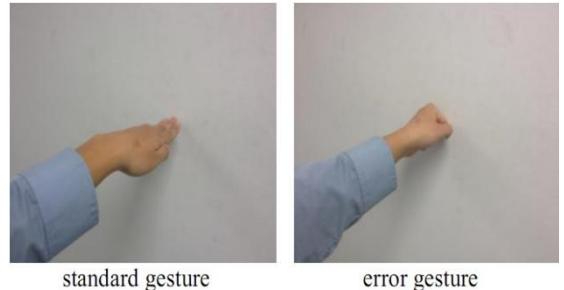


Fig.10 Different gesture

### ● perception for action

Four actions is designed: wave to the left, wave to the right, zoom in and zoom out. And we develop the photo browser for test the accuracy of actions, shown as Fig.11 and Fig.12. And the result of perception for position is shown as Table 2.

Based on this result, four actions can be recognized in most cases. The errors in wave are caused by the wrong orient and speed of movement. It can be improved by change the time (T), distance(D) and angle(A), offset(O), mentioned as chapter 3.2.

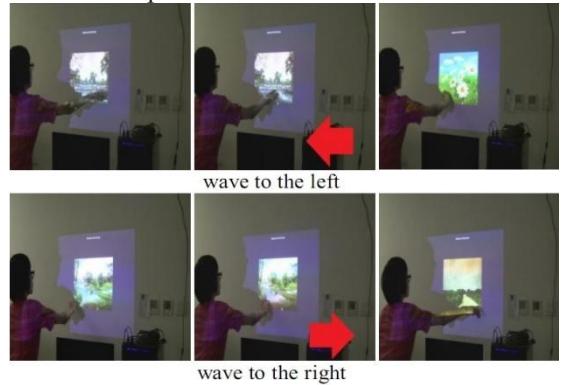


Fig.11 Wave action



Fig.12 Zoom action

Table 2: The result of action recognition

Action Type	Total	Correct	Accuracy
wave to the left	215	207	96.3%
wave to the right	233	218	93.6%
zoom out	242	227	93.8%
zoom in	236	218	92.3%

The errors in zoom are caused by the loss of depth in a moment, while multiple hands cross from the view of depth sensor.

## 7 Conclusion

The algorithm of touch perception based on depth information is proposed in the paper. The position and action can be acquired and recognized by calibration between perception and display. Besides, the size of perception area is expanded by distributed data fusion. For future work, we will study on robustness and accuracy for diverse hand gestures and different hand operation ways.

## References

- [1] Han, Jefferson Y, Multi-touch interaction wall. *International Conference on Computer Graphics and Interactive Techniques*, 2005: 71–75.
- [2] B Buxton, Multi-Touch system that I have know an loved. *USA: Microsoft Research*, 2009.
- [3] R Juan-Yao, P C P Chao and C Wei-Par, A multi-touch interface circuit for a large-sized capacitive touch panel, *Sensors 2010 IEEE*, 2010:309-314.
- [4] S Hodges, S Izaddi, A butler, A Rrustemi and B Buxton, Thin-Sight: Versatile Multi-Touch Sensing for Thin Form-Factor Displays, in *Proceedings of 20th annual ACM symposium on User interface software and technology*, 2007:259–268.
- [5] Li Q, Fang H L, Zhang P Y, et al, The ARM vased hardware system of infrared multi-touch screen design and research, in *Proceedings of 24th Chinese Control and Decision Conference*, 2011: 511–515.
- [6] A D Wilson. Play Anywhere: A Compact Interactive Tabletop Projection-Vision System, in *Proceedings of 18th annual ACM symposium on User interface software and technology*, 2005:83-92.