Vision-Based Face Tracking System for Window Interface: Prototype Application and Empirical Studies

Kotaro Kitajima[†] **Yoichi Sato**[‡] **Hideki Koike**[†]

[†]Graduate School of Information Systems,
University of Electro-Communications
1-5-1, Chofugaoka, Chofu
Tokyo 182-8585, Japan
+81-424-43-5651
{kita, koike}@vogue.is.uec.ac.jp

[‡]Institute of Industrial Science,
The University of Tokyo
7-22-1, Roppongi, Minato-ku
Tokyo 106-8558, Japan
+81-3-3401-1433
ysato@iis.u-tokyo.ac.jp

ABSTRACT

In this paper, we study the effective use of gaze information for human-computer interaction based on a stereo-based vision system which can track the 3D position and orientation of a user in real-time. We have integrated our face-tracking system into the X Window interface system, and conducted experiments to evaluate the effectiveness of our proposed framework for using gaze information for window interfaces.

Keywords

Human-computer interaction, gaze inputs, computer vision, real-time face tracking

INTRODUCTION

The gaze of a user is known to be closely related to the user's intention and attention. For this reason, in the past, there have been a number of studies regarding the use of gaze direction for human-computer interaction.

Unfortunately, however, the success of those previous attempts was rather limited due to several reasons. The main problem was that those attempts were mainly focused on the use of gaze inputs in place of other input devices such as a mouse or a keyboard. Therefore, they suffered from the common difficulty for determining when a user's gaze direction should be considered as a trigger to execute an input command. In this study, we investigate the use of a user's gaze tracked with a real-time vision system by a different approach. To achieve more natural interaction in window interface systems, instead of making users aware that their gaze directions are being used as active inputs, we investigate the use of gazing which is not consciously controlled by the user [1].

OUR FACE TRACKING SYSTEM

We have developed a vision-based face tracking system which can track the position and orientation of a user in real-time. The original configuration of the tracking system is similar to the one proposed by Matsumoto et al. [2]. This system runs on a PC (Pentium3-866MHz, Linux OS) equipped with two HITACHI IP5005 image processing boards. The rotational accuracy of face tracking with this system is approximately 1 degree of errors in a horizontal direction 2 degrees in a vertical direction. Currently, a user's gaze direction is determined based on the orientation of the user's face. We intend to improve our tracking system to the point where it will detect a user's irises in addition to other facial features.

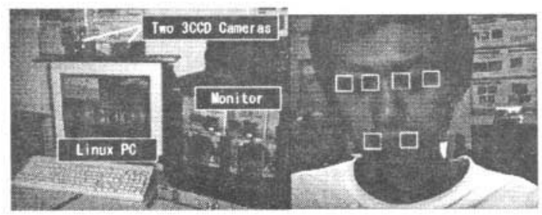

Figure1. The overview of our face tracking system, and face tracking with several facial features

USE OF GAZE DIRECTION FOR WINDOW INTERFACE SYSTEM

For basic operations in window interface systems such as switching a focus among windows, or moving a window, the key issue for using a user's gaze is determining timing for executing those operations. For instance, in some previous studies, users were required to hold their gaze directions still for a certain duration of time to achieve this purpose. This resulted in somewhat unnatural interaction for using window interface systems with a user's gaze.

In this study, we examine a different approach for using gaze directions. Instead of using only a user's gaze, we incorporate gaze inputs with keyboard inputs. More specifically, we use inputs from a keyboard to determine when window operations are executed, while a user's gaze is used for moving or selecting a window. In this way, users need not control their gaze directions intentionally, and their unconsciously moved gaze directions can be used for interaction.

One of the major advantages of the use of gaze inputs in comparison with the use of conventional 2D input devices such as a mouse is that 3D information can be directly controlled. For instance, zooming and scrolling can be directly controlled at the same time with the position and orientation of a user's face. If a 2D input device is used, several steps are required for controlling zooming and scrolling. Thus, we expect that the use of gaze input is particularly advantageous for applications that require 3D inputs.

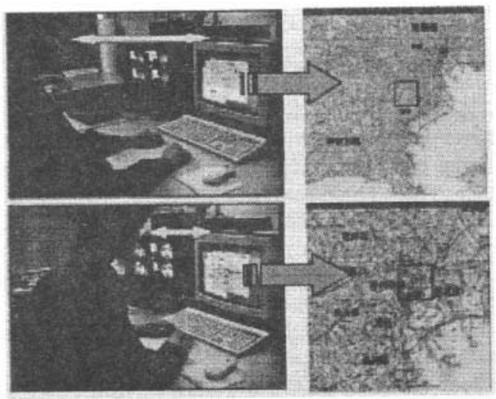

Figure2. Map application

In this study, we used a prototype application of using a map to examine how to use gaze information for 3D inputs in window interface systems. We have tested three different modes for controlling zoom and scroll of a map by using the position of a user's face. In mode 1, the position of a user's face is directly mapped to the scale of zoom of a map. To avoid unwanted change of scale, a user can fix the scale of zoom with a keyboard input. In mode 2, a user can control the scale of zoom in a similar way as with a joystick, as illustrated in Figure 3. If a user's face is closer than a certain threshold distance, the scale of zoom increases at constant speed. Mode 3 is similar to mode 2 except the speed of zoom change varies depending on the distance from a user's face to a monitor.

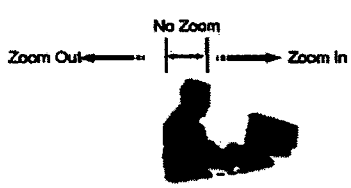

Figure3. Method 2 and method 3 for zoom control

EXPERIMENTS

We have conducted experiments to evaluate the effectiveness of the use of a user's gaze with our proposed approach, and we found the results to be very encouraging. We compared the use of gaze inputs with a mouse for three basic operations in window interface systems: switching a focused window, moving a window, and selecting a window from overlapped windows. For switching a focused window, the use of gaze with our proposed approach was 24.3% faster. Similarly, the test results were 5.4% faster and 29.6% faster than using a keyboard for moving a window and for selecting an overlapped window, respectively.

As for zooming and scrolling of a map application, we examined how long it takes to search for a target on a map by using the gaze inputs with those three different approaches described above. The result is shown in Figure 4 with the average time and the standard deviation for 30 trials with 3 different users.

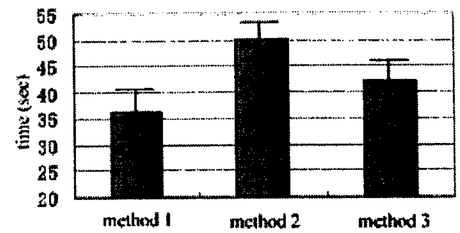

Figure 4. The experiment result

The result showed that method 1 was the fastest among those three modes. Also, test subjects reported that method 1 was the easiest for switching between two different scales of zoom, e.g., a detailed view of a map and an overview of an entire map. This indicates that zoom control with this method is suitable for applications such as a drawing tool where several discrete zoom scales are used frequently. Method 2 resulted in the least satisfactory performance. The main reason was the lack of fine control in this mode. This method was reportedly the most difficult for adjusting the scale of zoom. In this aspect, method 3 was preferred over method 2 since the speed of zoom scale change could be adjusted by the position of the user's face.

CONCLUSIONS

In this paper, we have examined the use of gaze information based on a stereo-based vision system for tracking the 3D position and orientation of a user in real-time. We have integrated our face-tracking system into the X Window interface system, and conducted preliminary studies to evaluate the effectiveness our proposed framework for using gaze information for window interface systems.

REFERENCES

1. R. J. K. Jacob, J. J. Leggettz, B. A. Myers, R. Paucsh, Interaction styles and input/output devices; Behavior and Information Technology, Vol12, No.2, 69-79(1993)
2. Y. Matsumoto and A. Zelinsky: Real-time Face Tracking System for Human-Robot Interaction; *Proc. 1999 IEEE International Conference on Systems, Man, and Cybernetics (SMC'99)*, pp.2-830-2-835 (1999).