MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

REAL-TIME DETECTION OF POINTING ACTIONS FOR A GLOVE-FREE INTERFACE

FUKUMOTO, Masaaki MASE, Kenji SUENAGA, Yasuhito

NTT Human Interface Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa-ken 238-03, JAPAN

ABSTRUCT

This paper presents a human pointing action recognizing system called Finger-Pointer. This system recognizes pointing actions and simple hand forms in real-time by an image sequence processing of stereoscopic TV cameras. The operator does not need to wear any special devices such as Data-Glove. Fast image processing algorithms employed in this system enable real-time processing on a graphic workstation without any special image processing hardware. The system can detect stable and accurate pointing regardless of the operator's pointing style.

INTRODUCTION

With the growing popularity of information systems, we want to develop new interfaces that are easier for everyone to use. The keyboard is the most common computer interface, but it requires a lot of practice to master. We think that a good Human-to-Computer interface for the general population should be so simple to use that no practice is required. It must also allow the operator to communicate in a manner similar to Human-to-Human interaction. Human-to-Human interaction is composed of verbal and non-verbal modes. The role of the non-verbal-mode, which encompasses posture, gesture, gaze, facial expression and so on, is as important as that of the verbal-mode. It is our belief human gesture is suitable for such a computer interface.

Human gestures can be classified into three groups (**Table 1**). The first group contains the simple pointing actions used to indicate location. We call this the 'locator' group. The next group, which we call 'valuator', contains gestures that indicate extents of quantity. For example "about *this* size" or "rotate *this* much". The last group comprises gestures for indicating general images such as "*triangle*" or "*running*". We call this group 'imager'. Some prototype systems of Human-to-Computer interfaces using pointing actions and hand gestures have been proposed [1][2]. In the above classification method, these actions correspond to locators and some valuators. But in these systems, the operator must wear special devices, such as Data-Glove or a magnetic-sensor. Some researches to recognize hand gesture use real-time image processing hardware[3][4].

In our study, we developed a human-pointing action recognition system called "Finger-Pointer"[5]. The system can recognize pointing actions and simple hand forms in real-time without forcing the user to wear any special device such as a Data-Glove. The system can detect stable and accurate pointing regardless of the operator's pointing style by a simple calibration. The operator can interact with the system by any combination of multimodal pointing messages such as gestures and voice commands.

In this report, our "Finger-Pointer" system is outlined first. Then, fast image processing methods for hand image detection are described. Next, new pointing direction determination method called "Virtual Projection Origin (VPO)" is proposed. Experiments have shown that VPO is very effective in various situations. Finally, remaining problems and possible applications for this system are discussed.

"FINGER-POINTER"

Finger-Pointer system

Fig.1 shows a block diagram of Finger-Pointer system. The operator's pointing actions are captured by two stereoscopic TV cameras, one mounted on the wall and the other on the ceiling. The system determines the coordinates of the operator's finger tip and the direction in which the operator's finger is pointing by analyzing these camera images. The system then determines the target location and displays a corresponding cursor on the screen.
The system works on a GWS[1] and processes 10 frames per second. The user-specified, single-word-type voice recognition unit [2] on a personal computer [3] is used for voice command recognition. Another GWS[4] is used as the application platform. The system uses a telescopic type microphone, making it unnecessary to wear even a headset.

[1]IRIS-4D/220GTX
[2]Voice Navigator
[3]Macintosh IIfx
[4]Personal IRIS

Table 1: Classification of gesture

Class	Content	Example
locator	Indicate location in space	pointing
valuator	Indicate extents (including **switcher**)	manipulation of 3D-CAD, hand sign
imager	Indicate general images	sign language, body language

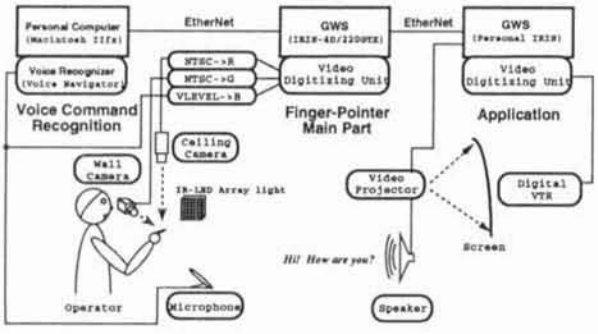

Fig. 1: System Diagram of Finger-Pointer

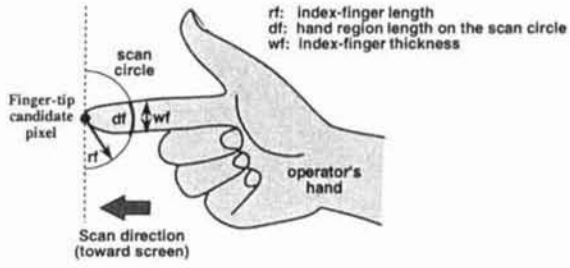

Fig. 2: Determining of finger tip location

Capturing multiple images

The system uses the two images captured by the ceiling and wall cameras but first they must be synchronized to accurately detect the location of the operator's finger tip. The video digitizing unit[5] in GWS can capture only one RGB color image at a time. In this system, we use two monochrome CCD cameras, and these cameras are synchronized by one sync signal. The ceiling camera image is converted to the "R" plane of the digitizing unit, and the wall camera image is converted to the "G" plane. These camera images are then digitized by the video digitizing unit. Thereafter, the two planes are easily separated by a simple memory operation, and the system can capture two separated camera images simultaneously. The operator's voice level is also recorded in "B" plane for voice and gesture synchronization.

Infrared LED array

Captured camera images are converted to binary images for finger tip detection. In this process, a fixed threshold is used for real-time processing. However binarization by using a fixed threshold is affected by lighting conditions. A strong visible light is effective for stable binarization, but such a light makes the operator feel uncomfortable. This problem can be solved by using an array of infrared LEDs whose light is unnoticeable to the operator. A filter for cutting visible spectrum is positioned in front of the CCD cameras. With this combination of infrared LEDs and filters, the system can stably binarize by a fixed threshold regardless of the lighting condition of the room.

IMAGE PROCESSING METHODS

Determining finger tip location

This system uses fast image processing methods to detect human pointing actions and simple hand forms in real-time. **Fig.2** illustrates the algorithm for determining finger tip location. The system makes one assumption; the operator's finger tip is the part of his body nearest the screen while he is pointing. First, the captured camera images are binarized with fixed threshold. Next, the system scans the binary images and determines the pixel that is closest to the screen , as the most likely candidate for the finger tip. The system then decides, based upon the length and thickness of the extracted region, whether that pixel represents the real finger tip. The length and thickness of usual human fingers are not drastically different. Therefore, no calibration is done when using this method.

[5]Live Video Digitizer

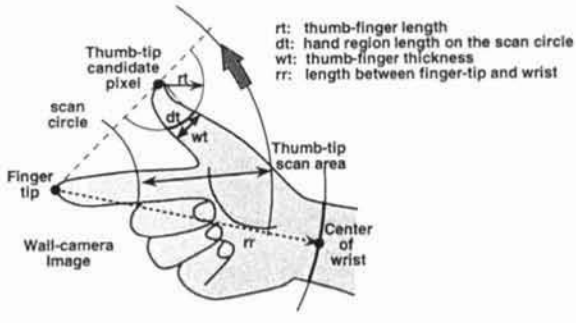

Fig. 3: Thumb-Switch Detection

Thumb-Switch detection

The system can detect thumb-up and thumb-down positions, which allows the thumb to be used as a switch. For example, the operator can use a click and drag function, similar to that in a one-button mouse.

The scanning method is similar to finger tip detection (**Fig.3**). The system scans the binarized wall-camera image in a motion similar to a spreading fan from the line determined by the operator's finger tip and wrist position. Then the system determines the thumb tip candidate pixel and decides whether the candidate pixel is the real thumb tip.

Finger-Number detection

People often use their fingers to indicate numbers. We call this "Finger-Number". The system can also recognize the number of outstretched fingers. The operator can select different commands by displaying different numbers of fingers. This feature allows more speedy selection than pointing to icons.

The recognition sequence is shown in **Fig.4**. The detection method first extracts the hand region on the scan circle of the binarized wall-camera image. Next, the system de-

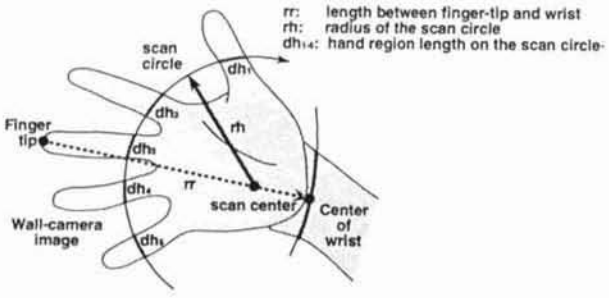

Fig. 4: Finger-Number Detection

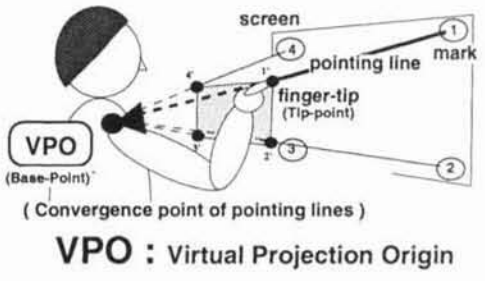

Fig. 5: VPO Calibration

cides how many fingers occupy each extracted region. Using this method, the system can detect the correct number of fingers, even if two or more fingers are held together.

Finger tip tracking

A finger-tip-tracking method is implemented in this system for speedy processing. The position where the finger will next appear is estimated by the two most recent finger tip positions. Then a small area (about 8% of captured image) is scanned, and if the operator's finger tip is included in this area, the system can detect the finger tip candidate pixel quickly.

VPO (VERTUAL PROJECTION ORIGIN)

Pointing direction

The operator's pointing direction is determined by a straight line which is defined by two points in 3-D space. These two points are called "Tip-Point" and "Base-Point". The Tip-Point corresponds to the operator's finger tip. However, the question is the location of the Base-Point. If a fixed Base-Point is located independent of the operator and the manner of pointing style, the system only needs to track two specified points in order to detect the operator's pointing direction. However, a preliminary experiment indicated that the position of the Base-Point is different for each operator. Even for the same operator, this point changes depending on the pointing gesture, for example, whether it is tense or relaxed.

"VPO": Virtual Projection Origin

To overcome the problem of differing Base-Point positions, we assume that the lines of the pointing direction converge at one point when an operator points at objects on a distant screen (**Fig.5**). We call this point the "VPO" – Virtual Projection Origin.

The VPO calibration procedure requires the operator to point to a few predetermined marks on the screen. This calibration decides pointing lines from the displayed mark and the corresponding finger tip position. The VPO is then estimated as the point where these lines converge. To put it concretely, the VPO is a center of sphere that has minimum radius and is intersected by all pointing lines (**Fig.6**). After the VPO is estimated, the operator's pointing direction is determined as the projection from the VPO to the operator's finger tip, and the target is determined as the intersection of the pointing direction and the screen.

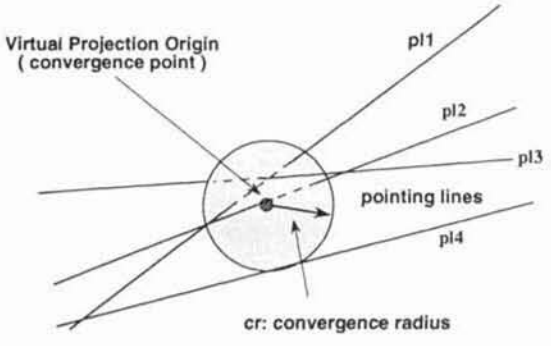

Fig. 6: Estimating convergence point

VPO distribution

Fig.7 illustrates typical VPO distributions. Each sphere indicates the VPO position for one operator. The radius of each sphere indicates the convergence rate. A small sphere means good convergence. This figure indicates the VPO position is different for each operator, and even for the same operator, the VPO position changes depending on the pointing style.

The experiment using 20 operators provided that the VPO for each operator converges within a 3.5-cm radius with a probability of 95%. By using the VPO method, the system has a pointing accuracy of 2.0° in the non cursor feedback mode, and 0.6° in the cursor feedback mode.

APPLICATIONS

Presentation system

There are many applications for the "Finger-Pointer" system. **Fig.8** illustrates a presentation system that uses a computer-based slide projector. Rectangular regions at the bottom of the screen serve as command buttons, for example, NextPage, PrevPage, ClearScreen, etc. The operator can select commands and emphasize the slide image by adding marks and lines. The operator can control the

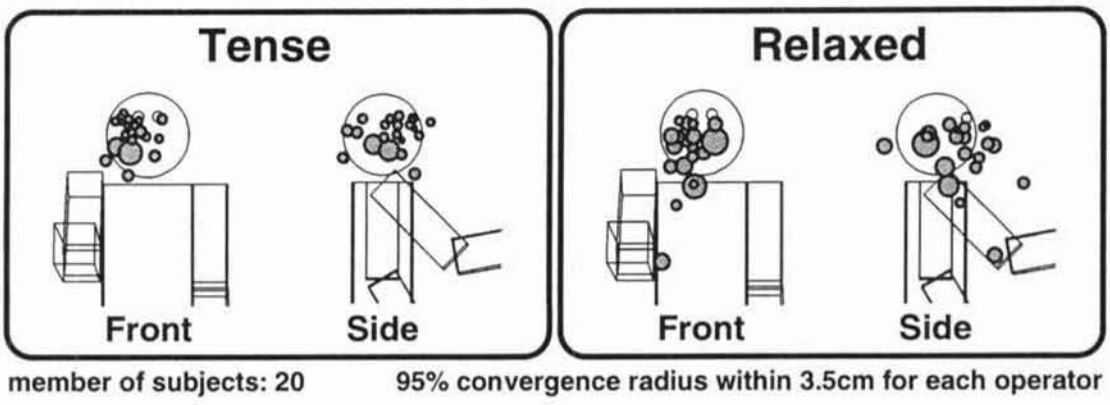

Fig. 7: VPO Distribution

system using any combination of multimodal pointing messages such as gestures and voice. Moreover, the operator can select commands using combinations of finger-number and voice commands. In this case, finger-number is used to indicate the argument of a command.

Another applications

"Finger-Pointer" can also be used as a video browsing system. The operator can use hand motions and thumb-switch actions to control a VCR. For example, Play, Stop, and some special search operations. In another application, the system can detect alphanumerics and graphic figures written in space by the operator. The pen-up/down operation is controlled by the operator's thumb-switch.

CONCLUSION

In this report, we proposed a new pointing action recognition system called "Finger-Pointer". The operators need not wear any special devices such as Data-Glove. The system can detect the 3D position of the finger tip, pointing action, thumb clicking and the number of shown fingers in real-time through a simple and fast image processing method. The operator can interact with the system by any combination of pointing messages, such as gestures and voice commands. By estimating VPO (Virtual Projection Origin) through a simple calibration step, the system can detect stable and accurate pointing regardless of the operator's pointing style.

The experimental Finger-Pointer system was implemented on an IRIS-4D/210GTX without any special image processing hardware. It processes 10 frames per second and has 2.0° (without cursor feedback) and 0.6° (with cursor feedback) pointing accuracies. This smooth and natural interface has been tested for some applications: a presentation system , a VCR browser and so on. This system is also useful as a real-time platform of a multi modal interface.

Improvement of processing speed together with pointing accuracy and the recognition of more complex hand gestures in any place are still remaining as future problems.

Acknowledgements

The authors wish to thank Dr. Takaya Endo, Dr. Kazunari Nakane, Dr. Yukio Kobayashi, and the members of the Human & Multimedia Lab and the Speech & Acoustics Lab for their encouragement and valuable discussions.

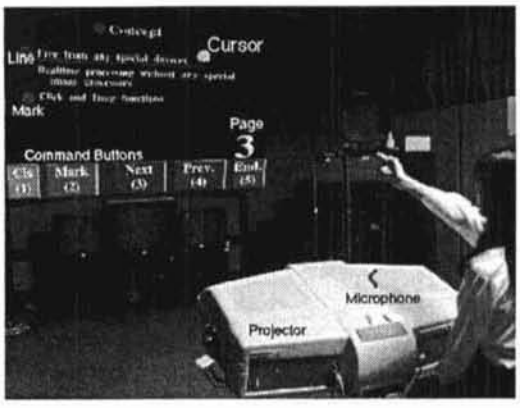

Fig. 8: Presentation System (imposed screen image)

References

[1] R.A.Bolt: "Put-That-There: Voice and Gesture at the Graphics Interface", ACM-SIGGRAPH, **Vol.14**, No.3, pp.262-270, April (1980).

[2] D.Weimer, S.K.Ganapathy: "A Synthetic Visual Environment with Hand Gesturing and Voice Input", CHI'89 Proceedings, pp235-240, May (1989).

[3] W.Wiwat, Ishizuka: "A Visual Interface for Transputer Network (VIT) and its Application to Moving Analysis", 3rd Int. OCCAM Conf., (1990).

[4] Ishibuchi, Takemura, Kishino: "Real-Time Hand Shape Recognition Using Pipeline Image Processor"(In Japanese), IEICE Technical Report, HC92-14, (1992).

[5] Fukumoto, Mase, Suenaga: "Finger-Pointer: A Glove Free Interface", CHI'92 Interactive Poster, Monterey CA, May (1992).