A Survey of Glove-based Input

David J. Sturman
Medialab

David Zeltzer
MIT Research Lab of Electronics

Clumsy intermediary devices constrain our interaction with computers and their applications. Glove-based input devices let us apply our manual dexterity to the task.

Peter Menzel

Our primary physical connection to the world is through our hands. We perform most everyday tasks with them. However, when we work with a computer or computer-controlled application, we are constrained by clumsy intermediary devices such as keyboards, mice, and joysticks. Little of the dexterity and naturalness that characterize our hands transfers to the task itself.

In an effort to change this, people have been designing, building, and studying ways of getting computers to "read" users' hands directly, free from the limitations of intermediary devices. The development of electronic gloves has been an important step in this direction. The commercialization and widespread availability of devices such as the VPL DataGlove and the Mattel Power Glove has led to an explosion of research and development projects using electronic gloves as interfaces to computer applications and computer-controlled devices. The applications span fields as diverse as telemanipulation, virtual reality, medicine, scientific visualization, puppetry, music, and video games.

In this article we provide a basis for understanding the field by describing key hand-tracking technologies and applications using glove-based input. The bulk of development in glove-based input has taken place very recently, and not all of it is easily accessible in the literature. We present here a cross-section of the field to date.

Hand-tracking devices

It could be said that the history of tracking devices for mechanically or electrically interpreting hand motions began with post-WWII development of master-slave manipulator arms, or even earlier during the Renaissance with the development of the pantograph. However, we begin with developments at the Massachusetts Institute of Technology in the 1970s. At that time, researchers at the MIT Architecture Machine Group were demonstrating general-purpose computer input based on direct interpretation of hand motion. The "Put-that-there" project[1] used the newly commercialized Polhemus 3-space tracking sensor to communicate the user's hand position to the computer. The Polhemus, now in widespread use, works by radiating a pulsed magnetic field from a stationary source. Companion sensors, which can be attached to any object (such as the hand), report their 3-space position and orientation relative to the source. By attaching the Polhemus sensor to the user's hand, the MIT researchers knew exactly where the user was pointing on a large wall display. They used this information to let the user indicate graphical elements of interest, move them from point to point on the screen, and query the contents.

Since then, a variety of technologies have been used to capture mechanical and gestural information from the hand. We've divided these into position tracking, which uses optical, magnetic, or acoustic sensing to determine the 3-space position of the hand, and glove technologies, which use an electromechanical device fitted over the hand and fingers to determine hand shape.

Position tracking

Hand position is characterized by the location of the hand in space and the orientation of the palm. The three technologies used predominantly to track the position of the hand are optically based, using cameras to examine the hand from a distance; magnetically based, such as the Polhemus described above; or acoustically based, using triangulation of ultrasonic "pings" to locate the hand.

Optical tracking

There are two common methods of optical tracking. The first puts small markers on the body, either flashing infrared LEDs or small infrared-reflecting dots. A series of two or more cam-

0272-17-16/94/0100-0030$03.00 © 1994 IEEE

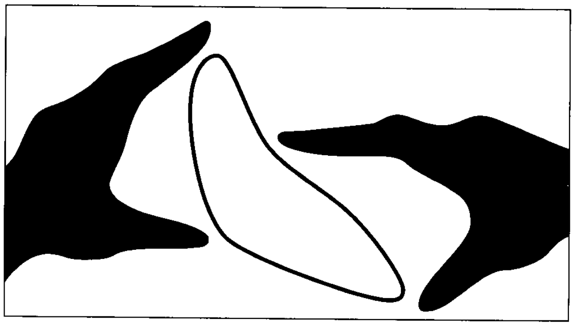


Figure 1. Manipulating graphics by hand: Fingertips control a spline curve. (Based on a drawing by Krueger.[2])

eras surround the subject and pick out the markers in their visual field. Software correlates the marker positions in the multiple viewpoints and uses the different lens perspectives to calculate a 3D coordinate for each marker. The second method uses a single camera to capture the silhouette image of the subject, which is analyzed to determine positions of the various parts of the body and user gestures.

Marker systems

Biomechanics labs and rehabilitation clinics have long used synchronized infrared LED systems (such Selspot, Op-Eye, and Optotrak) and reflective marker systems (such as Elite and Vicon) to analyze the motion of the body and limbs. One limitation of these systems is the processing time needed to analyze the several camera images and determine each marker's 3D position. Most of the systems operate in a batch mode, where the trajectories of the markers are captured live, followed by a period of analysis to calculate 3-space positions from the 2D images (the Optotrak is an exception, performing these two steps in real time). With LED systems, the LEDs are sequenced so that only one lights up at a time. However, the reflective marker systems require a middle stage of analysis to identify markers and resolve ambiguities when markers coincide in the visual field (the more cameras, the smaller this problem). The time that it takes to sequence all the LEDs or to perform the two stages of analysis limits the real-time capabilities of these systems and the number of markers that can be used simultaneously.

In the past several years, these systems have seen wide use in recording human motion for computer animation. However, their real-time limitations and an inability to resolve markers that are too close together restricts their use for tracking fingers in interactive applications.

Silhouette analysis

For more than two decades, Myron Krueger has been constructing systems to allow natural interaction with computers, free of encumbering equipment or interface devices.[2] By processing silhouette images with custom hardware, he can analyze complex motions in real time. His techniques successfully discriminate parts of the body such as head, legs, arms, and fingers.

In one example application, participants can draw figures with their fingers. When the computer sees that the thumb and index finger are outstretched on both hands, it draws a curve that inscribes the region between them (see Figure 1). Moving the hands or fingers changes the size and shape of the curve. A rapid pull away from the curve fixes it in place on the screen.

Krueger has developed a whole array of example interactions and games that he has integrated into his system, Videoplace. Videoplace is on permanent exhibit at the University of Connecticut and is occasionally featured in special exhibitions.

Borrowing ideas from Krueger's work, Pierre Wellner at Rank Xerox EuroPARC has developed DigitalDesk, a normal office desk with papers, pencils, coffee cups, and so forth, onto which a computer can project electronic documents and applications (such as calculators or spreadsheets).[3] A computer camera observes the worker's hands and fingers on the desk and determines when the user points to or gestures above a real or projected object. Thus, with their fingers users can operate a computer projected calculator, indicate an electronic text item to delete, or outline a paragraph on a physical document on the desk for the computer to scan. By integrating computer images with real objects and allowing the same free-form, "deviceless" interaction with both, DigitalDesk moves us toward a world of more natural interactions with computers.

Image-based visual tracking of the hands has several general problems:

1. The resolution of conventional video cameras is too low to both resolve the fingers easily and cover the field of view encompassed by broad hand motions.
2. The 30- (or 60-) frame-per-second conventional video technology is insufficient to capture rapid hand motion. (Infrared systems, such as Selspot or Optotrak, can operate above 300 Hz, and special-purpose high-speed video cameras are available, but conventional video cameras are limited to 60 Hz.)
3. Fingers are difficult to track, as they occlude each other and are occluded by the hand.
4. Computer vision techniques are not sufficiently mature to interpret complex visual fields in real time.

For these reasons, researchers have turned to glove-based and other mechanical systems for practical monitoring of hand motion.

If the performance of camera-based systems improves to the point that they can track individual fingers while maintaining a large visual field, operate in real time, and work without special clothing or encumbering devices, we think certain applications will return to this method of capturing hand motions.

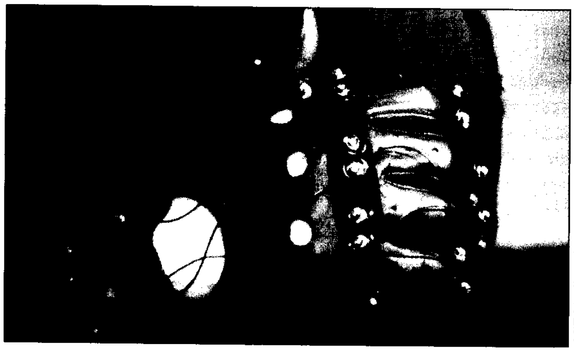
Magnetic tracking

As described above, magnetic tracking uses a source element radiating a magnetic field and a small sensor that reports its position and orientation with respect to the source. Competing systems from Polhemus and from Ascension Technologies provide various multi-source, multi-sensor systems that will track a number of points at up to 100 Hz in ranges from 3 to 20 feet. They are generally accurate to better than 0.1 inches in position and 0.1 degrees in rotation. Magnetic systems do not rely on line-of-sight observation, as do optical and acoustic systems, but metallic objects in the environment will distort the magnetic field, giving erroneous readings. They also require cable attachment to a central device (as do LED and acoustic sys-


Daniel J. Sandin

Figure 2. Sayre Glove, developed by Rich Sayre, Thomas DeFanti, and Daniel Sandin of the Electronic Visualization Laboratory at the University of Illinois, Chicago, in 1976.

Figure 3. MIT LED Glove, developed at the MIT Media Lab in the 1980s.


MIT Media Lab

tems). However, the current technology is quite robust and widely used for single or double hand-tracking.

Acoustic tracking

Acoustic trackers use high-frequency sound to triangulate a source within the work area. Most systems, like those from Logitech and the one used in the Mattel Power Glove (see below), send out pings from the source (mounted on the hand, for instance) received by microphones in the environment. Precise placement of the microphones allows the system to locate the source in space to within a few millimeters. These systems rely on line-of-sight between the source and the microphones, and can suffer from acoustic reflections if surrounded by hard walls or other acoustically reflective surfaces. If multiple acoustic trackers are used together, they must operate at nonconflicting frequencies, a strategy also used in magnetic tracking.

Glove technologies

Glove devices measure the shape of the hand as the fingers and palm flex. Over the past decade, especially in the last few years, many researchers have built hand and gesture measuring devices for computer input. We describe in roughly chronological order the more significant ones that have appeared in the literature or in the marketplace.

Sayre glove

Thomas DeFanti and Daniel Sandin at the University of Illinois at Chicago[4] developed an inexpensive, light-weight glove to monitor hand movements. Based on an idea from Rich Sayre, they used flexible tubes (not fiber optics) with a light source at one end and a photocell at the other. Tubes were mounted along each of the fingers of the glove (see Figure 2). As each tube was bent, the amount of light passing between its source and photocell decreased evenly. Voltage from each photocell could then be correlated with finger bending. They used this as an effective method for multidimensional control, such as to mimic a set of sliders. They did not use the glove as a gesturing device.

MIT LED glove

In the early 1980s researchers at the MIT Architecture Machine Group, and then at the MIT Media Lab, used a camera-based LED system to track body and limb position for real-time computer graphics animation, termed "scripting-by-enactment."[5] This work included a glove studded with LEDs (see Figure 3). By focusing the camera on just the hand, they captured finger motion that they then "grafted" onto the body motion. Unlike the Sayre glove, the LED glove was used for motion capture, not as a control device. The technology was not sufficiently developed to make a truly effective input device, and the glove was used only briefly.

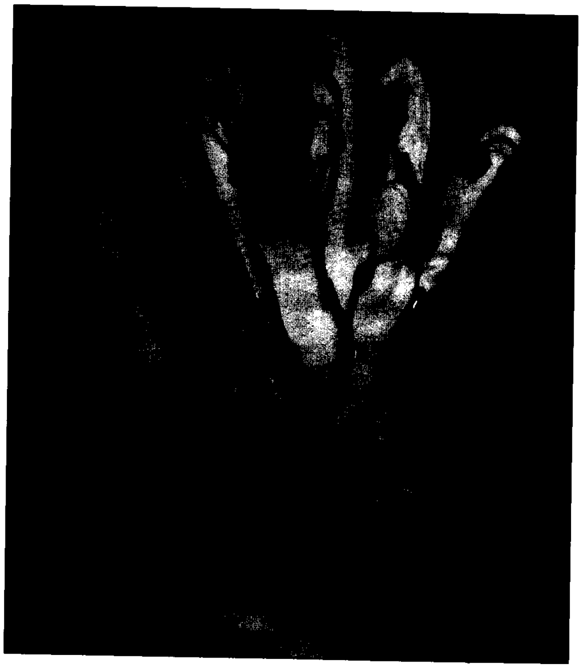
Digital Data Entry Glove

In 1983, Gary Grimes of Bell Telephone Laboratories developed a glove specially tailored to data entry using an alphabet of hand signs.[6] It consisted of a cloth glove onto which was sewn numerous touch, bend, and inertial sensors, specifically positioned so as to recognize the Single Hand Manual Alphabet for the American Deaf (see Figure 4). The circuitry was hard-wired to recognize 80 unique combinations of sensor readings to output a subset of the 96 printable ASCII characters. Grimes' glove was never put into actual use or commercially developed.

DataGlove

In 1987, Thomas Zimmerman and others developed a glove that monitored 10 finger joints and the six degrees of freedom of the hand's position and orientation.[7] The DataGlove was a clear improvement over the existing camera-based hand-monitoring techniques because it operated in real time and did not rely on line-of-sight observation. It was better than previous master-slave manipulators because it was light-weight, comfortable to wear, unobtrusive to the user, and general purpose. Commercialization of the DataGlove by VPL Research, at a reasonable cost to research institutions, lead to its widespread use around the world.

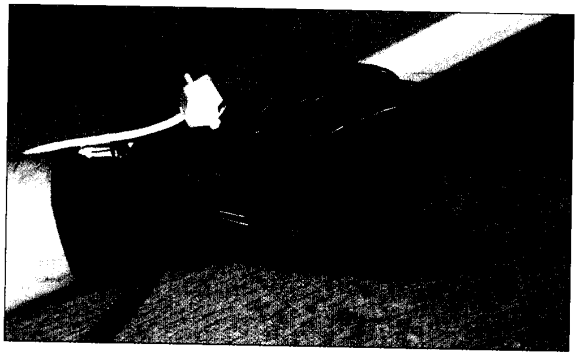
Physically, the DataGlove consists of a lightweight Lycra glove fitted with specially treated optical fibers along the backs of the fingers (see Figure 5). Finger flexion bends the fibers, attenuating the light they transmit. The signal strength for each of the fibers is sent to a processor that determines joint angles based on precalibrations for each user. Most DataGloves have 10 flex sensors, one for each of the lower two knuckles of the fingers and two for the thumb, but some have been made with abduction sensors that measure the angle between adjacent fingers. A 3-space magnetic tracker attached to the back of the hand determines position and orientation of the palm. VPL



Gary Grimes

Figure 4. Digital Data Entry Glove, developed by Gary Grimes at AT&T Bell Labs in 1983.

Figure 5. VPL DataGlove, the fiber-optic glove developed by VPL in 1987.

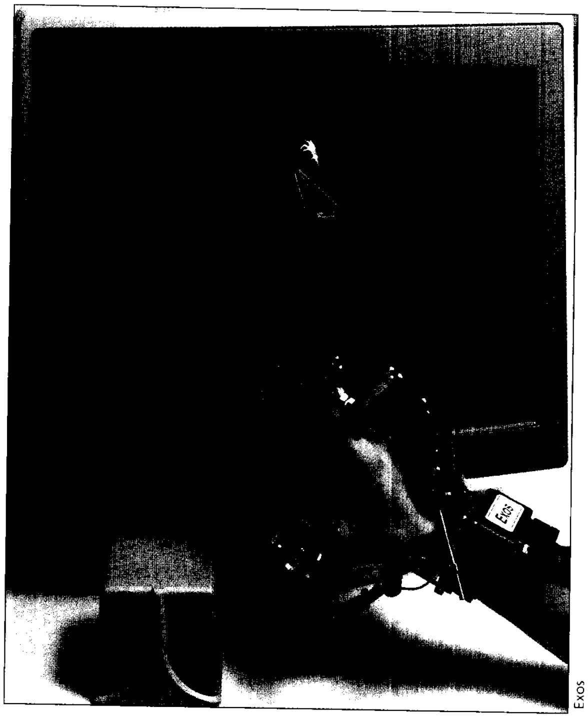


electronics combine the tracker readings with the flex sensor readings and send them out across a serial line.

The finger-flex accuracy of the DataGlove is rated at 1-degree joint rotation, but formal testing and informal observations have shown the actual flex accuracy to be closer to 5 or 10 degrees.[8] Although sufficient for general hand tracking and simple gestural input, this is not accurate enough for fine manipulations or complex gestural recognition. The speed of the DataGlove, approximately 30 Hz, is also insufficient to capture very rapid hand motions, such as might be used in time-critical applications or by untrained users.

Dexterous HandMaster

The Dexterous HandMaster (DHM) was originally developed as a master controller for the Utah/MIT Dexterous Hand robot hand by Arthur D. Little and Sarcos. Since then it has been redesigned and is now sold by Exos. DHM is an exoskeleton-like device worn on the fingers and hand (see Figure 6). Using Hall-effect sensors as potentiometers at the joints, it accurately measures the bending of the three joints of each finger as well as abduction of the fingers and complex motion of the thumb. The DHM measures 20 degrees of freedom of the hand—four for each finger and four for the thumb. The analog signals from the joint sensors are collected by a PC-compatible custom A/D board at up to 200 samples per second. Based on informal observation, the accuracy of the device is well within 1 degree of flexion. The DHM does not measure palm position or orientation, but a 3-space tracker can be attached for that purpose.



Exos

Figure 6. Exos Dexterous HandMaster, an exoskeleton with Hall-effect sensors, developed in 1989.

Although originally developed for robotics, the DHM has been successfully marketed as a tool for clinical analysis of hand function and impairment. Its highly accurate sensors make it an excellent tool for fine work or clinical analysis. The DHM is a little cumbersome to put on and take off, and requires some adjustment to fit the hand properly. Although light-weight, it has more mass than gloves and is less stable on the hand when the whole hand is shaken or moved rapidly. It is not an interface device suited for casual use. However, Exos has simplified and improved the technology, and it is available for measuring individual fingers and other body joints.

Power Glove

Inspired by the success of the VPL DataGlove, the Mattel toy company manufactured in 1989 a low-cost glove as a controller for Nintendo home video games. The Power Glove is a flexible molded plastic gauntlet with a Lycra palm (see Figure 7). Embedded in the plastic on the backs of the fingers are resistive-ink flex sensors that register overall bending of the thumb and in-

Figure 7. Power Glove, the low-cost glove developed by Mattel in 1989 for the home video game market.

Figure 8. CyberGlove, Virtual Technologies' 1990 glove with 18 sensors.

Virtual Technologies

dex, middle, and ring fingers with two bits of precision per finger. (This is a limitation of the A/D converters used, not the sensors themselves.) Mounted on the back of the hand are acoustic trackers that locate the glove accurately in space (to one-fourth inch) with respect to a companion unit mounted on the television monitor. The trackers also provide four bits of roll orientation for the hand (rotation of the wrist).

Although the least accurate of the whole-hand input devices, the Power Glove is also the cheapest by a factor of 100. It works with several Nintendo games, such as one where punching motions control the swing of an on-screen boxer. Some games have been especially designed for the Power Glove. One allows a player to "hit" or "grab and throw" a ball against tiles in a handball-like court imaged on the screen.

Unfortunately, after a two- or three-year run, Mattel stopped making the Power Glove, and now they are available only from stock or second-hand. The glove's low cost prompted many researchers to refit them for VR and glove-input systems. A general-purpose computer interface is not officially available for the Power Glove, but some people have reverse engineered the electronics necessary for connecting the Power Glove to a computer's serial port. One of the Usenet news groups, Sci.virtual-worlds, is a good source for this information. (See also a 1990 Byte article which has good descriptions and comparisons of the DataGlove, the DHM, and the Power Glove.[9]) The Power Glove is not particularly comfortable or accurate, but it serviceably provides a crude measure of hand position and shape.

CyberGlove

James Kramer developed the CyberGlove at Stanford University as part of his work to translate American Sign Language into spoken English.[10] It consists of a custom-made cloth glove with up to 22 thin foil strain gauges sewn into the fabric to sense finger and wrist bending (see Figure 8). A small electronics box converts the analog signals into a digital stream that can be read by a computer's standard serial port. As with the Dataglove and DHM, a 3-space tracker can be mounted on the glove to get hand position in space.

Informal experiments have found the CyberGlove's performance to be smooth and stable, with resolutions within a single degree of flexion. A useful feature of the CyberGlove is the capability to change the A/D hardware sensor offsets and gains from software, permitting the sensors to be tuned to use the full A/D range on a per-user basis. In our experience with the DataGlove, this was a persistent problem. To accommodate all hand sizes, the A/Ds were set such that the average DataGlove user exercised less than three-quarters of the full A/D range, reducing glove precision.

The CyberGlove is commercially available from Virtual Technologies. It is comfortable, easy to use, and has an accuracy and precision well suited for complex gestural work or fine manipulations.

Space Glove

W Industries, recently renamed Virtuality Entertainment Systems,based in Bristol, England, makes virtual reality arcade games. In 1991 the company released the Space Glove for use with their Virtuality system. The glove is made of soft molded plastic that fits over the back of the hand (see Figure 9). Rings around the fingers and a strap around the wrist hold the glove in place. One flex angle for each finger and two flex angles for the thumb are measured using sensors with 12-bit A/D converters. A 3-space magnetic tracker is incorporated into the back of the glove.

Personal experience in using the glove for a short time found it fairly responsive to finger bending and hand movement, but somewhat uncomfortable, as the plastic has little give and constricts the fingers. The stiffness of the plastic also makes it hard to get the rings over the finger joints when putting on or taking off the glove. The Space Glove only works with W. Industries products.

Applications and systems

With the commercial availability of hand sensing devices, research using the hand for computer input has blossomed. We've roughly categorized projects into the pursuit of natural inter-

Figure 9. Space Glove, developed in 1991 by W Industries for their Virtuality systems.

faces, systems for understanding signed languages, teleoperation and robotic control, computer-based puppetry, and musical performance.

Pursuit of natural interfaces

Since we manipulate the physical world most often and most naturally with our hands, there is a great desire to apply the skills, dexterity, and naturalness of the hand directly to the human-computer interface. A number of research projects in the past few years dealt with precisely this subject. Much of the work has been done in the context of developing virtual environments.

VPL

The developers of the VPL DataGlove were primarily interested in simulated environments or virtual realities and used the hand as the user's manipulative extension into those environments. Users wearing the DataGlove in the VPL system see a graphic hand that follows the motions of their own hand in the simulated environment. By pantomiming reaches and grabs, the user causes the graphic hand to reach and grab objects in the simulated environment. The viewer can move through the virtual space by pointing in the desired direction and "flying" to the destination. The actual implementations of the grab and flight behaviors are based on software that triggers events in response to recognized finger postures. (VPL uses look-up tables containing min/max values that define a range of finger sensor values for each posture. Following VPL's example, most researchers' DataGlove systems use similar methods, some with root-mean-squared (RMS) or other error reducing techniques.)

NASA Ames

Working with the VPL DataGlove in its initial stages of development, the Aerospace Human Factors Research Division of the NASA Ames Research Center used it for interaction with their Virtual Environment Display System.[11] Like VPL, they used the DataGlove as a tool for grasping and moving objects, indicating direction of motion, picking from menus, and invoking system commands. They also used the location of the hand as an event trigger for such things as drum beats on a virtual drum machine.

Later, in another Ames Research Center Laboratory, Steve Bryson and Creyon Levit[12] used the DataGlove in a virtual wind tunnel for visualizing the output of computational fluid dynamics programs run on supercomputers. With this system, aeronautic researchers can put their hands (and head) into a simulated fluid flow, "grab" onto one or more streamlines and move them about the model, and observe the changing airflow patterns in real time. In another mode of interaction, the user's fingers become sources of smoke trails that can be moved and positioned anywhere in the environment. The position and orientation of the synthetic camera can be changed by direct manipulation with the hands as well.

Point, reach, and grab

In many applications, the DataGlove is used similarly to its application at VPL and NASA. The hand's graphic image is displayed in an interactive computer environment and used as a tool for "point, reach, and grab" interaction. At the MIT Media Lab, we used the DataGlove as a master for a graphical hand in a virtual environment. The user could grab, move, and throw objects with the graphical hand, as well as use finger postures and motions to select from on-screen menus.[13] Arie Kaufman and Roni Yagel[14] used the DataGlove similarly in a modeling environment. The user could grab and manipulate objects on the computer screen. Steven Feiner and Clifford Beshers[15] and Haruo Takemura et al.[16] also used the DataGlove to allow users to touch, grab, and manipulate on-screen objects and recognize finger postures as event triggers (buttons), the former in a financial market simulator and the latter in a large-screen stereoscopic virtual environment.

The advantage of this model of interaction is naturalness—users' actions correlate closely with those that might be performed on physical objects. However, in each of these applications, the DataGlove functions as little more than a 3D joystick with several buttons.

At MIT we first considered implementing the virtual hand as a dynamic object in the simulated environment so that grabbing, pushing, and other interactions would be physically based. However, lacking the appropriate computing power to use this scheme in real time, we approximated the functionality with posture recognition.

In fact, in the MIT implementation, the DataGlove was occasionally replaced by a Spaceball—a six-degree-of-freedom force input device with eight buttons—since its software interface closely resembled that of the DataGlove, with button events substituting for posture recognition. Not surprisingly, many researchers and companies developing systems for virtual environments favor 3D joysticks over the more expensive glove devices.

Using more of the hand

More advanced use of the glove takes advantage of the extra capabilities of the hand over a 3D joystick. AT&T Bell Laboratories[17] used a DataGlove in the same way as the systems described above with the addition of two thumb-based gesture controls they called "clutch" and "throttle." They used clutching for incremental transforms, such as rotation. The screen object followed the rotation of the hand only when the thumb was brought against the index finger. Thus, object manipulations

Charles Painter/Virtual Technologies

Figure 10. Virtual Technologies' TalkingGlove, demonstrated by its creator, James Kramer, and ASL teacher Cathy Haas, herself deaf. Haas is signing with the instrumented glove. Her movements are translated to synthesized speech on the speaker pendant around her neck. Kramer responds by typing on a keypad whose readout is on Haas' right wrist.

could be ratcheted, instead of twisting the hand uncomfortably. Throttling was a variation of the clutch mechanism in which the angle of the thumb was used to scale the effect of a hand motion.

Thomas Baudel and Michael Beaudouin-Lafon[18] analyzed complete hand gestures captured by a DataGlove to control audio-visual presentations. By gesturing with the DataGlove, the presenter controls the sequencing of images projected from an Apple Macintosh onto a screen in front of a room. An important part of this work is the development of an icon-based notation for describing and documenting dynamic gestures. This allows gesture sequences to be concisely documented and potentially used by other systems. Their gesture recognition algorithm is a hybrid, using an extension of Dean Rubine's excellent method of feature analysis.[19] They've achieved high recognition rates for both trained and untrained users.

Interpreting sign language

One of the obvious applications of glove devices is the interpretation of signed languages, both for computer input and control, and for human communication. Several projects have investigated various levels of recognizing hand signs from simple finger spelling to analysis of American Sign Language (ASL).

Grimes Digital Data Entry Glove (described above) is one of the earliest of these projects. His approach to recognizing finger spelling postures relied on custom circuits, not software algorithms.

Soon afterwards, the MIT Media Lab used their LED glove as part of an experimental system for finger-spelling, using lookup tables in software to recognize finger postures.[20]

Kramer's system to translate ASL into spoken English (see CyberGlove, above) used a Bayesian decision rule-based pattern recognition scheme to map finger positions, represented as a "hand-state vector," into predefined letters or symbols. When the instantaneous hand-state lay close enough to a recognizable state, the corresponding ASL letter or symbol was put in an output buffer. When a word phrase was complete, a special sign caused the result to be spoken by a voice synthesizer. Hearing participants in conversations typed back answers on a hand-held keyboard. (See Figure 10.) His system also had the option of using a neural network approach to the hand shape recognition.

ATR Research Labs in Japan developed a coding scheme to allow computer recognition of the Japanese kana manual alphabet.[21] Their system used the DataGlove to capture hand posture. It recognized signs through a combination of principal component analysis (to determine the contributions of each finger joint to the differences between signs) and cluster analysis (to group hand configurations). Because of the difficulty of accurately measuring the lower thumb joint with the DataGlove, and because some of the signs have similar finger positions, they were able to discriminate only 30 of the 46 kana signs.

Interpreting hand signs that involve motion is a much more difficult problem than simple finger spelling, since pattern analysis must be performed on the moving hand. Researchers at the Salk Institute in La Jolla, California, identified more than 50 different linguistic processes in ASL.[22] However, they proposed that these processes differ along only 11 spatial and temporal dimensions. We believe that by using these same dimensions in gestural control, perhaps powerful yet manageable methods for gestural control can be developed. They also proposed various analytical techniques, including feature analysis and frequency analysis, from which to qualify the linguistically relevant features of signed language. As an interesting side note, they found that fingertip tracking was sufficient for human understanding of signed language.

Sidney Fels[23] used a DataGlove to interpret hand motion to drive a speech synthesizer. His particular approach used a three-stage back-propagation neural network trained to recognize gestural "words." He divided hand motions among 66 finger positions and 6 hand motions. Finger positions defined the root word, while hand motions modified the meaning and provided expression. These combined to form the 203 words of his "language," loosely based on conventional gestural languages. Fels reported a high recognition rate once the system was fully trained.

In his report, Fels included an interesting analysis of hand-to-language mapping at various levels of granularity, from using hand motions for the control of parameters of an artificial vocal tract to interpreting whole hand motions as words and concepts. The trade-offs, as Fels put it, are between extent of vocabulary—unlimited at the most granular level—versus ease of learning and speed of communication—highest at the word and concept level.

Although Fels' system demonstrates the viability of neural net techniques for interpreting finger position and hand motion, it is uncertain if these techniques realistically can be extended to include the added complexity of finger motions and complex hand trajectories necessary to interpret the full expression of signed languages. However, these methods might be adequate as a control structure for limited-vocabulary computer input.

Three of the methods of hand shape and motion recognition described above (and a method used by Martin Brooks for robotic control, below) are conceptually similar. Basically, Kramer, Takahashi, and Fels all analyze the hand-space-degrees-of-freedom vector for each posture or gesture and match it to a landmark hand-space vector representing the target posture or gesture. The match must occur within error tolerances (often Euclidean distance) weighted by the significance of each degree of freedom. In the Takahashi-Kishino method, principal component analysis determines the weighting of the degrees of freedom. Kramer's Bayesian analysis uses a similar algorithm. In Kramer's and Fels' neural nets, the process is hidden in the coefficients for each node. Brooks' neural-like net has few nodes, each with an *n*-space vector of coefficients. These coefficients contain the weightings, with the interaction between the nodes of the net determining the identity of a dynamic gesture.

Teleoperation and robotic control

Glove interfaces in teleoperation and robotic control are important for facile, dexterous control of the remote end. Two research projects have used the DataGlove to control a dexterous robot hand. AT&T constructed algebraic transformation matrices to map human hand poses to robot hand poses.[24] The transformation matrices compensated for the kinematic differences between the human hand, as measured by the DataGlove, and the robot hand. The user controlled the robot hand by mimicking the desired poses. In a similar project, New York University's Courant Institute resolved the kinematic differences between the human hand and the robotic hand by determining the position of the user's fingertips and driving the robot hand fingertip positions to match.[25]

The AT&T work was extended from the DataGlove to the DHM.[26] Since the DHM was kinematically similar to the robot hand, the transformation matrix scheme used for the DataGlove was not necessary. Instead, they transformed the raw sensor data into strings of 7-bit characters. Lexical recognition routines matched string patterns to autonomous manipulation functions for the robot hand (similar to the poses used previously with the DataGlove).

Brooks used a neural net to interpret DataGlove motion for robot control.[27] Unlike Fels, Brooks incorporated dynamic gestures into the control language. He used Kohonen nets[28] to recognize paths traced by finger motion in the *n*-dimensional space of the degrees of freedom of the digits. Each Kohonen net (typically on the order of 20 cells) was trained to recognize a single gesture. Operating several concurrently on the DataGlove input meant several gestures could be recognized. He achieved moderate success at simple gesture recognition, such as closing all the fingers, leading with the index finger; opening the thumb and first two fingers simultaneously; and moving from a neutral hand posture to a "pen" grasp posture. However, in his conclusion, Brooks stated that he had yet to show that his methods were sufficient for practical dynamic gesture recognition or that the DataGlove is an appropriate interface for robot control.

At the MIT Media Lab, we demonstrated the operation of a simulated construction crane with hand signals conventionally used on construction sites, implementing a gesture recognition system based on Rubine's feature analysis.[29] We also used this system for the simulated teleoperation of a six-legged robot. The robot's entire interface, including locomotion, point of view, manipulator control, and mode selection, was glove-based. We used both the VPL DataGlove and Exos DHM in our MIT work.

Stelarc, an Australian-based performance artist, wears two DataGloves, one on each hand, to control a third (robotic) hand mounted on his own right arm.[30] One DataGlove serves as a master to the mechanical hand, which mimics its behavior, while the other provides mode controls to the mechanical hand. Like many others, Stelarc's gloves use gesture recognition based on Rubine's method of feature analysis.

Computer-based puppetry

Most computer animation of characters uses a key-frame technique, much like conventional hand animation. Linear or spline interpolation generates the frames between keys. The relative smoothness of the interpolation tends to give these animations a subtly unnatural quality, not quite mechanical, but not quite living. Programmed (or procedural) animation yields motion that is occasionally life-like, but often too regular to be a product of life itself. To inject life into computer animation, and as a way to overcome the trade-off between animation/programming time and motion quality, production companies have turned to puppetry and body tracking for computer animation of characters. Putting a performer in direct interactive control of a character, as in puppetry, or capturing body motion for later application to animation, translates the nuances of natural motion to computer characters, making them seem very much alive.

The beginning of this work dates back to the late 1970's, when Thomas Calvert attached goniometers to people to track joint movement. His purpose was to combine this information with dance notation to drive computer animation.[31] A few years later researchers at MIT began a similar project, called the "graphical marionette," which included the MIT LED glove.

In 1989 Pacific Data Images collaborated with Jim Henson to produce a computer graphic character whose motion could be performed alongside the conventional puppets. They built a simple one-handed controller that allowed the puppeteer to move the character around on the computer graphics screen as well as control the character's mouth movements.

Following their lead, Geoff Levner, working at Videosystem in France, developed a real-time computer animation system he called PORC (Puppets Orchestrated in Real-time by Computer). Using DataGloves, joysticks, foot pedals, and other custom devices, puppeteers control the motion of characters generated in real time by high-end graphics workstations. For example, Poupidoo, a computer puppet who anchored a 24-hour animation marathon on French television, was controlled by three puppeteers. One used a glove to control the mouth shape and expression (each of three fingers controlled a facial parameter such as smile/frown). Another used a glove to control the expression and closing of the eyes, and a joystick to choose direction of the eyes. The third used two gloves and a set of Pol-

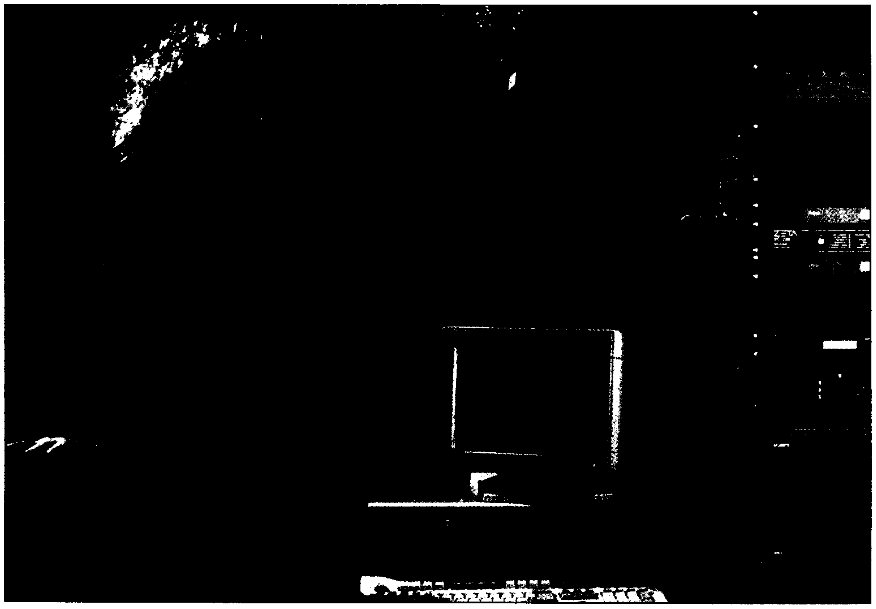


Figure 11. Composer Tod Machover conducting hyperinstruments at the MIT Media Lab using an Exos Dexterous HandMaster.

Peter Menzel

hemus trackers to control the upper body and arm motions. Videosystem, since renamed Medialab, uses similar setups on an ongoing basis with a multitude of real-time characters in client productions for film and television.

Musical performance

Tod Machover used an Exos DHM at the MIT Media Lab to control acoustic parameters in live musical performances.[29] In a piece called "bug-mudra" (see Figure 11), two guitarists and a percussionist provide input to a MIDI-based computer-music system that reshapes the guitar sounds and synthesizes new sounds based on the performance. In concert, the conductor wears a DHM on his left hand, using it to dynamically mix timbre and volume of various channels of the combined output. "Bug-mudra" premiered in Tokyo in 1990 and since has been performed in various venues in the United States and Europe.

Hideyuki Morita also uses a glove to conduct music.[32] In this case, a human conducts a synthetic orchestra. The system uses an infrared light on the end of the baton in the conductor's right hand and a magnetic tracker and DataGlove on the conductor's left hand. A CCD camera follows the trajectory of the baton using feature detection to extract tempo information from the motion. The magnetic tracker on the left hand indicates the hand's location and where the conductor is pointing, targeting a group of instruments for those instructions. The attitude and posture of the left hand as captured by the tracker and DataGlove are interpreted through a function table to determine commands of musical expression, such as vibrato, crescendo, sostenuto, and dolce. These are combined and used to control the playback of prerecorded MIDI scores, adding performance expression to the otherwise flat MIDI control. The result is a synthetic music system that can interpret a conductor by following that conductor's conventional method of communication, thus yielding more expressive results.

Conclusion

Interest in direct manipulation interfaces continues to grow, especially for immersive virtual environments. Many labs in the US, Japan, Europe, and Australia have purchased DataGloves, DHMs, or Power Gloves, or built their own hand devices in pursuit of natural interfaces. As research continues, hand- and finger-tracking devices will improve, along with gesture recognition and interface software.

Despite many advances in this area, glove-based input or, more generally, whole-hand input, remains in its infancy. For the most part, the user must still wear a device such as a glove, or work in a special environment such as a room brightly lit for video cameras. Achieving the goal of "deviceless" natural computer interaction with the hands and body requires advances in many areas, including freeing the user from electrical connecting cables, improving the speed and accuracy of tracking devices, lowering manufacturing costs, and developing more commercial applications for the technology. ❑

Acknowledgments

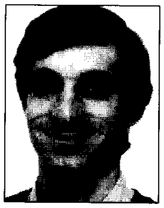
We wish to thank the members of the Computer Graphics and Animation Group of the MIT Media Lab and the members of the Research and Development Group of Medialab for their assistance.

This work was supported in part by NHK (Japan Broadcasting Company), Defense Advanced Research Projects Agency-RADC Contract #F30602-89-C-0022, and equipment grants from Hewlett-Packard.

References

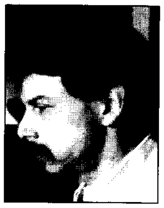
1. R.A. Bolt, "Put-that-there: Voice and Gesture at the Graphics Interface," *Computer Graphics* (Proc. Siggraph), Vol. 14, No. 3, July 1980, pp. 262-270.
2. M.W. Krueger, *Artificial Reality*, 2d ed., Addison-Wesley, Reading, Mass., 1990.
3. P. Wellner, "Interacting with Paper on the DigitalDesk," *Comm. ACM*, Vol. 36, No. 7, July 1993, pp. 87-96.
4. T.A. DeFanti and D.J. Sandin, "Final Report to the National Endowment of the Arts," US NEA R60-34-163, University of Illinois at Chicago Circle, Chicago, Ill., 1977.
5. C.M. Ginsberg and D. Maxwell, "Graphical Marionette," *Proc. ACM Siggraph/Sigart Workshop on Motion*, ACM Press, New York, April 1983, pp. 172-179.
6. G.J. Grimes, "Digital Data Entry Glove Interface Device," Bell Telephone Laboratories, Murray Hill, N.J., US Patent 4,414,537, Nov. 8, 1983.

7. T. G. Zimmerman et al., "A Hand Gesture Interface Device," *Proc. Human Factors in Computing Systems and Graphics Interface*, ACM Press, New York, April 1987, pp. 189-192.
8. S. Wise et al., "Evaluation of a Fiber Optic Glove for Semi-automated Goniometric Measurements," *J. Rehabiliation Research and Development*, Vol. 27, No. 4, 1990, pp. 411-424.
9. H. Eglowstein, "Reach Out and Touch Your Data," *Byte*, July 1990, pp. 283-290.
10. J. Kramer and L. Leifer, "The Talking Glove: An Expressive and Receptive 'Verbal' Communication Aid for the Deaf, Deaf-Blind, and Nonvocal," tech. report, Stanford University, Dept. of Electrical Engineering, Stanford, Calif., 1989.
11. S.S. Fisher et al, "Virtual Environment Display System," *Proc. 1986 ACM Workshop on Interactive Graphics*, ACM Press, New York, Oct. 1986, pp. 77-87.
12. S. Bryson and C. Levit, "The Virtual Wind Tunnel," *IEEE CG&A*, Vol. 12, No. 4, July 1992, pp. 25-34.
13. D. Zeltzer, S. Pieper, and D. Sturman, "An Integrated Graphical Simulation Platform," *Proc. Graphics Interface 89*, Morgan Kaufman, Palo Alto, Calif., June 1989, pp. 266-274.
14. A. Kaufman and R. Yagel, "Tools for Interaction in Three Dimensions," in *Work with Computers: Organizational, Management, Stress, and Health Aspects* (Proc. HCI 89), M.J. Smith and G. Salvendy, eds., Elsevier Science Publishers, Amsterdam, 1989, pp. 469-475.
15. S. Feiner and C. Beshers, "Visualizing *n*-dimensional Virtual Worlds with *n*-Vision," *Computer Graphics* (Proc. 1990 Symp. Interactive Graphics), Vol. 24, No. 2, March 1990, pp. 37-38.
16. H. Takemura, A. Tomono, and Y. Kobayashi, "An Evaluation of 3D Object Pointing Using a Fields Sequential Stereoscopic Display," *Proc. Graphics Interface 88*, Morgan Kaufmann, Palo Alto, Calif., 1988.
17. D. Weimer and S.K. Ganapathy, "A Synthetic Visual Environment with Hand Gesturing and Voice Input," *Proc. CHI 89*, ACM Press, New York, May 1989, pp. 235-240.
18. T. Baudel and M. Beaudouin-Lafon, "Charade: Remote Control of Objects Using Free-Hand Gestures," *Comm. ACM*, Vol. 36, No. 7, July 1993, pp. 28-35.
19. D. Rubine, "Specifying Gestures by Example," *Computer Graphics* (Proc. Siggraph), Vol. 25, No. 4, July 1991, pp. 239-337.
20. J.A. Hall, *The Human Interface in Three-Dimensional Computer Art Space*, MSVS thesis, Media Lab, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 1985.
21. T. Takahashi, and F. Kishino, "Hand Gesture Coding Based on Experiments Using a Hand Gesture Interface Device," *Sigchi Bulletin*, Vol. 23, No. 2, April 1991, pp. 67-74.
22. H. Poizner et al., "Motion Analysis of Grammatical Processes in Visual-Gestural Language," *Proc. ACM Siggraph/Sigart Workshop on Motion*, ACM Press, New York, April 1983, pp. 148-171.
23. S.S. Fels, "Building Adaptive Interfaces with Neural Networks: The Glove-Talk Pilot Study," Tech. Report No. CRG-TR-90-1, Dept. of Computer Science, University of Toronto, Toronto, Canada, Feb. 1990.
24. L. Pao and T.H. Speeter, "Transformation of Human Hand Positions for Robotic Hand Control," *Proc. IEEE Int'l Conf. Robotics and Automation*, Vol. 3, IEEE CS Press, Los Alamitos, Calif., 1989, pp. 1,758-1,763.
25. J. Hong and X. Tan, "Calibrating a VPL DataGlove for Teleoperating the Utah/MIT Hand," *Proc. IEEE Int'l Conf. Robotics and Automation*, Vol. 3, IEEE CS Press, Los Alamitos, Calif., 1989, pp. 1,752-1,757.
26. T.H. Speeter, "Transforming Human Hand Motion for Telemanipulation," Tech. Memorandum, AT&T Bell Laboratories, Holmdel, N.J., Sept. 19, 1989.
27. M. Brooks, "The DataGlove as a Man-Machine Interface for Robotics," 2d IARP Workshop on Medical and Healthcare Robotics, Newcastle upon Tyne, UK, Sept. 1989, pp. 213-225.
28. T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1984.
29. D.J. Sturman, *Whole-Hand Input*, doctoral dissertation, Media Lab, Massachusetts Institute of Technology, Cambridge, Mass., Feb. 1992.
30. M. Papper and M. Gigante, "Using Gestures to Control a Virtual Arm," in *Virtual Reality Systems*, R. Earnshaw, H. Jones, and M. Gigante, eds., Academic Press, London, 1993.
31. T.W. Calvert, J. Chapman, and A. Patla, "The Integration of Subjective and Objective Data in the Animation of Human Movement," *Computer Graphics* (Proc. Siggraph), Vol. 14, No. 3, July 1980, pp. 198-203.
32. H. Morita, S. Hashimoto, and S. Ohteru, "A Computer Music System that Follows a Human Conductor," *Computer*, Vol. 24, No. 7, July 1991, pp. 44-53.



David Sturman is currently a member of the Research and Development Group of Medialab, a Paris-based computer graphics production company specializing in computer animation, computer puppetry, special effects, and virtual reality. His research interests include motion specification and control for computer animation, and interaction devices for computer animation and virtual reality. He has a BS in computer science from MIT, an MS from Rutgers University, and a PhD from the MIT Media Lab.

Readers may reach Sturman at Medialab, 104 av. du President Kennedy, 75016 Paris, France, email sturman@dmi.ens.fr.



David Zeltzer is a principal research scientist with the Sensory Communication Group at MIT's Research Laboratory of Electronics. In addition to work in virtual environment technology, his research interests include human interface design, and biological and artificial motor control systems. He received his MS and PhD degrees in computer and information science from Ohio State University in 1979 and 1984, respectively. Zeltzer serves as associate editor of the MIT Press journal *Presence: Teleoperators and Virtual Environments*.

Readers may contact Zeltzer at Sensory Communication Group, MIT Research Lab of Electronics, Cambridge, MA 02139, USA, e-mail dz@irts.mit.edu.