Reviewing Design and Performance of natural mid-air gestures for virtual information manipulation

Pratik Gupta
Department of Computer Science
University of Auckland, New Zealand
pgup014@aucklanduni.ac.nz

ABSTRACT

Interacting with a computing device should be as similar as interacting with the real world. Humans have an incredible ability to reach out and interact with real objects which have always been intuitive to humans. This innate ability has been confused or unlearnt when interacting with a computing device by traditional peripherals such as keyboards, mice and touch screens. It is ideal for the user to interact with a computing device in a more natural way which doesn't require much learning. This is the point of a Natural User Interface (NUI). The user interface essentially becomes invisible when it is interacted with it in a natural way. This report will review various literature that discuss applications of natural gestures and real world metaphors for direct manipulation at a distance to explore large information spaces. The report will also review the design and performance comparisons of various natural interaction techniques discussed in the publications. Majority of the publications include a user study carried out with quantitative analysis on the input based interaction techniques. Such an analysis is reviewed to gather a conclusion from the literature. The commonalities across the publications are also discussed to draw the trend and future of natural gestural interfaces.

Author Keywords

Literature Review, NUIs, Gestures, Pan & Zoom, Wall-sized displays.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces. - Graphical user interfaces - Input devices and strategies.

General Terms

Human Factors; Design; Performance.

INTRODUCTION

Certain tasks such as manipulating 3D models, an expansive map or large sets of images are inefficient to perform with only ubiquitous and conventional peripherals such as keyboards, mice and touch screens. This is true especially with large displays [2, 4, 6]. The GUIs are traditionally interacted with the common peripherals through virtual elements such as windows, menus or buttons. These virtual elements and metaphors tend to contaminate the gestural interface fading out their naturalness [2]. Thus there is a need for an interaction technique that is efficient and easier. This is achieved by Natural User Interfaces (NUI) which give the users more control over the digital content while at the same time being easy and comfortable to use.

The ultimate goal of such interfaces is to enable the user to interact with the system in the same way as they would with another human. It is apparent that there is on-going effort in the human-computer interaction research community to investigate various natural interaction techniques with the end goal of designing most suitable ways for input and manipulation. The area of NUI research deserves a review as there is much proven potential in its application and effectiveness [2, 3, 4, 6, 7, 8].

This report will review the design and performance of natural mid-air gestures for interacting with virtual information, researched in various publications . The body of the report is divided into four sections, where each (based on the publications referenced at the end) discusses the design, application, technology, user study findings and problems of the natural interaction techniques. These techniques are classified as controller based and free hand interactions. Finally this literature review concludes with a summary and discussion on future trend of NUI, inspired from the publications. The review is believed to spark the reader's interest on NUI.

INTERACTION DESIGN AND APPLICATION

Natural interaction design is a crucial aspect in NUI research for a particular application and should follow guidelines for optimal usability [1]. There should be a iterative design process and experimentation with users. There is emphasis on gestural input being independent of location and should not require a mounted hard surface [2, 3, 6, 7]. This section will review various literature that report natural interaction designs classified as Controller based interaction and free hand interaction.

Controller based interaction

Nancel et al. [6] analyse related work on pan-and-zoom by a distant pointing interface for multi-scale navigation on wall-sized displays. They reveal three key factors for their design on mid-air interaction of pan-and-zoom techniques: uni- vs. bi-manual interaction, linear vs. circular movements and level of guidance that are movements in a 1D path, 2D surface and 3D free space to accomplish the gestures in mid-air. Their bi-manual interaction evaluates the effectiveness of coupling intuitive free hand gestures with device based interaction techniques to provide a natural experience in navigating digital information on a large scale. Panning is achieved by simple pointing and zooming controllers are mouse scroll wheel, click wheel and smart phone touch screen. The application of interaction techniques by Nancel et al. [6] is said to span across: astronomers, biologists, artists and crisis management centres.

Stellmach and Dachselt [8] also devise interactions for pan and zoom that is tested on navigating maps of cities in Google Earth. They design a gaze supported panning in combination with different zoom interfaces (similar to Nancel et al. [6]): a mouse scroll wheel, tilting a hand held device and a smartphone's touch screen. It is believed that only a few studies have examined gaze interaction in combination with handheld control devices [8]. Such a combination allows the user to focus on current point of interest by vision while zooming in simply using the scroll wheel or touch screen. Their potential applications are to browse medical data in sterile environments or navigate geographical information system on a distant display. But they do not indicate the range at which the eye gaze tracker can sense the user's eyes. This causes doubt on the function of gaze supported panning on a large display from a distance greater than 60cm as evaluated in the user study [8].

Francese et al. [2] presents a way to interact and navigate on 3D geographical maps using Bing by two gesture tracking methods via gaming devices: Nintendo Wiimote and Microsoft Kinect. The Wiimote is the main controller based interface held in one hand with the inclusion of an additional controller called Nunchuck held in the other hand. The Wiimote acts on forward and backwards motion when rotated along its longer dimension (Roll) and its inclination senses if navigation turns (Pitch). These two gestures are analogous to the throttle and handlebar on a motorcycle to go ahead and turn respectively [2]. Simultaneously, tilting the Nunchuck up and down controls the altitude of the map navigation.

Free hand interaction

The literature by Baudel et al. [1] composed ten years ago points out, that the known benefits of using free hand gestures are natural, direct, terse and powerful. But they also suggest that tasks requiring precision interaction such as drawing should not be performed by free hand gesture input. This suggestion could be due to the fact that the sensors at that time did not have high enough resolution to sense the instability of the hand in free space [1]. Now, higher resolution and advanced features of camera sensors (such as Microsoft Kinect) have capability to accurately sense the hand in free space [2, 4, 7].

Iacolina et al. [4] points out, controlling 3D virtual objects on general 2D displays tend to be a challenging task for a novel user. They discussed that exhibition centres such as museums benefit from interaction systems where there is a natural relationship between the user and content. This literature describes the fact of NUIs fading out the notion that a computing device is controlling a user's experience. They compare two innovative interaction techniques: multi-touch tables and free hand gesture recognition based on vision which both allow unrestricted manipulation. Their two interaction techniques allows casual users to manipulate virtual 3D objects, on an optimal display space, with a range of scaling, panning, rotation and zooming controls as intuitive two hand motions. Their technique frees the user from utilising a separate interaction device which reduces user interface complexity. They describe that the multi-touch interface is based on press/release which is similar to the free hand interaction of opening and closing the hand. This is analogous to the act of grasping a real object.

Song et al. [7] reveal a more robust free hand interaction design using a handle bar metaphor that has the ability to execute continuous transitions between Rotation, Translation and Scaling (RTS) operations on 3D objects without the need to switch manipulation modes. Their virtual handle bar is controlled by the user's two free hands which just gestures a grasp on a handle bar. The handle bar is simply a tool to not only manipulate a 3D object but also the viewpoint of the scene. The beauty of the handle bar metaphor is that multiple objects can be manipulated together by piercing through them. As seen from their user study results, this reduces the interaction time and effort [7]. Issues with their rotation operation and gestures are reviewed in the problem section of this literature review

IMPLEMENTED TECHNOLOGY

The mid-air gestural panning and zooming by Nancel et al. [6] was carried out at a distance on a very large display area constituting of thirty-two 30-inch tiled monitors which display 131 million pixels. Such a display size affords more physical rather than virtual forms of navigation and thus provides the researchers better chance to evaluate the effectiveness of natural interaction techniques. Other studies suggest that large displays are also beneficial for information visualising and analysing large sets of data [6, 10]. Studies compare the effects on users with the controller interaction, free hand motion tracking [2, 6].

Controller based interaction

The handheld controller device, often leveraged by the NUI research community is known as the Nintendo Wiimote [2]. This commercially available gaming device is primarily used with the gaming console Ninento Wii. Explained by Francese et al. [2], the Wiimote can connect to a computing device over wireless Bluetooth. It embeds a accelerometer sensing three axis motion, an Infra-Red (IR) sensor bar that determines where the device is point. It also offers a set of classic joypad buttons. Being a haptic device, it adopts a speaker, a vibrating motor and four light emitting leds as feedback [2]. Wiimote is expandable with several accessories. Francese et al. [2] use a second controller known as a Nunchuck that conveniently plugs into the Wiimote via a cable. The Nunchuck provides 2 buttons, an analog joystick as well as an independent accelerometer. The capability of these two controllers give users realistic experiences. The standard mouse scroll wheel, click wheel and touch screen devices are used by studies to mostly perform zoom operations [6, 8].

Free-hand interaction

Many free hand interaction techniques are functioned through the use of Microsoft Kinect which represents the first consumer full body motion with depth capture device embedded with an infra-red emitter, two video cameras and array of microphones [2, 3, 4, 7]. Kinect is primarily designed for augmenting gaming with the Microsoft Xbox 360 console. But Kinect can interface with desktop computers, as the NUI application programming index in the Kinect enables applications to access and control data acquired from the sensor [2]. Specialised interaction software drivers are used, such as PrimeSense's OpenNI, to recognise complex gestures [7]. This level of customisation would indeed require expert technical developers as evident with Song et. al [7]. There are some limited resolution issues with the Kinect 3D scene acquisition sensors, but gestural design choices can be made to circumvent issues caused by the limitations [4, 7]. To overcome the optimal distance restrictions of Kinect (from 0.8 to 4 meters) while still maintaining good screen readability, wall projectors were used for display [2, 3, 4].

Stellmach and Daschselt [8] implement a table mounted binocular eye tracker for gaze supported panning. Iacolina et al. [4] used a multitouch table for free hands 3D models manipulation. They used an improved sensor to allow increased robustness to change of lighting conditions. Even though the devices implemented are available off the shelf there is still a requirement to augment or modify them in order for them to function accurately for the interaction technique's purpose [3, 4, 7, 8].

FINDINGS AND PERFORMANCE

Francese et al. [2] evaluate their natural interaction techniques on subjective usability and deep perceived sense of Presence and Immersion. Such qualitative analysis is crucial as it proves effectiveness of the interaction technique. If other NUI interaction studies would have carried out this sort of analysis, they could have ensured viability on the range of applications they boast [3, 6, 7]. Of course to experience the presence and immersion the task choice of the studies needs to be suitable as well. In the case of Nancel et al. [6], interaction with target circles was enough to gauge speed and accuracy of their techniques but insufficient to gain feedback on the dynamic experience. They were interested to know if gestures performed freely in space work better than the input through devices operated in mid-air. User studies that carry out quite an extensive analysis on findings, gain the most thorough insight on the techniques performance [6,8]. Various user study findings are analysed mostly on: task completion time, overshoots, qualitative results and individual techniques [2, 3, 6, 7, 8]. Also all participants in the user studies have no or slight experience to the interaction techniques which give fair and non-biased results.

Research groups which do not carry out or present any user studies fail to give any evidence on the performance of their technique [4, 10].

Controller based interaction

Nancel et al. [6] find that in terms of task completion time, controller based gestures with 1D path are fastest (avg. 9511ms) than 2D touch surface (avg. 10894 ms) and 3D (avg. 11934 ms) free gestures. They also found that bi-manual (two handed) gestures are significantly faster than uni-manual (one handed) gestures (avg. 9690ms vs. 11869ms) . These findings support the hypothesis of two-handed techniques being faster than one-handed techniques and gestures being performed in free space are less efficient and easily susceptible to tiredness [6].

But it is discovered that some hypothesises such as linear and controller based gestures should be slower because of clutching contradict the outcomes [6]. Majority of participant's subjective comments suggest: accuracy is much better with more haptic guidance to input gestures [6]. Also that linear gestures have higher efficiency (avg. 9384 ms) than circular clutch-free gestures (avg. 12175 ms) in 3D free space.

The user study from Stellmach and Dachselt [8] evaluates participant's feedback of their experience with their techniques, improvements and if these techniques could replace the traditional mouse input for situations in which a mouse may not be suitable. They found that task completion time for panning by gaze with scroll wheel and touch screen zooming was fastest at around 15 seconds. While the panning by gaze and tilt zooming (no touch guidance) was much worse at about 21 seconds. This also matches findings by Nancel et [5] that there is lot more efficiency and accuracy with interfaces offering higher interaction guidance.

Free-hand interaction

To navigate 3D maps, Francese et al. [2] find that average overall evaluation of 5.78 on the Kinetic for free hand motion is higher than that of Wiimote controller (5.13). This suggested their interaction technique with Kinect has better system usefulness, information and interface quality. Also the involvement and control factor of 5.89 and 6.14 respectively from free hand interaction is higher than that from Wiimote controller (5.39 and 5.87). This as well as the qualitative results proves that their Kinect free hand gesture technique with the virtual paper plane metaphor is more effective for sense of presence and immersion.

Hespanhol et al. [3] user study results show median time taken to learn gestures: pushing to place an item and lassoing to select an item are both 45 seconds first time. In second attempt pushing is 14 seconds and lassoing is 30 second. Compared to improvement rates of other gestures, it seems that pushing and lassoing are not highly intuitive. This means that the design of interactions should be made more suitable for the task at hand.

Song et al. [7] find that participants in their user studies are able to increase the times an object is manipulated correctly within a set duration in successive attempts, where at first attempt average number of times is 4.6 and is doubled at sixth attempt with 9.3 times. Also with their unique handle bar metaphor they found that average time taken to manipulate objects one by one at 34.5 s was significantly slower than multi-object manipulation at 10.5 s. These findings secures the fact that their interaction design improves the efficiency of free hand manipulation of 3D models.

PROBLEMS

Baudel et al. [1] discuss some of the known pitfalls of using any gestural communication such as fatigue which require more muscle usage than mouse, keyboard or speech interaction since the wrist, fingers, hand and arm together express commands. They suggest that gestural commands must thus be terse and fast to issue in order to have minimal effort. Other pitfall they point out is non-self-revealing, which means the user must remember the set of gestures that the system recognises. So the gestural commands should be simple and consistent yet natural with importance on suitable feedback to the user. Zigelbaum et al. [10] present video content interaction technique on large displays with a vast set of gestures that are bound to be difficult for the user to remember and perform.

Controller based interaction

With controller based interaction, user's controlling experience is limited by the need grasping the controller with hands [2, 6]. But evident from the findings instability of the hand in free space still proves a need for tactile feedback [1].

Free-hand interaction

Hespanhol et al. [3] point out, that gestures involving free hand movements that do not provide visual feedback need to be augmented by modifying the virtual object's behaviour. With free hand interaction, the interactive display can communicate to the user only by visual and auditory senses [3, 9]. Further, the opposite channel of communication from the user to system can only track physical movements performed by the user in front of the display [9]. So for clear and unambiguous communication in both directions, visual and audio cues must be presented to aid the user performing the task while not being distracted [3]. The interfaces that provide visual feedback allow a better sense of control on the virtual scene or object's behaviour, there is evidence to support this in the user study results [2, 3, 7].

The literature by Baudel et al. [1] which was published in the early nineties, further point out that due to limitations in computer vision technology there is lack of comfort. This is referring to their gestural interface of wearing a glove wired to a system which is an obsolete technique considering the advancement of capturing gestures with modern vision sensors. But despite the present capability of vision sensing systems, there is still another issue called Immersion Syndrome or Midas touch where every motion of the user is captured and constantly interpreted by the system [1, 7, 8, 9]. This causes undesired operations from misinterpreting user's unintended hand gestures.

There are interaction techniques which circumvent this problem by entering into a “Neutral state” or toggling the sensing mode by brief gestures, touch on a screen or clicking the scroll wheel [7, 8]. Such a problem is crucial to address especially when integrating a bi-manual free hand gesture interaction in a medical and sterile environment application. But in regards to 3D content manipulation by natural hand gestures, Iacolina et al. [4] did not address or point out this problem.

Furthermore, with camera sensors there is an issue of occlusion with hand gestures. Song et al. [7] utilise a constrained or incremental rotation to alleviate the problem of rotating the virtual handle bar about the x-axis in high angles where the hand in the front occludes the hand at the back, resulting in an undetermined 3D pose of the handle bar.

The virtual handle bar interaction from Song et al. [7] use the point and open hand gestures that are sensitive to orientation of the hand, making it less robust to recognize than the close gestures. This is the reason why they are responsible for interactions that require less complicated gestures and used less often such as manipulating and browsing the handle bar.

But since the close fist gesture is orientation independent and therefore more robust, it is utilised in interactions for object manipulation that frequently need the user to perform bi-manual motion gestures that have a high degree of freedom. The interaction technique by Hespanhol et al. [3] faced a technical impediment when implementing the grabbing and enclosing gesture to select or rearrange virtual items. They point out that open/close fist movement with blob tracking is not trivial and believe it will hinder the user sessions. So in order to quickly test if the gesture is relevant from a usability perspective, they utilised a quick solution called the Wizard-of-Oz prototype where the activation of the gesture is simulated by the researcher clicking a specific key on the computer running the application. This ensured the illusion of total smoothness when opening and closing fist for the gesture regardless of its orientation or magnitude. Hespanhol et al. [3] and Song et al. [7] both make sensible decisions to implement the recognition of their interaction technique.

SUMMARY

It is apparent that there is no standard for developing gesture interfaces, and each research may come up with unique solutions or design but there is much less synergy and effectiveness of interaction techniques if standard guidelines are not followed [1, 9]. There are interaction techniques that are specifically suitable for certain tasks and applications only. Such is the case of 3D map navigation [2] and 3D model manipulation [7] with free hand gestures.

FUTURE WORK

There is discussion that future work will be aimed at exploring new applications for the techniques reviewed that match the way people interact with objects and people [2, 3, 4]. The publications further conclude that social and collaborative dimensions will be explored in the future for interactive scientific, corporate, leisure and learning environment [2, 3, 4, 5, 6, 7]. An interesting discussion is on the possibility of multi modal interfaces where the advantages of different interaction techniques can be amalgamated to deliver an effective and efficient NUI experience [5].

REFERENCES

1. Baudel, T. and Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Commun. ACM* 36, 7 (July 1993), 28-35. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/159544.159562

2. Francese, R., Passero, I., and Tortora, G. (2012). Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. *In Proc. of the International Working Conference on Advanced Visual Interfaces (AVI '12).* Genny Tortora, Stefano Levialdi, and Maurizio Tucci (Eds.). ACM, New York, NY, USA, 116-123. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/2254556.2254580

3. Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J. (2012). Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. *In Proc. of the 2012 International Symposium on Pervasive Displays (PerDis '12).* ACM, New York, NY, USA, , Article 6 , 6 pages. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/2307798.2307804

4. Iacolina, S. A., Soro, A., and Scateni, R. (2011). Natural exploration of 3D models. *In Proc. of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity (CHItaly), Patrizia Marti, Alessandro Soro, Luciano Gamberini, and Sebastiano Bagnara (Eds.).* ACM, New York, NY, USA, 118-121. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/2037296.2037326

5. Jain, J., Lund, A., and Wixon, D. (2011). The future of natural user interfaces. *In CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11).* ACM, New York, NY, USA, 211-214. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/1979742.1979527

6. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. (2011). Mid-air pan-and-zoom on wall-sized displays. *In Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11).* ACM, New York, NY, USA, 177-186. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/1978942.1978969

7. Song, P., Goh, W.B., Hutama, W., Fu, C., and Liu, X. (2012). A handle bar metaphor for virtual object manipulation with mid-air interaction. *In Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12).* ACM, New York, NY, USA, 1297-1306. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/2207676.2208585

8. Stellmach, S., and Dachselt, R. (2012). Investigating gaze-supported multimodal pan and zoom. *In Proc. of the Symposium on Eye Tracking Research and Applications (ETRA '12), Stephen N. Spencer (Ed.).* ACM, New York, NY, USA, 357-360. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/2168556.2168636

9. Steven C. Seow, Dennis Wixon, Ann Morrison, and Giulio Jacucci. (2010). Natural user interfaces: the prospect and challenge of touch and gestural computing. *In CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10).* ACM, New York, NY, USA, 4453-4456. DOI=10.1145/1753846.1754172 http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/1753846.1754172

10. Zigelbaum, J., Browning, A., Leithinger, D., Bau, O., and Ishii, H. (2010). g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. *In Proc. of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10).* ACM, New York, NY, USA, 261-264. http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/1709886.1709939