Human Computer Interaction via the Human Hand: A Hand Model

James J. Kuch* and Thomas S. Huang

Beckman Institute, University of Illinois
405 N. Mathews Urbana, IL 61801 USA
*now with TouchVision Systems, Inc. Chicago IL, 60631

Abstract

In this paper, a new method for building lifelike hand models which articulate in a realistic manner is presented. This method has distinct benefits over previous methods, since the fitting to a particular person's hand is quick, simple and very accurate. Following the calibration process which fits it to a particular persons hand, this hand model can be used in numerous HCI scenarios.

Calibration is based on anatomical studies of the human hand and on the specific method of recognition to be employed in the HCI scenarios. The calibration method is done visually and requires only three views of the hand to be modeled. The calibration system is designed to be accurate, to be easy to use and to allow for a short calibration time. These characteristics are all desirable when one is working in the realm of a human computer interfacing.

Introduction

Human computer interaction (HCI) of the future will entail speech and vision as a means of communicating with machines. Using vision provides a natural, non-invasive approach to HCI for tomorrow's world. This paper presents a new approach to human hand modeling for the task of HCI. Since HCI involves both analysis and synthesis, the model is therefore general and accurate enough to be used in computer vision and computer graphics alone.

Model based tracking of a real human hand using the presented hand model has been successfully demonstrated with the results reported in [1].

Previous Work

Modeling of the human hand has been approached by a number of disciplines: computer graphics, medicine and computer vision. The models resulting from these methods all have their merits and have been designed to serve a specific function. In computer graphics, the primary goal is the most realistic looking hand model possible for use in computer animation. Skin deformations at the joints, lifelike texture and realistic joint movements are all critical issues, with each having a great deal of research devoted to them [2] - [6]. For the more sophisticated models that incorporate such advanced features such as skin bulging, computational complexity prohibits real-time animation.

Typical schemes for acquiring the 3-D data used in computer graphic models include electromagnetic sensors, such as the Polhemus 3SPACE [7], and direct physical measurement of a particular person's hand [8]. The former results in a very accurate 3-D reconstruction, yet usually requires a plaster cast of the person's hand due to the length of the data acquisition process. In acquiring the 3-D data, the user is required to hold his hand still in a single pose for several minutes. This is quite a difficult task, even for professional actors, therefore, the use of plaster casts is necessary.

The latter of the two schemes in computer graphic modeling, direct measurement, lends itself to large surface errors due to the physical contact of the measuring device made with the skin during the measurements. In addition, there is usually a lack of 3-D data points (relative to the Polhemus method) to provide an accurate 3-D model. Inaccurate models or long acquisition times are clearly short comings of this modeling method.

A majority of the medical models, on the other hand, are concerned with accurate bone and tendon models as well as accurate joint articulation [9] - [13]. Most 3-D data used in these models, and subsequently their renderings, are acquired by very sophisticated and expensive means such as X-rays or computed tomography (CT) [12], [13], with others acquired from cadavers [11], [12]. Here, modeling of the skin, if used at all, is not critical, since most models are used primarily for clinical and educational applications in studying hand biomechanics [11], [13]. Still, other medical researchers model only the quantitative aspects of the hand such as tendon displacement and the range of motion (ROM) of joints [14]. Therefore, the resultant models are rich in the motion and articulation of the hand, yet they have no physical 3-D information of the hand. Often computer graphic modelers will simplify these quantitative results and apply them to their models [15].

1058-6393/95 $4.00 © 1995 IEEE

Hand modeling in the case of computer vision is rather unique. Here, the goal is to understand or track what a real hand is doing from one or more cameras. Typically, the model is chosen based on what features in the image will be used to track the hand. Thus, models vary greatly in this area of research and usually are not related to those in computer graphics. In [16], the authors used voxels to model the human hand. They used up to ten images of a fixed hand (plaster cast) at different viewpoints to reconstruct a volumetric model. The model was acquired by back projecting the 2-D images onto one another. The difficulty of this method was in the need for many cameras to acquire images simultaneously. This is acceptable for a research scenario, but in the realm of HCI, it is quite impractical. Most other computer vision modeling of the hand has not explicitly had a 3-D model other than a stick figure. This method is similar to the quantitative research found in medicine. Accurate hand articulation is used to constrain the movement of the stick figure and to aid in searching for key features in the images of the hand to be tracked. Such an example can be found in [15]. The authors used a special glove with seven color coded marking delineating such features on the hands as fingertips. Easily finding these color markings in the images, or image features, along with the constraints on the articulation of the human hand allows recognition of several hand configurations or gestures.

Hand model building in this paper is the marriage of all three preceding methods. Again, the underlying objective is to recognize and track what a real hand is doing, as viewed by a video camera (see [1]). Thus the model should correspond to what a video camera would see. This correspondence entails using a computer graphics quality hand model that is tailored in some aspects in order to achieve real-time recognition. Also, the model should be easy to calibrate to any person's hand in order that we do not restrict or limit anyone from using the HCI system. These issues and other important modeling issues will be addressed in the next sections.

Model Requirements

As stated in the introduction, the hand model developed here is intended to be used in a model based recognition system for HCI. Since during the tracking portion of the system we will be comparing two 2-D projections, one received from the camera and one generated from projecting the 3-D hand model, an important aspect of our model is that it renders accurate projections (see [1]). This rendering requires that the model have accurate surface characteristics in all areas that can be seen in a projection. The concavity and the fleshy bulges of the palm, for example, need not be modeled very accurately, since they will not be seen in a projection. Other requirements of the model are accurate joint location and motion which include interdependent finger movement and range of motion limitations. These constraints on the model will not only result in realistic movement of the model but will also reduce our search space when we perform our tracking.

If we desire to have the HCI system work with a diverse group of people, it is necessary to have a generic model which is sufficiently flexible to fit any hand. Since we also desire real-time rendering of the hand model for tracking and displaying purposes, such as in video teleconferencing, time consuming features such as skin deformations will not be addressed. Cost is also an important issue and should be kept to a minimum during the calibration procedure, which implies avoidance of special hardware such as multiple cameras or 3-D space digitizers. Last is the issue of calibration ease. If the model is to be used for HCI, a new user should be able to start interacting with the computer immediately and not have to wait for a long calibration process. Therefore, the hand model must lend itself to quick calibration. This requirement indirectly limits the number of calibration steps as well as the computational intensity of each step. With the above list of requirements established, the model can now be described in detail.

Model Description

To solve the problems of fitting the model to an arbitrary hand as well as having accurate surface characteristics, cubic B-splines are used to represent the individual surfaces of the palm, fingers and thumb. The use of B-splines allows the rendering of smooth surfaces, while allowing the calibration system to keep track of a smaller set of control points versus every vertex in the model.[1] The current model uses 300 control points for an entire hand opposed to over several thousand vertices using standard polygonal mesh techniques. Even with the reduced complexity that B-splines afford a programmer, they can now also be rendered in real time, a necessary condition for our hand tracking system.

The reduction on the number of controlling points in the hand model also greatly reduces the calibration time, since fewer points will be available for modification during the calibration process. By making simple adjustments to the control points, we can obtain many smooth variations in the generic model to adjust for any given human hand. Realistic bending of the fingers is also achieved without increase in the model size, calibration time or rendering time. In addition, as computer speeds increase, skin bulging can be easily incorporated into the model by the

[1] Development is being done on Silicon Graphics machines utilizing a NURBS surface function which renders a tessellated surface based on the given set of control points.

movement of additional control points as described by Komatsu [17].

Within the hand model, there are a total of 23 degrees of freedom (DOF) which are all based on medical and anatomical analysis of the hand [9] - [11], [18].

Each of the four fingers is given four DOF, one each at the distal interphalangeal (DIP) and the proximal interphalangeal (PIP) (refer to Figure 1 for hand notation). The remaining two DOF are located at the metacarpophalangeal (MCP). With five DOF, the thumb has two at the trapeziometacarpal (TM), also referred to as the carpometacarpal, and two at the metacarpophalangeal. The remaining DOF of the thumb is at the interphalangeal (IP). The palm is given two internal DOF located at the base of the forth and fifth (ring and pinky) metacarpals. These last two DOF reflect the ability of the palm to fold or curve, as when one brings the pinky across the palm toward the thumb [18]. In addition to the 23 DOF internal to the hand, the base of the palm has three DOF. These DOF determine the overall orientation in space of the entire hand. The palm's DOF thus brings the total DOF for a given human hand to 26.

The initial location of each joint and its axis of rotation is roughly approximated from anatomical analysis of a typical hand and is applied to the generic model. Later we will refine these locations and axes as well as detail the shaping of each of the surfaces of the hand to match that of a real human hand.

Articulation of the hand's internal 23 DOF is achieved through a combination of flexions, extensions, abductions and adductions. Flexion and extension are used to describe rotations toward and away from the palm, respectively. Flexion and extension occur at every joint within the hand. Abduction is the movement of separation (e.g., spreading fingers apart) and adduction is the movement of approximation (e.g., bringing fingers together). These movements occur at each finger's MCP and also at the MCP and TM of the thumb. All the above rotations are easily implemented in the model by simply rotating the B-spline control points which correspond to that particular link. Within each finger and thumb, there exists a set of control points which enable the model to bend in a realistic fashion. These additional points are located around each PIP and DIP and give the effect of knuckles.

We are now ready to describe the constraints which control the above DOF and to allow our model to articulate in a manner similar to that of a real human hand.

Model Constraints

The results from both medical [9] - [11], [18] and computer graphics research [15] are applied directly to our model to satisfy the model requirements mentioned earlier and others.

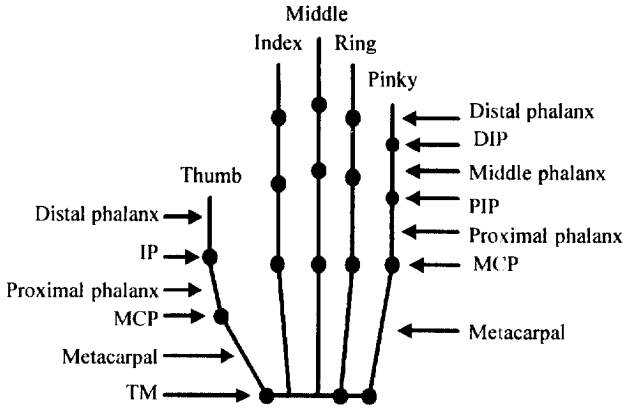

Figure 1. Notational diagram for the human hand.

The set of constraints on the hand can be subdivided into static and dynamic constraints. Static constraints are independent of the hand's pose. Static constraints include all joint ROM limits, joint length and location, and finger MCP flexion convergence angle. All these preceding constraints are set interactively during the calibration procedure for each human hand the HCI system is to track.

Dynamic constraints have to be updated every time a joint on the hand is moved. Dynamic constraints are derived from the tendons within a given hand, which is in contrast to static constraints which are primarily derived from the hand's bone structure and are fixed in the calibration procedure. An example of such a dynamic constraint is the convergence of the fingers upon flexion [15], [18] (see Figure 2).

As the fingers flex downward toward the palm, their ability to abduct or adduct is reduced. This ability can be mathematically described by the following equation:

$$MCP^{lim}_{(a/a)} = \frac{MCP(f/e)}{90} * (MCP^{converge} - MCP^{s\_lim}_{(a/a)}) + MCP^{s\_lim}_{(a/a)}.$$

Qualitatively, as the $MCP_{(f/e)}$ is flexed, its abduction/adduction angle, $(MCP_{(a/a)})$, linearly approaches $MCP^{converge}$, the MCP flexion convergence angle, one of the static constraints mentioned earlier. Each MCP has two $MCP^{lim}_{(a/a)}$, an upper and a lower, which are also static constraints. These two values limit the abduction and ad-


Figure 2. MCP flexion convergence angle.

duction at a given MCP. In the extreme case, when the MCP is flex past 90 degrees, the MCP can no longer abduct or adduct. Conversely, when the MCP is not flexed, zero degrees, the MCP can fully abduct and adduct up to their predetermined static limits, $MCP^{s\_lim}_{(a/a)}$.

One other dynamic constraint utilized is the relationship between each finger's PIP and DIP. Both anatomical and empirical studies [18] show a near linear relationship between these two joints. The following equation was used to represent this dynamic constraint:

$$PIP = \frac{3}{2} DIP.$$

This equation follows from [15] and results in a natural curling motion of the distal portion of the fingers. The application of this constraint to the hand reduces the number of internal DOF by four, down to 19.

In addition to the above dynamic constraints of the fingers, one other constraint was added which related each finger's MCP flexion to its PIP flexion:

$$MCP(f/e) = \frac{1}{2} PIP.$$

This constraint simply states that the MCP(f/e) cannot flex without the PIP in turn flexing.

The result of combining the two finger flexion constraints can be seen in Figure 3. These two constraints provide lifelike flexing of an individual finger. Also, as with the PIP/DIP constraint, the MCP/PIP constraint reduces the number of internal DOF on the hand by four to a total of 15.

Constraints on the thumb to account for natural motion and in turn to reduce the total number of DOF are similar to those applied to the fingers.

With the generic model fully described, we can now address its calibration from real images to that of the human hand we wish to animate or track.

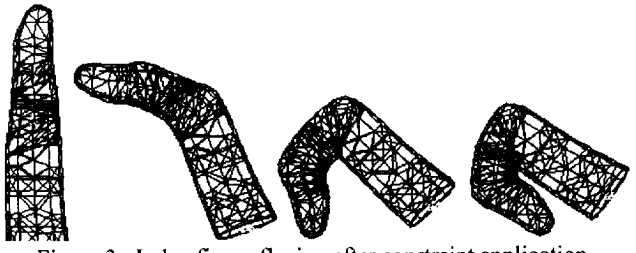

Figure 3. Index finger flexion after constraint application.

Hand Model Calibration

Calibration of the generic model to a real hand is done visually. A clear benefit of this method of calibration is the fact that we are already using a camera for tracking [1], thus there is no additional cost incurred. More important is the premise that a 3-D model acquired visually from 2-D images will yield accurate 2-D projections, or images. The

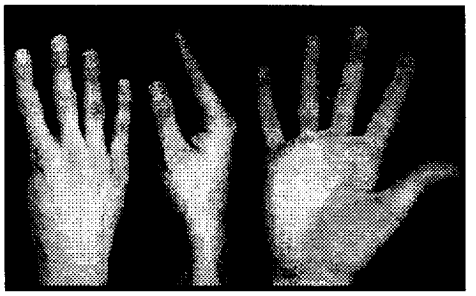

Figure 4. The three views of a hand needed for calibr ation.

calibration procedure requires only three specific views (Figure 4) of the real hand and does not place extreme pose restrictions on the person being modeled; neither does it require multiple cameras or other special hardware. The reason only three views are required lies in the richness of the underlying model described earlier.

The calibration step is broken into two sections: interactive selection and automatic fitting. The interactive selection portion is used to locate all the joints and to delineate the portion currently being fitted from the background. Performing this step automatically would be an extremely difficult vision task in itself, since most joint locations are determined from the cutaneous markings on the hand [18]. Such slight creases in the skin would be very difficult to detect accurately from a computer vision perspective. Thus, we leave this step as interactive for now. During the interactive selection, two static constraints are updated automatically: joint length and joint location. The remaining static constraints are applied at the end of the calibration procedure. There are four interactive sections, each followed by an automatic fitting stage, which accounts for the smooth contours which make up the surface of the final hand model. Calibration starts with the palm.

Following the interactive marking of the hand, the system then automatically fits the palm and then the fingers. Figure 5 shows the sequence of events in the complete dorsal plane calibration.

The palm and fingers now must be fitted in the radial plane. Since the palm has roughly the same thickness at the ulnar (pinky) side as it does on the radial (thumb) side, excluding the contribution due to the thumb itself, the calibration values for depth obtained are used across the entire

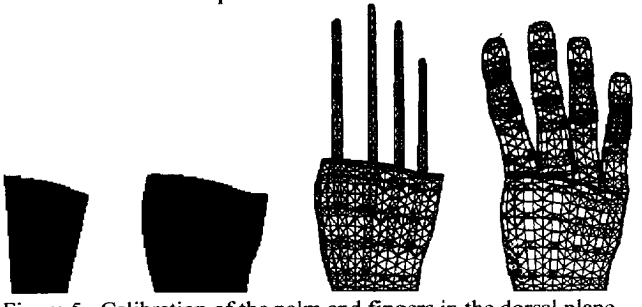

Figure 5. Calibration of the palm and fingers in the dorsal plane.

palm. Again, recall that modeling the concave, fleshy center portion of the palm is not important, since it will not show up in the projections of the real hand. After calibration, the palm can then be curved or bent at the forth and fifth metacarpals to reflect the true curvature of the palm. Before the automatic calibration step, the user aligns the palm and index finger model to the real image and delineates the palm border from the thumb. The hand model is calibrated for depth using the same technique as for dorsal calibration with one modification: only the palm and index finger are calibrated directly. From there, each finger uses the calibrated index finger's depth values as a basis and is scaled accordingly. This generalization is justified by the fact that all the fingers are relatively similar in depth, as well as the fact that a majority of occlusions will involve the depth of a finger being occluded by either other fingers, the thumb, the palm, or any combination of the three.

Calibration of the thumb is done in a similar manner as the individual fingers.

With the hand model completely calibrated, the two remaining static constraints are applied to the model: joint range of motion and MCP flexion convergence angle. Since these constraints are very similar from one person to the next and are well documented in medical literature [12], [18], these values typically do not change from one person to another. The fully calibrated hand model for the author is shown from several views in Figure 6.

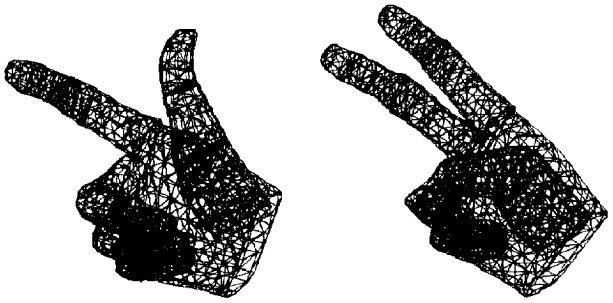

Figure 6. Two views of the fully calibrated hand model.

Conclusions

There are numerous areas to pursue, which could greatly improve the current system. Improvements to the hand model include the addition of finger webbing to the model. Other improvements include additional finger constraints which incorporate interfinger dependency [15], [18] and the addition of texture to the model's surface.

A fully automated calibration system also is under consideration. Given a new hand to track, this system would start with the best matching pre-calibrated model already in the system. From there, the system will modify the existing model using the standard three views and the addition of motion in order to accurately obtain the joint locations.

Acknowledgments

This work was supported in part by National Science Foundation Grant IRI-89-08255 and in part by a Grant from Sumitomo Electric, Inc.

References

[1] J. Kuch and T. Huang, "*Virtual Gun:* A Vision Based Human Computer Interface Using the Human Hand," *Proc. IAPR Workshop on Mach. Vision App,* Tokyo, 1994

[2] N. Badler and M. Morris, "Modeling flexible articulated objects," *Proc. Comp. Graphics '82*, On-line Conference., pp. 305-314, 1982.

[3] E. Catmull, "A system for computer-generated movies," *Proceedings ACM Annual Conf.*, vol. 1, pp. 41-52, 1972.

[4] J. Gourret, N. M. Thalmann and D. Thalmann, "Simulation of object and skin deformations in a grasping task," *Computer Graphics*, vol. 23, no. 3, pp. 21-30, July 1989.

[5] N. Magnenat-Thalmann, R. Laperriere and D. Thalmann, "Joint-dependent local deformations for hand animation and object grasping," *Proc. Graphics Interface '88*, Edmonton, Canada, pp. 26-33, 1988.

[6] H. Rijpkema and M. Girard, "Computer animation of knowledge-based human grasping," *Computer Graphics*, vol. 25, no. 4, pp. 339-348, July 1991.

[7] J. Foley, A. van Dam, S. Feiner and J. Hughes, *Computer Graphics: Principles and Practice.* Reading, MA: Addison-Wesley Publishing Company, 1990.

[8] J. Rehg and T. Kanade, "DigitEyes: vision-based human hand tracking," CMU-CS-93-220, Carnegie Mellon University, Pittsburgh, PA, December 1993.

[9] J. Agee, P. Brand and D. Thompson, "The moment arms of the carpometacarpal joint of the thumb: Their laboratory determination and clinical application," *Journal of Hand Surgery*, vol. 7, no. 4, pp. 412-413, 1982.

[10] J. Agee, A. Hollister and F. King, "The longitudinal axis of rotation of the finger MP joint," *Journal of Hand Surgery*, vol. 11A, no. 5, p. 767, 1986.

[11] W. Buford, L. Myers and A. Hollister, "3-D computer graphics simulation of thumb joint kinematics," *Proceedings Annual International Conference IEEE Engineering in Medicine and Biology Society*, November 1989.

[12] W. Cooney, M. Lucca, E. Chao and R. Linscheid, "The kinesiology of the thumb trapeziometacarpal joint," *The Journal of Bone and Joint Surgery*, vol. 63-A, no. 9, pp. 1371-1381, December. 1981.

[13] D. Thompson, W. Buford, L. Myers, D. Giurintano and J. Brewer, "A hand biomechanics workstation," *Computer Graphics*, vol. 22, no. 4, pp. 335-343, August 1988.

[14] D. Thompson, "Biomechanics of the hand," *Perspectives in Computing*, vol. 1, no. 3, pp. 12-19, October 1981.

[15] J. Lee and T. Kunii, "Constraint-based hand animation," in *Models and Techniques in Computer Animation*, Tokyo: Springer-Verlag, pp. 110-127, 1993.

[16] E. Petajan, D. Baraff and J. Weil, "Human hand model calibration using 3D reconstruction from multiple silhouette views," *SPIE*, vol. 902, pp. 87-91, 1988.

[17] K. Komatsu, "Human skin model capable of natural shape variation," *The Visual Comp.*, no. 3, pp. 265-271, 1988.

[18] R. Tubina, *The Hand*, vol. 1. Philadelphia, PA: Sanders, 1981.