A HAND GESTURE INTERFACE DEVICE

Thomas G. Zimmerman, Jaron Lanier,
Chuck Blanchard, Steve Bryson and Young Harvill

VPL Research, Inc.
656 Bair Island Road, Suite 304
Redwood City, CA 94063

ABSTRACT

This paper reports on the development of a hand to machine interface device that provides real-time gesture, position and orientation information. The key element is a glove and the device as a whole incorporates a collection of technologies. Analog flex sensors on the glove measure finger bending. Hand position and orientation are measured either by ultrasonics, providing five degrees of freedom, or magnetic flux sensors, which provide six degrees of freedom. Piezoceramic benders provide the wearer of the glove with tactile feedback. These sensors are mounted on the light-weight glove and connected to the driving hardware via a small cable.

Applications of the glove and its component technologies include its use in conjunction with a host computer which drives a real-time 3-dimensional model of the hand allowing the glove wearer to manipulate computer-generated objects as if they were real, interpretation of finger-spelling, evaluation of hand impairment in addition to providing an interface to a visual programming language.

Resume

Cet article presente le developpement d'un systeme main-machine qui fournit en temps reel les positions, orientations et gestes de la main de l'utilisateur. L'element de base du systeme est un gant, l'ensemble du systeme utilisant differentes technologies: Des capteurs analogiques places sur le gant mesurent les flexions des doigts. La position et l'orientation de la main sont mesures, soit a l' aide d'un systeme ultrasonique, qui fournit cinq des degres de liberte du systeme, soit a l'aide d'un systeme magnetique, qui fournit es six degres de liberte. Des elements piezoceramiques places sur le gant sont controles pour fournir a l'utilisateur differentes sensations tactiles. Ces differents capteurs sont montes sont un gant de faible poids, connecte au systeme de controle par un cable de faible dimension.

Les applications du systeme comprennent son utilisation relie a un ordinateur contenant un modele temps reel en trois dimensions de la main permettant a l'utilisateur de manipuler des objets virtuels generes par l'ordinateur comme s'ils existaient reelee-ment, l'interpretation de signes alphabetiques pour mal-entendants, l'evaluation clinique de troubles fonc-tionnels de la main, ainsi que son utilisation comme interface a un langage de programmation visuelle.

Keywords: Human Interface, User Interface, Motor Interface, Tactile Interface, Gesture Recognition.

1. INTRODUCTION

The hand gesture input devices presented here, the Z-Glove™ and the DateGlove™, are lightweight cotton gloves containing flex sensors which measure finger bending, positioning and orientation systems, and tactile feedback vibrators.

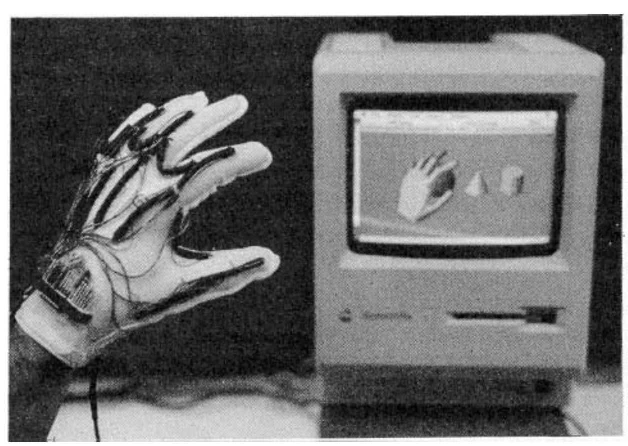

Figure 1.
DataGlove with outer glove removed to show sensors

Different orientation and positioning systems of the Z-Glove and the DataGlove distinguish the two models. The Z-Glove employs an ultrasonic positioning and orientation system controlled by a Commodore 64™ computer and is less costly and more limited in application. The DateGlove, using a magnetic positioning and orientation system controlled by an Apple Macintosh™ computer is more expensive and of wider application (see Fig. 1).

2. ELEMENTS OF THE SYSTEM

2.1 FLEX SENSORS

On both gloves, the flex sensor is a patented [Zim85] optical goniometer that can be manufactured in a variety of dimensions suited to the application. Flex sensors measuring two inches long by one quarter inch in diameter are glued to a tight-fitting, stretchable inner glove. Multiplexing electronics are mounted on top of the inner glove and protected by a loose-fitting outer glove.

Five to fifteen flex sensors are mounted on a glove, depending on the application. The finger joints measured are the metacarpophalangeal (MP or inner) joints, the proximal interphalangeal (PIP or middle) joints and their abductions, and the palm (see Fig. 2).

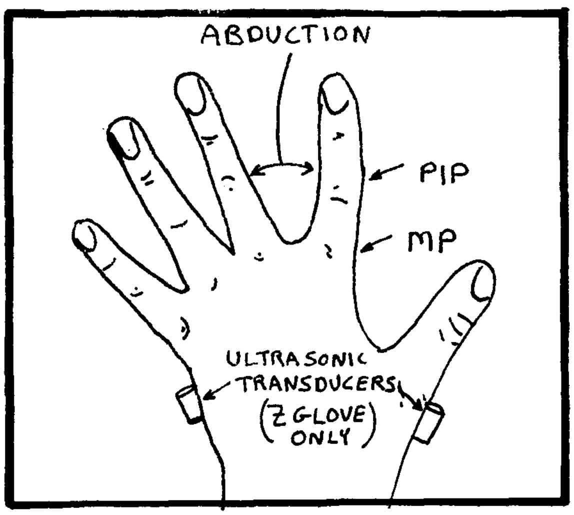

Figure 2
Flex and ultrasonic transducer placement

On the DataGlove, analog signals from the flex sensors are multiplexed to data-acquisition hardware which interfaces to an Apple Macintosh computer. On the Z-Glove, analog flex signals are fed directly to the internal converters of the Commodore 64 computer.

2.2 POSITIONING AND ORIENTATION

2.2.1 ULTRASONIC POSITIONING AND ORIENTATION

The positioning and orientation system on the Z-Glove consists of two ultrasonic transducers attached to opposite sides of the metacarpal (between the wrist and the fingers), a hardware timer, three ultrasonic receivers mounted around the perimeter of a monitor, and triangulation software running on a Commodore 64 computer. Processing overhead is reduced by using triangulation approximations. Visual feedback is provided by an icon which tracks hand movement, minimizing the impact of tracking errors.

Roll and yaw of the hand is determined from the position of the transmitter pair. The ultrasonic transmitters need a direct line-of-sight for their acoustic signals to reach the receivers. If one transmitter is blocked (by the fingers or part of the hand), orientation information is lost. If both transmitters are blocked, the tracking icon freezes.

Position jitter due to detection error is reduced, with minimum impact on response time, by means of a software filter that decreases the filtering with increasing hand velocity.

2.2.2 MAGNETIC POSITION AND ORIENTATION

On the DataGlove, a 3SPACE™ (Polhemus Navigation Sciences Division, McDonnell Douglas Electronics Company, Esses Junction, Vermont) position and orientation system is used. The 3SPACE uses low frequency magnetic fields to measure six degrees of freedom. The small 3SPACE sensor is mounted on the dorsal side of the hand between the glove's two layers. The 3SPACE is connected to one of the serial ports on the Apple Macintosh computer. The 3SPACE requires no filtering of data.

2.3 TACTILE FEEDBACK

Tactile feedback is used to add realism to interactions between computer-generated (virtual) objects and the virtual hand. Tactile cues are being experimented with to simulate object contact, hardness and surface texture.

Piezoceramic benders driven with a 20-40 Hz sine wave, (below peak sensitivity of 250 Hz [Sherrick82] to prevent audible sound generation) are mounted underneath each finger. The sensation produced is "tingling" or "numbness".

Benders are used because of their small size, low cost and low operating voltage. Frequency modulation is used to vary the intensity of the tactile sensation and to minimize the finger "numbing" sensation. Tactile stimulation is increased by driving benders closer to peak tactile sensitivity frequency. Object contact is cued when virtual fingertips touch the surface of virtual objects. Contact is signaled by oscillating finger benders, which produce a "buzzing" sensation at the fingertips.

3. GESTURE RECOGNITION AND CALIBRATION

Representing and recognizing human hand gestures is a deep problem analogous to the recognition of human speech or handwriting. We present here some of the basic methods of gesture recognition.

It is possible to identify two types of gestures: object manipulations (eg: pick up, rotate, throw, squeeze), and commands (eg: draw a line, produce a sound, set a color). Gestures can also be classified as static (eg: a "victory" or a "peace" sign), or dynamic (eg: waving good-bye).

Z-Gloves and DataGloves are manufactured in several sizes taking into account the wide variety of human hands. Individual gesture styles further complicate gesture recognition. Two techniques of calibration have been developed to compensate for these variations.

3.1 MANUAL CALIBRATION

Manual calibration is a method in which a subject wearing the Z-Glove performs tasks. One manual calibration method requires the Z-glove wearer to grab a series of objects placing finger joints at known angles. The joint bending angles recorded in this manual calibration are used in a clinical study of hand rehabilitation.

A finger spelling interpretation application uses manual calibration to form gesture templates. Gesture recognition is performed by differentiating finger values with values from such gesture templates. The absolute value of these differences is summed for each gesture template. The gesture with the minimum sum is chosen if that sum is less than a confidence threshold. If the gesture does not match any of the templates closely enough then the gesture is ignored.

3.2 AUTOMATIC CALIBRATION

Gesture recognition using automatic calibration scales flex sensor values by recording mimimum and maximum joint extension values. The maximum values are decremented over time to prevent an extreme joint extension from skewing the scaling of recorded values. Hysteresis thresholds unique to each finger are applied to reduce the position of the fingers to quantized states. These finger states are then compared to gesture templates containing permissable finger states and the first successful gesture template match is chosen, or "recognized". Threshold values and gesture templates are user-independent and are empirically derived.

Unintentional passing (transitional) gestures are "debounced" by requiring a valid gesture to be held for a period of time.

Dynamic gesture recognition techniques are being investigated. As in speech recognition, dynamic gesture recognition is able to take advantage of context in order to limit the number of gestures to be distinguished at a given time.

4. A VISUAL PROGRAMMING LANGUAGE INTERFACE

A Z-Glove based user interface is explored in a visual programming environment known as Grasp™ (originally called Mandala) under development at VPL Research, Inc. [Lanier84]. The Grasp system runs on microcomputers and uses the ultrasonic positioning and orientation system.

A moving hand-shaped icon tracks the user's hand. When a valid gesture is detected, the hand image is shown performing that gesture. This produces a discrete animation of a hand, rather than a continuous representation of the hand.

In most mouse (or other pointing device) based interfaces, there are two phases to each user action: selection of an object, and selection of the operation to be performed on the selected object. In Grasp, object selection and operation are accomplished simultaneously by gesturing over an object.

In order to make gestures easier to learn and remember, gestures in Grasp are analogous to real-world gestures. A "grab" (all fingers closed like a fist) picks up an object. Once picked up, the object can be carried around the screen and "dropped" (all fingers opened) at a new location. If the user merely opens a few fingers, a copy of the object is put down. If the user picks up an object with only the thumb and index finger, a value is "plucked" from the object. Opening fingers quickly over an objects opens or expands the object.

5. APPLICATIONS

Among the present applications of the DataGlove and the Z-Glove are a gesture recognition device, a clinical tool for evaluating hand function, a three-dimensional hand model controller, an interface to a visual programming language, a music and sound synthesis controller, a finger spelling interpreter, and a computer-generated object manipulator. Future projections for DataGlove and Z-Glove applications lie in the fields of robotics, human factors and ergonomics research.

5.1 THREE-DIMENSIONAL HAND MODEL

A flexible, three-dimensional, articulating hand model is constructed using local coordinates for each finger joint, represented by a linked list of records, with each finger having a full range of angular motion. The hand model (virtual hand) is generated in real time using the DateGlove and the Apple Macintosh computer (see photo).

5.2 A CLINICAL HAND IMPAIRMENT MEASURING TOOL

Hand impairment measurements are classified as anatomical and functional. A goniometer obtains anatomical measurements by measuring the range of motion of a given joint. Functional tests measure the time it takes to perform common unilateral tasks like stacking checkers, turning over cards, or putting small objects in a can [Jebsen69].

The process of measuring the range of motion of a patient's hand by a skilled therapist with a mechanical goniometer can take one to two hours and is only repeatable to within five degrees if the same physical therapist with a mechanical goniometer performs the measurements [Rosen86]. The DataGlove has the capability of measuring a patient's range of motion in a fraction of the time, under the supervision of a less skilled assistant, with more repeatable results. Tests are now being conducted to examine this application.

Functional tests like the Jebsen Hand Function Test take a considerable amount of time for the patient and the physical therapist. Dynamic data recorded on video tape is also time consuming to analyze. The DataGlove, with a positioning and orientation system, is being investigated as a means of logging and analyzing functional tests.

5.3 THREE-DIMENSIONAL OBJECT MANIPULATION

Manipulating three-dimensional virtual objects with two-dimensional controllers such as digitizing tablets, touch pads and mice are awkward since these objects are capable of six-dimensional movement. Both the Z-Glove and DataGlove (with a positioning and orientation system) allow users to interact with virtual objects much as they do with real objects. Virtual objects can be picked up, grabbed, twisted, squeezed, thrown, and set down.

A virtual object can be moved (translated and rotated) with either the Z-Glove or the DataGlove while it is being operated on (eg: squeezed and twisted). And since simultaneous actions are usually faster than sequential ones [Buxton86], the entire function can be performed in a more natural, more efficient manner.

The Z-Glove and the DataGlove address the question of "scoping", the selection of a subset from a set. This is significant when several objects appear near or on each other. Consider a computer-generated image of cherries in a glass bowl viewed from the side. When using a mouse to point at a cherry, it is ambiguous whether one is pointing to the cherry or the bowl. In contrast, one could use the DataGlove or the Z-Glove to close one's fingers around and pick up the desired item, without ambiguity. A handful of cherries, for that matter, could be picked up, the number of cherries in the handful being a function of the size of the grasp.

A virtual environment project at NASA Ames Research Center [Fisher86] uses a head-mounted, three-dimensional display [Sutherland68], a DataGlove, and a 3SPACE TRACKER in current research directed toward eventual control of a remote robot hand. The hand performs operations in an environment hostile to human life (in space or under conditions which, due to radiation or other factors, is dangerous), while the operator remains protected and comfortable (inside a space capsule or safe enclosure).

6. CONCLUSION

Two systems, the DataGlove and the Z-Glove, have been presented both of which allow the direct manipulation of computer-generated objects.

The best type of interface device performs its task unobtrusively. A joystick is a controller that a user acts upon. The DataGlove and the Z-Glove, on the other hand, are articles of clothing which instrument the user's actions. The user's hand is the controller, as is natural for it to be.

Just as speech is our natural means of communication, the human hand is our natural means of manipulating the physical world. As computer systems begin to simulate the physical world, the technologies presented in this paper suggest a broad spectrum of possibilities to a wide variety of users. It is increasingly important that we shape the simulated world of the computer in ways which reflect our human universe, rather than allow ourselves to be shaped by our machines.

ACKNOWLEDGEMENTS

The authors would like to thank VPL Research, Inc. for its support, Sam Wantman for early brainstorms on the Z-Glove, Joseph M. Rosen, M.D. of the Stanford University Medical Center, and Scott Fisher of NASA Ames Research Center for their research contributions, and Marvin M. Zauderer for reviewing drafts of this paper.

REFERENCES

[Buxton86] Buxton, William and Myers, Brad A. A Study in Two-Handed Input, ACM SIGCHI 86 Conference Proceedings, pp.321-326, April 1986

[Fisher86] Fisher, Scott S. Telepresence master glove controller for dexterous robotic end-effectors, SPIE Cambridge Symposium on Optical and Optoelectronic Engineering, 26 October 1986, Cambridge, Mass.

[Jebsen69] Jebsen, R.H. et al, Archives of Physical Medicine and Rehabilitation 50: pp. 311-316, June 1969

[Rosen86] Rosen, Dr. Joseph M., Personal conversation, September 1986

[Lanier84] Lanier, Jaron, Scientific American, cover, September 1984

[Sherrick82] Sherrick, Carl E., and Craig, James C., Tactual Perception: A Sourcebook, (Schiff, W. and Foulke, E., Editors), Cambridge: Cambridge University Press, 1982

[Sutherland68] Sutherland, Ivan E. A Head-Mounted Three-Dimensional Display, Fall Joint Computer Conference 1968, pp. 757-764

[Zim85] Optical Flex Sensor, September 17, 1985, United States Patent Number 4,542,291