Natural Interaction in Augmented Reality Context

John Aliprantis[1], Markos Konstantakis[1], Rozalia Nikopoulou[2],
Phivos Mylonas[2] and George Caridakis[1] ✉

[1] University of the Aegean,
81100 Mytilene, Greece
{jalip, mkonstadakis, gcari}@aegean.gr

[2] Ionian University
49100 Corfu, Greece
rnikopoulou@gmail.com, fmylonas@ionio.gr

Abstract. In recent years, immersive technologies like Virtual and Augmented Reality have been accelerating at an incredible pace, building innovative experiences and developing new interaction paradigms. Current research has widely explored gesture interaction with Augmented Reality interfaces, but usually requires users to manipulate input devices that could be cumbersome and obtrusive, thus preventing them from interacting efficiently with the 3D environment. Therefore, Natural User Interfaces and freehand gesture interaction are becoming more and more popular, improving the user's engagement and sense of presence, providing more stimulating, user-friendly and non-obtrusive interaction methods. However, researchers argue about the impact of the interaction fidelity in usability and user satisfaction, questioning the level of naturalness that should characterize the interaction metaphors. Current paper proposes different gesture recognition techniques for three basic interaction categories (translation, rotation and scaling) in a Leap Motion Controller - Augmented Reality framework. A prototype is implemented in order to evaluate efficiency and usability of the proposed architecture. Finally, experimental results are discussed.

Keywords: Natural Interaction · Augmented Reality · Leap Motion Controller · Gesture Recognition

1   Introduction

Over the last few years, Augmented Reality (AR) has developed into a cutting edge technology, providing new ways to interact with computer – generated information. By removing the boundaries between physical and virtual, AR has been able to create more engaging experiences, enhancing user's enjoyment and satisfaction. Meanwhile, interaction in AR applications now requires users to manipulate the AR virtual content in a 3D interface, adding a different perspective in the Human – Computer Interaction (HCI) research field. However, according to Bowman [1], 3D interaction

is more complicated than nontraditional systems as it requires new sets of interface components (new devices, new techniques, new metaphors) that offer unlimited options in designing new interactive methods based on user experience. 3D interaction can also be challenging for users to manipulate these innovative systems efficiently and perform as desired.

Natural User Interfaces (NUIs) seem to be in a position similar to that occupied by the GUIs (Graphical User Interfaces) in the early 1980s. NUIs promise to reduce the barriers to computing still further, while simultaneously increasing the user's power, and enable computing to access still further niches of use [2]. NUIs allow users to use the interface with little or no training, based only on their existing knowledge and can be characterized as intuitive, flexible and fluid, as they enable users to easily customize the interface to better suit their needs and also use it without any interruption [3]. With the emergence of NUIs, HCI aims in evolving in a regime where interactions with computers will be as natural as interactions between humans, and to this end, incorporating gestures in HCI is an important research area.

Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with computers [4]. The hand is extensively used for gesturing compared with other body parts because it is a natural medium for communication between humans and thus the most suitable tool for HCI [5]. With the increasing performance of computational and graphics hardware and the emergence of low-cost sensor technologies such as Leap Motion or Intel RealSense, interaction in 3D environments is now more natural, stimulating and user - friendly for users who perform gestures or spoken commands without any other peripheral equipment. This is also crucial for AR applications in fields like Cultural Heritage (CH), where state-of-the-art technology can serve as an on-demand service for users / visitors, who can fully enjoy the cultural heritage application without being distracted by the technology itself.

The challenge of generating natural and intuitive user interfaces while keeping user experience in mind has been characterized as an important area for future research. This work aims in taking a small step towards understanding the importance and the potential of natural interaction and free-hand gesture recognition in designing 3D interfaces. We discuss the current research and NUI literature, focusing on gesture approaches and naturalism levels on the design of 3D User Interfaces (UIs). Furthermore, we describe the implementation procedure of gestures using the Leap Motion sensor in an AR framework and evaluate their efficiency and ease-of-use. Finally, we summarize with the results of our experiment and our future plans.

2   Related Work

In recent years, many research activities have been carried out with the goal of designing a fully featured, interactive augmented reality NUI. Authors in [6] presented a novel concept and prototypical implementation of a Leap Motion Controller in a hybrid AR interface approach which allowed for correct mutual occlusions and interactions in a finger-based interface. The "VoxelAR" concept can be applied in modified ways to any video see-through AR system and enables users to

interact with a virtual environment (VE) in a hand-controlled interface, allowing for correct mutual occlusions between interacting fingers and the VE.

In this work [8], authors proposed a free-hand interaction system with Leap Motion controller for stroke rehabilitation by modifying the Fruit Ninja game to use the Leap sensor's hand tracking data. The combination was prepared for patients with stroke to practice their fine motor control. In another study [7] in CH field, a prototype of a wearable, interactive AR system for the enjoyment of the CH in outdoor environments is presented. By using a binocular see-through display and time-of-flight (ToF) depth sensor, the system provides users with a visual augmentation of their surroundings and they also can use touchless interaction techniques to interact with synthetic elements overlapping with the real world. Furthermore, authors in [9] investigated the potential of finger tracking for gesture-based interaction by presenting two experiments in which they evaluated canonical operations such as translation, rotation, and scaling of virtual objects with respect to performance (time and accuracy) and engagement (subjective user feedback).

Finally, recent research has explored free-hand gesture interaction with AR interfaces, but there have been few formal evaluations conducted with such systems. Authors in [10] introduced and evaluated two natural interaction techniques: the free-hand gesture-based Grasp-Shell, which provides direct physical manipulation of virtual content; and the multi-modal Gesture-Speech, which combines speech and gesture for indirect natural interaction. These techniques support object selection, 6 degrees of freedom movement, uniform scaling, as well as physics-based interaction such as pushing and flinging.

3   User Interfaces in 3D Virtual Environments

With the advent of Virtual Reality (VR), AR, ubiquitous and mobile computing and other "off-the-desktop" technologies, a new term has been introduced to cover the interaction in three-dimensional (3D) environments. 3D User Interface is a human–computer interaction in which the user's tasks are performed directly in a real or virtual 3D spatial context [1]. 3D interaction research field is still at its infancy, and its potentials are limited only by the imagination of researchers and developers.

3.1   Interaction in Augmented Reality

AR technologies create immersive experiences that embed interactive digital content which enhances the user's field of view. 3D interactions' major characteristic is the relevance to real-world tasks, thus users can rely upon their experience from their daily life movements to interact with virtual objects. However, users usually struggle to understand and perform actions in 3D spaces, as the physical world contains many more cues for understanding and constraints and affordances for action that cannot currently be represented accurately in a computer simulation [1]. Nevertheless, there are multiple categories of 3D UIs based on input methods and devices, and interaction techniques that may require multiple users' skills that may or may not be familiar with [25]:

- *Information Browsers* use the mobile device's camera to align and display virtual content in the real world. Users can rely on their knowledge of traditional mobile user interfaces to navigate through the physical environment.
- *3D Interaction* with virtual objects by using 3D spatial input devices such as 3D mouse, wand-type pointing devices and 3D joysticks. This method can be challenging due to the fact that users are familiar with manipulating physical objects with their hands and not through devices.
- In *Tangible User Interfaces* users interact with virtual objects by manipulating physical objects with similar characteristics, in order to bridge the physical and digital world.
- *Natural User Interfaces* no longer require users to manipulate input devices as they use body motion and gestures to interact with the 3D UI. By using natural skills, users are able to perceive how to perform the required actions in the 3D environment and anticipate the corresponding outcomes.
- *Multimodal User Interfaces* - Combining different modalities of input is considered to provide a richer and complete 3D interaction.

In recent years, new methods and approaches have been introduced for the natural interaction in a 3D environment, taking advantage of the emergence of low cost sensors and depth cameras, which can track the spatial movement and positioning of user's body and use this data for virtual object manipulations. Furthermore, the above technologies present many advantages regarding the 3D interaction, such as the absence of additional cumbersome devices like head-mounted displays or gloves that may annoy users, and the constant and easy swapping to the real world that facilitates a collaborative interface, supporting an immersive viewing mode.

3.2   Natural interaction

Based on [19], a natural interaction interface *allows users to interact in a way similar to real life, and enables them to learn, acquire and master shape modeling quickly with the least mental load and training*. NUIs are interfaces that enable users to interact with computers in the way they interact with the world. When people refer to NUIs they are often talking about interaction modes such as speech or touch. But if the focus is on combinations of input and output that are experienced as natural, the collection of natural user interfaces includes modes such as gesture and body language, proximity and location, eye gaze and expression, and biometrics on the input side, and the full spectrum of audio and visual output, smell, tactile and object location, and other experiences on the "output" side (leveraging the full range human senses).

Natural user interfaces aim to provide a seamless user experience where the technology is invisible. Experience and action are integrated in the natural world and typically involve a combination of multiple modalities such as voice recognition, gesture, touch, AR etc. In recent years, there has been a tremendous interest in introducing various methods for gesture and speech input into AR that could help overcome user interaction limitations in an AR environment.

Despite the increasing prevalence of AR interfaces, there is still a lack of interaction techniques that allow full utilization of the medium. Natural hand interaction has the potential to offer these affordances however, as yet, has not been

well explored. Freehand interaction has been explored to deliver natural, intuitive and effective interaction. For a natural user interface, traditional input devices such as keyboard and mouse are not appropriate. Previous works [9, 11] require fiducial markers or digital gloves to track hand gestures. Other works leverage Microsoft Kinect to detect hand poses and movements for freehand menu selection [12] and object manipulation [13]. However, these methods suffer from drawbacks. Instrumented gloves are encumbered and prone to induce fatigue. Fiducial markers and ambient sensors require delicate set-ups and calibrations. Image-based methods [14, 15] have been proposed to detect and recognize hand gestures using image processing technology, which are suitable for both closed and public environments [16]. As mobile devices become more powerful, these methods are promising for mobile devices as built-in cameras can be used without resorting to additional devices or sensors.

3.3   Gesture Recognition

Research in hand gesture recognition aims in design and development of such systems that can identify explicit human gestures as input and process these gesture representations for device control through mapping of commands as output. Creation and implementation of such efficient and accurate hand gesture recognition systems are aided through two major types of enabling technologies for human computer interaction namely contact based and vision based devices (shown in Figure 1).
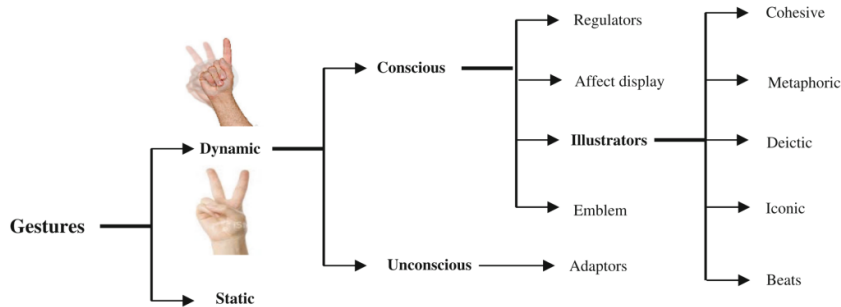

Fig. 1. Vision based hand gesture taxonomies [4]

The main challenge of vision-based hand gesture recognition is to cope with the large variety of gestures. Recognizing gestures involve handling a considerable number of degrees of freedom (DoF), huge variability of the 2D appearance depending on the camera view point (even for the same gesture), different silhouette scales (i.e. spatial resolution) and many resolutions for the temporal dimension (i.e. variability of the gesture speed). Moreover, it needs also to balance the accuracy-performance-usefulness trade-off according to the type of application, the cost of the solution and several criteria such as real-time performance, robustness, scalability and user-independence [18]. Finally, gesture recognition alongside other natural interaction methods such as speech recognition (multimodal interaction) improves the

efficiency and accuracy of the interactions, while also reduces the time learning and error rate [20, 21, 22].

3.4   Naturalism in 3D User Interfaces

Natural interfaces are built on users' existing knowledge and skills, thus the actions required are corresponding to real-world experiences. One of the key challenges that designers face is the level of naturalism that characterizes the interaction methods. Hyper natural design approach offers realistic interactions and enhanced abilities that avoid some unwanted constraints of the real world, while natural interactions replicate the real-world experience exactly [23]. "Magic" techniques are intentionally less natural in order to give users abilities that cannot have in the real world, thus making tasks in the 3D environment easier and less cumbersome. However, these interactions are not corresponding to user's real world experience, and thus requiring users to be trained in order to perform efficiently their interaction methods. On the other side, natural interactions don't require any training phase as users manipulate 3D objects exactly the way they would use the real ones, but this method could result in the same if not more effort from users as the equivalent movements in real world.

Designers have to adjust the design of interaction methods based on the balance between fidelity and usability / performance. High levels of naturalism can be achieved if users are familiar with the actions required to interact, but it is important to highlight that some tasks may have not real-world counterpart that can be exploited to efficiently design a natural UI. Additionally, performance and usability can be achieved by providing users with enhanced abilities that can be still familiar to users, even though they don't replicate the real world. For example, the HOMER (Hand-Centered Object Manipulation Extending Raycasting) technique combines the ray-based selection with hand-centered object manipulation in a hybrid interaction method that allow users to manipulate virtual objects that are not in reach of their hands, but they still use their natural skills for object manipulation [1].

4   Proposed architecture

For the implementation of our work, we used the Leap Motion Controller (LMC) alongside the Unity3D platform. Unity is a popular cross-platform game engine that can create 2D and 3D virtual environments. LMC is a camera sensor developed by Leap Motion [24] that senses natural hand movement and finger positioning, allowing the interaction with virtual objects through hand gestures like pinch, grab, swipe, rotate. It can translate hand movements into computer commands, thus enabling users to interact with a Virtual Reality environment displayed through Head-Mounted displays like the Oculus Rift, or in desktop mode (Unity environment).

In our prototype, we design two different gestures (high and low level of naturalism each) for these three basic interaction categories (Figure 2):

- *Translation*: When users move a physical object, they usually grab it with their fingers (one or two hands depending on its size), move it in midair and place it in the desired target. This movement has been translated to the Unity – Leap

Motion application by applying the interaction "grasp event", which tracks user's fingers and when they come close enough to the "collider box", the virtual object is "grabbed" (it is released when user's fingers move away). The second movement that is tested is not so natural, but users are familiar with it as they use it to pick and move objects in 2D surfaces (traditional UIs). In our Unity – Leap Motion implementation, users point with their index finger the virtual object they want to move, and then in the same way point to the target position. Both movements are one-handed.

- *Scaling*: The unique characteristic of this movement is that there is no natural equivalent. Users can scale objects only in the 3D environment. However, there is a quite common gesture about scaling that users are familiar with, especially in devices with touch screen, the pinch-to-zoom gesture. In our framework, the Unity – Leap motion software tracks user's both hands, and when user grasps a virtual object at an angle and moves away or closes his hands to each other, the grabbed object is scaled up or down respectively. In the second movement that is not so "natural", the Leap Motion controller tracks users' palm based on its relative positioning, and if the palm moves towards the controller, the virtual environment is "zooming in" the desired direction / virtual object (zooming out when the palm moves backwards).
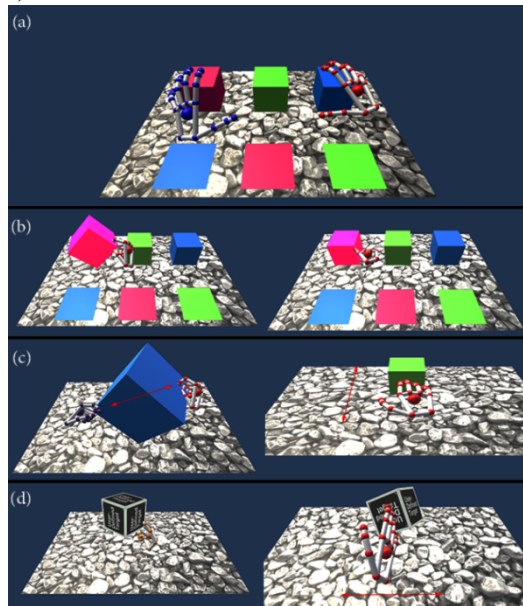
Fig.2. Gestures screenshots

- *Rotation*: For the rotation of a virtual object, we take a similar approach to the translating movement. With the grasp event, user can grab and manipulate the virtual object with his fingers, turning it around and rotating it based on any axis he desires. For the second movement, the Unity – Leap Motion application tracks user's palm and depending on its angle it rotates the virtual object (which is in fixed position) to the same direction on a single axis. The rotation stops when user's palm changes to fist gesture. Both movements are also one-handed.

4.1 Experimental Setup

In this section, we test the usability and effectiveness of our prototype and evaluate the feedback received from the testing users. Furthermore, we analyze how the level of interaction fidelity impacts user performance, and we argue about the balance between naturalness and effectiveness in 3D interactions techniques.

Participants were invited to complete a series of tasks regarding the translation, scaling and rotating of virtual boxes with the Leap Motion controller. Our prototype was set up using Unity's desktop mode alongside the Vuforia Software Development Kit (SDK) for the AR framework. 10 users were recruited from outside of the university, with no previous experience in 3D interaction but all of them were familiar with touch screen devices. They also were required to use their right hands for the one – handed gestures (they were all right-handed). Participants were asked to perform the three basic interaction movements described in the previous section, (a) move three virtual cubes in the desired positions using the grasp interaction and then the "point and select" interaction method, (b) scale a virtual box using both hands to drag its corners away and then scale the same object using their palm to zoom in and out, and (c) rotate a virtual box grabbing and moving it around with one hand, and then rotate it by rotating their palm at the desired direction. The order of the tasks was the same for all participants who also were explained what and how to perform these tasks during the experiment. We didn't measure user's performance such as time completion or accuracy, but after the experiment we asked them to answer a few questions about usability and user satisfaction, while also they asked to rate and compare the two different approaches of each interaction category (Figure 3).

Fig.3. Part of evaluation questionnaire

4.2 Evaluation and Results

In this section, we evaluate the feedback by the participants and their answers to our questionnaire, and we present the results of our experiment.

Regarding the background of our participants, they had no previous experience with 3D interaction and Leap Motion interfaces, and so it was challenging for them to perceive and perform effectively in the tasks we gave them, but after a few failed attempts and with our guidance they were able to complete the experiment. One of the major challenges was the correct alignment of their hands regarding the Leap Motion controller and the virtual objects, with the tracking occlusion issue between the virtual objects and user's hands to be also a difficulty. Furthermore, previous experience with touch devices helped users to perform well in gestures that were already familiar with, such as the "point and select" or the scale gestures, and therefore they preferred to use these gestures rather than the more natural ones. Thus, UIs that do not achieve high levels of fidelity may actually improve usability, if their design approach is familiar to users.

Participants struggled more performing the gestures with high natural approaches, with the accuracy of their movements to be more challenging. However, many of them admitted that even if these gestures were more enjoyable to perform, they would still choose the less natural approaches for their future interactions, which were also more easy to use and learn. Also, a few participants felt frustrated with the complexity of the high natural gestures (especially the translation movement). Additionally, feedback sounds and graphics helped users to understand how to perform the gestures.

Our prototype was designed to play specific sounds and display color changes when a collision of hands with virtual objects was detected, and these functions helped users to align their hands in the 3D environment and understand how to complete the tasks required.

Finally, the majority of the participants commented that ease of use, completion time and accuracy were more important factors than fun and enjoyment, for their decision in their preferred gestures. They also argue about the importance of the guidelines provided during the experiment to perform the tasks needed.

5. Discussion - Future Scope

In this paper, we analyzed our approach in gesture-based natural interaction for 3D User Interfaces, and presented our prototype alongside an evaluation test of basic gestures. We asked participants to perform different types of gestures for three basic interaction categories (translation, rotation and scaling of a virtual object), and then we analyzed their feedback and answers to our questionnaire. Our experiment focused on clarifying the desired level of naturalism in the proposed gestures, and its impact to user's performance.

For our future work, we will implement an AR natural (gesture and voice) interface with the Leap Motion controller integrated in a head-mounted display (for example a Google Cardboard). Furthermore, we aim in designing gestures with different levels of fidelity for more interaction categories that will also include menu management, manipulation of many virtual objects simultaneously and voice recognition interaction. Finally, we need to evaluate our system rating user's performance, such as accuracy, time completion and error rate, in order to analyze in depth user's feedback.

References

1. Bowman, D., Kruijff, E., LaViola Jr, J. J., & Poupyrev, I. P.: 3D User interfaces: theory and practice. In: CourseSmart eTextbook. Addison-Wesley (2004)
2. Wigdor, D., & Wixon, D.: Brave NUI world: designing natural user interfaces for touch and gesture. In: Elsevier (2011)
3. Steinberg, G.: Natural user interfaces. In: ACM SIGCHI conference on human factors in computing systems (2012)
4. Rautaray, S. S., & Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. In: Artificial Intelligence Review, 43(1), 1-54, (2015)
5. Hassanpour, R., & Shahbahrami, A.: Human computer interaction using vision-based hand Gesture recognition. In: Journal of Computer Engineering, 1, 3-11 (2009)
6. Regenbrecht, H., Collins, J., & Hoermann, S.: A leap-supported, hybrid AR interface approach. In: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (pp. 281-284). ACM (2013)
7. Khademi, M., Mousavi Hondori, H., McKenzie, A., Dodakian, L., Lopes, C. V., & Cramer, S. C.: Free-hand interaction with leap motion controller for stroke rehabilitation. In: CHI'14 Extended Abstracts on Human Factors in Computing Systems (pp. 1663-1668). ACM (2014)
8. Caggianese, G., Neroni, P., & Gallo, L.: Natural interaction and wearable augmented reality for the enjoyment of the cultural heritage in outdoor conditions. In: International Conference on Augmented and Virtual Reality (pp. 267-282). Springer, Cham (2015)
9. Hürst, W., & Van Wezel, C.: Gesture-based interaction via finger tracking for mobile augmented reality. In: Multimedia Tools and Applications, 62(1), 233-258 (2013)
10. Piumsomboon, T., Altimira, D., Kim, H., Clark, A., Lee, G., & Billinghurst, M.: Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR), (pp. 73-82). IEEE (2014)
11. Billinghurst, M., Clark, A., & Lee, G: A survey of augmented reality. In: Foundations and Trends® in Human–Computer Interaction, 8(2-3), 73-272 (2015)
12. Cui, J., Kuijper, A., & Sourin, A.: Exploration of natural free-hand interaction for shape modeling using leap motion controller. In: 2016 International Conference on Cyberworlds (CW) (pp. 41-48). IEEE (2016)
13. Ni, T., Bowman, D., & North, C.: AirStroke: bringing unistroke text entry to freehand gesture interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2473-2476). ACM (2011)
14. Guimbretière, F., & Nguyen, C.: Bimanual marking menu for near surface interactions. In: Proceedings of the SIGCHI conference on human factors in computing systems (pp. 825-828). ACM (2012)
15. Song, P., Goh, W. B., Hutama, W., Fu, C. W., & Liu, X.: A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1297-1306). ACM (2012)

16. Wilson, A. D.: Robust computer vision-based detection of pinching for one and two-handed gesture input. In: Proceedings of the 19th annual ACM symposium on User interface software and technology (pp. 255-258). ACM (2006)

17. Benko, H.: Beyond flat surface computing: challenges of depth-aware and curved interfaces. In: Proceedings of the 17th ACM international conference on Multimedia (pp. 935-944). ACM (2009)

18. Huang, Z., Li, W., Hui, P.: Ubii: Physical World Interaction Through Augmented Reality. In: Proceedings of the 23rd ACM international conference on Multimedia, 341-350 (2015)

19. Lee, M., Billinghurst, M., Baek, W., Green, R., & Woo, W.: A usability study of multimodal input in an augmented reality environment. In: Virtual Reality, 17(4), 293-305 (2013)

20. Lv, Z., Halawani, A., Feng, S., Li, H., & Réhman, S. U.: Multimodal hand and foot gesture interaction for handheld devices. In: ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 10 (2014)

21. Heidemann, G., Bax, I., & Bekel, H.: Multimodal interaction in an augmented reality scenario. In: Proceedings of the 6th international conference on Multimodal interfaces (pp. 53-60). ACM (2004)

22. Bowman, D. A., McMahan, R. P., & Ragan, E. D.: Questioning naturalism in 3D user interfaces. In: Communications of the ACM, 55(9), 78-88 (2012)

23. Leap Motion, Homepage URL: https://www.leapmotion.com/