LITERATURE REVIEW: GLOVE BASED INPUT AND THREE DIMENSIONAL VISION BASED INTERACTION

Submitted in partial fulfilment
of the requirements of the degree of

BACHELOR OF SCIENCE (HONOURS)

of Rhodes University

João Lourenço

Grahamstown, South Africa
29 October 2010

Abstract

This literature review takes a brief look at the research into human computer interaction with gloves, followed by an investigation into vision-based interaction, gesture recognition and some of the hardware used in the Wii3D System. A look at related work will then conclude the review.

Contents

Chapter 1

Literature Review

The Association for Computing Machinery defines human computer interaction as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them"[15]. It also states that the design and implementation of a human computer interactive system should draw on the relevant aspects from both humans and computers.

1.1 Human Computer Interaction using Glove Systems

Historically, gloves have been an interesting focus of research for human computer interaction. The first of the gloves started appearing in the late 1970's[25, 28].

The process of tracking a hand generally involves calculating some of the properties of the hand - position, orientation and pose. There are several documented methods for position tracking when using glove-based input[25]:

- Optical tracking (using marker systems or silhouette analysis)
- Magnetic tracking
- Acoustic tracking
- Circuitry tracking

1.1.1 Optical tracking

There are two main ways to achieve optical tracking - markers and silhouettes. The marker system uses multiple cameras to detect the markers. These markers are either infra-red LEDs that can either be constant or flash in a pattern. The silhouette method uses edge detection to extract the silhouette of the gesture. An analysis of the silhouette is used to determine the position and orientation of the elements of the hand[25].

The marker approach hinges on the ability to triangulate the three dimensional position of the markers in real time. The substantial processing power needed to make the necessary calculations at a rate that is sufficient for real time applications makes this approach difficult. The accuracy of the detection of the markers depends on the number of cameras used. However, as the number of cameras increases, the complexity of the calculation increases as there are more components in the linear algebra system[25].

A great deal of work has been done on natural gestures that are free from gloves, allowing the user a more liberated experience with computer interaction[19]. However, the silhouette approach still has several inherent issues[25]:

- The resolution of conventional video cameras is not high enough to capture the detail required for each individual finger and cover the field of view necessary for large motions
- The framerate of conventional video cameras (30 or 60 frames per second i.e. 33.33Hz or 16.67Hz) does not allow for the capture of rapid movements, while other devices, such as the Wii Remote have frequencies of up to 100Hz[31] and Selspot or Optotrak can operate at above 300Hz
- Parts of the hand can occlude other parts - the occlusion problem cannot be solved with a single camera
- Matchmaking (the ability to map objects in a scene to a three dimensional model) is an inexact science that is still in its infancy

These problems detract from the silhouette's freedom and often ensure researchers taking an alternate approach.

1.1.2 Magnetic Tracking

The magnetic tracking approach uses a source device that generates a magnetic field. A sensor reports its position and orientation in relation to the source. Multiple source and multiple sensor configurations allow the tracking to be more accurate. The magnetic systems that have been developed support polling frequencies of up to 100Hz. The primary advantage of this approach is that line of sight is not necessary for tracking purposes. However, objects with much higher magnetic permeability, such as metals, can cause interference in the magnetic field or fields that are generated by the source device or devices. This can lead to inaccuracies with the positions and orientations of the sensors[25].

1.1.3 Acoustic Tracking

Acoustic tracking is achieved by sending ultrasonic sounds from a source device, mounted on points of interest on the hand, and using receivers in the environment to measure the time taken for the sound to reach them. From this data, the tracking system can triangulate the position of the sources. Unfortunately, this approach requires line of sight from the sources to the receivers and acoustically reflective surfaces can cause interference with the system[25].

1.1.4 Circuitry Tracking

This technique uses hard-wired circuitry to detect touches, bends, and inertia etc. changes by monitoring the sensors built into the glove[25].

1.1.5 Glove Systems

There have been many gloves developed for use with a computer, each with its own merits and downfalls for a wide variety of applications [25, 28].

1.1.5.1 Sayre Glove

Richard Sayre postulated that a glove that used flexible tubes (not fibre optics), with a light source at one end and a photocell on the other could measure the extent to which a finger is bent. The reported voltage across the photocell is correlated to the finger bending[25, 28]. Thomas DeFanti and Daniel Sandin developed such a glove, which is both inexpensive and lightweight[29].

1.1.5.2 MIT LED Glove

The MIT Architecture Machine Group used a camera focused on an LED-studded glove to track limb position for real-time computer graphics animation. This glove, however, was designed and used for motion capture rather than a control device[25].

1.1.5.3 Digital Data Entry Glove

This glove used hard-wired circuitry that consisted of bend, touch and inertial sensors. Although not commercially developed, this system was developed to recognize 80 unique combinations of sensor readings mapped to a subset of the 96 printable ASCII characters from the gestures defined in the Single Hand Manual Alphabet for the American Deaf[10, 25, 28].

1.1.5.4 DataGlove

The DataGlove, developed by Thomas Zimmerman, used optical fibres to measure the angle of the bend in 10 of the finger joints of the hand to give a description with 6 degrees of freedom (position and orientation) of the hand gesture. The glove was constructed from Lycra and optical fibre that ran along the back of the finger joints. Each glove was calibrated at a per-user level, and this calibrated glove would then solve the configurations of the fingers using the analogue attenuation of the light in the optical fibres in the 10 flex sensors. The glove used a magnet that detected the orientation of the hand in three dimensional space. The glove uses a serial cable to transmit the flex and positional information to a computer[5, 25, 28].

This glove had several key advantages over its predecessors - it operated in real-time, did not require line of sight to a camera, and was lightweight and unobtrusive to the user. This glove was made commercially available by VPL Research at a reasonable cost, and resulted in quite widespread use around the world[25]. The accuracy of the glove's flex sensors was rated at $1^{\underline{o}}$, but research showed that it was closer to $5^{\underline{o}}$ or $10^{\underline{o}}$[4]. Furthermore, the operating frequency of 30Hz was not sufficient for precision gestures[25].

1.1.5.5 Dexterous Handmaster

This input device, developed by MIT for the control of Dexterous Hand robot, was far more accurate than the data glove[28], with its 20 degrees of freedom measured (4 per finger) by Hall Effect sensors as potentiometers at the joints[25]. However, the accuracy gained by the glove was at the expense of comfort, made of an intricate aluminium exoskeleton that was attached to the joints throughout the hand[28]. The glove accurately measures the bend of each of the joints in the fingers, the relative rotation of the each finger in relation to the hand and the complex motion of the thumb at 200Hz within 1° of flexion[25].

1.1.5.6 Power Glove

Nintendo, inspired by the VPL DataGlove, designed a glove for its gaming consoles that was constructed from moulded plastic and Lycra to allow flexible movement, with one resistive ink sensor per finger for flex detection. The glove used an acoustic unit mounted on the hand to track the glove in three dimensional space to one quarter of an inch using a television mounted acoustic sensor, with further trackers to determine the rotation of the hand[25].

The Power Glove was not particularly accurate, but its crude gesture recognition abilities were sufficient for the gaming application for which it was designed[25, 28].

1.1.5.7 CyberGlove

The CyberGlove was designed to translate American Sign Language into verbal English. It was constructed from 22 thin foil strain gauges sewn into thin fabric. The analogue signal are processed and converted into a digital streaming signal that is sent to a computer using a serial connection. The observed performance of the glove was smooth and stable, while retaining accuracy within 1° of flexion[25].

1.1.5.8 Space Glove

Virtual Entertainment Systems, a company in the development of arcade games, developed a glove for use with its arcade games that measure the flexion of the fingers using sensors that measure 1 degree of freedom per finger and 2 degrees of freedom on the thumb. This, in conjunction with the magnetic tracker in the back of the glove that tracks the gloves position in three dimensional space, is used for the gaming interface with other Virtual Entertainment Systems (previously W. Industries) arcade games[25].

1.2 Vision-Based Interaction

Gunnar Grape presented a system for computer vision which maps a hierarchy of features using a two dimensional prototype. The prototype models that were used were various projections of the three dimensional objects as a camera views them from a several different locations. The items in the scenes were limited to planar faced, convex objects. The objects that are recognized are then matched to generalizations of the prototype models [9]. This paper deals with many of the concepts of two dimensional vision-based interaction.

1.2.1 Stereoscopic Depth Perception

The ability of an organism to perceive depth is achieved using the composition of multiple perspectives of the same scene. Human vision has two perspectives that are processed by the brain to resolve an estimate of the third dimension[30], which is important, but not essential, to a human's function. Since the mid twentieth century, researchers have documented investigations into depth perception and judgments of distances in the real world[8, 16].

The stereoscopic depth perception problem is solved by deriving the points in three dimensional space relative to some predefined point in the space. In the world of computer vision, the methods for resolving points using stereo vision are[26]:

1. Production of a camera model: the position and orientation of cameras in three dimensional space
2. Position of matching point pairs: loci of corresponding features in the two pictures
3. Computation of the point in three dimensional space for each point pair
4. Presentation of the resultant depth information

A great deal of research into matching areas in stereo images has been done. Marsha Hannah[13] discusses measures of match which are suitable for the stereo vision matching of areas. She goes on to describe several methods for pruning the search space.

1.2.1.1 Triangulation

The linear triangulation of a point in $\mathbb{R}^3$ from two projections, whose views are known, is simple in a non-noisy environment using geometry[14]. Using Figure 1.1, for triangulation in 2 dimensions, let:

- A be Observer 1
- B be Observer 2
- S The observed point
- l The line between A and B
- d The perpendicular distance from the line between A and B and the S
- α The angle between the l and $\overrightarrow{AS}$
- β The angle between l and $\overrightarrow{BS}$

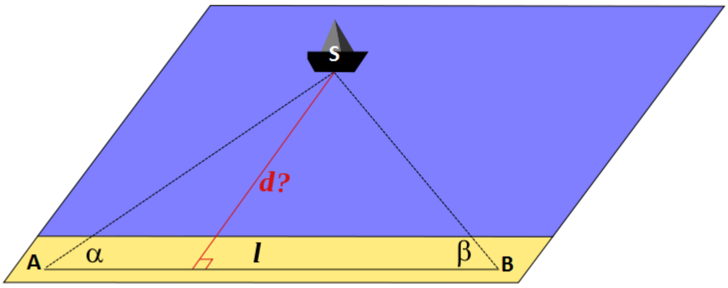

Figure 1.1: Triangulation[34]

Using Pythagoras' theorem:

$$l = \frac{1}{\tan\alpha} + \frac{1}{\tan\beta}$$

Therefore, d can be defined as:

$$d = \frac{l}{\left(\frac{1}{\tan\alpha} + \frac{1}{\tan\beta}\right)} = \frac{l \cdot \sin\alpha \cdot \sin\beta}{\sin(\alpha+\beta)}$$

This can be extended into the third dimension using both horizontal triangulation and vertical triangulation to calculate an estimate of an object's position in $\mathbb{R}^3$.

1.2.2 Matching Point Pairs

Hannah [13] described several ways to find the loci of two corresponding features in the stereo images. In the event of few similar features in each image, it is sufficient to simply do a permutation of pairwise operations on the points and choosing the pairs based on the minimization of the distance between the corresponding points.

1.2.3 Point Tracking

The study by Tziritas[11] shows methods of predictive interpolation by estimating the motion and structure of three dimensional objects from a sequence of images. The study discusses the estimation by using a recursive predictor based on the velocity vectors of a point in the three dimensional space. This predictor is based on a mixture of the previous velocities of the point in question.

In another paper[27], Tziritas goes on to investigate the problem of discontinuity detector which deals with occlusion and algorithmic discontinuities. This detection is attained by checking the errors between the predicted values and the observed values. The detection of such a discontinuity error results in the resetting of the system to start tracking the points from a newly initialized state. The point tracking system is summarized in Figure 1.2.

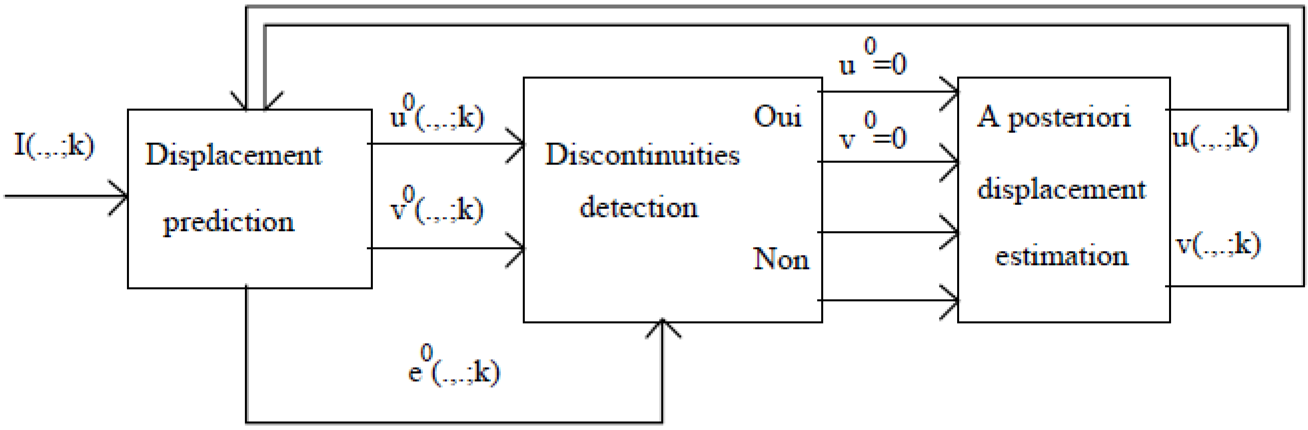

Figure 1.2: Predictive Interpolation

1.3 Gesture Recognition

Humans use gestures, especially hand gestures, for day to day communication. The gestures that humans use are ingrained from childhood, and gestures therefore have an inherently high level of intuitiveness. The use of gestures in computer software gives the user the ability to interact with a computer in a more natural and intuitive fashion.

Fu[7] states that:

> The problem of pattern recognition usually denotes a discrimination or classification of events.

A gesture recognizer uses the spatiotemporal changes as the gesture progresses for its discrimination/classification process[25]. A recognizer generally has three components:

- Encoding - the representation of the gesture
- Classification - the injection of the supported gestures into the recognizer using ideal situations and randomization or by example
- Recognition - the matching of observations to gestures

1.3.1 Encoding

The encoding (representation) of a gesture is important as the optimal transformation can result in a very efficient system. The recognition of shapes in gesture recognition can be achieved by encoding the data in such a way that the resultant encoding is a pattern that matches a specific gesture. The separation of gestures into categories, or families, allows a system to represent that family of gestures based on common features of the family[7].

1.3.1.1 Size Functions

Frosini[6] proposed a theory that sign language could be recognized by representing the shapes using size functions. The recognition of sign language has been successfully implemented by these representations[21].

A size function is generated by mapping the observations to some measurement system. The importance of choosing the correct measuring system is evident in the possibility of mapping different gestures to the same size function. The possibilities of observed gesture data are inherently infinite and the ability of the classifier to recognize a gesture depends on the size function's ability to map the 'same' gestures to the same encoding. Due to the physical nature of the system, the 'same' gesture can vary, globalized and localized, in its displacement, rotation and scale. These variances need to be taken into account by a process of normalization of the gesture[21]. A simple example of a size function is a mapping of all of the gesture's coordinates' distances from a reference point.

In Figure 1.3, an example of a decomposition of a gesture into a size function is shown. Figure 1.3(a) is the graph of some measuring function, φ. The shaded regions in Figure 1.3(b) and (c) identify the set of points with $\varphi \leq x$ and $\varphi \leq y$ in each of their graphs. In Figure 1.3(d), the darker shaded regions identify the set of points where $\varphi \leq x$ and $\varphi \leq y$ - the union of graphs (b) and (c). Figure 1.3(e) shows the resultant size function for all possible values of x and y. The labelled regions in the size function identify the value of the size function within the underlying region[21].

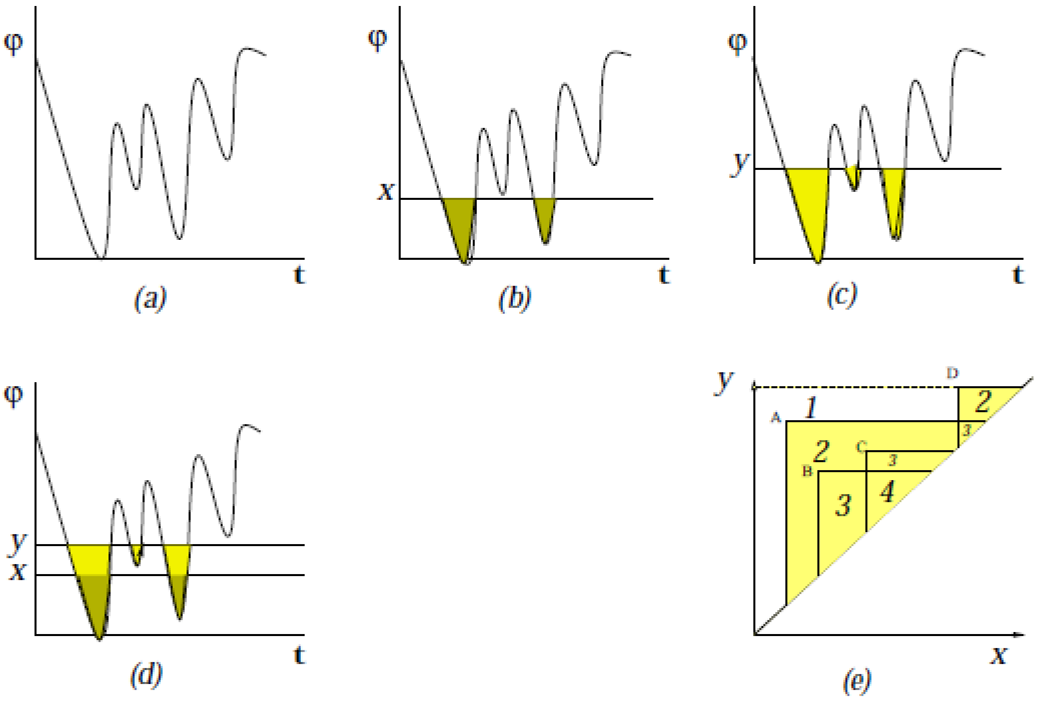

Figure 1.3: Size Function Example

1.3.1.2 Approximate Directional Vectors

The approximate directional vectors are simply the directional vectors between the point at times p_t and p_{t-1} rounded to the predefined principal directions.

1.3.2 Finite State Automata

The use of Finite State Automata for template matching is the simplest approach to recognizing gestures[28]. The gestures are recognized by simply comparing the observed values with the template values, and transitions between the states to a known output results in a gesture being recognized.

1.3.3 Hidden Markov Models

A three dimensional gesture may be recognised and processed using a Hidden Markov Model[18], which is a simple dynamic Bayesian network. This spatiotemporal model, which is defined below, reduces the 3 dimensional complexity of the hand gesture into a two dimensional problem, and analyses and categorizes these gestures using a state machine. This model has been employed, with success, in speech [24] and handwriting recognition [12].

The HMM can be defined as such[24]:

- states: $S = \{s_1, s_2, \ldots, s_N\}$ where the state at time t is q_t and N is the number of states in the model
- symbols: $V = \{v_1, v_2, \ldots, v_M\}$ where M is the number of distinct observation symbols per state
- probability vector (state transition probability distribution): $A = \{a_{ij}\}$ where $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$ $1 \leq i, j \leq N$
- observation state probability distribution in state j: $b_j(k) = P(v_k\ at\ t \mid q_t = s_j)$ $1 \leq j \leq N$ $1 \leq k \leq M$
- initial state distribution: $\pi = \{\pi_i\}$ where $\pi_i = P(q_1 = s_i)$

The resultant Hidden Markov Model can be visualized as shown in Figure 1.4, where:

- x states
- y possible observations
- a state transition probabilities
- b output probabilities

Sets of Hidden Markov Models are grouped into statistical classifiers. These are statistical methods that map n-feature vector to a point in that same n-space.

1.3.3.1 Learning

The learning task of the Hidden Markov Model is an intractable problem that uses a maximum likelihood approach to determine the best set of state transition and output probabilities, given an output set of sequences. However, the Baum-Welch algorithm is often used to efficiently derive a local maximum likelihood[17].

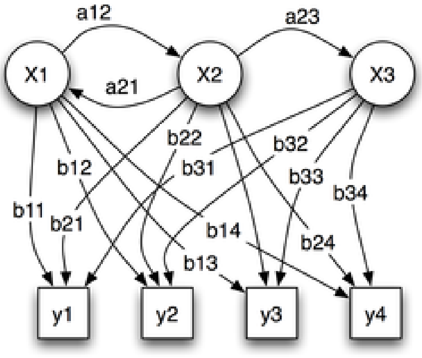

Figure 1.4: Hidden Markov Model[33]

1.3.3.2 Recognition

The recognition process tackles the problem of deciding whether an observed set can be described by the Hidden Markov Model. This is achieved by calculating the probability of the observed set, given the parameters of the model. If this probability is above a predefined threshold, the gesture is recognized as part of the model[17].

1.3.4 Other Methods

1.3.4.1 Artificial Neural Networks

Artificial neural networks are computational models that simulate aspects of a biological neural network. As shown in Figure 1.5, artificial neural networks are built up as a collection of node layers[1]:

- An input layer
- A series of hidden layers
- An output layer

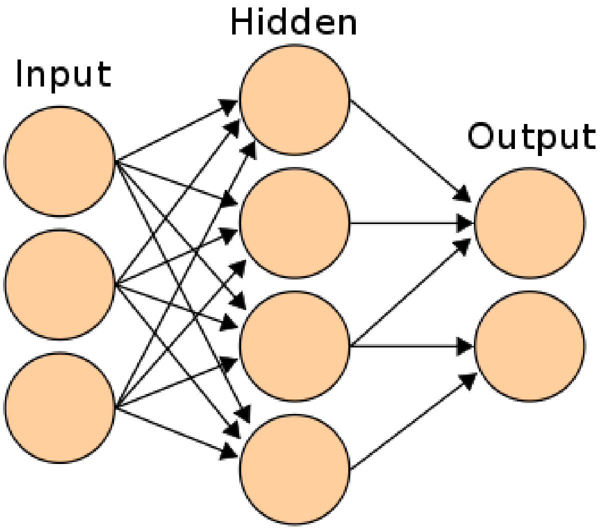

Figure 1.5: Artificial Neural Network[32]

The outputs between states are based on weights are learnt using back propagation through examples. The neural network does not generally perform as well as the more specialized Hidden Markov Models, but this approach has been used with some success in the recognition of sign language gestures in a paper demonstrating the use of size functions used an artificial neural network approach[21].

1.3.4.2 Statistical

Statistical methods of gesture recognition use classifiers, just as the Hidden Markov Models do. Other approaches include:

- Bayesian classifiers
- Hidden Markov Model with Gaussian distributions

Due to the requirement of large training sets and the inflexibility of Hidden Markov Models classification, the Bayesian approach might be preferable, as a study in head gesture recognition showed[35]. The sparse classification model used in that study demonstrated the flexibility of the Bayesian approach.

The Hidden Markov Models can be modified to emit continuous distributions, which would be useful for continuous data rather than transforming the observed values into discrete observations[23, 17].

1.4 Hardware

1.4.1 Nintendo Wii

Although the Nintendo Wii is a proprietary product, and Nintendo has attempted to keep much of its hardware information from the general public, there has been some detailed information released about the products since the game console's inception[31].

1.4.1.1 Technical Specifications

The Nintendo Wii console's hardware has a PowerPC-based "Broadway" CPU, clocked at 729 MHz accompanying an ATI "Hollywood" GPU with a clock speed of 243MHz[2]. The console has 88MB of main memory, of which 24MB is internal 1T-SRAM integrated into graphics package and 64 MB is external GDDR3 SDRAM[3]. It connects to up to 16 Wii Remotes using Bluetooth[22].

The Nintendo Wii Remotes have a 16KB EEPROM chip that allows data and calibration information to be stored, accessed and modified. The Wii Remote camera contains a Charged Coupled Device and an infra-red filter with a hardware resolution of 128x96 pixels, with a reported virtual resolution of 1024x768 pixels at 100Hz[2]. The camera tracks up to four infra-red blobs, whose information, including the blob coordinates and relative sizes, is sent to its paired device using Bluetooth, instead of inefficiently transmitting the data for the entire resolution. Any light that the camera detects whose wavelength is greater 800nm is detected. The camera best detects light sources that have high luminescence and small active areas. The removal of the infra-red filter allows the camera to pick up any high luminescence (bright) objects[36].

1.5 Related Work

Since the release of the Nintendo Wii, there has been a great deal of experimentation into what can be accomplished with the Wii Remotes.

1.5.1 Johnny Chung Lee's Wii Projects

Johnny Chung Lee has developed several applications that use the Nintendo Wii Remotes for interaction with a computer using a Bluetooth connection[20]:

- Tracking Your Fingers with the Wiimote
- Low-Cost Multi-point Interactive Whiteboards Using the Wiimote
- Head Tracking for Desktop VR Displays using the Wii Remote

The projects developed are built using the WiimoteLib library.

1.5.1.1 Tracking Your Fingers with the Wiimote

The tracking of fingers in two dimensions allows a user to have a pointer interface with which interaction with a computer can be achieved by using the Wii Remote's ability to track infra-red blobs, as shown in Figure 1.6.

Figure 1.6: Finger Tracking with the Wiimote

This project uses one Nintendo Wii Remote, an array of infra-red LEDs and infra-red reflectors which are attached to the fingers. The infra-red array is positioned behind the Wii Remote's camera, enabling the users to simple have reflectors attached to their fingers[20].

1.5.1.2 Low-Cost Multi-point Interactive Whiteboards Using the Wiimote

Custom made pens with infra-red lights installed in the tips of the pens are used to track the pens' positions using one Nintendo Wii Remote. This project uses the Wii Remote's two dimensional infra-red blob tracking capabilities. A possible schematic for the pens was suggested, as shown in Figure 1.7[20].

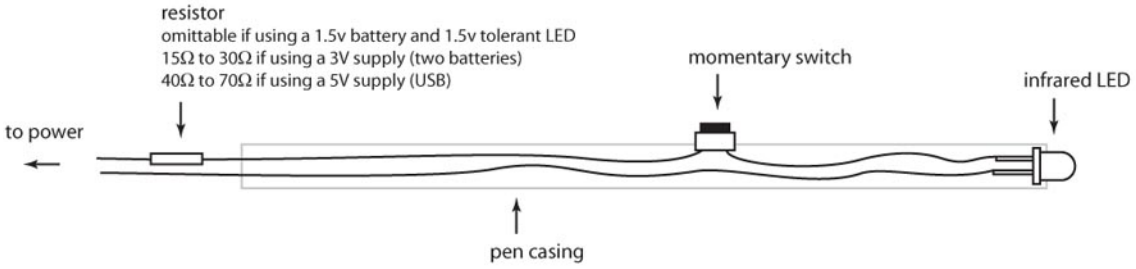

Figure 1.7: Whiteboard Pen

The pen allows a user to use the momentary switch to turn the infra-red light in the pens on or off.

1.5.1.3 Head Tracking for Desktop VR Displays using the Wii Remote

This implementation uses the Wii Sensor bar, mounted on the user's head, to track the user's head. The demonstration application renders targets on the screen based on the position of the Nintendo Wii Remotes, resulting in a more immersive experience, as shown in Figure 1.8.

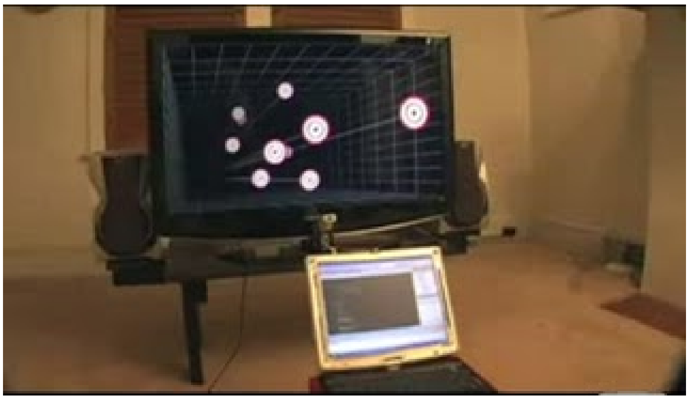

Figure 1.8: Desktop VR

The illusion of depth and space that is created by the rendering engine could add to the quality of computer games and other perspective applications[20].

1.5.2 Design and Implementation of a Hand Tracking Interface using the Nintendo Wii Remote

This project investigated the use of two Nintendo Wii Remotes to track hands in three dimensions for a specific task - molecular visualization applications. The Wii Remote's camera properties were explored, and a hand tracking interface with six degrees of freedom was implemented and tested. The investigator found that this technique was an acceptable method for the visualization of complex molecules, and could be extended to other Computer Aided Design (CAD) applications[36].

1.5.2.1 Camera Details

Wronski[36] experimentally determined the Wii Remote's intrinsic properties necessary for the stereoscopic triangulation calculation by measuring the changes in the horizontal and vertical viewing ranges as the distance from the camera. The results of the experiment are summarized in Figure 1.9

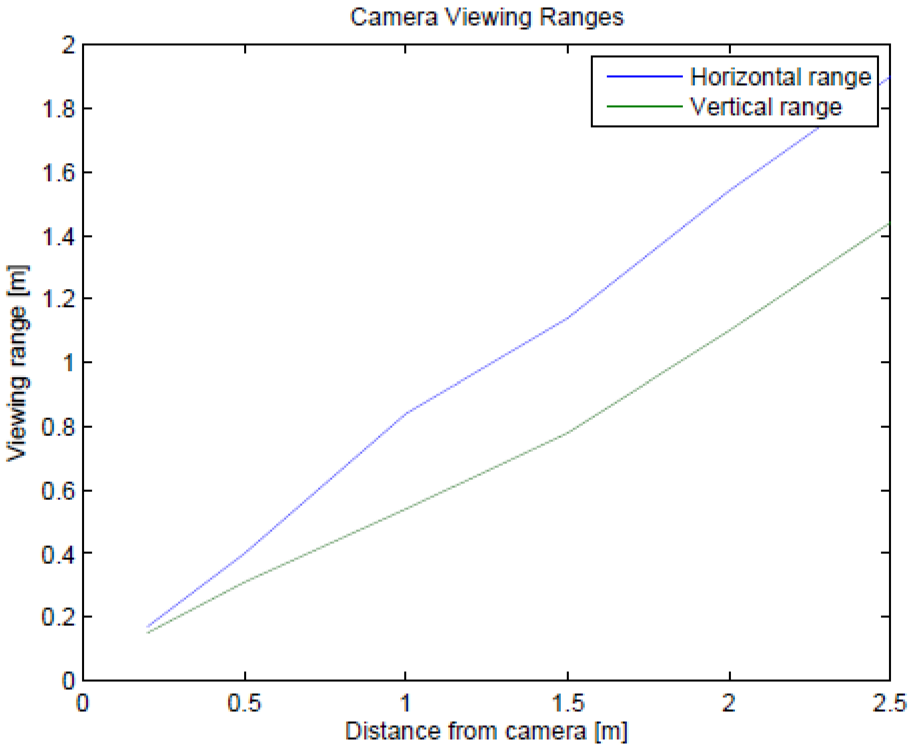

Figure 1.9: Wii Remote Camera Viewing Ranges

Wronski[36] found that viewing angles were 41º horizontally and 31º vertically. The detected near viewing plane and far viewing plane were found to be 10cm and 3m for the infra-red LEDs used in his investigation. Furthermore, he inferred that due to the relatively linear graphs, that there was negligible lens distortion.

1.5.2.2 Implementation

The configuration of the system developed by Wronski entailed a single finger glove, as shown in Figure 1.10, and a layout of Wii Remotes that were placed parallel to each other and 20cm apart[36].

Figure 1.10: Wronski Glove

Wronski's[36] implementation made use of the molecular visualization software package, Avogadro, was used in the investigation due to its easily extensible plugin architecture. The plugin that was developed received tracking data from the two Wii Remotes and modifies the translation, rotation and scale of the viewport in the Avogadro software. He defined several gestures:

- Moving hands closer together/further apart zooms into/out of the molecule
- Moving both hands up, down, left or right simultaneously pans horizontally and vertically
- Moving the left hand down and right hand up, or vice-versa rotates the molecule about z-axis (roll)
- Moving left hand forward and right hand backward, or vice-versa rotates the molecule about y-axis (yaw)
- Moving both hands forward or backward simultaneously rotates the molecule about x-axis (pitch)

1.5.2.3 Outcome

The implementation of the hand tracking interface was found to be an acceptable interface for the manipulation of the Avogadro application viewport. The gestures were successfully recognized and the user experience when using the interaction technique was improved. Wronski investigated the cost of such a setup, and discovered that the hardware, excluding a computer, comes to approximately ZAR2500 (about US$350) [36].

1.6 Summary

There are a multitude of methods for tracking the hand in space, each with its own pros and cons. A great deal of work has been invested in the algorithms involved in hand tracking and their associated approaches to the recognition of gestures. The array of choice in this respect allows researchers and developers to choose an approach that best suits their particular strengths and their systems inherent function and design, while still giving a wide scope in terms of resources required.

The low cost configurations generally yielded lesser quality devices with low accuracy, and those with a higher cost yielded high quality devices with high accuracy. The Wii Remote configuration appears to balance cost, quality and accuracy.

Bibliography

[1] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[2] Matt Casamassina. Revolution's horsepower. Website, March 2006.

[3] Hiroki Eda. Ps3 vs wii, comparisons of core lsi chip areas. Website, November 2006.

[4] S. Wise et al. Evaluation of a fiber optic glove for semi-automated goniometric measurements. *Journal of Rehabilitation Research an Development*, 27(4):411–424, 1990.

[5] T. G. Zimmerman et al. A hand gesture interface device. In *Proceedings of Human Factors in COmputing Systems and Graphics Interface*, pages 189–192, New York, April 1987. ACM Press.

[6] P. Frosini. Measuring shapes by size functions. In *Proceedings of SPIE on Intelligent Robotics and Computer Vision*, volume 1607, pages 3–26, Boston, Mass., 1991.

[7] K. S. Fu and T. S. Yu. *Statistical Pattern Classication using Contextual Information.* Recognition and Image Processing Series. Research Studies Press, 1980.

[8] A. S. Gilinsky. Perceived size and distance in visual space. *Psychological Review*, 58:460–482, 1951.

[9] Gunnar Rutger Grape. Model based (intermediate-level) computer vision. Technical Report AD0763673, Stanford University, May 1973.

[10] G. J. Grimes. Digital data entry glove interface device, November 1983.

[11] G.Tziritas. Estimation of motion and structure of 3d objects from a sequence of images. In *1st IEEE Conference on Computer Vision*, pages 693–697, London, June 1987.

[12] J. Y. Ha, S. C. Oh, J. H. Kim, and Y. B. Kwon. Unconstrained handwritten word recognition with interconnected hidden markov models. *IWFHR*, 3:455–460, May 1993.

[13] Marsha Jo Hannah. Computer matching of areas in stereo images. Technical Report AD0786720, University of Stanford, July 1974.

[14] Richard I. Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, November 1997.

[15] Card Carey Gasen Mantei Perlman Strong Hewett, Baecker and Verplank. Acm sigchi curricula for human-computer interaction. Website, July 2009.

[16] J. W. Philbeck S. S. Fukusima J. M. Loomis, J. A. Da Silva. Visual perception of location and distance. *Current Directions in Psychological Science*, 5:72–77, 1996.

[17] Yangsheng Xu Jie Yang. Hidden markov model for gesture recognition. Technical Report CMU-RI-TR-94-10, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 1994.

[18] In-Cheol Kim and Sung-Il Chien. Analysis of 3d hand trajectory gestures using stroke-based composite hidden markov models. *Applied Intelligence*, 15(2):131–143, 2001.

[19] M. W. Krueger. *Artificial Reality*. Addison-Wesley, 2nd edition, 1990.

[20] Johnny Chung Lee. Johnny chung lee - projects - wii. Website, February 2009.

[21] D. Ziou M. Handouyahia and S. Wang. Sign language recognition using moment-based size functions. In *Proceedings of the International Conference on Vision Interfaces*, pages 210–216. Ker, CRC Press, 1999.

[22] Nintendo. *Wii operations manual*. Nintendo.

[23] Valery A. Pterushin. Hidden markov models: Fundamentals and applications (part 2 - discrete and continuous hidden markov models). *Online Symposium for Electronic Engineers*, 1:1, 2000.

[24] Lawrence R. Rabiner. *Readings in Speech Recognition*, chapter 18, pages 267–296. Morgan Kaufmann Publishers Inc., 1990.

[25] David J. Sturman and David Zeltzer. A survey of glove-based input. *IEEE Computer Graphics and Applications*, 14(1):30–39, January/February 1994.

[26] Clark Thompson. Depth perception in stereo computer vision. Technical Report STAN-CS-7 5-5 2 1, Stanford University, 1975.

[27] G. Tziritas. Displacement estimation for image predictive coding and frame motion-adaptive interpolation. In *Proceedings of SPIE Visual Communications and Image Processing*, pages 936–941, 1988.

[28] Richard Watson. A survey of gesture recognition techniques. Technical Report TCD-CS-93-11, Trinity College, Dublin, July 1993.

[29] P. Wellner. Interacting with paper on the digital desk. *Communications of the ACM*, 36(7):87–96, July 1993.

[30] C. Wheatstone. XXXVI. Contributions to the physiology of vision. Part the First. On the some remarkable, and hitherto unobserved, phænomena of binocular vision. *Philosophical Magazine Series 4*, 3(18):241–267, 1852.

[31] Wiire.org. Main page - wiire. Website, May 2007.

[32] Wikipedia. Artificial neural network. Website, October 2010.

[33] Wikipedia. Hidden markov models. Website, October 2010.

[34] Wikipedia. Triangulation. Website, October 2010.

[35] Shu-Fai Wong and Roberto Cipolla. Continuous gesture recognition using a sparse bayesian classifier. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 1084–1087, Washington, DC, USA, 2006. IEEE Computer Society.

[36] Michal Piotr Wronski. Design and implementation of a hand tracking interface using the nintendo wii remote. Master's thesis, University of Cape Town, 2008.