Discussion of a Research Paper

CA1034A: e-Business Challenges

Nicolas Lanquetin
0604918@abertay.ac.uk

University of Abertay Dundee
School of Computing & Creative Technologies

Discussion of a Research Paper

Nicolas Lanquetin – 0604918@abertay.ac.uk

University of Abertay Dundee

March 26, 2007

1 SUMMARY OF RESEARCH PAPER

To interact in virtual environments, users currently have to rely on tethered interfaces. In their paper, Tollmar, Demirdjian and Darrell (2004) describe an untethered interface based on real-time tracking of body poses and gestures. It is implemented through a stereo camera and a special algorithm. More importantly, the study includes an evaluation of different body interaction models for virtual navigation in a 3-D game world.

The study tries to find an alternative to tethered interfaces, such as keyboard, mouse or game pads. The first interactive untethered interfaces go back to 1993 and a lot of work has been done since then. However there has been a lack in these implementations: The latter were not in real-time and had many restriction in gesture use. Tollmar, Demirdjian and Darrell therefore implemented an own system based on the Iterative Closest Point (ICP) Algorithm. Its goal is to enable real-time navigation in games and 3-D environments using gestures. Beside enhancing untethered interfaces, the study also tries to find out how this interface is best used.

To analyse the untethered interface, a two-phase experiment was set up. The first phase consists of a Wizard of OZ (hereafter WOz) experiment, where twelve participants had to accomplish different navigation task in a 3-D game, using their own invented gestures[1]. The tasks involved navigation in open and indoors areas. The results of the WOz experiment were used to classify the different gestures into categories of body interaction models. Based on these models, two interaction modes were implemented. In a second phase, ten participants were asked to navigate through the same virtual world, using the two new interaction models and a classical keyboard. After the experiment, the participants were also asked their opinion about the use of the different models.

The experiment results were obtained by direct observation and also by analysing video recordings. For the first phase of the experiment, the analysis showed that the users have individual variants of interaction styles. Three significant gesture models resulted from the first phase: (1) State driven[2], (2) gesture driven and (3) direct body interaction. The results of the second phase were measured in the time required for a participant to accomplish a task and how often the participant lost orientation or hit objects. The user were faster and more accurate using a keyboard, followed by the gesture pointing (see 2) and direct relative pose (see 3) interaction model. The results also showed that spoken instructions are not popular amongst participants and only used as last resort. Additionally, the adaptive algorithm must be tweaked to handle individual variants. It became also clear, that a mix of different models could be of interest. Finally, the feedback questions showed that direct relative pose (see 3) has the highest presence, even though it is very tiring.

The research study described in this paper has been very well organised and conducted. Every aspect of the research is well argumented. Particulary noteworthy is the carrying out of the WOz experiment to find out about users' instinct gesture and pose preferences. However, the results could have been more significant with a greater amount of participants. The

[1] See section 4 on page 3

[2] cf. section 2.2 on page 2

study is nevertheless of great interest in the research field of untethered interfaces. The latter might very well replace traditional interfaces in many applications and especially in the gaming industry. By the time this paper was written, some implementations already existed, such as the successful EyeToy for the consoles Playstation 2/3 and XBox 360 (Backman, 2007, slide 49). According to EyeToy's success, untethered interfaces surely will have a place in future applications.

2 TERMS AND DEFINITIONS

2.1 'perceptual interface' (abstract)

In Human Computer Interaction (HCI), the 'interface' is what allows the user to interact with the computer. The term 'perception' (lat. [*per*-] *capere*) means 'to seize' or 'take'. Therefore, the computer will 'take' information of the user. The whole term 'perceptual interface' in return, is more specific: It is used to describe an interface based on poses, gestures and voice.

2.2 'stateful interface' (p. 114 para 1)

Three models of interactions resulted from the WOz experiment[3], one being the 'stateful interface' (Tollmar, Demirdjian and Darrell, 2004, p. 117). As the name implies, the interface is based on the state of the user or object in question. Technically, the interaction depends on the user's or obecjt's pose relative to a fixed point.

2.3 'planar interaction' (p. 114 para 6)

The word 'planar' (lat. *plānāris*) refers to surfaces, having a 2-dimensional characteristic. In the context of tracking techniques, this means that objects can only be recognised as 2-D planes. In the research paper, two applications of planar interaction are cited: Hall, Gal, Martin, Chomat, Kapuscinski and Crowley (1999) and Oka, Sato and Koike (2002) use this technique to track poses and gestures of hand and fingers. For these tasks, the recognition of simple planes were apparently enough. If the application should support 3-D recognition, planar interaction would however not be sufficient anymore.

[3]See section 4 on page 3

2.4 'canonical configuration' (p. 114 para 6)

The adjectives 'canonical' or 'canonic' refers to something being 'reduced to the simplest and most significant form possible without loss of generality' (TheFreeDictionary, 2007). Tollmar, Demirdjian and Darrell (2004) use the term 'canonical configuration' to describe the hand positioning in an initialisation step of a real-time tracking system for hand gesture recognition. According to the mentioned definition, the hands position ought to be very simple or basic, but still very significant. We can assume that this step is required to analyse the hand anatomy, in order to calibrate the tracking software. The hands and fingers must therefore be clearly visible in this step.

3 THE ICP ALGORITHM

The Iterative Closest Point algorithm (henceforth ICP algorithm) is widely used in 3-D shape matching or heuristic movement tracking. In the tracking software described in the research paper, all important limbs of the user's body are subdivided into a manageable amount of points forming a 3-D cloud. The ICP algorithm is then used to iteratively calculate the minimum distance between all the points of a cloud, and consequently estimate a movement of the concerned limb.

An iteration of the ICP algorithm is divided into three separate steps, as described by Tollmar, Demirdjian and Darrell (2004, p. 115): Firstly, the algorithm has to match all the points of the virtual 3-D model to the points of the 3-D data retrieved from the real (physical) model. If the real model moves, the spatial points locations of both models will differ. For every two points belonging together, a vector can be calculated, representing the displacement of the two points in a vector space. Once all the vectors has been obtained, a global motion displacement can be calculated in a second step. This displacement is calculated for every individual limb of the the 3-D model. In the last step, all the displacements are applied to the entire 3-D model. Provided that there were no errors, these three steps are repeated, which gives the algorithm the iterative characteristic. Finally, an additional joint constraint correction was added by the authors.

4 THE WIZARD OF OZ EXPERIMENTAL METHOD

The wizard of oz experiment (henceforth WOz experiment) got its name from the famous book 'Wonderful Wizard of Oz' (Baum, 1972). In that story, a character known as the wizard appears as a terrifying floating head. In reality however, the head is controlled by the wizard hidden behind a curtain.

Analog to the story, there are also three elements in a WOz experiment: A system, the user interacting with that system, and the 'wizard' (experimenter). The latter is hidden from the user and controls the system, depending on the user's actions. The user should believe that the system reacts on her actions, whereas it is actually the wizard who undertakes that task[4]. The WOz is ideal to test a not yet fully implemented system or to observe and evaluate how a user would instinctively use an interface. In the research paper, the latter is the case, since the aim of the experiment is to evaluated which gestures or poses a user would instinctively use to navigate in a 3-D environment.

By using that technique, it became rapidly clear that users have different expectation of how the interaction should be implemented. Without the WOz experiment, this procedure could have become a much longer process.

5 PURPOSE OF AN INTERACTION MODEL

An interaction model describes in which manner a user can interact with a system. There can be different interaction model for a same task. In the research paper, the task consisted in navigating in a 3-D environment. According to the WOz experiment results, user haexd individual preferences in gestures and poses, to accomplish this task. The different gestures and poses form different ways of interacting with the system. For each of the latter, an interaction model was designed.

[4] In rare cases the user can have knowledge of the experiment technique and the hidden wizard.

References

Backman, Anders. 2007. "Interaction in Virtual Reality." Lecture 'Interaction', VRlab HPC2N, at Umeå University Sweden, Department of Computing Science.

Baum, L. Frank. 1972. *Wonderful Wizard of Oz.* Dutton Childrens Books.

Hall, D., C. Le Gal, J. Martin, O. Chomat, T. Kapuscinski and J. Crowley. 1999. Magicboard: A contribution to an intelligent office environment. In *Proceedings of the International Symposium on Intelligent Robotic Systems, 1999.*

Oka, Kenji, Yoichi Sato and Hideki Koike. 2002. "Real-Time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems." *fgr* 00:0429.

TheFreeDictionary. 2007. "Online Dictionary, Encyclopedia and Thesaurus.". Available at http://www.thefreedictionary.com.

Tollmar, Konrad, David Demirdjian and Trevor Darrell. 2004. Navigating in virtual environments using a vision-based interface. In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction.* New York, NY, USA: ACM Press pp. 113–120.