iGBL 2017
http://www.igbl-conference.com/abstract-details/?abstract_id=107

Assessment of Gesture-Based Natural Interface Systems for Serious Games

David Murphy* and Karina Dubé

* d.murphy@cs.ucc.ie

Computer Science Department
University College Cork

Introduction

Since the 1990s computer based learning and its application to various fields of medicine and surgery have become popular in medical education (Gorman, 2000; Bradley, 2006; Murphy, 2008). With respect to surgery, the emphasis in computer-based learning has been on the acquisition of skills – where students learn new procedures and skills, and practice them in a safe environment – and described by Kneebone as a "Zone of Clinical Safety" (2003). Furthermore, the area of medical and surgical training lends itself to more immersive forms of computer based learning, such as game-based learning and Virtual Reality based learning (Gallagher, 2005; McCloy, 2001; Sliney, 2011).

The goal of the current study is to examine interaction designs and devices for natural gesture-based interaction in a 3D serious game for surgical training. The surgical procedure of interest is the Laparoscopic Cholecystectomy, more commonly known as Gall Bladder Removal. While this procedure is very common, it has a relatively high error rate associated with it, hence it is the focus of a lot of training simulators. As the procedure is laparoscopic, that is minimally invasive and conducted through an endocscopic camera and instruments, it is an ideal candidate for serious game-based learning (Murphy, 2008). The current version of our surgical training system is built in Unreal Engine 4 (UE4), incorporating the PhysX engine and Flex Particle system, with a First-Person perspective. For the current phase of research, we wanted to examine the feasibility of incorporating gesture-based natural interface input methods. As part of this investigation, we compared the use of a conventional computer mouse, non-haptic data glove (5DT Data Glove), and non-contact gesture input system (Leap Motion sensor) in the selection and manipulation of a simple deformable virtual object.

The 3D model used for interaction was a particle-based deformable object using NVIDIA Flex, which is a particle based simulation system for real-time visual effects. FleX uses a unified particle representation for all object types, which facilitates new effects where different simulated substances can interact with each other seamlessly. FleX is designed to take advantage of GPUs acceleration making these effects possible in real-time applications, such as games and VR. The interface was constructed in C++ in the Unreal Engine 4 game engine. The Unreal Engine 4 is an advanced set of software tools for the development and compilation of sophisticated games and VR experiences. The interface was designed to measure the movement of the wrist around a virtual object.

Previous Work

Today we are seeing the widespread availability and commoditisation of consumer-grade virtual reality hardware, however the interaction and design concepts are not yet established and are an active area of research.

As we move toward total immersiveness, we need to adopt new modes of interacting within that immersive space. The mouse model is very effective when working with 2D interfaces and spaces, and can be mapped neatly to the physical action of moving a mouse on a table. In 3D space, the naturalness is hindered and limited by the 2D constraints of the mouse. Since the planes of motion in 3D can be dissimilar to those available in 2D, the mouse-driven techniques to compensate for this tend to be complicated, cumbersome and somewhat contrived (Herndon, 1994). To address these inherent difficulties, alternative techniques have been proposed by looking at how we interact in the real world and developing corresponding tangible interfaces with appropriate affordances. While speech can be used in some aspects of virtual interaction, dexterous tasks require more direct, manipulative interactions as associated with hand based interfaces. This can be challenging as the computer has to track and interpret the movements, positions and intentions of the user. Interactions have been classified into contact and non-contact means. Contact refers to devices using the "sensors and some kinds of hardware to construct the gesture capturing mechanism" (Weng 2015) and includes devices such as gloves, wands, and touch screens. Non-contact refers to the "vision-based technique which offers more natural interaction without using wearable devices in the hands," (Weng 2015) and include devices such as the Leap Motion and Kinect.

Gestural interfaces for interacting in 3D space, and in particular techniques for navigating and interacting in virtual space have not been tied down and are active areas of research. In 3D user interfaces several tasks are common to most systems - navigation, selection, and manipulation are three of the most common (Bowman, 2004). LaViola, as part of a course given at SIGGRAPH 2011, outlined the various means of these interactions (LaViola, 2011). For selection, defined as "the process of accessing one or more objects in a 3D virtual world", several implementation aspects need to be considered. Triggering of the selection event, feedback regarding the selected object, and efficiency in detecting selectable objects. The most common selection technique is a virtual hand, which selects based on collision with virtual objects, followed by the ray-casting technique which casts a vector (ray) in to a virtual scene and selects the object that it first intersects with. The first method is more akin to real life explorative manipulation while the second allows for more direct and greater accuracy and efficiency.

Manipulation can be seen as an extension of the selection process, and so careful integration between the two techniques is important. Additionally, the selection mode must be turned off when switched to manipulation mode and consideration of the outcome of releasing the selection must be given.

In the current study we have implemented a version of Hand-Centered Object Manipulation Extending Ray-Casting (HOMER – Lawrence, 1997). The metaphor of the laser pointer is adopted for selecting objects, as it allows the user to see what is actually being selected without “physically” touching the model in virtual space. However, once the selection is made, it switches to the “virtual hand” method for manipulating the object. The system modifies both of these methods by having the ray-casting operate on the area of a hemisphere for selection, while the virtual hand representation is simply the pull and push, indicated by the disk, with the intent of the user is inferred by the direction they move their hands. This adapted hybrid approach, especially in the context of Virtual Reality, was first recommended by Pavlovic (Pavlovic, 1997), and further explored by Mitra et. Al (Mitra, 2007).

Methods

To evaluate the veracity of using natural interfaces as a form of interaction as idendified by Satava (Satava, 1993), in our surgical training system we developed a prototype where particular input methods were evaluated and compared. In designing the experiment, we wanted to evaluate the accuracy and discoverability of the input methods. The first criterion was measured by the error from an ideal target, while the second was measured by the time between starting the task and making a selection.

The experiment utilizes three different devices (mouse, glove, and Leap Motion) along with the keyboard to manipulate a virtual bar and disk used to push and pull the spherical particles of a softbody (Nvidia Flex) object in the world. The finalized softbody object was a slightly rounded cube of Flex particles.

Shared keyboard inputs for the three device interactions include the A and D keys to rotate around the model, the space bar to lock a selection and switch from selection mode to manipulation mode, and the enter key as the marker for start and stop of a task.

The A and D keys were chosen as they are the familiar and standard input navigation keys in computer games (WASD). However, the mapping is slightly different as the movement is not translational nor changing the viewport according to the user, but rather rotational around a point. The A and D keys were also chosen instead of the arrow keys because it would ease the separation between the two hands, allowing the right hand to focus on movement while the left hand focused on other selection tasks.

The space bar was chosen to maintain consistency among the different devices. Original versions of the system included different methods for locking a selection for each device. The mouse selection involved clicking and holding down the left mouse button, a standard and intuitive way of making a selection. The Leap selection utilized the pinch gesture available through UE4. However, the glove had no particular method for making a selection, since the original design utilized the flex of the fingers as the area of the selection. Because there was no method of selection for the glove, and the Leap selection gesture was unstable, the decision was made to use the spacebar as the locking mechanism for all devices.

The enter key was chosen as it was farther away from the A and D keys, requiring deliberate thought to select the beginning and end of the trial. The mouse was used specifically for training purposes, familiarising subjects with committing selections using the space bar, locating the targets, and pressing the Enter key on start and end of the task. The mouse utilized the 2D Cartesian coordinates of the mouse in the viewport and mapped the values to the pitch and yaw of the target. The glove and Leap used the roll to control the yaw, and the pitch to control the pitch of the target. The difference was that the glove used the wrist sensors located on the back of the wrist, while the Leap used the palm orientation. The effect of this is negligible as they are in approximately the same position on the hand.

A combined user study was conducted to compare the

A combined user study was conducted to compare the usability of contact (5DT Glove) and non-contact (the Leap) gesture-based input devices for use in the system. The combined study incorporated both User Experience and functional usability techniques. Eleven right-handed adult volunteers were recruited. Each volunteer was asked to perform four tasks on the deformable model. These tasks involved pushing and pulling certain points and faces on the 3D model.

To account for possible learning bias, subjects were randomly assigned their first device. After each device, the subjects completed a System Usability Scale questionnaire.

The experiment collected information on the (a) accuracy, (b) discoverability, and (c) ease of use of the two devices.

- Accuracy was measured based on the distance between the user's selected location and a predetermined target point.
- Discoverability was based on the time taken to reach the selection point.
- Ease of use was determined by survey.

The users were asked to complete four tasks related to virtual object manipulation. They were:

1. Push the front bottom right corner of the model
2. Pull the back top right corner of the model
3. Push the front face of the model
4. Pull the top face of the model

Subjects started with the mouse as the input method for training, and were then randomly assigned a first device by coin toss. Out of 11 subjects, six were assigned the glove first and five were assigned the Leap first.

Subjects were read a script designed to give only cursory information about how to make selections and manipulate the target; they were given very little instruction on the mechanism for using the wrist for manipulation. This was intentional as the discoverability of the method was something that we wanted to establish.

The user completed the four tasks with their first assigned device, completed a questionnaire, completed the four tasks with the other device, and then filled out the questionnaire in relation to the second device. The questionnaire was a relation to the second device. The questionnaire was a version of the System Usability Scale, chosen for ease of evaluation and its standardization.

While the user was performing the tasks, the system was capturing the input data and writing it to a file. To measure accuracy, the pitch, yaw, and roll of the manipulation object was compared with a pre-determined ideal position. The time taken to reach this ideal position, was captured as a measure of how easy it was for the user to manipulate around the object.

The subjects were polled on whether they play video games, how many hours per week they play, and their primary mode of gaming (virtual, i.e. mobile phones; keyboard-and-mouse, i.e. PC; or console, i.e. controller).

The questions asked in the User Experience (UX) part include:

- Which is more accurate: the Glove or the Leap?
- How discoverable is movement designed to mimic the motion of the users wrist?
- Which device provides a better experience for the user?

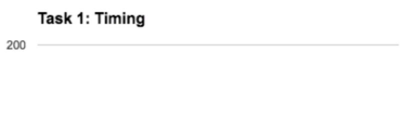
Findings & Results

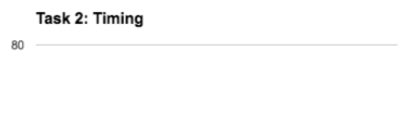
Standard descriptive statistics are employed (mainly means and standard deviations).

Discoverability

Discoverability was based on the time it took for the user to commit to a location for their selection. This measures the "playing around" time in order to achieve the goal.

The first three tasks show a preference for the glove in speed, which when taken independently of accuracy can indicate users' preference for it. Only the fourth task indicates a choice for the Leap. Figure 1 shows the four results. Note that timing should not be compared between different tasks, only within the devices for the particular task.





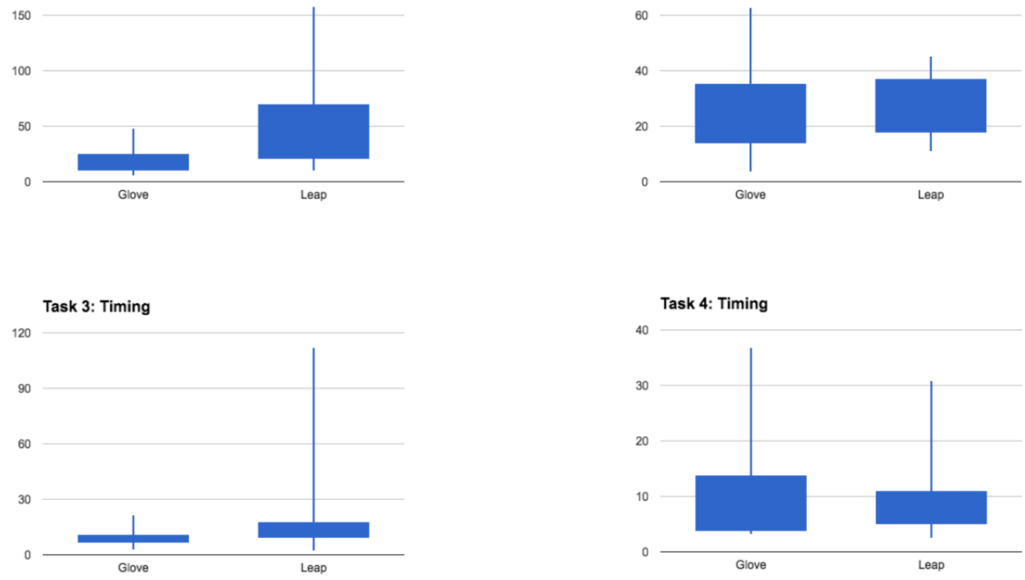


Figure 1. Spread of timing values from start to most accurate selection. Glove has better averages in all tasks except for task 4

Error Rate

(Comparing accuracy between the glove and the Leap)

The first task involved the user pushing the bottom-right corner of the model. From the results, it seems that the glove (mean of 18.37) was more accurate overall than the Leap (mean of 30.10). A possible explanation for the greater accuracy could be that the user had more freedom to move around with the glove and eventually hit the correct spot, even if the mental model was not well discoverable.

The second task involved pulling the back top-right corner of the model. Once corrected for some ambiguity in the instruction, the result is a mixed picture in which neither the Leap (mean of 6.78) nor the Glove (mean of 5.41) appears to be more accurate.

The third task involved pushing the front face of the model. Again the users were slightly better using the Glove (mean of 11.08) than they were with the Leap (mean of 16.74).

Finally, the fourth task was pulling the top face of the model. The results were comparable for the glove (mean of 49.08) and the Leap (mean of 43.76) with the Leap slightly better.

User Experience

Out of the 11 subjects, six chose the Leap as the preferable interaction method while five subjects chose the glove as the preferable method. The SUS results give a slightly different picture. The SUS results show five subjects with a preference for the glove while only four subjects prefer the Leap and two rated the two devices as equal. Additionally, the average score for the glove of those who preferred the glove is much higher than the average score for those who preferred the Leap.

Comments on preferences seem to mostly fall into the category of "device X seemed more accurate" or "device X seemed easier to use". Glove-subjects tend to describe the glove as more accurate, along with easy to use, while Leap-subjects cite ease of use more often than they mention accuracy. For those who preferred the Leap, a major factor was the limits of the glove and how clunky it felt, whether citing the glove as negative or the Leap as positive. That being said, this preference has the possibility of being eliminated as lighter and more advanced gloves are available.

We can also split users up into various demographics. Four subjects self-identified as non-gamers, three of whom preferred the Leap. Seven subjects self-identified as gamers and they were also close, with a majority toward the glove (4-3). Among the gamers, there seems to be no correlation between how long individuals play video games and their preference.

Discussion

The study undertaken comprised a small number of subjects. A greater sample size is necessary to yield results regarding any possible correlation between gaming and device usage. That said, even with a small sample size, the results highlight interesting questions. Primarily they indicate that while non-contact based devices might be perceived as having greater ease of use due to the lack of restrictions and cabling when compared to the glove, the glove was more accurate. The level of preference for the glove among those who preferred the glove was higher than the level of preference for the Leap among those who preferred the Leap.

The significance of this is dependent upon the type of learning scenario employed in the serious game. If the emphasis within the serious game is on task training, e.g.

emphasis within the serious game is on task training, e.g. dexterous tasks associated with surgical instrument manipulation, then the results suggest the use of the Glove is more appropriate due its higher accuracy. However, if the focus in the serious game is on naturalness of interaction then the Leap is more highly preferred.

There has been relatively little research undertaken to date on this problem, despite it being identified by Schijven (Schijven, 2003) and Halvorsen (Halvorsen, 2005). This probably has more to do with the emphasis on traditional surgical instrument inputs.

These insights may help serious game designers and educational content creators in adopting the appropriate input device for natural gesture-based interaction.

Conclusion

In this study we have undertaken a preliminary evaluation of the feasibility of incorporating gesture-based natural interfaces in to a serious game. The results indicate that the non-contact gesture input device offered a more natural interaction, while the more encumbered contact gesture interaction device (glove) was more accurate.

Recommendations

We plan to further explore the suitability of these two different (contact vs non-contact) gesture based input devices in more dexterous tasks in the next version of the surgical training system. There is scope for further studies to compare the learning against more established simulation systems. We also plan to evaluate these gesture inputs with medical and surgical practionners, and finally we will have to evaluate the 'transferability' of the learning to the actual surgical procedure itself.

References

Gorman, P.J., Meier, A.H., & Krummel, T.M., 2000. Computer-

Gorman, P.J., Meier A.H., & Krummel, T.M., 2000. Computer assisted training and learning in surgery. *Computer Aided Surgery, 5,* 120–130.

Bradley, P., 2006. The History of Simulation in Medical Education and possible future directions. *Medical Education, 40,* 254-264

Bowman, D., Kruijff, E., LaViola Jr, J.J. and Poupyrev, I.P., 2004. 3D User Interfaces: Theory and Practice, CourseSmart eTextbook. Addison-Wesley.

Murphy, D., Doherty, G., &Luz, S., 2008. Differentiating between novice and expert surgeons based on errors derived from task analysis., 15th European Conference on Cognitive Ergonomics. ACM, 32.

Pavlovic, V.I., Sharma, R. and Huang, T.S., 1997. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on pattern analysis and machine intelligence*, *19*(7), pp.677-695.

Mitra, S. and Acharya, T., 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *37*(3), pp.311-324.

Satava, R.M., 1993. Virtual reality surgical simulator. *Surgical endoscopy*, *7*(3), pp.203-205.

Schijven, M. and Jakimowicz, J., 2003. Virtual reality surgical laparoscopic simulators. *Surgical Endoscopy And Other Interventional Techniques*, *17*(12), pp.1943-1950.

Halvorsen, F.H., Elle, O.J. and Fosse, E., 2005. Simulators in surgery. *Minimally Invasive Therapy & Allied Technologies*, *14*(4-5), pp.214-223.

Gallagher, A. G., et al., 2005, Virtual Reality Simulation for the Operating Room: Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training. Annals of Surgery, 241(2), p.364

McCloy, R., Stone, R., 2001. Virtual Reality in Surgery. British Medical Journal, 323, p912.

Sliney, A., Murphy, D., 2011. Using serious games for assessment.

Serious Games and Edutainment Applications. Springer London

Kneebone, R., 2003. Simulation in surgical training: Educational issues and practical implications. *Medical Education, 37*(3), 267–77.

Herndon, K.P., Van Dam, A., Gleicher, M., 1994. “The challenges of 3D interaction”. ACM SIGCHI Bulletin 26(4) pp. 36–43.

Weng, S., Hsieh, C., Chu, Y., 2015. A Vision-based Virtual Keyboard Design. Intelligent Systems and Applications: Proceedings of the International Computer Symposium. pp. 641–647.

Lawrence, D., Cutler, L.D., et al., 1997. Two-handed direct manipulation on the responsive workbench. Proceedings of the 1997 symposium on Interactive 3D graphics, pp. 107–114.

LaViola, J.J., Keefe, D.F., 2011. 3D spatial interaction. ACM SIGGRAPH ’11, pp. 1–75.