Hand Gesture Modeling, Analysis, and Synthesis

Thomas S. Huang and Vladimir I. Pavlović
University of Illinois
Beckman Institute
405 N. Mathews Ave.
Urbana, IL 61801
U.S.A.
E-mail: huang@ifp.uiuc.edu

Abstract

Hand gestures are a form of communication among people. Yet we still limit human-computer interaction to cumbersome mice movements. The use of hand gestures in the field of human-computer interaction has attracted new interest in the past several years. Special glove-based devices have been developed to analyze finger and hand motion and use them to manipulate and explore virtual worlds. To further enrich the naturalness of the interaction, different computer vision-based techniques have been brought into use. At the same time the need for more efficient systems has resulted in new gesture modeling approaches. In this paper we present a review of the most recent work related to hand gesture modeling, analysis and synthesis. We describe four major classes of hand gesture interface techniques: glove-based techniques, vision-based techniques, techniques that use drawing gestures, and other gesture analysis techniques. A brief description of more than thirty gesture-interface systems is presented with special attention focused on the systems that use computer vision techniques.

1 Introduction

Human society lives through the communication among its entities. We communicate through our speech, our bodies and our eyes. Ceremonies, dances, and religious rituals combine all three aspects. However everyday life only seemingly moves the focus of the communication to our eyes and speech. Think about body movements for a moment. We shrug our shoulders when in doubt, we wave at familiar faces, we point at things we want to explore. These are all the body motions we simply call *gestures*. Even though they are expressed through the whole human body, gestures are still mostly related to an exceptional tool - the human hand. We use it to act on the world, to grasp and explore objects, to express our ideas. We communicate through *hand gestures*.

Computers and computerized machines have become a new element of our society. They increasingly influence our day-to-day lives. Yet we still interact with them using "stone-age" devices - keyboards, mice and joysticks. Ever since the early days of computers we have been attempting to make them understand our speech. But only in the last several years has there been an increased interest in trying to introduce the other means of human-to-human interaction to the field of human-to-computer interaction. This coincides with a growing interest in a closely related field - virtual reality. It became obvious that it is very acceptable to point at virtual space using the index finger and to explore virtual objects using one's hands [1]. It is also easier to understand other people in the same virtual environment if we can see them manipulate objects in it. Today, teleconferencing systems provide us with an opportunity to see and hear other people thousands of miles away, but they leave us short of real collaborative work. That is why virtual reality has grown into another new dimension: *televirtuality*, a virtual world that truly allows us to communicate with the others [2, 3, 4, 5, 6].

This paper describes recent advancements in hand gesture modeling, analysis and synthesis, and their applications. Particular attention is focused on techniques that use computer vision since these, in the authors' opinion, result in a more natural interaction between the humans and computers.

2 Gesture Modeling

"Every gesture is the physical expression of a mental concept" [7]. A gesture is motivated by an intention to perform a certain task: indication, rejection, grasping, drawing a flower, or something as simple as scratching one's hand. From the initial intention to the final performance, gesture follows a characteristic motor pattern in space and time. Kendon distinguishes three motion phases that comprise a single gesture: preparation, stroke and retraction [8]. Properties of this pattern are universal and permanent, and can, therefore, be used to describe any particular gesture. Quek [9] developed the following set of rules for gesture segmentation based on the gesture pattern: 1. Gestures are contained in movements that start with a slow initial move from the rest position, continue with a phase with substantially increased speed (the stroke), and end by returning to the rest position. 2. The hand assumes a particular configuration during the stroke. 3. Slow motions between resting positions are not gestures. 4. Hand gestures should be constrained within a certain volume - workspace. 5. Static hand gestures require a finite period of time to be recognized. 6. Repetitive movements can be gestures. He also suggests that, except for a few signs in the American Sign Language (ASL), the movement of individual fingers can have a gestural meaning only when the hand is still. This leads us to one way hand gestures can be modeled: Static hand gestures are characterized by the hand posture which is determined by a particular finger-thumb-palm configuration. Dynamic hand gestures are on the other hand characterized by the initial and final stroke hand configuration and the general stroke motion. Hand configuration can change during the motion but the change does not contain any gestural information and can therefore be disregarded.

Many of the gesture analysis systems developed in the past use the simpler, stationary gesture model. However, as it will become clear in the next section, more and more systems developed today try to explore the full potential of human gestures through the more complex dynamic hand gesture model.

3 Gesture Analysis

From the discussion of hand gesture modeling it should be quite obvious that the gesture analysis is not a trivial task. Gesture recognition should involve both temporal and spatial analysis. In addition to this, even though it is sometimes simple, the task of segmenting the hand from the other still or moving objects in the environment should not be neglected. Finally, interaction between the hand, virtual objects and speech should also be allowed.

There are three main approaches in the hand gesture analysis: glove-based analysis, vision-based analysis, and analysis of drawing gestures. There have also been attempts to use some other techniques for this purpose. We follow with the discussion of all of the mentioned approaches in more detail.

3.1 Glove-Based Analysis

Analysis of hand gestures using glove-based devices has been around since the late 1970's. Glove-based devices employ some kind of sensor (mechanical or optical) attached to a glove that transduces finger flexion and abduction into electrical signals for the purpose of determining the hand posture. The relative position of the hand is determined by an additional positional sensor (magnetic or acoustical) attached to the glove. A detailed survey of glove-based input devices can be found in [10].

One of the most widely used glove input devices today is the *DataGlove* by VPL Research. *DataGlove* uses optical fiber technology for flexion detection and a magnetic sensor for position tracking. The total number of degrees of freedom that the *DataGlove* allows is sixteen: ten flexions and six positional data. Some models of *DataGlove* also include sensors for finger abductions. A look-up table software toolkit is provided with the glove to be used for hand posture recognition. However, not all of the sensor data has to be used at the same time in order to recognize a certain vocabulary of gestures. David Quam has shown that for positive identification of fifteen ASL gestures, only four flexion sensors were needed [11]. Some of the systems that use the *DataGlove* are briefly described below in their chronological order.

Sturman, Zeltzer and Pieper developed a whole-hand interface to their virtual environment graphics system [12]. Based on the taxonomy they defined, the hand was used as a button, valuator, locator or pickup device. The system operating speed was between 3 and 6 fps[1].

An implementation of a multi-user *Virtual World* that combines hand gestures, speech, sound, stereoscopic graphics and head-motion parallax was presented by Codella et.al. [13]. They used the system to simulate a room containing flexible objects (*Rubber Rocks*) that can be created, grabbed, hit or shot by two users in real-time.

Glove-Talk is an interface between a user's hand and a speech synthesizer. It was developed by Fels and Hinton [14]. Using five neural networks they defined a 203 gesture-to-word vocabulary and used *Glove-Talk* to map complete hand gestures to complete words.

One of the most famous glove-based systems is *Charade* by Baudel and Beaudouin-Lafon [15]. *Charade* used hand gestures to control browsing in a hypertext presentation system. It ran in real-time and recognized sixteen gestural commands. All the commands consisted of three phases: start posture, dynamic phase, and end posture. The discrimination among different commands was based on the start posture and the dynamic phase.

Su and Furuta designed the *VirtualPanelArchitecture* (VPA) framework [16]. VPA combined elements of both a physical panel and a computer-based panel and enabled users to turn virtual buttons, move virtual sliders and point on virtual screens. Gesture Server of VPA used up to six temporal-spatial pieces of information to describe any hand gesture.

The *Virtual End-Effector* pointing system by Wang and Cannon was used to train and direct robots via hand gestures [17]. Using neural-network-based skeleton transform, they applied the system to workpiece inspection for surface flaw identification.

Figueiredo, Böhm and Teixeira's *GIVEN* (Gesture-based Interaction in Virtual ENvironments) system introduced techniques that enable the user to grab and surround virtual objects and thus provides very precise interaction [18]. They proposed the use of tactile sensors on *DataGlove* for an enhanced feeling of virtual space.

Responsive Workbench is a more recent concept developed by W. Krueger and Froehlich as an alternative to multimedia and virtual reality systems as they are known [19]. The system represents a virtual working environment that locates virtual objects and control tools on a real workbench and promotes collaboration between users working on the same project.

3.2 Vision-Based Analysis

Vision-based analysis of hand gestures is a most natural way of constructing a human-computer gestural interface, more so than glove-based analysis. It is based on the major way humans perceive information about their surroundings. Yet it is also the most difficult one to implement in a satisfactory manner because of the limitations in machine vision today. Several different approaches have been tested so far. The first and the most straight forward one is simply the use of a single video camera or a pair of cameras to acquire visual information about a person in some virtual environment and try to extract the necessary gestures. However, this approach faces several difficult problems: segmentation of the moving hand from sometimes complex environment, analysis of hand motion, tracking of hand position relative to the environment, and, finally, recognition of hand postures. To lower the burden of this analysis some of the systems rely on passive or active

[1] frames per second

markers or marked gloves. The others use restrictive setups: uniform background, very limited gesture vocabulary, or just a simple static posture analysis.

We now present an overview of several vision-analysis-based systems that have been developed and experimentally implemented. They include 3D hand-model-based analysis, analysis using markers and marked gloves, image-property-based analysis, and other vision-based techniques. Table 1 summarizes the most important aspects of these systems.

3.2.1 Three Dimensional Hand-Model-Based Analysis

One approach that has been used in hand gesture recognition is to build a three dimensional model of the human hand. The model is matched to images of the hand obtained by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. The parameters are then used to perform gesture classification.

Downton and Drouet developed a system for the human limb movement tracking in a sequence of images using an articulated generalized cylindrical human model [20]. Images were obtained from a single monochrome camera in a uniform background setup. Model-to-image matching was based on perspective projection of the cylindrical model and comparison with collinear edges obtained from the image. The output of the system was a kinematic specification of joint axis angles that could be used for sign language analysis.

The generalized cylinder model was also used by Etoh, Tomono and Kishino of ATR Research Labs [21]. They used stereo cameras and developed an algorithm that partitions an object into a hierarchical set of generalized cylinders. Using contours and a scale-space expansion technique, a set of local contour extrema points was found. These points were then used to find the axial and the contour lines associated with the cylinders. The system was used to model the human hand, but no specific gestural analysis was performed at that time.

A complete hand gesture recognition system named *DigitEyes* that uses a three dimensional cylindrical kinematic model of the human hand with 27 degrees of freedom was designed by Rehg and Kanade [22]. Finger tips and links were chosen as the model matching features and were extracted from either single or stereoscopic images using the edge-based analysis in the restrictive background. For feature tracking and model-parameter estimation they used a modified Gauss-Newton minimization of the residual feature error. The system was tested in a 3D graphical mouse application (single camera) and for the hand joint angle estimation (stereo cameras).

A hand gesture analysis system based on a three dimensional hand skeleton model with 27 degrees of freedom was developed by Lee and Kunii [23]. They incorporated five major constraints based on the human hand kinematics to reduce the model parameter space search. To simplify the model matching, specially marked gloves were also used. This model was used for the analysis of 16 ASL symbols in an iterative fashion and produced very small errors. However, computational time was considerably high.

Kuch formulated a 300 point NURBS-based hand model with 26 degrees of freedom that includes six constraints from the real hand kinematics [24]. The model can be fit to a particular user and then used to track complex hand gestures in a long sequence of single-camera images. This system was used for the Virtual Gun and the ASL tracking, and can also be used for hand gesture synthesis.

3.2.2 Analysis Using Markers and Marked Gloves

The human hand as a geometric shape is a highly non-convex volume. Trying to detect the hand configuration from camera images is therefore a difficult, if not an impossible, task. To overcome this problem some of the hand gesture recognition techniques use a system of markers. Markers are usually placed on the fingertips. They are colored in a manner easily detectable through the image histogram analysis. Once the markers are detected and tracked, the gesture recognition can be accomplished using several different classification techniques.

Torige and Kono designed a gesture recognition system that specifies the hand's moving direction [25]. Using stereoscopic cameras and black gloves with colored markers on fingertips, wrist, elbow and shoulder, they calculated the three dimensional finger position and motion parameters and used them to control a robot manipulator.

Another vision-based system that employs marked fingertips was developed by Davis and Shah [26]. By tracking the fingertips in multiple frames with uniform background, the system computed motion trajectories and used them to determine the start and the end position of the gesture. Each gesture was then modeled by a set of start-end vectors. At 4 fps, the system performed a successful time segmentation of seven predefined hand gestures. Davis and Shah also developed a technique for identification of fingertips based on the cylindrical fingertip model [27]. They used this model to determine the three dimensional hand motion.

The use of a specially marked glove for hand tracking was present in the system designed by Maggioni [28]. The glove had two slightly off-centered, differently colored circular regions. Using single camera images the system computed several image geometry parameters based on the first and second moments and used them to estimate hand position and orientation. Maggioni used this system as a module for *DIVE* (DIstributed Virtual Environment) by the Swedish Institute of Computer Science, and developed several object moving and exploring applications.

Cipolla, Okamoto and Kuno also used a glove with marked fingertips to determine hand motion and position [29]. From the motion parallax they calculated the translation and rotation of the human hand and used it in a feedback system to generate changes in the viewing position and orientation of a virtual object.

3.2.3 Image Property-Based Analysis

Several of the developed hand gesture recognition systems are based on the extraction of some of the properties that are associated with the images of hand postures. The analyzed properties range from basic geometric properties (the analysis of image moments) to the ones that are the result of a more complex analysis (Zernike moments and neural networks). What is common to all the approaches is that they do not result in the estimation of the real hand parameters (like joint angles). The systems that use this analysis are used for both simple hand tracking and more complex gesture classification.

A system that used simple sets of views to model hand gestures was developed by Darrell and Pentland [30]. Every hand gesture was represented by its own set of different views that were later matched to the gesture image sequence using temporal correlation and dynamic time wrapping. The speed of 10 fps was achieved using special-purpose correlation hardware.

The use of edge-based techniques to extract image parameters from simple silhouettes was utilized by Segen [31]. His system was able to discern ten distinct postures in real-time.

Ahmad and Tresp dealt with a general classification tasks in the case of missing and uncertain inputs [32]. They proposed a closed-form Gaussian neural network approximation to the Bayesian solution to the problem. The neural network was used to classify hand postures described by fingertip polar coordinates and the center of mass of the hand image. Based on seven different postures, the system managed to classify new inputs with the error rate of about 5% in the case of all features being present and about 50% in the case where 50% of the input information were missing. Ahmad also developed a real-time three dimensional hand tracker in a complex environment based on histogram segmentation [33]. A simple tracker based on three image geometry moments as well as the more complex 19 degree of freedom tracker were designed. The 19 degree of freedom tracker used a coarse hand model to extract fingertip locations and, based on that, the hand joint angles.

Another system that used image geometry parameters was employed by Starner and Pentland for ASL recognition [34]. Image geometry parameters were extracted from a sequence of images of a uniformly colored hand. A hidden Markov model five-state topology was selected for the gesture classification. Using the additional grammar a recognition rate of 85% was reported.

M. Krueger's *VIDEOPLACE*, *VIDEODESK* and *VIDEOTOUCH* are specialized virtual world systems based on the user's hand-silhouette geometry analysis [35]. The systems analyze the user's image, identifies the parts of his/her body, and let the user explore, modify or point at objects using his/her hands.

Schlenzig, Hunter and Jain used the Zernike moments as their image features [36]. Images were obtained by a single camera in a restrictive background. After the feature extraction, a hidden Markov model was used to recognize sequences of hand gestures. A vocabulary of six gestures was designed to remotely control a robotic gopher at a rate of 1/2 Hz.

Kjeldsen designed a system for the hand gesture control of a computer window interface [37]. Based on single-camera image-histogram segmentation, the system used the ALVINN neural network for posture differentiation and a set of rules for the gesture recognition. Pointing as well as several control gestures were used as an alternative to a standard mouse input device.

Recently, Freeman and Roth presented a simple and fast posture recognition system that used histograms of local orientations [38]. The system showed robustness to local changes in the lighting, but still required a uniform background. A computer graphics crane control was used as an application example.

3.2.4 Other Vision-Based Techniques

Finally, novel and specialized analysis approaches are employed in several other systems. The approaches are either specifically developed for the purpose of being used in the gesture-based systems or are general ideas that can be successfully applied in the hand gesture recognition arena.

Fukumoto, Suenaga and Mase of NTT Human Interface Laboratories designed a system for three dimensional pointing and simple gesture recognition that also supports a speech command interface [39]. The system used two camera views and simple algorithms to extract fingertips and the virtual projection origin (VPO), usually located near the user's shoulder. The VPO together with the index finger location is used to determine the pointing direction. The authors also developed a system of timing tags to synchronize speech commands with the pointing. Three applications were developed that employ the above system: a presentation system, a video browser and the *Space Writer*.

Hand posture modeling using the conjunction of local shape properties is an approach suggested by Cho and Dunn [40]. They developed a property-based learning algorithm for the conjunction learning that can also classify instances with partial information and a large number of classes. Straight line segments that can be extracted from the edge analysis were chosen as the local shapes. The system performed well in the classification of five different hand postures.

Furthermore, Kervrann and Heitz presented a general framework for the modeling and unsupervised training of deformation modes of nonrigid dynamic objects [41]. Using statistically described deformations of prototype shapes, it is possible to estimate them based on some optimization techniques. This system was used for hand tracking in both uniform and complex backgrounds.

3.3 Analysis of Drawing Gestures

Drawing gestures can be regarded as those gestures aimed at commanding a computer through a sequence of hand strokes. It usually involves the use of a stylus or computer mouse as an input device, but extends their use beyond the common "clicks" and "drags". Analysis of drawing gestures can also lead to on-line recognition of a written text.

GRANDMA (Gesture Recognizers Automated in a Novel Direct Manipulation Architecture) is a drawing-gesture application tool designed by Rubine [42]. It used the *Sensor Frame* as a multifinger input

Author	Task	Method Used	Camera	Speed
S. Ahmad	Hand tracking	Fingertip detection model	Mono	10-30fps
S. Ahmad	Posture recognition	Gaussian NN	Mono	n.av.
R. Cipolla	Gesture recognition	Marked fingertips	Mono	25 fps
K. Cho	Posture recognition	Local shape property learning	Mono	n.av.
T. Darrell	Gesture recognition	Set-of-views model	Mono	10 fps
J. Davis	Hand Tracking	Fingertip detection model	Mono	n.av.
J. Davis	Gesture recognition	Marked fingertips	Mono	4 fps
A. Downton	Limb tracking	3D cylindrical limb model	Mono	n.av.
M. Etoh	Hand tracking	3D cylindrical hand model	Stereo	n.av.
M. Fukumoto	Pointing	Finger detection and virtual projection origin	Stereo	real-time
W. Freeman	Gesture recognition	Orientation histograms	Mono	real-time
C. Kervrann	Hand tracking	Stochastic deformable model	Mono	n.av.
R. Kjeldsen	Gesture recognition	ALVINN	Mono	n.av.
M. Krueger	Object manipulation	Silhouette	Mono	30 fps
J. Kuch	Gesture recognition	3D NURBS hand model	Mono	3-30 s/frame
J. Lee	Gesture recognition	3D hand skeleton model	Mono	40-80 min/frame
C. Maggioni	Gesture recognition	Marked glove	Mono	25 fps
J. Rehg	Gesture recognition	3D cylindrical hand model	Stereo	10 fps
J. Schlenzig	Gesture recognition	Zernike moments	Mono	1/2 fps
J. Segen	Gesture recognition	Silhouette edges	n.av.	real-time
T. Starner	Gesture recognition	HMM image geometry parameters	Mono	5 fps
A. Torige	Pointing	Marked glove	Stereo	30 fps

Table 1: Vision-Based Hand Gesture Systems.

device for a drawing program where, for example, a two finger gesture automatically rotated, translated and scaled any chosen on-screen object. Rubine designed another system called *GSCORE* for musical score editing [43].

Yang, Xu and Chen developed a hidden Markov model based system for recognition of drawing gestures [44]. They use a computer mouse as the input device and recognize nine gestures corresponding to nine written digits.

3.4 Other Hand Gesture Analysis Techniques

Several other techniques have been tested in human-computer gesture type interaction applications. One such technique that provides potential for further exploration involves the use of electromyograms (EMG). An important advantage of EMG-based systems is that the time resolution exceeds the one of both glove- and vision-based systems. Putnam and Knapp constructed a real-time computer control system that uses EMG signals for neural-network-based gesture classification [45]. The system was used as an alternative input system for a graphical user interface.

4 Gesture Synthesis

Synthesis of hand gestures has, until recently, mostly been confined to computer graphics and animation fields. However, emerging new applications like the televirtuality and model-based video conferencing require both analysis and synthesis of hand gestures. Most of the systems gesture analysis in virtual environments use very simple three dimensional hand models, usually based on cylindrical approximations. Kinematics of these models is frequently disregarded and the whole model movement in a virtual space is based on several estimated parameters from the gesture analysis stage. Such models are probably sufficient for experimental applications, but will definitely have to be improved if they are to be used in the next generation of virtual environments.

The computer animation field requires a high degree of realism in any object movement. A project in the human body movement animation (*HUMAN FACTORY*) paid considerable attention to many details in order to produce a realistic virtual actor [46, 47, 48]. Rounding of joints, muscle inflations, and skin deformations were all considered and successfully modeled and resulted in natural grasping actions. Further work on the grasp analysis and modeling is continued in the *HUMANOID* project [49].

We should also mention here the hand gesture analysis system developed by Kuch [24]. The system is also capable of synthesizing almost natural hand movements based on its three dimensional NURBS model of the human hand that includes six kinematic constraints and can be calibrated to represent a particular user's hand.

5 Conclusion

Virtual worlds are already becoming a part of our present time. The necessity to outgrow old cumbersome interaction devices and manipulate and explore new virtual environments results in new approaches to human-computer interaction. The use of hand gestures in such tasks is influenced by a very common way humans communicate and interact with their natural environment. In this paper we have presented a review of the methods and systems that have been used to model, analyze and synthesize the hand gesture-based interfaces. Hand gesture models used in interface systems today incorporate not only hand posture estimation but also knowledge of the inherent dynamics of human hand gestures. The analysis of gestures increasingly relies on computer vision techniques as an unencumbered alternative to glove- and stylus-based approaches. Still, the present systems provide us with merely sufficient experimental laboratory tools. Future advances in computer vision together with the implementation of gesture dynamics will pave a path to the new world of human-computer interaction.

6 Acknowledgments

The authors would like to thank Karin McCutcheon, Jay Geistlinger and James Kuch for their suggestions. This work was supported in part by National Science Foundation Grant IRI-89-08255 and in part by a grant from Sumitomo Electric Industries.

References

[1] A. G. Hauptmann and P. McAvinney, "Gesture with speech for graphics manipulation," *International Journal of Man-Machine Studies*, vol. 38, pp. 231–249, Feb. 1993.

[2] P. Quéau, "Televirtuality: The merging of telecommunications and virtual reality," *Computers and Graphics*, vol. 17, no. 6, pp. 691–693, 1993.

[3] A. Nakajima, T. Sakairi, F. Ando, and M. Shinozaki, "A two-way dual-view teleteaching system conveying gestures and chalkboard contents," *IEICE Transactions on Information and Systems*, vol. E77-D, pp. 1335–1343, Dec. 1994.

[4] H. Ishii and M. Kobayashi, "Integration of interpersonal space and shared workspace: Clearboard design and experiments," *ACM Transactions on Information Systems*, vol. 11, pp. 349–375, Oct. 1993.

[5] H. Ishii, K. Arita, and M. Kobayashi, "Toward seamless collaboration media: From teamworkstation to clearboard," *NTT Review*, vol. 5, pp. 24–29, Jan. 1993.

[6] J. C. Tang and S. L. Minneman, "Videodraw: A video interface for collaborative drawing," *ACM Transactions on Information Systems*, vol. 9, pp. 170–184, April 1991.

[7] S. Thieffry, "Hand gestures," in *The Hand* (R. Tubiana, ed.), pp. 488–492, Philadelphia, PA: Sanders, 1981.

[8] A. Kendon, "Current issues in the study of gesture," in *The Biological Foundations of Gestures: Motor and Semiotic Aspects* (J.-L. Nespoulous, P. Peron, and A. R. Lecours, eds.), pp. 23–47, Lawrence Erlbaum Assoc., 1986.

[9] F. Quek, "Toward a vision-based hand gesture interface," in *Virtual Reality Software and Technology Conference*, Aug. 1994.

[10] D. J. Sturman and D. Zeltzer, "A survey of glove-based input," *IEEE Computer Graphics and Applications*, vol. 14, pp. 30–39, Jan. 1994.

[11] D. L. Quam, "Gesture recognition with a dataglove," in *Proc. of the 1990 IEEE National Aerospace and Electronics Conference*, vol. 2, 1990.

[12] D. J. Sturman, D. Zeltzer, and S. Pieper, "Hands-on interaction with virtual environments," in *Proc. of the ACM SIGGRAPH Symposium on User Interfaces*, (Williamsburg, VA), pp. 19–24, 1989.

[13] C. Codella, R. Jalili, L. Koved, and al., "Interactive simulation in a multi-person virtual world," in *ACM Conference on Human Factors in Computing Systems - CHI'92*, vol. 35, pp. 329–334, 1992.

[14] S. S. Fels and G. E. Hinton, "Glove-talk: A neural network interface between a data-glove and a speech synthesizer," *IEEE Transactions on Neural Networks*, vol. 4, pp. 2–8, Jan. 1993.

[15] T. Baudel and M. Beaudouin-Lafon, "Charade: Remote control of objects using free-hand gestures," *Communications of the ACM*, vol. 36, no. 7, pp. 28–35, 1993.

[16] S. A. Su and R. Furuta, "Virtual panel architecture: a 3d gesture framework," in *1993 IEEE Annual Virtual Reality International Symposium*, pp. 387–393, IEEE, 1993.

[17] C. Wang and D. J. Cannon, "A virtual end-effector pointing system in point-and-direct robotics for inspection of surface flaws using a neural network based skeleton transform," in *Proc. of IEEE International Conference on Robotics and Automation*, vol. 3, pp. 784–789, May 1993.

[18] M. Figueiredo, K. Böhm, and J. Teixeira, "Advanced interaction techniques in virtual environments," *Computers and Graphics*, vol. 17, no. 6, pp. 655–661, 1993.

[19] W. Krueger and B. Froehlich, "The responsive workbench," *IEEE Computer Graphics and Applications*, vol. 14, pp. 12–16, May 1994.

[20] A. C. Downton and H. Drouet, "Image analysis for model-based sign language coding," in *Progress in image analysis and processing II: Proc. of the 6th International Conference on Image Analysis and Processing*, pp. 637–644, 1991.

[21] M. Etoh, A. Tomono, and F. Kishino, "Stereo-based description by generalized cylinder complexes from occluding contours," *Systems and Computers in Japan*, vol. 22, no. 12, pp. 79–89, 1991.

[22] J. M. Rehg and T. Kanade, "Digiteyes: Vision-based human hand tracking," Tech. Rep. CMU-CS-93-220, School of Computer Science, Carnegie Mellon University, 1993.

[23] J. Lee and T. L. Kunii, "Constraint-based hand animation," in *Models and techniques in computer animation*, pp. 110–127, Tokyo: Springer-Verlag, 1993.

[24] J. J. Kuch and T. S. Huang, "Vision based hand modeling and tracking," to appear in *Proc. of International Conference on Computer Vision*, (Cambridge, MA), June 1995.

[25] A. Torige and T. Kono, "Human-interface by recognition of human gestures with image processing recognition of gesture to specify moving directions," in *IEEE International Workshop on Robot and Human Communication*, pp. 105–110, 1992.

[26] J. Davis and M. Shah, "Gesture recognition," Tech. Rep. CS-TR-93-11, Department of Computer Science, University of Central Florida, 1993.

[27] J. Davis and M. Shah, "Determining 3-d hand motion," in *Proc. of the 28th Asilomar Conference on Signals, Systems, and Computer*, 1994.

[28] C. Maggioni, "A novel gestural input device for virtual reality," in *1993 IEEE Annual Virtual Reality International Symposium*, pp. 118–124, IEEE, 1993.

[29] R. Cipolla, Y. Okamoto, and Y. Kuno, "Robust structure from motion using motion parallax," in *Proc. of International Conference on Computer Vision*, pp. 374–382, IEEE, 1993.

[30] T. Darrell and A. Pentland, "Space-time gestures," in *Proc. of Computer Vision and Pattern Recognition Conference*, 1993.

[31] J. Segen, "Controlling computers with gloveless gestures," in *Proc. of Virtual Reality Systems*, April 1993.

[32] S. Ahmad and V. Tresp, "Classification with missing and uncertain inputs," in *Proc. of 1993 International Conference on Neural Networks*, vol. 3, pp. 1949–1954, 1993.

[33] S. Ahmad, "A usable real-time 3d hand tracker," in *Proc. of the 28th Asilomar Conference on Signals, Systems, and Computer*, 1994.

[34] T. E. Starner and A. Pentland, "Visual recognition of american sign language using hidden markov models," to appear in *International Workshop on Automatic Face- and Gesture-Recognition IWAFGR95*, June 1995.

[35] M. W. Krueger, "Environmental technology: Making the real world virtual," *Communications of the ACM*, vol. 36, pp. 36–37, July 1993.

[36] J. Schlenzig, E. Hunter, and R. Jain, "Vision based hand gesture interpretation using recursive estimation," in *Proc. of the 28th Asilomar Conference on Signals, Systems, and Computer*, 1994.

[37] R. Kjeldsen, "Visual hand gesture interpretation," in *IEEE Computer Society Workshop on Non-Rigid and Articulate Motion*, (Austin, TX), Nov. 1994.

[38] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," to appear in *International Workshop on Automatic Face- and Gesture-Recognition IWAFGR95*, June 1995.

[39] M. Fukumoto, Y. Suenaga, and K. Mase, "Finger-pointer: Pointing interface by image processing," *Computers and Graphics*, vol. 18, no. 5, pp. 633–642, 1994.

[40] K. Cho and S. M. Dunn, "Learning shape classes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 882–888, Sept. 1994.

[41] C. Kervrann and F. Heitz, "Learning structure and deformation modes of nonrigid objects in long image sequences," to appear in *International Workshop on Automatic Face- and Gesture-Recognition IWAFGR95*, June 1995.

[42] D. Rubine, "Integrating gesture recognition and direct manipulation," in *Proc. of the Summer 1991 USENIX Technical Conference*, pp. 281–298, June 1991.

[43] D. Rubine, "Combining gestures and direct manipulation," in *ACM Conference on Human Factors in Computing Systems - CHI'92*, pp. 659–660, 1992.

[44] J. Yang, Y. Xu, and C. Chen, "Gesture interface: Modeling and learning," in *Proc. of the 1994 IEEE International Conference on Robotics and Automation*, pp. 1747–1752, IEEE, 1994.

[45] W. Putnam and R. B. Knapp, "Real-time computer control using pattern recognition of the electromyogram," in *Proc. of the 15th Annual International Conference on Engineering in Medicine and Biology Society*, vol. 15, pp. 1236–1237, IEEE, 1993.

[46] N. Magnenat-Thalmann, *Computer animation: theory and practice.* Tokyo; New York: Springer-Verlag, 2nd rev. ed., 1990.

[47] N. Magnenat-Thalmann, R. Laperrière, and D. Thalmann, "Joint-dependent local deformations for hand animation and object grasping," in *Proc. of Graphics Interface '88*, pp. 26–33, June 1988.

[48] J.-P. Gourret, N. Magnenat-Thalmann, and D. Thalmann, "Simulation of object and human skin deformations in a grasping task," *Computer Graphics*, vol. 23, pp. 21–30, July 1989.

[49] R. M. Sanso and D. Thalmann, "A hand control and automatic grasping system for synthetic actors," in *Proc. of EUROGRAPHICS '94*, vol. 13, pp. C–168–C–177, Eurographics Association, 1994.