Natural Interaction for 3D Manipulation in Telemedicine:
A Study Case developed for Arthon Video Collaboration Tool

Rafael Henrique Assis de Castro
Department of Informatics UFPB
Informatics Centre
João Pessoa, Brazil
rafaelhenriquec@gmail.com

Anna Carolina Soares Medeiros
Department of Informatics UFPB
Informatics Centre
João Pessoa, Brazil
linnamedeiros@gmail.com

Tatiana Aires Tavares
LAVID – Digital Video
Application Lab
Informatics Institute UFPB
João Pessoa, Brazil
tatiana@lavid.ufpb.br

Abstract—The main feature desired in a good user interface is naturalness. Natural activities are the ones that humans were meant to do, implicitly written in their body structures and mind. Natural interaction turns it possible for the development of user interfaces. This feature is very useful in controlling complex systems that requires user attention. This paper discusses the application of Natural Interaction (NI) techniques for 3D environment manipulation in Telemedicine systems. The developed study case evaluated the Kinect use for manipulating 3D objects with Arthon, which is a video collaboration tool designed for Telemedicine.

Keywords: Natural Interaction; 3D Manipulation; Telemedicine Systems.

I. INTRODUCTION

Human communication always used gestures, movements and expressions to support oral language. Certain gestures are so commonly used around the world that they are understood throughout different cultures and times, such as a wave or thumbs up [01].

Furthermore, human communication mediated by computers is not so different. Computer applications try to "emulate" reality by creating devices that allow the use of domain of our world knowledge, thus facilitating the learning process.

Making a brief retrospect we can identify the evolution of these devices through the command-based languages, through the graphical user interfaces (GUIs), and finally the direct manipulation with the advent of the use of the mouse. Touch screen interfaces, the use of cameras to recognize users allow us to use technology easily. Kinect is an example that make possible the development of more sophisticated user interfaces, enabling an ever greater natural interaction [14,15].

The Natural Interaction (NI) studies ways that men and device can interact through humans five senses, be that with gestures, voice commands, corporal expressions or human body parts detection and identification, and others [02].

The advent of games powered by gestures is a popular trend like the Nintendo Wii, which controls enable an interaction different from the "old" consoles, not just pushing buttons, but pointing, moving, rotating, offering a world of new interactions. From the innovation of Nintendo, variations have emerged as the Sony Play Move that works with the Playstion3. Microsoft has brought innovation with Kinect, a natural interaction device for XBOX360 that does not make use of peripheral devices, just use your own body and movement it to interact with the equipment. Furthermore, Kinect can be connected to the computer and thus be programmed to control and manipulate systems in general [08,10,12].

Telemedicine is one area that could benefit from this kind of technology. The widespread use of computer systems in healthcare is now a reality. Systems that can connect doctors, patients and students or make possible the evaluation of diagnostic tests from distance are good examples. Trends in telemedicine are systems that are capable of adding more advanced modes of interaction, such as three-dimensional interfaces and natural interaction.

In this paper we discuss a development experience that combines features of Natural Interaction with 3D objects in telemedicine systems. More specifically, we demonstrate the use of natural interaction as an interface for control and manipulation of 3D object inside Arthon [18] tool, dedicated to video collaboration in Telemedicine. Arthon is a tool developed by GTAVCS (Working Group for Video Collaboration in Health) of RNP for control and management concurrent streams in Telemedicine transmissions.

II. INTERACTION STYLES EVOLUTION

In the late '70s, the great majority of computer users were information technology professionals and enthusiasts with financial resources. That changed dramatically in the early '80s with the advent of the personal computers, which were much more comfortable than conventional ones.

Personal computing, including personal software (word processors, spreadsheets and interactive computer games) and personal computer platforms (operating systems, programming languages and hardware), made everyone in the world become a potential computer user, highlighting the shortcomings of computers regarding usability.
Concurrently with the advent of personal computing, consolidated the cognitive science, where part of its program was to articulate systematic and scientifically informed applications to be known as "cognitive engineering" [8]. The need for computer interfaces improvement, along with the emergence of a new science that studies the learning process, initiated the study of the Human-Computer Interaction (HCI).

The evolution of technology has been causing a revolution in contemporary human communication through its several multiprocessor information systems and the development of open systems and standards. The technological convergence integrates telecommunications, computer and information sciences. The convergence provides users information anywhere, on any network, by any communication channel, that is, ubiquity, mobility and interactivity.

Communication is no longer a simple phenomenon where information is transferred from one agent to another and starts being characterized by behaviors marked out by technological convergence. It promotes the integration of video, data, voice and image within advanced techniques that allow mobility, portability of applications and content, interconnectivity, and interoperability between platforms and operators.

Technological development convergent promotes constant access to information and imposes profound changes in all areas of knowledge presenting urgent requirements of research that show the evolution of human-computer interface in order to monitor this development. In a critical area such as Health, human performance is determinative. An example is the surgical area that often requires of the health care professional the option of communication with certain technological artifacts, from a distance.

Admittedly, the physical and mental capabilities of the human interface user are considered in any project, because the user profile influences the development of the interface.

Generally, people naturally communicate through gestures, expressions and movements Developing applications that deal with Natural Interaction means creating systems that are capable of understanding such users' actions while they are interacting naturally with the environment around them. In this sense, people don't need to wear or use any gear, much less learn specific instructions, the interaction is intuitive. Natural interface follows a new paradigm in order to respect the human perception [03].

The success of a system that uses natural interaction depends on how it affects those who experience it. Users should be delighted, have fun, be satisfied with the obtained result, have the illusion that they are dealing with something alive, something almost magical experiencing. The term "experiencing" is used by its broad sense.

A natural interface enables the cognitive and cybernetic dynamics which are commonly experienced in real life, thus persuading the user that it is not dealing with abstract and digital media, but with real objects. These results in a reduction of the cognitive load, thereby increasing the amount of attention to content. Of course the key to achieving this goal is the synthesis of a number of aspects, such as non-obstructive perception, visualization, response time and cognitive load [08].

The Kinect is a technology for natural interaction developed by Microsoft that uses a depth camera and gesture recognition software to allow people play videogames using natural body movements instead of handling a particular controller [04]. This technology has the primary purpose of supporting the Microsoft Xbox 360 console to enable this new way of playing in different types of games [09].

Recently, research has shown that the Kinect is already being used for other purposes, such as navigate in Google Earth using natural movements [05] and games for physical therapy [11,12,13,17]. Kinect is also used in operating rooms for control programs or see images of the surgery without touching. This feature guarantees the patient protection against infections [06].

The current trend is having user interfaces whose use natural formats to act as main point, resulting in a device that better integrates people. Over time, this progression will result in tools that are part of the human life naturally [9].

According to Alessandro Valli [3], the first years of every person's life defines what is natural and what is not. This definition is important because this work has the proposal of using technologies for natural interaction. Although subjective, the definition exposed previously justifies the following sentence: "it is natural to manipulate colored balls with your hands, but it is not, using the keyboard". Natural activities are those that humans were meant to do, implicitly written in their body structures and mind. Natural can be considered synonymous with something common in real life.

The terms "natural interfaces", "natural interaction" or "natural interactivity" are being used by several researchers in a very general perspective where they want to describe something that is somehow different from the common interfaces [03].

Applying the concept of natural interaction in the development of user interfaces is a current trend, since usually results in an interface very well accepted. The success of NI depends on how the user is influenced when using the system. People should be impressed, have fun and have the illusion that they are dealing with something alive. The NI activates cognition and cybernetic dynamic of people, what usually happens when they experience something in real life.

Natural User Interface [10] or NUI, is the common language used by designers and developers of computer interfaces to refer to a user interface that applies concepts of natural interaction in its construction. NUI is considered a new interaction paradigm, since the old and known WIMP (Window, Icon, Menu, Pointer) use concepts of interaction with devices themselves, i.e., there is the use of keyboard or mouse to access any part of the graphic interface, what is not considered as natural interaction, since using such hardware is not part of the naturalness of human being.
III. DEVELOPING A NUI FOR 3D MANIPULATION ON ARTHON

This paper presents the experience of developing and evaluating a natural user interface for manipulation 3D objects in Arthon tool [18]. Arthon is a tool for digital video collaboration in Telemedicine that enables in addition to media streams (such as video), using three-dimensional features.

A. The Arthon

The GTAVCS [19] developed Arthon tool which is a remote management tool for capturing and distributing multiple simultaneous media streams to provide support for video-collaboration scenarios. To this end, we present experiences in the context of Health and Telemedicine, as well as the transmission of surgeries.

The Arthon tool main feature is to offer the user a simple interface for handling different sources streams of media simultaneously. Therefore the user can remotely add, remove, configure the presentation format and schedule the exhibition in time and space of media streams.

The tool consists of the following components: Articulator (Manager), Encoder Agent (Encoder), Decoder Agent (Decoder), Reflector (Distributor of stream), Video Server (VideoServer), User Management Server (WebService) and Videoconference Agent (VideoRoom). These streams can be generated in geographically spread locations and are manipulated by the Arthon tool, which is responsible for capturing, managing, transcoding, transmitting and decoding the media in real-time. These media can be sent in high, medium and low resolution, simultaneously, both to specific decoders on the network and the Internet.

B. 3D Objects on Arthon

An innovation of Arthon tool is to provide the possibility to manipulate 3D objects, especially human anatomical structures, while viewing other streams, such as video. The addition of these 3D models is especially useful as a didactic resource focused to distance training and learning. Through this feature the professors or mentors can show students live videos integrated to 3D models that demonstrate the normal functioning of organs, tissues or structures of the human body. In Figure 01 we present the screenshot that demonstrated the use of Arthon for training. In this demonstration 3D models are integrated with video streams at Arthon tool.

C. Arthon 3D Objects Manipulation using Kinect

To provide a better way to manipulate 3D models we are proposing the use of natural interaction with the Kinect. This feature makes it easier and allows the use and manipulation of the 3D models while the user is viewing other streams managed by Arthon. Besides, in Telemedicine using hands and touch prevent simultaneous execution of medical procedures due to sterilization. In this sense the interaction without touching devices to manipulate 3D objects makes it more safety. Professors and mentors can control and manipulate three-dimensional models during surgery just using their own movements.

In this work we use the natural interaction as a way to access the main control functions used for 3D manipulation presented in Figure 02.

Figure 1. 3D models with video streams at Arthon.

Figure 2. UML Use Case Diagram.

In this project, it was used Java Swing, where an instance to the application’s main class - Screen - is initialized, and where the classes Renderer and HandTracker are started.

HandTracker is the class responsible for the use of the libraries provided by OpenNI framework [16]. It is where the Kinect and the hand tracking are initialized and is performed the exhibition of the image depth stream. To begin the tracking is created an event that recognizes the gesture of waving, which activates the tracking of the used hand.
initializing other events, that create, update and destroy the trace. These last events use an instance of the HandTrackerHandler object to pass all the updates needed from the trace, which are: when the a tracking is created (and is initial information), the updates of the position of the hand and when the tracking has been destroyed.

![Software Architecture Overview](image)

The Renderer class implements the HandTrackerHandler interface and has an instance passed as parameter at a HandTracker object's instantiation in the Screen class, thus making that the object Renderer receive updates related to the traces.

As the Renderer object receives the information of the creation of a new tracking, it saves the initial point, and if there is not another point being tracked defines it as primary hand, in case there is already a primary hand, it is defined as secondary hand.

IV. VALIDATION

In order to validate the use of NI as interface to 3D objects manipulation, we did user tests. The tests were very important because its feedback was used in order to scale the acceptance of an interface that uses NI.

A. Test Methodology

The test methodology involved three phases, as illustrated in Figure 04. During Phase I, one evaluator (HCI specialist) performed a cognitive walkthrough on the interface. This phase detected user interface primary errors. So, the user interface was redesign to incorporate the upgrades suggested by the evaluator (HCI specialist).

Phase II involved empirical evaluation through usability tests, redesigning the application as according to its results. Finally, Phase III tested with end users the latest version of the interface using usability tests again.

The participants which helped us to test user interface were recorded while they performed the tasks using the interface. The participants also answered a post-test questionnaire where they asked closed and opened questions.

For guiding the end users tests we adopted a task model, as follows:
1. Choose a 3D model
2. Enable tracking of both hands
3. Rotate the chosen 3D model
4. Zoom the chosen 3D model

5. Pause the 3D model manipulations using the "Stop" button
6. Request access to a new model using the "New Model" button

![Test Phases](image)

After the initial screen the user will see the interface shown in Figure 05. Following the instructions, the user will choose a hand to first interact with the Kinect, then the system will track the both hands. If the user desires he can interact with both hands simultaneously, thus allowing more efficient use of the application.

The first hand tracked is responsible for handling the option chosen by the second hand. That is, if the user chooses with his second hand the "Zoom" button what will define how the object will approach or move away from the screen are the movements of approach or departure from first hand tracked, in relation to the Kinect.

Following the task model, the user must moves the cursor designated to his tracked second hand to the button "Rotation", and with movement controlled by the cursor designated to his first hand tracked, the user can rotate the 3D model chosen. For the second movement described by the task model, the second hand will select the "Zoom" button and modify the 3D model visualization according to the handling of his first hand tracked. The selection of the "Pause" button is independent of the first hand tracked movement, because the function of this button is only to keep the exact position of the model at the time that this button was activated. When you choose the "New Model" button it returns to home screen for choosing a new 3D model.

B. Results

Initially participants were invited to answer the following questions:
- "Was the interaction by using Kinect intuitive?"
- "Did you feel comfortable using the Kinect to manipulate 3D objects?"
- "If you need to use an application like this, you see yourself using it?"
- "If you could change any aspect of the application, what would it be?"
In general, users have defined the use of Kinect for manipulating 3D objects very useful when asked about a situation in which they could use it. Each participant had a different level of knowledge of the Kinect tool, but all completed the test successfully.

After that participants had to answer a complete usability evaluation questionnaire (as shown in Figure 08). The results obtained were used to validate specific user interface objects. For example, the figure of a person waving his hand used to show how to enable the tracing of hands (as shown in Figure 05) was changed to an animation in the last user interface version.

![HandTracking indication animation](image1)

Figure 5. HandTracking indication animation.

Figure 07 represents the current version of the user interface. The most notable changes were: relocating the animation panel, relocating the instructions panel and changing the animations design. At this way the user interface highlights the contrast between the video stream window and the 3D functions panel. It was very important to show for users that they are in the control.

V. CONCLUSION

This work described the development of an application that uses natural interaction to manipulate a 3D objects. This paper investigated possibilities of development with Natural Interaction, investigating libraries for developing with Kinect and developed and verified a pilot application using the Kinect as a way to improve Natural Interaction.

We noticed a great advantage in using the Kinect, which allows affordable access to depth image, available in existing libraries, image treatments that use little processing power and provide information that facilitates the development, as body parts recognition.

As a result of this study, it is clear that the developed application aggregates value to the project GTAVCS with the addition of a new form of interaction that enables natural interaction for Telemedicine.

As future works it is suggested to improve the application and implement more detailed tests so that it might actually be used to assist professionals in Health. Furthermore, it is expected that progress in the Kinect APIs allows the development of a future application with detection of more refined movements than the ones used in this present work, thus improving the application.
ACKNOWLEDGMENT

We would like to thank colleagues from Project GTAVCS and Lauro Wanderley Academical Hospital for their participation in the project.

REFERENCES

http://old.sigchi.org/cdg/cdg2.html#2_1.


![Figure 8. Usability Evaluation Questionnaire.](image)