Development of Mouse Cursor Control System Based on Face Direction Using Kinect

Masayoshi Tabuse
Graduate School of Life and Environmental Sciences, Kyoto Prefectural University
1-5 Hangi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail:tabuse@kpu.ac.jp

Kaori Tamura
ISI Software Corporation
4-6-17 Honmachi, Chuo-ku, Osaka 541-0053, Japan

Abstract

We propose a method of operating mouse cursor and controlling television based on face direction using Kinect. Using our system anyone can control television by motion of a user easily and cheaply. Our system measures face direction and controls mouse cursor using 4 face direction using Kinect. It changes a television's channel and volume by controlling mouse cursor based on face direction using Kinect.

Keywords: Mouse Cursor Control System, Kinect, Face detection, Face Direction, Television Tuner Software

1. Introduction

In recent years, development of many new kinds of input devices is performed, for example, video and audio input devices. Gaze measurements are used as operation of a mouse cursor. For gaze measurement, input devices are classified into a contacted type and a non-contact type. For a contacted type, detection of a motion is easy and high accuracy. However, equipment and removal of it and adjustment according for a user are required whenever it is used. For a non-contact type, removal of it is not required. However, in order to acquire high accuracy, it is necessary to fix a user's head or to restrict user's head movement, and adjustment according for a user is required. Therefore, the burden on a user is heavy when a user uses a gaze measurement device.

In the present study, we propose a method of operating mouse cursor and controlling television based on face direction using Kinect. Kinect consists of a RGB camera, a depth sensor, a multi-array microphone and software (Kinect for Windows SDK), released by Microsoft. It can recognize a position, a voice, a face, face direction and skeleton of a user. Kinect was made for the purpose of a game and entertainment, but is applied to the support of medical care and a person with a disability. By using our system, anyone can control television by motion of a user easily and cheaply. It is not necessary to fix a user's head or to restrict user's head movement. Therefore, it can ease a user's burden compared with conventional devices. Furthermore, a person with trouble in a hand can control television easily.

In our system, Kinect is connected to a personal computer. A television tuner is installed in the personal computer. Television tuner software is controlled by a mouse cursor and a mouse wheel. Our system measures face direction and controls mouse cursor using 4 face directions of rightward, leftward, upward and downward directions using Kinect. In the case of a rightward face the system turns a television channel to the forward direction. In the case of a leftward face it turns a television channel to the opposite direction. In the case of an upward face it raises a television's volume. In the case of a downward face it lowers a television's volume in the downward face direction. Our system changes a television's channel and volume by controlling the mouse cursor and the mouse wheel based on face direction using Kinect in a range with distance less than 4 meters between a user and Kinect. It is also applicable to operation of other application software controlled by mouse cursor, such as a web browser.

2. Outline of Our System

Fig.1 shows the flowchart of our system.

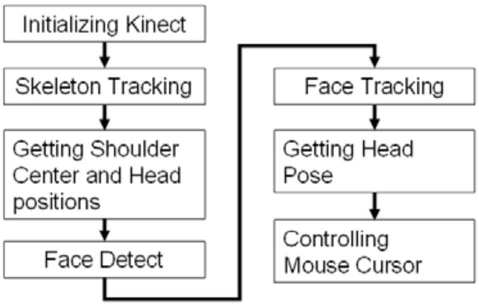

Fig. 1. The flowchart of our system.

3. Face Tracking and Face Direction

Microsoft Face Tracking SDK together with Kinect for Windows SDK enables us to track human face in real time.

3.1. *Human Body Detection and Skeleton Tracking*

Kinect for Windows SDK detects human bodies using depth data acquired from the depth camera in Kinect, estimates 3D joint positions and tracks skeletons[1]. Fig.2 shows joint data.

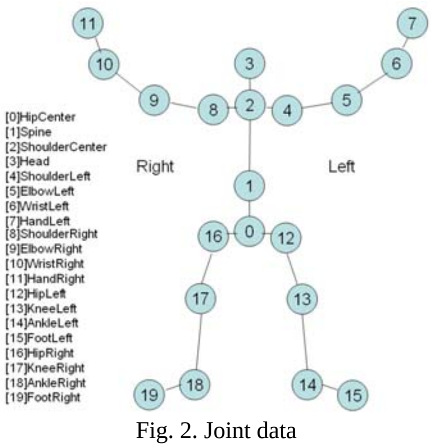

Fig. 2. Joint data

3.2. *Face Detection and Face Tracking*

Face Tracking SDK extracts human face region in RGB image using Shoulder-Center and Head positions in 3D joint data, depth data and RGB data[2]. It fits a deformable face model in face region and obtains 2D/3D position of 100 landmark points on human face. It tracks these points based on a feature of landmark points in the previous frame.

3.3. *Face Direction*

Face Tracking SDK tracks the 100 points on human face and outputs tracking status, 2D position of points, 3D head pose. The head pose is expressed by three angles: pitch, roll and yaw, shown in Fig.3. A pitch angle φ ranges from -90 degrees (looking down towards the floor) to 90 degrees (looking up towards the ceiling). A roll angle θ ranges from -90 degrees (horizontal parallel with right shoulder) to 90 degrees (horizontal parallel with left shoulder). A yaw angle ψ ranges from -90 degrees (turned towards the right shoulder) to 90 degrees (turned towards the left shoulder). Therefore the face direction is determined by these tree angles. In our system we divide face directions into 5 directions: facing up, facing down, facing right, facing left, facing the front. For $\varphi \geq 15°$ and $-15° < \theta < 15°$ human faces up. For $\varphi \leq -15°$ and $-15° < \theta < 15°$ human faces down. For $-15° < \varphi < 15°$ and $\theta < -15°$ human faces right. For $-15° < \varphi < 15°$ and $\theta \geq 15°$ human faces left. Otherwise human faces the front.

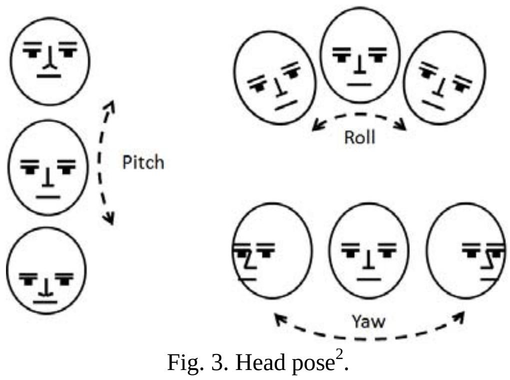

Fig. 3. Head pose[2].

3.4. *Mouse Cursor Control by Face Direction*

In our system we install a television tuner and television tuner software in a personal computer. The television tuner software is controlled by a mouse cursor and a mouse wheel. In the right region of the software window we roll the mouse wheel, so that the software turns the television channel. In the left region of the software window we roll the mouse wheel, so that the software changes the television's volume. Therefore, our system moves a mouse cursor in response to a face direction and it controls the television tuner software.

4. Experiments

4.1. *Experimental Environment*

In our experiments we have used a PC (Dell Optiplex 790; CPU : Core i7-2600, Memory : 4GB) , Kinect for Windows and a television tuner and software (Buffalo DT-F120/U2). For programming, we have used Microsoft Visual C++ 2010, Kinect for Windows SDK v1.7, Kinect for Windows Developer Toolkit v1.7. Fig. 4 shows mouse cursor control system.

Fig. 4. Mouse cursor control system.

4.2. *Operation Screen*

We show a computer operation screen in Fig.5. When our system detects and tracks user's face, it displays image of user's face in the upper right on the window. If our system detects one of 4 face directions of a user (facing up, facing down, facing right, facing left), a region of face direction turns into white, shown in Fig. 6. Then if the user keeps the same direction more than two seconds, the region of face direction turns into yellow, shown Fig 7.

Fig. 5. Compuer operation screen.

Fig. 6. A region of face direction turns into white.

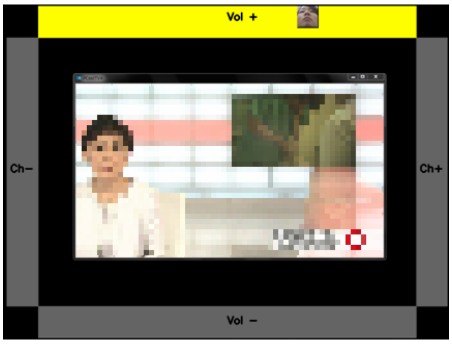

Fig. 7. A region of face direction turns into yellow.

4.3. *Experiments of television control*

We instruct a subject to operate the television and let the subject perform the operation according to the instructions. Subjects are 10 (4 males and 6 females). The evaluation (5 points are perfect) of subjects is as follows.

(i)Arrangement of the region and the design

The evaluation for arrangement of the region and the design is shown in Table 1.

Table 1. The evaluation for arrangement of the region and the design .

Point	5	4	3	2	1	Mean
Frequency	8	4	0	1	0	4.6

(ii)Operability
The evaluation for operability of our system is shown in Table 2.

Table 2. The evaluation for operability.

Point	5	4	3	2	1	Mean
Frequency	3	3	2	1	1	3.6

Comments of subjects are as follows.
(i)Arrangement of the region and the design
(a)Good points

- Reaction of the display when changing the direction of the face is easy to understand.
- It's easy to see, and it's easy to understand.
- It's colored, so it's easy to understand.
- When seeing television screen, a button of the left, right, up or down isn't annoying.

(ii)Operability
(a)Good points

- When a face is recognized once, it reacts accurately.
- It's simple, so it's easy to understand.
- It reacted by a small movement more than I thought.
- A face was turned to the side, but it was possible to look at a screen, so it was good.

(b)Improvement

- The system doesn't detect a face of looking up easily.
- A reaction of looking up wasn't a little good.
- When facing down, it's difficult to look at a screen.

5. Discussions

In our experiments every subject could do operation of a television stably. But after beginning to experiment, a subject wearing glasses couldn't sometimes detect a face.

For a male subject, even if the subject is turning to the front Kinect has often detected it with declining. We setup Kinect which indicates a pitch angle $\varphi = 0°$ and a roll angle $\theta = 0°$ when a female of the height of 160cm faces to the front of Kinect. Therefore a taller male has a high location of the head to a RGB camera of Kinect.

We improved our system as a result of the television operation experiments. The calculation of pitch and roll angles was changed as follows.

We calculate average values of pitch and roll angles, $\overline{\varphi}$ and $\overline{\theta}$, in 30 frames after detecting a face, respectively. We redefine pitch and roll angles, φ' and θ', as follows.

$$\varphi' = \varphi - \overline{\varphi}$$
$$\theta' = \theta - \overline{\theta} \qquad (1)$$

where φ and θ are pitch and roll angles of a face measured by Kinect. For $\varphi' \geq 15°$ and $-15° < \theta' < 15°$ human faces up. For $\varphi' \leq -15°$ and $-15° < \theta' < 15°$ human faces down. For $-15° < \varphi' < 15°$ and $\theta' < -15°$ human faces right. For $-15° < \varphi' < 15°$ and $\theta' \geq 15°$ human faces left. Otherwise human faces the front. Two subjects (1 male, 1 female) operate the television using the improvement system. Subjects' comments are as follows.

- It is easy to operate the television.
- Reactions for faces become good.

As a result of the improvement of our system, a subject operates a television whenever the location of the face to a RGB camera of Kinect was different.

6. Conclusion

We have developed a mouse cursor control system based on face direction using Kinect and applied it to a television operating system. Furthermore, by using the average angle of a subject facing the front, a subject operates a television whenever the location of the face to a RGB camera of Kinect was different.

At present, our system operates only channel and volume changes of a television. In the future, we will add various operations, such as on-off of a power supply and recording operation, to our system. It is also applicable to operation of other application software controlled by mouse cursor, such as a web browser.

References

1. *Skeletal Tracking*, Kinect for Windows SDK v1.7, http://msdn.microsoft.com/en-us/library/hh973074.aspx.
2. *Face Tracking*, Kinect for Windows SDK v1.7, http://msdn.microsoft.com/en-us/library/jj130970.aspx.