Using gaze for 3-D direct manipulation interface

Takaki, Kazuya
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Yonemoto, Satoshi
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

出版情報：Proceedings of the 11th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.282-286, 2005-01
バージョン：
権利関係：

Using gaze for 3-D direct manipulation interface

Kazuya Takaki†, Daisaku Arita†, Satoshi Yonemoto ‡ and Rin-ichiro Taniguchi†

†:Kyushu University, {kazuya_t,arita,rin}@limu.is.kyushu-u.ac.jp
‡:Kyushu Sangyo University, yonemoto@is.kyusan-u.ac.jp

Abstract Use of 3-D human motion sensing without physical restrictions is the most promising approach to realize seamless coupling between virtual environments and the real world. Motion capturing without any specific markers by computer vision techniques is the most appropriate for such purposes. As the first step, we have developed an avatar motion control by user body postures, and we have applied it to 3D object manipulation in virtual environments[1]. However, the biggest problem here is that the intention of the user can not be fully recognized only by human body postures, and that, sometimes, unintentional human motions cause unintended object manipulations and incorrect selection of the target object.

In this paper, we introduce the user's gaze to control 3D Direct Manipulation User Interface. Using the gaze it becomes possible control the user interface more accurately and efficiently.

1 Introduction

Use of 3-D human motion sensing without physical restrictions is the most promising approach to realize seamless coupling between virtual environments and the real world. Motion capturing without any specific markers by computer vision techniques is the most appropriate for such purposes. As the first step, we have developed an avatar motion control by user body postures, and we have applied it to 3D object manipulation in virtual environments[1]. However, the biggest problem here is that the intention of the user can not be fully recognized only by human body postures, and that, sometimes, unintentional human motions cause unintended object manipulations and incorrect selection of the target object. Here, we introduce the user's gaze to control 3D Direct Manipulation User Interface. Using the gaze it becomes possible control the user interface more accurately and efficiently.

2 3D Direct Manipulation User Interface

The motion capture based on computer vision technique requires a lot of PCs and cameras, and we can use the system in the limited area where the system is located. In order that we can use the motion capture in the various places, we have developed the new motion capture system which uses a small number of human body features which can be detected stably. It requires only a PC and two IEEE1394-based digital cameras, and works anywhere. Then we have developed 3D Direct Manipulation User Interface which can make us easy to manipulate the virtual environment, using the motion capture system. Here we introduce the 3D Direct Manipulation User Interface.

2.1 Observing human body postures

From the viewpoint of human posture observation, our method is based on motion synthesis from a limited number of perceptual cues, which can be stably estimated by vision process. We employ skin color regions of a face and both hands in the image as visual features. Fig.2 shows examples of 2D blob tracking result. When a 2D blob is detected in two views, or in multiple views, the 3D position of the blob can be calculated by stereo vision. From the 3D positions of user's head and both hands, using model-based analysis, the user's body posture is estimated and interpreted as an object manipulation command, such as grasping, moving, etc, in the system. Fig.1 shows an outline of our system.

In the virtual environments, basic postures of the avatar, which are acquired by observing human body postures, are represented and interpreted as manipulations of the virtual objects (see Fig.3). In addition, we use virtual scene context as a priori knowledge. We assume that virtual objects in the virtual environment can afford avatar's action by simulating the idea of affordance in the real world. In other words, the virtual environments provide action information for the avatar, such as proper-

Fig. 1: Outline of our system

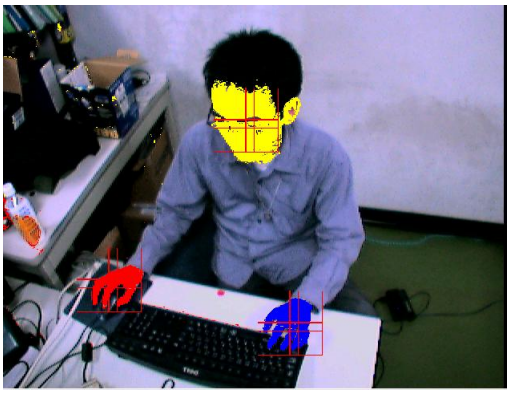

Fig. 2: Detecting the visual features

ties of the virtual objects. An important point here is that we can consider scene constraints in the virtual scene to generate more realistic motion beyond the limitation of the real world sensing. Every task in the manipulation is strongly related to objects in the virtual environments.

3 Interface with Gaze

The user interface uses only three visual features of a human body, it can produce a little and simple manipulation such as grasping, moving etc. To use the gaze information in a user interface, detecting gaze itself is quite important, the gaze information is said to express user's intention generally. Therefore, we discuss using the gaze information for the new input information of the user interface, in order to perform more complicate and advanced manipulations. Here, we introduce how to estimate user's gaze direction and how to use the gaze information in the user interface.

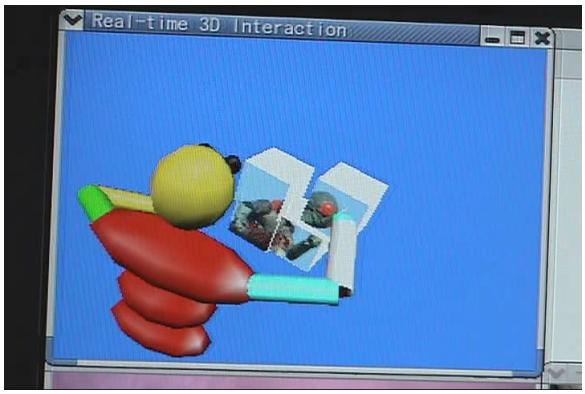

Fig. 3: Interaction scene

3.1 Gaze detection by vision processing

We develop a gaze measuring system based on vision processing, which does not require any attached sensors and which provides a natural way of gaze sensing. We use a stereo camera only for the gaze detection, the processing step of gazing sensing is summarized as follows:

1. Face direction estimation based on image features[2], which is used to transform of face-based local coordinates into the world coordinates for gaze calculation.

2. Detection of irises, which is realized by extracting circular features around possible eye regions .

3. Matching the detected irises with our eyeball model, the rotational parameters of the eyeballs are calculated and, then, the gaze is estimated.

Here, we introduce a method of the gaze detection of our system.

3.1.1 Face direction estimation

We use mouth ends and inner corners of user's eye for the facial feature points, because they are detected easily and stably . At first we detect some ellipses as eye and mouth area from stereo face images. The feature points are located eye or mouth ellipse areas, and we detect them using template matching in the ellipse corners (see Fig.4). The red ellipses in the figure indicate the position of the eye area, and the circles in the figure indicate the position of the irises (white), the mouth ends (blue) and the center of the mouth area (yellow). Next, we calculate the 3D positions of the feature points using

Fig. 4: Detect the feature points

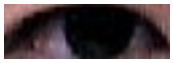

Fig. 5: Detected eye area

Fig. 6: Detected iris

calibrated camera information. The facial planes are constructed by the 3D feature points, and the user face direction is determined to be the normal vector of the facial plane.

3.1.2 Gaze direction and Use's viewpoint estimation

Irises are circular shaped, and we detect irises in the eye areas using ellipse detection at first (see Fig.5). We get edges of dark area using canny filter, and the center of irises from ellipse estimation (see Fig.6).

Then, we estimate the user's gaze direction and viewpoint. We estimate eyeball's 3D positions by fitting a simple eyeball model (see Fig.7). Here, we assume as follows:

- the eyeball is sphere, and its radius is 12mm.
- the center of the eyeball is located on the back of face direction from the center of eye area.
- the center of iris is located on the surface of eyeball.

We can calculate the distance between the center of the eyeball and the center of the eye area easily using a triangle which is made by the center of the eye area, the center of the eyeball and the center of the iris, and we can estimate the center of the eyeball easily. We decide the direction to the center of the iris from the center of the eyeball as the user's gaze direction.

Finally, we estimate user's viewpoint on the display and also view point in the virtual environment, referring to the gaze direction and the position of the display measuring previously. We estimate the virtual viewpoint on the virtual projection plane, and determine the direction to virtual viewpoint from the position of virtual camera as the virtual view direction (see Fig.8).

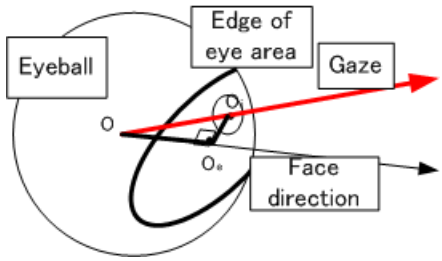

Fig. 7: Eyeball model

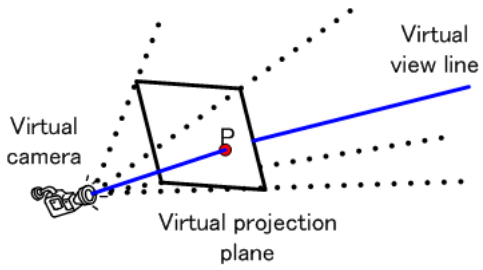

Fig. 8: User's view line on virtual environment

3.1.3 Experimental results

Here, we show some experiments of detecting the gaze direction. We have used two cameras and a PC in this experiment. Cameras used here are IEEE1394-based digital cameras, Sony DFW-V500. We have used the 19 inch LCD display, its resolution is 1280×1024. The user was sitting about 50 cm from the display, he/she have shown nine spheres drawn on the display.

If the accuracy of estimating the gaze direction is higher, the accuracy of estimating the user's intention is more higher, so we have to decrease these error. The reasons why the errors occur are follow:

the error of the 3D position estimation
: It is about 2mm though the radius of the eyeball is about 10mm - 12mm. So, the error occur in the processes of estimating face direction and the position of the eyeball.

the error of the position of the eyeball
: Estimated the face direction is different from the true face direction a little, and decided the radius of the eyeball is also different from true radius of it a little. So, estimated the position of the eyeball is different from true the position of it a little.

3.2 Gaze utilization

To make the interface more efficient and more natural, we consider the following three kinds of gaze

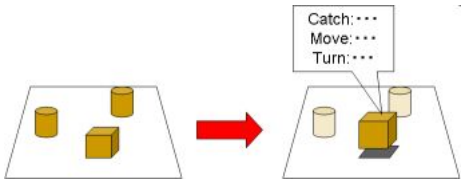

Fig. 9: User help

utilization:

Assistance of manipulation

When we can detect the target object which the user look at, we can attach annotation to the object, which helps the user's manipulation. The possible annotation is as follows (see Fig.9):

- Enhance the visibility of the target object, such as high lighting, enlargement, or translation (pop-up) of the object. When the user look a object, the system recognize as the user want to manipulate the one. And the system enhance the visibility of the target object to the grade with which virtual environment does not fails, in order that the user can manipulate more easy.
- Display supplemental information of the target object, such as properties, possible manipulations, etc. When the system recognizes that the user is confusing by referring to his/her gaze movement, it also displays some assistance information. For example, when the user don't look the objects and the user is confusing (the user's gaze is moving frequently), the system displays the information such as the objects he/she can manipulate. When the user look a object (the user's gaze is not moving) and he/she doesn't perform any action, the system displays the information such as how to manipulate the target object.

Disambiguation

In the interface by human posture, the target object is basically decided by the positions of the human hands. However, only by the hand position, it is not easy for the system to distinguish whether the hand is intentionally approaching the object for manipulation or the hand is accidentally approaching. Usually, the target object is located along the line of sight, and the gaze can give information whether the user is going to manipulate the object. Combining the human body motion and the gaze, we can disambiguate possible interpretation of human actions, and as a result, the number of objects and manipulations which can be handled efficiently is increased.

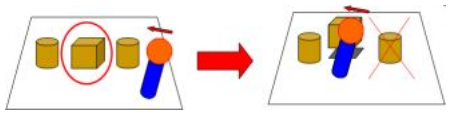

Fig. 10: Disambiguation

The examples of the disambiguation are as follows:

The intention of the target object

When there are the number of objects on the virtual plane, the user may manipulate unexpected objects. If the user's intention what he/she expects is detected, the system targets only the one (see Fig.10).

The intention of the manipulation

When the user want to manipulate an item, the way to manipulate it change by the purpose of the manipulate. For example, when the user want to have a cup of tea, he/she grasps the grip of the cup, but when the user want to see it in detail, he/she may grasp the side of the cup. When the system recognize the intention of the user, he/she can grasp the cup suitably without other recognition such as hand shape.

Concealment of system delay

One of the problems of the vision-based interface is its delay due to vision processing, and the feedback of the human action displayed on a monitor is a bit delayed. This sometimes causes unnatural feeling in using the interface. If we can recognize the user's intention referring to the gaze and if we can predict the user's action, we can compensate the feedback, which can virtually hide the system delay (see Fig.11).

4 Conclusion

Here, we have introduced the user's gaze to a virtual object manipulation interface in order to control the user interface more accurately and efficiently. The gaze was measured by a computer vision technique, which provides a natural way of sensing. Using the

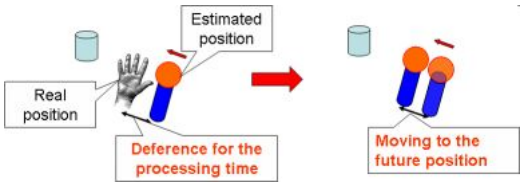

Fig. 11: Process time

gaze, we can achieve (a) disambiguation of objects and manipulations, (b) assistance of input, (c) concealment of the system delay based on the action prediction. Currently, we have just developed a prototypical user interface, and we have to sophisticate it by improving algorithms of vision processing and the user's psychological state recognition.

References

[1] S. Yonemoto et al, "Direct Manipulation Interface with Vision-based Human Figure Control," Proc. of HCI International 2003, pp.811-815, 2003.

[2] S. Ishii et al, "Real-time Head Pose Estimation with Stereo Vision," Proc. of the 9th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.79-97, 2003.

Kazuya Takaki: He received the B.E. degree from Kyushu University, Japan, in 2003. Currently he is a student in the master's course of Kyushu University.

Daisaku Arita: He received B.E. from Kyoto University and M.E. and Ph.D from Kyushu University in 1992, 1994 and 2000 respectively. He is a research associate of the Department of Intelligent Systems in Kyushu University from 1998. His research interest includes real-time parallel computer vision, distant communication and virtual reality.

Satoshi Yonemoto: He received B.E., M.E., PhD from Kyushu University, Japan in 1994, 1996, 1999 respectively. In 1999, he became an research associate of Graduate School Information Science and Electrical Engineering, Kyushu University. In 2000, he joined Kyushu Sangyo University, where he became a lecturer. Since 2004, he has been an associate professor of Department of Intelligent Informatics, Kyushu Sangyo University. His research interests covers image processing, computer vision, computer graphics and human computer interaction.

Rin-ichiro Taniguchi: He received B.E., M.E., PhD in computer science and communication engineering from Kyushu University, Japan in 1978, 1980, 1986 respectively. He became an associate professor of Interdisciplinary Graduate School of Engineering Sciences, Kyushu University in 1989. In 1996, he became a professor of Department of Intelligent Systems, Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests covers image processing, computer vision, human computer interaction and parallel processing.