A Logical Hand Device in Virtual Environments

S. Augustine Su
Richard Furuta

February 1994
TAMU-HRL 94-002

A LOGICAL HAND DEVICE IN VIRTUAL ENVIRONMENTS[a]

S. AUGUSTINE SU[b]
Department of Computer Science, University of Maryland
College Park, Maryland 20742, USA

and

RICHARD FURUTA
Department of Computer Science, Texas A&M University
College Station, Texas 77843-3112, USA

ABSTRACT

The human hands are the major means through which we gain our primary connection to the world. The modeling of human hands is a very important issue in virtual environments, however little research has been done to support higher levels of abstraction of using hands beyond that of just capturing raw data.

In this paper we present an alternative view of hand modeling, i.e., a point-based hand model, and then investigate 3D static hand gestures in detail. Thereby, we develop a device-independent and general-purpose logical hand device, which supports the use of comprehensive 3D gestural input in virtual environments. Based on our logical hand device, not only can the implementation of "point, reach, and grab" interaction be facilitated, but also American-Sign-Language-like static gestures can be conceived easily.

KEYWORDS: virtual environments, hand models, 3D gestures, logical hand device, American Sign Language

1. Introduction

Virtual Reality (VR) is the next quantum leap of Human-Computer Interaction (HCI), taking the area beyond a currently popular HCI technique—Graphical User Interface (GUI, i.e., two-dimensional-window programming, e.g., the X windows, Macintosh, and Microsoft Windows).

This brand new frontier is young and most work has focused on the development of hardware technology and the custom implementation of specific applications. There is currently very little research into the higher levels of abstraction that

[a] This material is based in part upon work supported by the National Science Foundation under grant number IRI–9496187.

[b] The author's current address is Department of Computer Science, Texas A&M University, College Station, Texas 77843-3112, USA, su@bush.cs.tamu.edu

facilitates the composition of new systems. It is informative to contrast this with GUI research, where the study of abstraction is very mature.

One of the GUI abstractions that is useful for graphics-package programmers is the logical device: locator, keyboard, pick, choice, and stroke devices [Foley *et al.*, 1990]. The goal of this paper is to propose a logical hand device that is general enough to serve most of the needs of the development and implementation of virtual environment systems in use of 3D gestures.

Sturman [1992] has discussed the use of whole hands as an input device in his thesis. He suggested three ways of interpretation of hand actions: direct, mapped, and symbolic interpretation. The direct interpretation refers to "point, reach, and grab" interaction; and the mapped interpretation refers to fiddling virtual input devices: these two ways of interpretation can be categorized as the manipulation paradigm. By this paradigm, most VR systems use the index fingertip as a 3D pointer [Fisher *et al.*, 1986] [Weimer & Ganapathy, 1989] [Sturman *et al.*, 1989], which is able to select items in menus floating in 3D space, to fiddle virtual devices, etc. This approach is the most natural way to use hand gestures, and is available in our daily life using computers with operations, like key pressing and mouse button pressing. However, it ignores the freedom that general hand gestures provide, and can be reduced into a device whose function can be achieved by using a 3D mouse.

The symbolic interpretation is to recognize hand gestures as a stream of tokens that are similar to signs in sign languages. This way of interpretation can be categorized as the sign language paradigm. Fels and Hinton [1993] designed their own gestures by this approach. Baudel and Beaudouin-Lafon [1993] used a visual language to define their three-stage gestures in the Charade presentation system. This visual language can be used to record gestures analytically. The major critique of this paradigm is that the devised gestures are not easy to recall. We think gestures have to be defined by users incrementally, and changed at users' will, rather than mandatorily defined by systems. This paper provide a foundation to fulfill the above goal. There was an excellent survey done by Sturman and Zeltzer [1994] about the current status of the hand-tracking hardware, and the applications and systems using the hardware.

Our proposed logical hand device has the ability to accommodate these two major paradigms to serve the need of using hand gestures in virtual environments. This logical hand device is the front-end processing for the Virtual Panel Architecture [Su & Furuta, 1993], which provides support for the rapid prototyping of virtual environments systems.

Inspired by the manipulation paradigm, we present an alternative hand model in this paper. The new model focuses on fingertips and thumb tip. We also proved that the new hand model is equivalent to a popular hand model that is constructed by joint angles. Then we investigate the handshapes of American Sign Language (ASL) in detail, and devise a set of digit-oriented information descriptions that can be used to define static gestures textually. Finally, we will discuss the possibilities of implementation of this logical hand device.

2. Hand Models

In real life, hands are distinct from people to people, due to the disparities of the length and the thickness of hand digits, etc. Therefore, we need common grounds to describe the hand, that is, abstract hand models, which extract universal and salient attributes of the hand. Based on the hand models, hand gestures can be described precisely.

2.1. An Angle-based Hand Model

Let us take a look at the anatomy of the hand first. There are 27 bones in each hand [Alexander, 1992] [Spence & Mason, 1987]. The bones are connected by a number of movable joints. Some of these joints are hinge joints allowing flexion and extension only, i.e., one degree of freedom (DOF) of movement for each. Some are biaxial joints allowing flexion, extension, abduction, adduction, and circumduction, i.e., two DOFs for each. The metacarpals of the index and middle fingers can hardly move at all relative to the carpals but those of the ring and little fingers move a little while we are making fists. The total number of DOFs of joints on each hand is 22. If we count the position and orientation of the hand itself, the number is 28 (the other six DOFs are *x, y, z, pitch, roll*, and *yaw*).

Further simplifications are needed to construct hand models. We ignore the two little-moving DOFs of the joints between metacarpals and carpals of the ring and little fingers to make hand modeling simple. Therefore, the number of DOFs of our hand modeling is 26. As a result, the abstract hand is a flat rectangle representing the palm, plus three inter-connected line segments for bones of each finger and the thumb (see Figure 1).

The first hand model of interest is an angle-based hand model, which consists of 26 parameters corresponding to 20 DOFs of joints (4 DOFs for each finger and thumb), and 6 DOFs of the position and orientation of the hand. This hand model is popular in human hand modeling, like in [Rijpkema & Girard, 1991].

The angle-based hand model has enough descriptive power to satisfy the need of hand modeling in virtual environments. However, it does not provide any further convenient features to support any approach of utilizing gestures.

2.2. A Point-based Hand Model

We have observed that pointing (touching), grabbing, and shooting are favorite gestures in virtual environments. Inspired by the above observation, the second model is designed as a point-based hand model, which consists of six points on a hand: each of these points is an origin of a reference frame, that is, carrying the position and orientation information to support the precise manipulation of 3D objects by gestures.

The construction of the point-based hand model is based on the angle-based model: the 4 DOFs of each digit of the hand are replaced by one reference frame

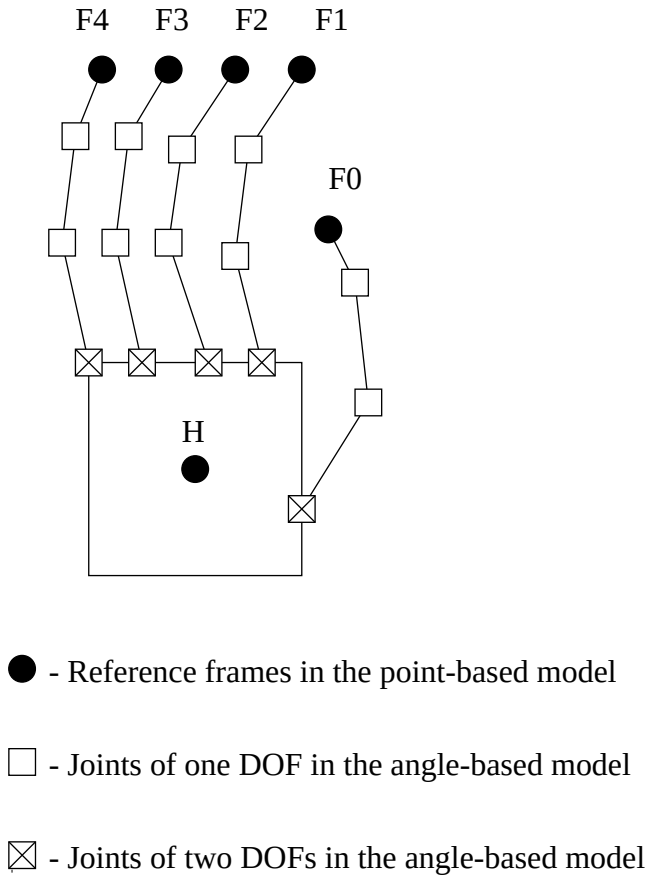

Figure 1. The angle-based hand model and point-based hand model.

whose origin is attached to the tip of the digit. In summary, the point-based hand model is comprised of six reference frames: five of their origins attached to the tips of four fingers and the thumb, the other to the center of the back of the hand (the points F0, F1, F2, F3, F4, and H in Figure 1). Since the 4 DOFs of a digit are replaced by the reference frame (6 DOFs), the 6 DOFs are not fully inter-independent with respect to the center of the hand, i.e., some values of the 6 DOFs are not available.

The two hand models are almost interchangeable. The computation of values of the six reference frames in the point-based model is straightforward: the step-by-step translations and rotations of these points with respect to the point on the hand center, based on the values of the angle-based model. Conversely, given a set of values for the point-based model, we can compute the solution to the angle-based model as follows: At first, track back the position and orientation of the point on a distal joint from the reference frame on a fingertip (the answer is unique). Since the points of the distal, middle, and knuckle joints will be on a plane due to the fact that the middle joint is a hinge joint allowing one DOF only, there exist two solutions for each digit to the angles of middle joint in the angle-based model.

Fortunately, if we consider the structural factor to exclude unnatural configurations of the hand, there exists only one solution to the angles in the angle-based hand model. Therefore, these two models are equivalent in the sense of naturalness. All of the above computations can be found in an earlier paper [Su, 1993].

The implication of this is that if we have a hand tracking device able to measure all of the DOFs of the angle-based hand model, then the information of the point-based model can be computed. To our knowledge, all of the current hand tracking devices only measure the angles of the joints. In our opinion, point-based hand tracking devices should be seriously considered.

3. Static Gestures

In order to devise a universal mechanism to define gestures, a big pool of gestures have to be studied. Our target is a real living gestural language, American Sign Language. Linguistically, Stokoe defined three parameters that were realized simultaneously in the formation of a particular sign of ASL [Stokoe *et al.*, 1976]: (1) DEZ (designator) for handshapes, (2) TAB (tabulation) for locations, and (3) SIG (signation) for motions. A fourth parameter, orientation, which refers to the orientation of the palm, was added later by Battison [Wilbur, 1987]. Initially, we have focused on static gestures; in other words, the motion attributes are not considered here. The location attributes can be solved easily by associating them with specific 3D objects. The orientation information can be carried by point H of the point-based hand model (see Figure 1). The remaining problem here is how to describe handshapes.

Stokoe categorized handshapes used in ASL into 19 major handshapes (hand configuration prime) and their 21 variants (sub-primes) (the total is 40 handshapes). Each prime or sub-prime is a configuration of a whole hand. Based on these 40 handshapes, we decomposed these handshapes into some digit-oriented attributes, which will be discussed in the following.

Each finger has three joints: metacarpophalangeal (MP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints. Only two of them are major factors controlling the configuration of a finger: the bending angle of the MP joint (the proximal bend, PB) and the angle of the PIP joint (the middle bend, MB) (The angle of the DIP joint is approximately two third of the angle of the PIP joint [Rijpkema & Girard, 1991]). Although the values of PB range roughly from π to $\pi/2$, we found that only three values occur in Stokoe's primes and sub-primes: π, $\pi/2$, and their in-between (near $3\pi/4$). For the MB, there are only two values: π and $\pi/2$. Therefore, we have extracted $2 * 3 = 6$ configurations for each finger (see Figure 2): **straight**, **slant**, **flat**, **hooked**, **curved**, and **fully curved**. For adjacent fingers having the same configuration, one more parameter, clustering information, has to be specified: **adducted** (gathering) or **abducted** (separated) when the fingers are not **fully curved**. For the thumb, there are four configurations: **inward** (toward the palm), **outward** (away from the palm), **neutral**, and **lowerly neutral**. If we do not distinguish the cases for the adducted and the abducted, then we have

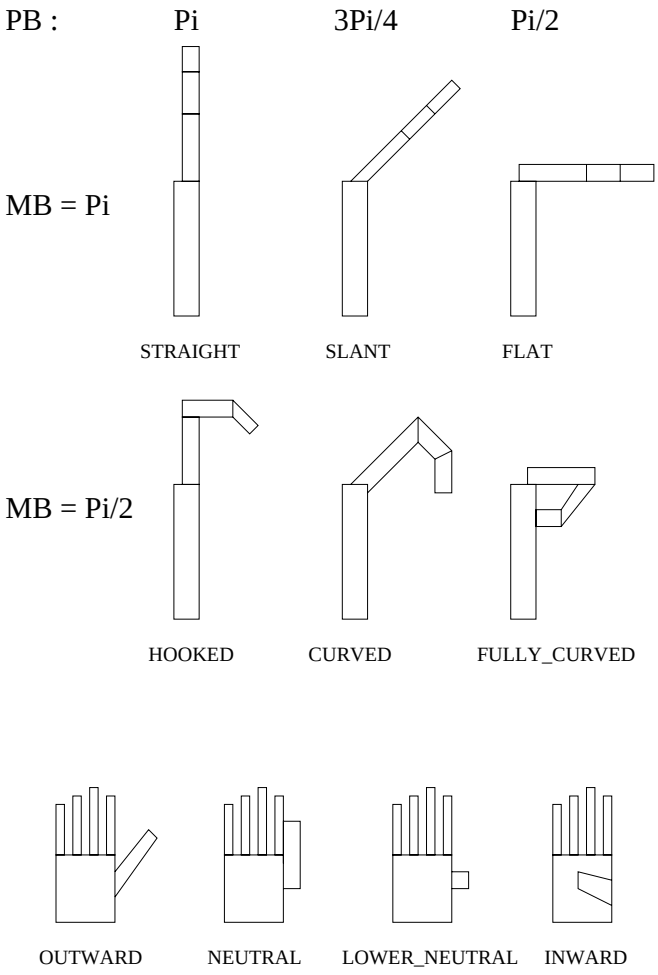

Figure 2. The abstract digit-oriented information of fingers (the upper six diagrams) and the thumb (the lower four diagrams).

$4 * 6 * 6 * 6 * 6 = 5184$ handshapes. However, it is obvious that there are some of them which cannot be gesticulated, or cannot be gesticulated naturally.

Besides the above parameters, we need information from the point-based model to help distinguishing the interrelations of digits. We call this information as *modifiers*. Only one modifier involves fingers only: **twist** for the index and middle fingers in letter R of the American Manual Alphabet. The other modifiers are for the thumb interacting finger(s): **no contact** for the tips of the thumb and fingers in letter C, **tip contact** for the thumb and middle finger in letter D, **enclosing fingers** in letter I, **enclosed by fingers** in letter X, **between12** for the thumb between F1 (the index finger) and F2 (the middle finger) in letter T, **between23** for the thumb between F2 (the middle finger) and F3 (the ring finger) in letter N, **between34** for the thumb between F3 (the middle finger) and F4 (the little finger) in letter M, and **below** fingers in letter E.

In summary, the static handshapes of ASL can be described as four types of parameters: (1) the configuration of each finger, (2) the clustering information of fingers, (3) the configuration of the thumb, and (4) modifier(s). That forms the basis of our gesture description language to tokenize static handshapes. For tokenizing gestures beyond ASL, we may need to define new modifiers to describe new situations.

The most tricky problem for the above setup is that we have to carefully specify the typical values of the configurations and their error tolerances. Also, some configurations between two adjacent configurations, say, between **straight** and **slant** for fingers, will be rejected as illegal configurations, or be accepted as ambiguous configurations between the two. Sometimes these ambiguities are necessary for specifying gestures.

4. Hand Device

Based on the work of last two sections, a complete logical hand device is defined by three parameters: (1) the six reference frames of the point-based hand model, (2) the abstract digit-oriented handshape information discussed in the last section, and (3) time stamp. This logical hand device is not only able to support the manipulation paradigm by providing 3D position and orientation of the tips of hand digits as in (1), but also able to support the sign language paradigm by providing sufficient information as in (2) to compose user-defined gestures. The time stamp is used to provide temporal information.

In the Virtual Panel Architecture, the logical hand device is extracted by the Gesture Server. The tips of hand digits are used to manipulate 3D objects maintained in the Panel Server by the way similar to 2D cursor manipulating 2D X windows [Scheifler & Gettys, 1986].

The following examples of gesture definition will demonstrate the descriptive power of the abstract digit-oriented handshape information of last section.

The victory sign, or the letter V in the American Manual Alphabet is defined as follows:

V :: **F3, F4: fully_curved;**
F1, F2: straight, abducted;
F0: inward, over_fingers;
Palm: superior, anterior.

where the capitalized words are tokens defined by the gesture description language, F0 is to denote the thumb, F1 the index finger, F2 the middle finger, F3 the ring finger, and F4 the little finger; the orientation of fingers' side of the palm is pointing up (**superior**), and the orientation of the palm is pointing away from signer (**anterior**).

Another example is the sign for "I love you", which is the combination of three letters, "I", "L", and "Y" in the American Manual Alphabet, is defined as follows:

I_Love_You :: **F1, F4: straight;**
F2, F3: fully_curved;
F0: outward;
Palm: superior, anterior.

where the orientation is the same as that of letter V.

The definition of the sign for number four follows:

4 :: **F1, F2, F3, F4: straight, abducted;**
F0: inward;
Palm: superior, anterior.

where the orientation is the same as that of letter V.

The definition of the sign for letter C follows:

C :: **F1, F2, F3, F4: curved, adducted;**
F0: lowerly_neutral, no_contact;
Palm: superior, medial.

where the orientation of the palm is pointing to the centerline of the user's body (**medial**).

5. Discussion

Although the angle-based and point-based hand models are equivalent mathematically, they are not equivalent pragmatically. Equipped with hand tracking devices measuring hand joint angles, we are able to calculate the six reference frames of the point-based model by matrix multiplication in real time, based on the raw data acquired. However, the accumulated errors occurring in the matrix computation may jeopardize the usefulness of the 6 DOF values of fingertips. Furthermore, the difficulty of measuring the lowest joint of the thumb may invalidate the 6 DOF values of the thumb tip. Also, the delay induced by the matrix computation may deteriorate the response time of gestural commanding.

Directly acquiring the 6 DOF values of the point-based hand model may relieve the above problem greatly. In this case, the abstract digit-oriented handshape information can be solved by table lookup. Some optics-based hand trackers trace the tips of the hand digits. However, they cannot solve the problem of the obstruction of fingertips.

We use a VPL DataGlove [Zimmerman *et al.*, 1987] and a Polhemus ISOTRAK on the glove to verify our notion of the logical hand device. Body Electric software running on a Mac has been used to develop routines to reconize gestures. Early

experiences tell us that the DataGlove may be good for extracting the digit-oriented handshape information except for that of the thumb. Therefore, one more point tracker, like an ISOTRACK, may be needed to put on the thumb tip. Developing a fully point-based hand tracking device is being considered since it will be more precise to provide the point-based information that the manipulation paradigm needs, and it will not be hard to compute the abstract digit-oriented information by table lookup.

6. Conclusions

Virtual Reality environments need more software support at all levels of the development and implementation than currently available. Also, we need to devise new concrete concepts for devices and tasks of virtual environments. In this paper we contribute to this area. We have proposed a general-purpose and device-independent logical hand device that supports a comprehensive use of 3D gestural input. The abstract digit-oriented handshape information lays a concrete foundation for gesture description language to define 3D static gesture sets used by other systems. We believe that our abstract digit-oriented handshape information also has some linguistical significance in the research on American Sign Language.

References

Alexander, R. M. [1992] *The Human Machine* (Columbia University Press, New York) Chap. 2, pp. 18–34.

Baudel, T. & Beaudouin-Lafon, M. [1993] "Charade: Remote control of objects using free-hand gestures," *Communications of ACM* **36**(7), pp. 28–35, July.

Fels, S. S. & Hinton, G. E. [1993] "Glove-talk: A neural network interface between a data-glove and a speech synthesizer," *IEEE Trans. Neural Network* **4**(1), pp. 2–8, January.

Fisher, S. S., McGreevy, M., Humphries, J. & Robinett, W. [1986] "Virtual environment display system," *Proc. 1986 Workshop on Interactive 3D Graphics*, pp. 77–87, Chapel Hill, NC.

Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F [1990] *Computer Graphics: principles and practice* (Addison-Wesley) Chap. 8, 2nd ed.

Rijpkema, H. & Girard, M. [1991] "Computer animation of knowledge-based human grasping", *Proc. ACM SIGGRAPH'91*, pp. 339-348.

Scheifler, R. W. & Gettys, J. [1986] "The X window system", *ACM Trans. Graphics* **5**(2), pp. 79-109, April.

Spence, A. P. & Mason, E. B. [1987] *Human Anatomy and Physiology* (Benjamin Cummings, Menlo Park, California) Chap. 7, 3rd ed.

Stokoe, W. C., Casterline, D. C. & Croneberg, C. G. [1976] *A Dictionary of American Sign Language on Linguistic Principles* (Linstok Press, Silver Spring, Maryland) New ed.

Sturman, D. J., Zeltzer, D. & Pieper, S. [1989] "Hands-on interaction with virtual environments", *Proc. ACM User Interface Software & Technology*, pp. 19-24.

Sturman, D. J. [1992] *Whole-hand Input*, PhD thesis, Media Arts & Sciences, MIT, Chap. 6.

Sturman, D. J. & Zeltzer, D. [1994] "A survey of glove-based input", *IEEE Computer Graphics & Applications* **14**(1), pp. 30–39, January.

Su, S. A. [1993] *Hand Modeling in Virtual Environments*, M.S. scholarly paper, Department of Computer Science, University of Maryland, College Park, Maryland.

Su, S. A. & Furuta, R. [1993] "The virtual panel architecture: a 3D gesture framework", *Proc. IEEE Virtual Reality Annual Int. Symp.*, pp. 387-393.

Weimer, D. & Ganapathy, S. K. [1989] "A synthetic visual environment with hand gesturing and voice input", *Proc. ACM CHI'89*, pp. 235–240.

Wilbur, R. B. [1987] *American Sign Language: Linguistic and Applied Dimensions* (Little, Brown & Co., Boston) Chap. 2, 2nd ed.

Zimmerman, T., Lanier, J., Bryson, S., Blanchard, C. & Harvill, Y. [1987] "A hand gesture interface device", *Proc. CHI+GI*, pp. 189–192.