ON NATURAL GESTURES FOR INTERACTING IN VIRTUAL ENVIRONMENTS

Radu Daniel VATAVU
Ştefan-Gheorghe PENTIUC
"Stefan cel Mare" University of Suceava
str.Universitatii nr.13, RO-720225 Suceava
raduvro@yahoo.com, pentiuc@eed.usv.ro

Christophe CHAILLOU
Laboratoire d'Informatique Fondamentale de Lille

Abstract. *Several considerations with regards to the use of natural gestures for interacting in virtual environments are discussed, taking into account current technologies for gesture acquisition, existing interaction metaphors and the usability, ease of use and advantages/disadvantages of gestures with rapport to other interacting technologies. Discussion is equally conducted on gestures being used for the three main tasks commonly encountered in virtual environments: selection, manipulation and travel. References to gesture definition and classification according to different criteria and their relevance for interacting in virtual environments are equally taken into account.*
Keywords: *natural gesture, gesture interaction, interaction metaphors, virtual environments.*

Introduction

Virtual environments have proven to be of an effective usefulness for a variety of applications such as medical training and exploration [1-6], military simulations [7], architecture [8], phobia therapy [9], etc. A great variety of devices have been developed for effectively interacting with such environments and many interaction techniques have been introduced [10, 11]. Many of them prove to be cumbersome and more or less obstructing the user and hence affecting his/her virtual experience. Natural gestures as a mean of human computer interaction have the great advantage of being the ideal interface as natural, efficient and intuitive means of communication (people use gestures everyday in a natural manner).

The paper addresses the viability of using gestures in virtual environments by taking into account current technologies for capturing human gestures, existing interaction techniques and various aspects dealing with gesture and gesticulation functionalities.

Overview of current interaction technologies

Non-traditional immersive devices for interacting with virtual environments have been very rapidly proliferating. They include spatial input devices (or trackers), pointing devices and whole hand devices that allow for hand gestures input. One main property of input devices is the number of DOFs they posses.

In accordance with the type of events input devices generate, one may distinguish between:

- **discrete** input devices that generate one event at a time according to the user need (events are fired for example when the user presses a button). Examples include the traditional keyboard, the pinch glove or the virtual tool belt. In the case of Fakespace Labs Pinch Gloves [12], the user will pinch two or more fingers for the device to signal an event. The gloves detect when two or more fingers are in contact and, after contact verification, a signal is fired. Elapsed time between gestures is also recorded. A complex variety of actions based on these

simple pinch gestures with natural meaning to the user can be programmed into applications (a pinching gesture can be used to grab a virtual object, and a finger snap between the middle finger and thumb can be used to initiate an action)

- **continuous** input devices that generate a stream of events. Common examples are position / orientation trackers and data gloves
- **hybrid** devices that combine both discrete and continuous events. Examples include the ring mouse such as Pegasus FreeD and digital pen based tablets [13].

With regards to tracking devices, one can discriminate between several technologies: magnetic, mechanical, acoustic, inertial, vision/video camera based or hybrid.

Magnetic trackers (such as Ascension's Flock of Birds [14]) have as main disadvantage the distortion of the magnetic field that metal or conductive metals will produce. Also, the magnetic field will interfere with nearby monitors. Otherwise, the work range is good and prices are moderate. Mechanical trackers (such as Fakespace's BOOM Tracker [12]) have as advantages accuracy and low latency. They may however be big or bulky with reduced mobility. Also, they may exhibit expensive prices. Acoustic trackers (such as Logitech's Fly Mouse [15]) prove to be relatively inexpensive and with no interference with metals but they exhibit line of sight issues as well as sensitiveness to noise. Inertial trackers (InterSense IS300 [16]) have as advantages the long working range, non interference with metals as well as no need for a transmitter. The sensors prove however to be bigger than magnetic based ones and they are likely subject of error accumulation. Camera based or vision trackers make use of video information and video based processing to achieve object tracking. The main advantage is their allowance of unobtrusive tracking. They demand a lot of processing power and depend on computer vision algorithms very much currently under development. Total or partial occlusion of the tracked object proves to be an additional problem. Employed techniques include: motion detection and motion flow following, color detection (and particularly skin color detection as a preprocessing stage in hands/face detection), pattern recognition methods, etc. Hybrid trackers (InterSense IS600 [16]) have the advantage of combining multiple technologies for improving on accuracy, reducing latency on a complexity cost however.

With respect to the main forms of feedback, one can classify devices into:

- **ground** referenced (such as Phantom devices [17])
- **body** referenced (for example CyberGrasp [18] which is a lightweight force-reflecting exoskeleton that fits over a CyberGlove data glove wired version and adds resistive force feedback to each finger. Grasp forces are produced by a network of tendons routed to the fingertips via the exoskeleton)
- **tactile** (CyberTouch [18], a tactile feedback option for Immersion's wired CyberGlove instrumented glove. It disposes of small vibrotactile stimulators on each finger and the palm of the CyberGlove system)

Natural gestures, a physiological view

Prior to anything, gestures can be defined as a physical movement of hands, arms, face and body with the intent of conveying information and meaning. From a biological and sociological perspective, gestures are loosely defined and thus researchers are free to visualize and classify gestures as they see fit. For example, biologists define gestures broadly, stating "the notion of gesture is to embrace all kinds of instances where an individual engages in movements whose communicative intent is paramount, manifest, and openly acknowledged" [19].

Cadoz [20] identifies 3 types of gestures in accordance to their function:

- **semiotic** or gestures that produce informational messages for the environment, communicate meaningful information and which result from commonly shared cultural experience
- **ergotic** or gestures that are associated with the idea of work and ability of modeling and manipulating the environment. The gesture acts directly on the environment by altering its form and properties
- **epistemic** or gestures that offer information with regards to environment revealing: temperature, pressure, surface quality for a given object, shape, orientation, weight. The environment gets revealed through tactile experience or haptic exploration.

Semiotic gestures are further classified by McNeill [21] according to their role in communication as:

- **iconics** or gestures that describe an actual concrete object or event and that bear a close relationship with the semantic content of speech
- **metaphorics** that are gestures similar to iconics but referring to abstract objects or events, depicting a general abstract idea
- **deictics** (pointing gestures)
- **beat-like** which gestures that accentuate the meaning of a word or a phrase.

Kendon [22] describes a gesture continuum as follows:

- **gesticulation** or spontaneous movements of hands and arms that take place during speech and always accompany speech
- **language-like gestures** that represent gesticulation actually integrated into speech that replaces a word or a phrase
- **pantomime** are gestures that depict objects, events or actions that may be or not accompanied by speech
- **emblems** or familiar gestures (for example the "V" sign for victory)
- **sign languages** that are sets of gestures and postures that define linguistic communication systems (for example ASL, the American Sign Language)

Starting from gesticulation to sign languages, the association with speech gets more and more reduced, spontaneity decreases and social regulation increases.

Nespoulos & Lecours [19] consider a more detailed approach and propose a 3-level classification. With respect to the universality of gestures, the following types were identified:

- **arbitrary gestures**: uncommon gestures that need to be learned
- **mimetic gestures**: more common, usually encountered within culture
- **deictics**: similar to McNeill's classification. They include **specifics** (gestures that point to a particular object), **generics** (gestures that point to a class of objects) and **function indication gestures** (that point to an object and equally indicate an action).

Considering gestures' functionality, [19] have identified:

- **quasilinguistic expressions**: gestures that occur in the absence of any verbal behaviour
- **coverbal exporessions**: gestures that occur in the presence of speech. They may be **illustrative** (gestures that depict objects which are described verbally), **expressive** (gestures that express emotions during verbal behavior) and **paraverbal** (gestures that accentuate verbal behaviour)
- **social interaction** that relate to interaction strategies: **phatic** (involving gestural or visual activity) or **regulatory** (gestures used to indicate attention to others for example)
- **meta communication** which are gestures that modulate the speaker's own verbal behaviour

- **extra communication** or gestures that do not posses any semiotic value

Cassell [23] considers 2 categories of gestures: **autonomous** (or independent) and **natural** (or spontaneous). Autonomous gestures are not necessary associated with verbal communication, possess fix spatio-temporal properties and are speaker independent. Natural gestures are usually associated with speech in a conscientiously or unconscientiously manner. They are speaker dependent and much influenced by educational and cultural factors as well as by the actual situation at the moment they are produced.

Interaction techniques for virtual environments

Virtual environments are impoverished versions of the physical world with incomplete sensory cues and simplified and inconsistent world models. The virtual experience is influenced by experiential, cognitive, perceptual and motor differences between users.

Distinction must be made between input devices and interaction techniques or metaphors. Input devices prove to be just the physical tools used for implementing various interaction techniques.

We can distinguish three most important tasks with virtual objects and inside virtual environments:

- selection
- manipulation
- navigation

Selection is the action of specifying one or more objects from a set. Usual goals of selections are: indicate action on an object, make an object active, travel to object's location, set up manipulation. Selection performance is affected by several factors: the object distance from the user, object's size, density of objects in area. Commonly used selection techniques include: the virtual hand metaphor [24], ray casting [25], occlusion/framing [26], naming or indirect selection

Selection tasks may be classified according to: feedback (graphical, tactile and audio), object indication (object touching, pointing, indirect selection) and the indication to select (button, gesture, voice).

Manipulation is the task of modifying object properties (for example color, shape, orientation, position, behavior, etc). The goals of manipulation may be object placement, tool usage, etc. Navigation and travel are the tasks to control viewpoint and camera location via movement and way finding [27].

The use of gesture as a natural interface

Gestures are perceived as a natural mean of interacting and conveying information hence a gesture based interface would prove to be ideal. However, gestures may also be described as imprecise, not self revealing and also non ergonomic.

Although it is very natural to gesture in the real world in order to interact with and model real objects or simply to transmit information, performing the same activities in the virtual environment is not self revealing. This in contrast with standard graphic user interfaces such as buttons, menus or selection lists. Gesturing comes in natural but in what regards virtual environments guidance must be available such as visual remainders of the interaction techniques.

Selection is a frequently occurring task and thus should be implemented to maximize efficiency. Gesture is an ideal interaction technique for selection tasks as it may be implemented in a way that closely mimics real-world interactions. Ray casting techniques make use of deictic or pointing gestures. A ray directed by the user's hand is used to indicate referent objects within the scene (first application Bolt's put-that-there

interface, 1980 [25]). Bowman and Hodges [27] found that ray-casting performance was more efficient that arm extension over a range of object distances, sized and densities.

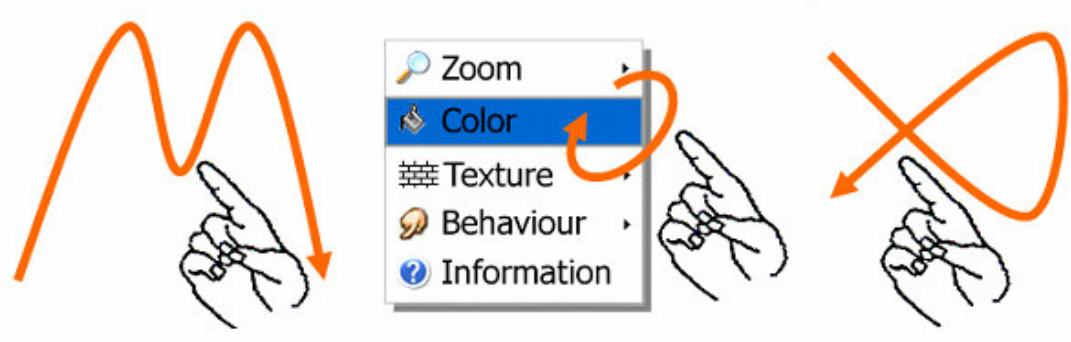

Figure 1. Sample hand gestures for working with an application menu (activate menu / select an option/ close menu)

Several guidelines have been proposed when using gestures for interacting in virtual environments. Cerney et. al [28] consider the use of gestures for virtual environments taken into account selection, manipulation and travel aspects providing recommendations for the design of natural interactions:

- guidelines must be provided to the user as to how the gesture recognition has been implemented
- continuous feedback to the user
- the beginning of gesture should be associated with tension and the end with releasing tension
- ability of canceling unwanted actions
- the dictionary of commands must be composed of easy and quick to executed gestures with limited degreed of precision and an according level of distinctiveness
- combining gestures with existing traditional commands

Matthew Turk [29] also considers a few suggestions for gesture based system design:

- inform the user (the type of allowed gestures and their meaning must be clear to the user)
- user feedback
- use the uniqueness of gesture
- take into account the limits of a particular technology
- use of non-fatiguing gestures
- low cognitive user memory load
- limit the precision of required gestures
- use of existing gestures commonly used instead of inventing new unnatural gestures for various commands.

Gesture dictionaries

A particular problem relates to finding the right gestures that would feel comfortable and natural from the user's experience point of view. Several attempts have been made on defining gesture dictionaries for application specific needs [30-35]. Both hand postures and hand gestures are considered. Nielsen et al. [33] conducted an ergonomic study for selecting the appropriate gesture commands for operations such as: selection, move, scale, copy, confirm, yes / no, undo using both single hand and two hands gestures in a video camera based top view of a working table. Similar gestures for habitual operations such copy / paste / menu activation are considered by Holte and Stoerring [34] using mainly single hand postures and simple gestures.

A real problem arises from the fact that gesture commands have to be identified (or designed) with the particularity of assuring a natural and

comfortable user experience, all this considering the existing GUIs and interaction paradigms, for example WIMP (Windows, Icon, Menu, Pointing). In this particular case, which would be the most suitable gesture for activating / closing a menu or for maximizing / minimizing an application window? A simple proposal for working with an application menu (activate, select, close) is presented in figure 1, but the question still remains: are the proposed gestures natural or are we looking for a compromise between natural and new to be learned gestures?

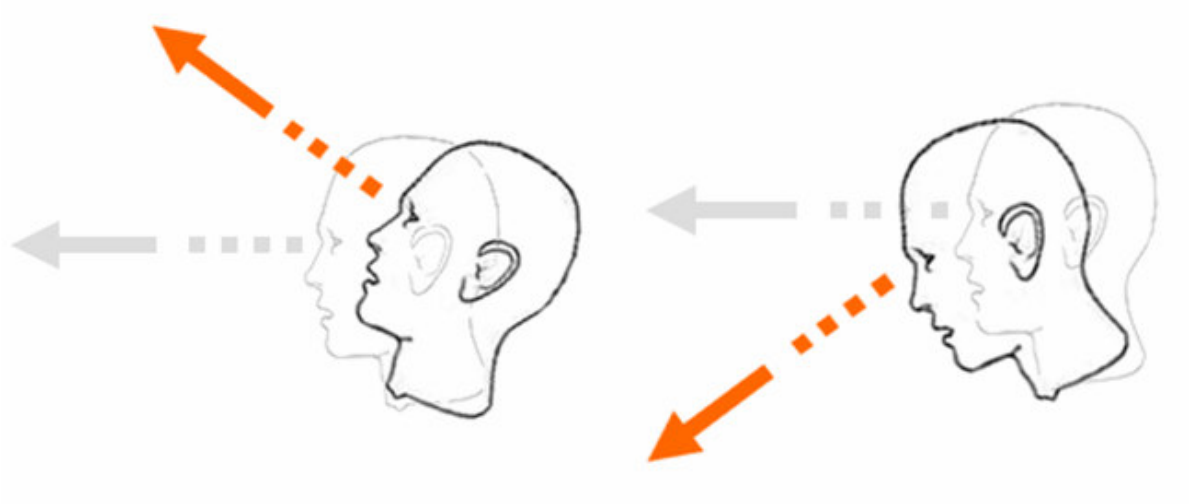

Figure 2. Are head gestures appropriate for navigation in virtual environments?

Of equally importance, not only hand but head gestures can be successfully exploited for getting the feel of an even more natural interaction. Face detection [38] and head tracking [39-42] have been given a lot of attention for various purposes. Again, questions arise on what would feel more natural: traveling in a virtual scene using head movements or simply considering hand postures would prove more appropriate (see figure 2)?

Conclusions

Considerations on the use of natural gestures for interacting in virtual environments have been presented together with an overview of existing interaction technologies as well as interaction metaphors.

References to gesture definition and classification according to different psychological criteria have been equally taken into account. Existing guidelines for using gestures in the human computer dialog are mentioned.

The paper equally questions on the construction of an appropriate dictionary for gesture interaction that would include gestures that feel natural, comfortable and allow for a successful user experience.

References

[1] P. Meseure, C. Chaillou (2000) *A Deformable Body Model for Surgical Simulation*, Journal of Visualization and Computer Animation, 11, 4, September 2000, pp 197-208

[2] P. Mesure, J. Davanne, L. Hilde, J. Lenoir, L. France, F. Triquet, C. Chaillou (2003) *A Physically Based Virtual Environment dedicated to Surgical Simulation*, Surgery Simulation and Soft Tissue Modeling (IS4TM'03) – LNCS, June 2003, pp. 38-47

[3] J.D. Westwood, H.M. Hoffman, G.T. Mogel, R.A. Robb, D. Stredney, eds. (2000)

Medicine meets virtual reality 2000. Envisioning healing: interactive technology and the patient-practitioner dialogue. Amsterdam: IOS Press

[4] L. Raghupathi, L. Grisoni, F. Faure, D. Marchal, M.-P. Cani, C. Chaillou (2004) *An Intestine Surgery Simulator: Real-Time Collision Processing and Visualization*, IEEE Trans. On Visualization And Comp. Graphics, 10(6), pp. 708-718

[5] J. Lenoir, P. Meseure, L. Grisoni and C. Chaillou (2004) *A Suture Model for Surgical Simulation*, Proc. of International Symposium on Medical Simulation, Cambridge, Massachusetts (USA), June 2004, pp. 105-113

[6] A. Chaudhry, C. Sutton, J. Wood, R. Stone, R. McCloy (1999) *Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface*, Ann R Coll Surg Engl 81, pp. 281-286

[7] Jeffrey R. Hogue, Steve Markham, Arvid Harmsen, Jerry MacDonald, Cliff Schmucker, *Parachute Mission Planning Training and Rehearsal Using a Deployable Virtual Reality Simulator* (2003), the 17th AIAA Aerodynamic Decelerator Systems Conference and Seminar, May 19-22, 2003, Monterey, California

[8] Whyte J (2003) *Industrial applications of virtual reality in architecture and construction*, ITcon Vol. 8, Special Issue Virtual Reality Technology in Architecture and Construction, pg. 43-50

[9] C. Botella, R. M. Banos, C. Perpina, H. Villa (1998) *Virtual reality treatment of claustrophobia: A case report*, Behaviour Research and Therapy, Oxford, 239-246

[10] William Sherman, Alan Craig (2003), *Understanding Virtual Reality: Interface, Application and Design*, Morgan Kauffman

[11] Michitaka Hirose (EDT) (2001) *Human-Computer Interaction - INTERACT'01*, IOS Press

[12] Fakespace Labs, http://www.fakespacelabs .com

[13] Pegasus Technologies Ltd. http://www. pegatech.com

[14] Ascension Technology Corporation (2005) http://www.ascension-tech.com/products/flockofbirds.php

[15] Logitech Inc., www.logitech.com

[16] Intersense, http://www.isense.com/products

[17] Sensable Technologies, the Phantom devices,http://www.sensable.com/products/phantom_ghost/phantom-omni.asp

[18] Immersion Corporation, the Cyber Grasp, Cyber Touch, Cyber Glove devices, http://www.immersion.com/3d/products/cyber_grasp.php

[19] Jean-Luc Nespoulous, Paul Perron, and Andre Roch Lecours (1986) *The Biological Foundations of Gestures: Motor and Semiotic Aspects*. Lawrence Erlbaum Associates, Hillsdale, MJ

[20] Claude Cadoz (1994), *Le geste canal de communication homme/machine*, Technique et Science Informatique, Vol. 13, No 1, pp. 31-61

[21] David McNeill (1992) *Hand and mind: What gestures reveal about thought*, University of Chicago Press

[22] Adam Kendon (1986) *Current issues in the study of gesture. In The Biological Foundation of Gestures: Motor and semiotic Aspects*, Lawrence Erlbaum Assoc., Hillsdale, pp. 23-47

[23] Justine Cassell (1996) *A Framework for gesture generation and interpretation*, Computer Vision in Human-Machine Interaction, R. Cipolla and A. Pentland Editors, Cambridge University Press, 1996

[24] I. Poupyrev, M. Billinghurst, S. Weghorst, T. Ichikawa (1996) *The go-go interaction technique: Non-linear mapping for direct manipulation in VR*, Proceedings of the ACM Symposium on User Interface Software and Technology, ACM Press, pp. 79-80

[25] Richard A. Bolt (1980) *"Put-That-There": Voice and Gesture at the Graphics Interface*, Proceedings of SIGGRAPH' 80, ACM SIGGRAPH, NY, pp. 262-270

[26] J. Pierce, A. Forsbert, M. Conway, S. Hong, R. Zelezink, M. Mine (1997) *Image plane interaction techniques in 3D immersive environments*, Proceedings of the ACM

Symposium on Interactive 3D Graphics, ACM Press, pp. 39-44
[27] D.A. Bowman, E. Davis, L. Hodges, A. Badre (1999), *Maintaining spatial orientation during travel in an immersive virtual environment*, Presence: Teleoperators and Virtual Environments, 8(6), pp. 618-631
[28] Melinda M. Cerney, Judy M. Vance (2005) *Gesture Recognition in Virtual Environments: A Review and Framework for Future Development*, Iowa State University Human Computer Interaction Technical Report ISU-HCI-2005-01, 28 March 2005
[29] Matthew Turk, *Gesture Recognition*, chapter 10, http://ilab.cs.ucsb.edu/projects/turk/TurkVEChapter.pdf
[30] Axel Mulder, *Hand Gestures for HCI*, Hand Centered Studies of Human Movement Project, Technical Report 96-1, School of Kinesiology, Simon Fraser University, February 1996
[31] Jose Marque Soares (2004) *Contribution a la Communication Gestuelle dans les Environments Virtuels Colaboratifs*, PhD dissertation sustained 9th July 2004 at the Institut National des Telecomunications and Universite de Technologie de Troyes
[32] Alan Wexelblat (1995) *An Approach to Natural Gesture in Virtual Environments*, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 2 , Issue 3 (September 1995)
[33] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum (2003) *A procedure for developing intuitive and ergonomic gesture interfaces for HCI*, Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, Genova, Italy, April 15-17, 2003
[34] M. B. Holte and M. Stoerring (2004) *Pointing and command gestures under mixed illumination conditions: video sequence dataset*, FGNET Face and Gesture Recognition Working Group, http://www-prima.inrialpes.fr/FGnet/data/03-Pointing/index.html
[35] Rochelle O'Hagan, Alexander Zelinsky, Sebastien Rougeaux (2002), *Visual Gesture Interfaces for Virtual Environments*, Interacting with Computers, Special Issue
[36] Mathias Kölsch (2004) *Vision Based Hand Gesture Interfaces for Wearable Computing and Virtual Environments*, Ph. D. Dissertation, August 2004
[37] Mathias Kölsch, Matthew Turk, Tobias Höllerer (2004) *Vision-Based Interfaces for Mobility*, Intl. Conference on Mobile and Ubiquitous Systems (MobiQuitous), August 2004
[38] Paul Viola, Michael Jones (2004), *Robust Real-Time Face Detection,* International Journal of Computer Vision**,** Volume 57, Issue 2, May 2004, pp. 137 – 154
[39] C. Lee, N. Lesh, C. Sidner, L.P. Morency, A. Kapoor, Trevor Darrell (2004), *Nodding in Conversations with a Robot*, CHI extended abstracts, Vienna, Austria, 2004, pp. 785-786
[40] L.P. Morency, Trevor Darrell (2004), *From Conversational Tooltips to Grounded Discourse: Head Pose Tracking in Interactive Dialog Systems*, International Conference on Multimodal Interfaces, College State, PA, October 2004, pp. 32-37
[41] L.P. Morency, C. Sidner, C. Lee, Trevor Darrell (2005), *Contextual Recognition of Head Gestures*, Proceedings of the 7th International Conference on Multimodal Interactions (ICMI'05), Torento, Italy, October 2005, pp.18-24
[42] L.P. Morency, Trevor Darrell (2006), *Head Gesture Recognition in Intelligent Interfaces: The Role of Context in Improving Recognition*, Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI'06), Sydney, Australia, January 2006