Design and Evaluation of Freehand Gesture Interaction for Light Field Display

Vamsi Kiran Adhikarla[1,2], Grega Jakus[3(✉)], and Jaka Sodnik[3]

[1] Holografika, Budapest, Hungary
[2] Pazmany Peter Catholic University, Budapest, Hungary
[3] University of Ljubljana, Ljubljana, Slovenia
{grega.jakus,jaka.sodnik}@fe.uni-lj.si

Abstract. The paper reports on a user study of freehand gesture interaction with a prototype of autostereoscopic 3D light field display. The interaction was based on a direct touch selection of simple objects rendered at different positions in space. The main goal of our experiment was to evaluate the overall user experience and perceived cognitive workload of such freehand interaction in 3D environment and compare it to the simplified touch-based interaction in 2D environment. The results of the experiment confirmed the hypothesis that significantly more time is required for the interaction in 3D than the interaction in 2D. Surprisingly, no significant difference was found in the results of the assessment of cognitive workload when comparing 3D and 2D. We believe the interaction scenario proposed and evaluated in this study could represent an efficient and intuitive future interaction technique for the selection and manipulation of content rendered on autostereoscopic 3D displays

Keywords: 3D display · Light field display · Free-hand interaction · Direct touch · Leap Motion Controller · Human-machine interface (HMI) · Human-computer interface (HCI)

1 Introduction

Projection-based light field displays [1] represent a novel technology in the field of 3D rendering, which provides a more sophisticated solution for visualizing 3D content than any other glasses-free 3D technology. The illusion of 3D perception is generally created by providing two slightly offset views of a scene, one for each eye. Figure 1 illustrates how different views are delivered in traditional stereoscopic 3D technology (S3D), multiview 3D and light field technologies. In S3D, the two views are captured by two cameras simultaneously and synchronized to the left and right eye of the viewer respectively (Fig. 1a). View isolation is achieved by using special eyewear.

In a glasses-free system, the process of view separation is generally achieved by producing light beams with directionally-dependent colour and intensity of emitted light from each pixel on the display, which enables a more realistic 3D visualization compared to S3D technology. As opposed to S3D, view separation is inherent to the hardware itself, which is why such displays are often called *autostereoscopic displays.*

The view separation in the existing autostereoscopic displays is often achieved by directing the light using lenticular lenses and parallax barriers. Such approach enables the projection of multiple views of a scene (Fig. 1b). However, due to its discrete nature, it requires the user to be located at certain predefined positions in order to perceive 3D comfortably. Any user entering the space between the views can see the light barrier, which can impair the overall 3D quality perception. The limited number of available views directly affects the angular resolution and the effective field of view (FOV) of the multiview autostereoscopic display.
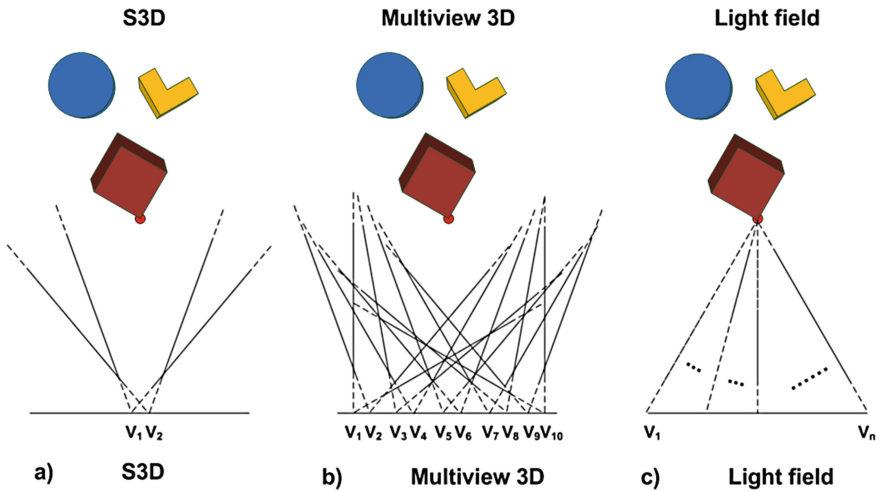
Fig. 1. Displaying a sample scene and a point in the scene (shown in red) in 3D using S3D, multiview 3D and light field technologies (Color figure online).

Additionally, S3D and multiview autostereoscopic displays do not take into account the positions of 3D scene points and display only a limited number of perspectives. These perspectives are actually 2D projections of the 3D image, which, when combined, define the scene. Light field displays, in contrast, treat each scene point differently resulting in a much more realistic and accurate 3D visualization. This is achieved by defining the scene by using a set of direction-dependent light rays emitted from various optical modules as if they were emitted from real scene points (Fig. 1c). A scene point is created at the point of crossing of two light rays emitted from two optical modules.

A light field display consists of a holographic screen, a group of optical modules and two mirrors along the sidewalls of the display (see Fig. 2). The screen is a flat hologram while the optical modules are arranged densely and equidistantly from the screen. The so-called *light field* is created when the light beams emitted from the optical modules hit the holographic screen and disperse in different directions. The separation of views created by the holographic screen provides a continuous-like horizontal motion parallax without blocking zones in the FOV of the display [2].
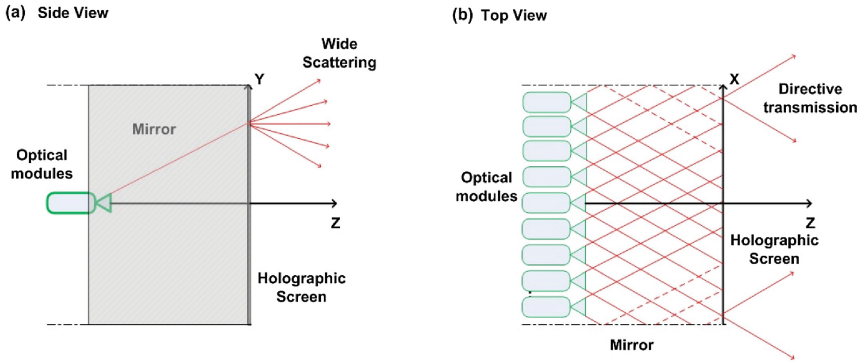
Fig. 2. Main components of a light field display: geometrically aligned multiple optical modules, a holographic screen and side mirrors.

The possibility of manipulating virtual objects in 3D environment through natural interfaces, such as freehand gestures, brings new opportunities for more useful 3D applications. As light field displays represent an innovative technology in the field, novel interaction technologies and techniques need to be designed and evaluated in order to enable such interaction. The main goal of the experiment described in this paper was to evaluate the performance, perceived user experience and cognitive workload while interacting with 3D content on a light field display. The interaction was based on the selection of objects by touching their virtual position. Leap Motion Controller (LMC) was used as the tracking device as it enables continuous tracking of hands with millimetre accuracy [3], thus allowing the interaction with individual 3D voxels of a light field display.

The remainder of this paper is organized as follows: the following section presents the related work on free-hand interaction with 3D content. The study design is described in Sect. 3, while the results of the study are presented and analysed in Sect. 4. The paper concludes with discussion and key conclusions.

2   Related Work

The input devices that enable free-hand interaction can be categorized into wearable and hands-free devices. As traditional wearable devices generally obstruct the use of hands while performing activities, it is more suitable to track hand movement visually, for example, by means of optical tracking systems. Such systems operate by tracking reflective markers attached to the user's hand and were used as the tracking device for the interaction with the 3D content in various experiments, including [4] and [5].

Optical tracking can also be applied for true hands-free tracking where users do not need to wear markers. However, as body surface reflects less light compared to highly-reflective markers, this generally results in a smaller interaction space. A number of studies with hands-free tracking for 3D interaction have been conducted using

various input devices and setups (e.g. [6–8]) including Microsoft Kinect sensor, one of the commercially most successful hands-free tracking devices in the market [7, 9–11].

Another important contribution to the affordable desktop hands-free motion tracking devices was the release of Leap Motion Controller [12] in 2013. The device uses three LED emitters to illuminate the surrounding space with infra-red (IR) light which is reflected back to the device from the nearby objects and captured by two IR cameras. The determined positions of recognized hands, fingers and other objects as well as detected gestures can be obtained through API (Application Programming Interface). The device and its coordinate system are shown in Fig. 3.
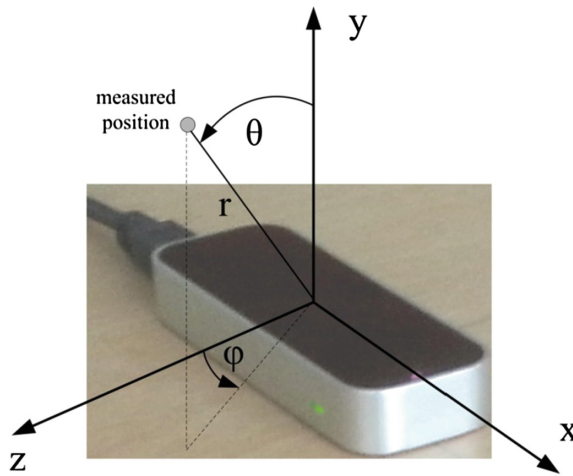
Fig. 3. Leap Motion Controller and the coordinate system used to describe positions of recognized objects.

A study on the LMC performance [3] revealed that the effective range of the device extends from approximately 3 to 30 cm above the device (y axis), approximately 0 to 20 cm behind the device (negative z axis) and 20 cm in each direction along the device (x axis). Standard deviation of the measured position of a static object was shown to be less than 0.5 mm.

A variety of studies on interaction involving LMC have been performed including studies with 3D gestures [13, 14] and pointing tasks [15, 16]. However, to our knowledge, we report on the first study involving LMC and a 3D display.

3   User Study

The main goal of the user study was to evaluate the performance, perceived user experience and cognitive workload while interacting with 3D content. The interaction consisted of selecting the tile-like objects using the "direct touch" method.

3.1   Study Design

The interaction with the light field display was evaluated through a comparison of 2D and 3D displaying modes representing two different experimental conditions:

- in 2D mode, the displayed tiles were distributed on a plane in close proximity of the display surface;
- in 3D mode, the displayed tiles were distributed in a space in front of the display (the distance between the tiles and the screen surface varied from 0 to 7 cm).

The 2D mode provided a control environment which was used to evaluate the specifics of this particular interaction design: the performance and properties of the input device, display dimensions, specific interaction scenario (e.g. touching the tiles), etc.

The study design was within-subject. Each participant was asked to perform 11 trials within each condition. In each trial, three tiles of the same size were displayed simultaneously and the participant was asked to touch the surface of the red tile as perceived in space (Fig. 4). The positions of the tiles varied from trial to trial to cover the FOV of the display.
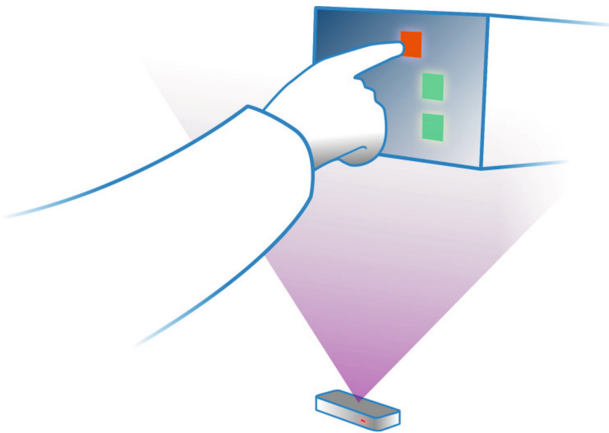
Fig. 4. The interaction with the content rendered on the light field display (Color figure online).

The main variables measured in the experiment were:

- task completion times;
- cognitive workload (measured through NASA TLX questionnaire);
- UEQ (User Experience Questionnaire) on the perceived user experience.

3.2   Test Subjects and Groups

A total of 12 test subjects (5 female and 7 male) participated in the study. The average age of the participants was 27 years and ranged from 20 to 36 years. All participants reported to have normal or corrected to normal sight.

To avoid learning effects, the participants were equally distributed among two experimental groups which differed in the order of the two conditions.

3.3   Technical Setup

A small-scale prototype light field display developed by Holografika was used for the study. The FOV of the display was comparable to the tracking volume of the LMC. The light field display was placed at the participants' eye-level at a distance of approximately 50 cm from the viewer. The gesture input was tracked by LMC placed on the table in front of the display.

Figure 5 shows the technical setup. The controlling PC hosts two applications. The frontend OpenGL application renders content for 2D LCD display and also communicates with LMC in real-time in order to receive and process user interaction commands. The second application (backend wrapper) generates a modified stream for light field rendering and tracks the commands in the current instance of OpenGL (frontend application).
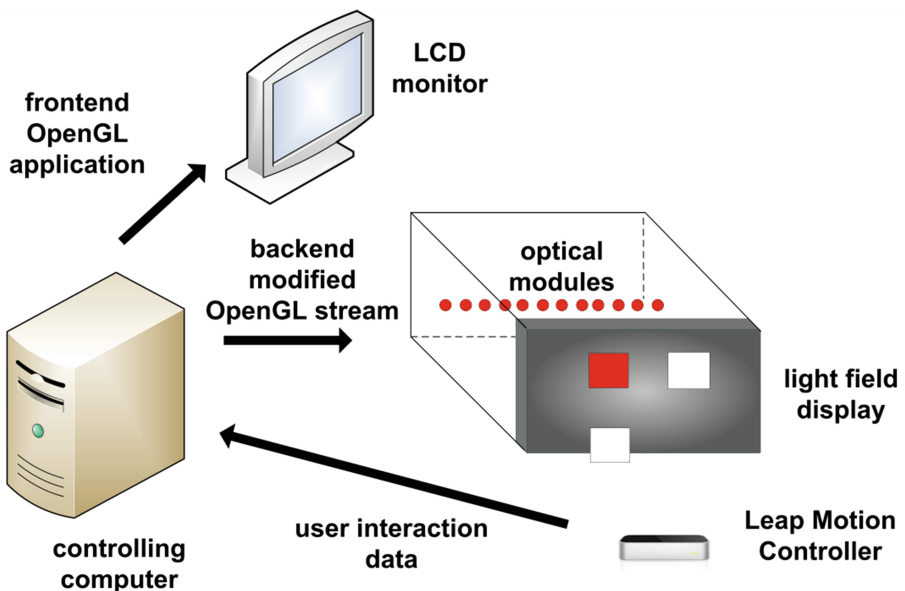

Fig. 5. Technical setup

3.4   Procedure

Prior to the experiment, the participants were informed about the nature and structure of the study. They were then asked to complete a pre-study questionnaire (age, gender, sight disabilities, and prior experience with LMC and autostereoscopic 3D displays). The participants were then thoroughly introduced to their tasks, the interaction scenario

and methods. Finally, they were given five minutes to familiarize themselves with their tasks and the interaction.

Participants were asked to perform the given tasks as quickly as possible. Task completion times and interaction activity were recorded automatically for each task. The measurements started automatically when the participants reached the distance of 15 cm from the display. After successfully completing the task, the participants had to remove their hand from this area and place it on their knee before proceeding with the next task. For the purpose of post-evaluation of the interaction performance, the entire user study was recorded with a digital video camera.

After completing each of the experimental conditions, the participants were asked to fill the NASA TLX [16] and UEQ questionnaires [17] to evaluate their perceived workload and user experience. After both conditions, the participants were asked to fill in a short post-study questionnaire on their overall perception of the interfaces, the design and realism of the display, and the complexity of the given tasks.

4   Results and Discussion

Figure 6 shows mean task completion times in both conditions. The mean object selection time in 3D condition was approximately half of a second longer than the selection time in 2D condition. The results of the T-test (t(22) = 2.521, p = 0.019) showed this time difference to be significant (p < 0.05). Such results of task completion times were expected since the additional dimension implies extra time to cognitively process the visual information and to physically locate the object.
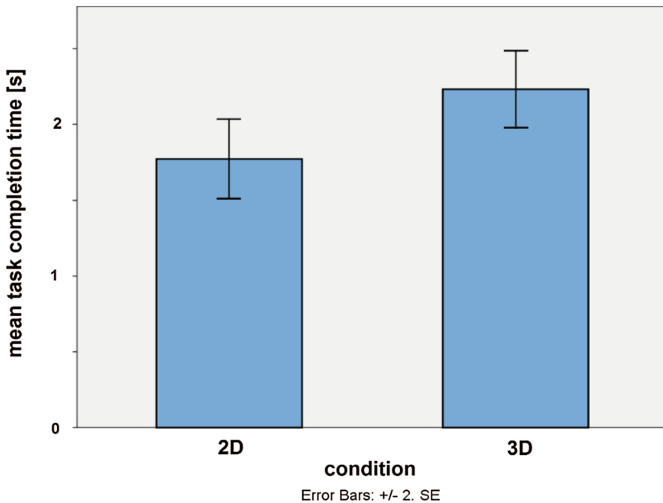
Fig. 6. Mean task completion times

When analysing the approach path of the finger towards the rendered object, some typical patterns of finger movement can be observed. Figure 7, for example, displays the fingertip positions measured along the z-axis (i.e. the depth; see Fig. 2 or Fig. 3 for coordinate system orientation). In 2D selection (Fig. 7a), the pattern is generally straightforward with direct approach towards the object (rendered just in front of the display in 2D) and then holding the finger still to trigger the selection. On the other hand, the trajectory of the finger in 3D typically includes a confident initial move towards the perceived position of the object, followed by a set of small corrections. Finally, the selection is triggered, often while still slowly moving the finger (Fig. 7b).

Figure 8 displays the results of the NASA TLX test. The T-test analysis (t(22) = −0.452, p = 0.655) revealed no significant difference in total cognitive workload between the conditions (p > 0.05) as well as no significant difference on any of the subscales.
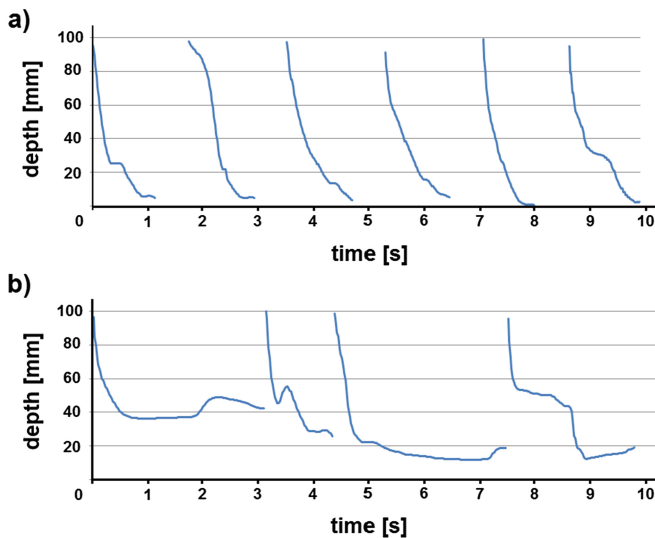
Fig. 7. Typical traces of fingertip positions in (a) 2D and (b) 3D conditions measured along z-axis (depth).

Similarly, the results of the UEQ (Fig. 9) also did not reveal any significant difference between both conditions (t(22) = −0.708, p = 0.486), except for the "novelty" subscale where a tendency towards higher preferences for the 3D mode can be observed.

When asked about their preference, two thirds of the participants chose the 3D mode as their favourite. The participants generally found the objects rendered on the light field display more realistic and reported that all the objects were seen clearly. The tasks were also found realistic and the participants reported they could be solved in a very straight-forward manner. The input device was also reported to be easy to use in the combination with the 3D display.
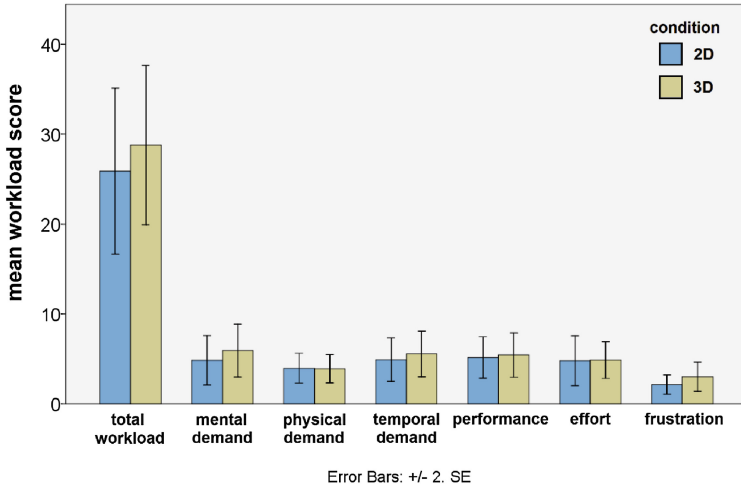
Error Bars: +/- 2. SE

Fig. 8. Mean workload scores
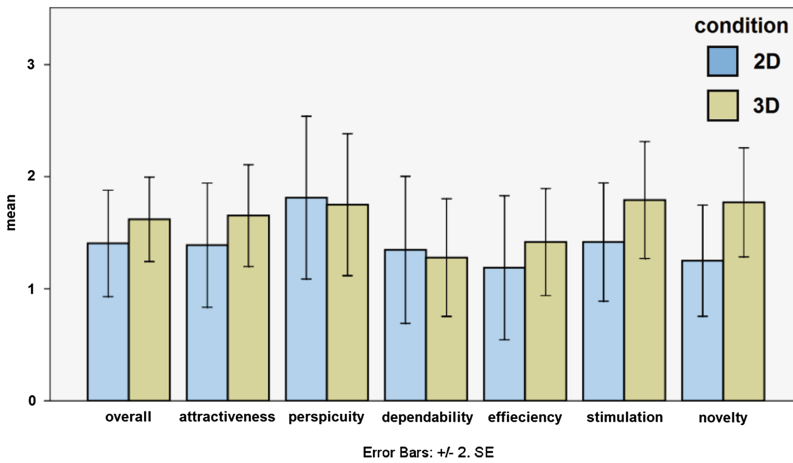

Error Bars: +/- 2. SE

Fig. 9. Mean user experience scores

5 Discussion and Conclusions

The statistical analysis of the selection times revealed that less time is needed to select the object in 2D than in 3D condition. Such result was, in fact, expected, as the interaction in an additional dimension undoubtedly requires extra time needed to cognitively process the visual information and then to physically locate the object in space.

When analysing the time course of the selection gesture, some interesting observations are also worth mentioning. The depth component of the gesture (the distance

from the display surface) can, in both conditions, be generally divided into two sections: the initial direct and relatively confident approach towards the target and the final ("triggering") part of the gesture. While the pattern of the first section is similar in both conditions, this is generally not true for the "triggering" part of the gesture. In 2D condition, the targeted tile was always displayed directly in front of the display with a permanent depth component of its location. Therefore, after identifying the longitudinal and transverse components of the tile location, all the users were required to do was to simply hold their finger at constant distance for a certain amount of time in order to trigger the tile selection. On the other hand, in 3D condition the users were not entirely certain of the exact displaying depth of the tile as it varied from trial to trial. Therefore, when they reached the approximate depth at which they perceived the tile, the participants began locating it more precisely by making small and slow corrections in order to trigger the selection. Longer task completion times in 3D condition can, therefore, be contributed precisely to such triggering attempts, which were based more or less on trial and error.

The somewhat longer and uncertain last stage of the selection gesture can be attributed to the lack of a tactile feedback due to the intangibility of the tiles that are being displayed and selected. The problem of object intangibility in 3D interaction is not insignificant and has already been addressed in similar studies (see for example [8]). A logical solution (and also the probable subject of our future experiments in the field) would be to provide the user with some sort of feedback when the selection gesture reaches close proximity of the targeted object.

In the present study, we have addressed the problem of human-computer interaction in 3D environment and also the transition between 2D and 3D modalities. As the old-fashioned but very effective computer mouse has been successfully replaced by various direct-input mechanisms in touch-screen environment, a similar evolutionary step will be required when upgrading displaying technologies with an additional dimension. We believe the freehand interaction setup proposed in this study could represent the next logical step in this transition as it enables very intuitive selection and manipulation of content in 3D environment. Simple and affordable tracking devices such as LMC will play a key role in the future human-machine interaction, either as primary input devices or as a part of a complex multi-modal interface. The results of UEQ questionnaires in our study revealed a high preference of users for freehand interaction setups resulting in a fast and straightforward adoption process of such technologies. Additionally, the NASA TLX test revealed a relatively low cognitive demand comparable to a simplified 2D environment. We believe this reflects the efficiency and the intuitiveness of the proposed interaction setup and the freehand interaction with 3D content in general.

In the future, we are planning on extending the freehand interaction paradigm in 3D environment by including both hands and more complex gestures. The "selecting" scenario should additionally be upgraded to free manipulation of objects in 3D space by changing their positions and orientations. Additionally, some form of audio or tactile feedback should also be considered offering a more realistic perception of 3D space and an even better user experience.

References

1. Agocs, T., Balogh, T., Forgacs, T., Bettio, F., Gobbetti, E., Zanetti, G., Bouvier, E.: A large scale interactive holographic display. In: IEEE Virtual Reality Conference (VR 2006), p. 57. Washington, DC, USA (2006)
2. Balogh, T., Kovács, P.T., Megyesi, Z., Barsi, A.: HoloVizio – true 3D display system. In: 3DTV Conference, pp. 1–4 (2007)
3. Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., Sodnik, J.: An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors **14**, 3702–3720 (2014)
4. Bruder, G., Steinicke, F., Sturzlinger, W.: To touch or not to touch?: Comparing 2D touch and 3D mid-air interaction on stereoscopic tabletop surfaces. In: Proceedings of the 1st Symposium on Spatial User Interaction, pp. 9–16. ACM (2013)
5. Grossman, T., Wigdor, D., Balakrishnan, R.: Multi-finger gestural interaction with 3D volumetric displays. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, pp. 61 – 70. ACM, New York (2004)
6. Wang, R., Paris, S., Popović, J.: 6D hands: markerless hand-tracking for computer aided design. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp. 549–558. ACM (2011)
7. Butler, A., Hilliges, O., Izadi, S., Hodges, S., Molyneaux, D., Kim, D., Kong, D.: Vermeer: direct interaction with a 360 viewable 3D display. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp. 569–576. ACM (2011)
8. Chan, L.W., Kao, H.S., Chen, M.Y., Lee, M.S., Hsu, J., Hung, Y.P.: Touching the void: direct-touch interaction for intangible displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2625–2634. ACM (2010)
9. Vogel, D., Balakrishnan, R.: Distant free hand pointing and clicking on very large, high resolution displays. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, UIST 2005, pp. 33–42. ACM, New York (2005)
10. Hilliges, O., Kim, D., Izadi, S., Weiss, M., Wilson, A.: HoloDesk: direct 3D interactions with a situated see-through display. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2421–2430. ACM (2012)
11. Ren, Z., Yuan, J., Meng, J., Zhang, Z.: Robust part-based hand gesture recognition using kinect sensor. IEEE Trans. Multimedia **15**, 1110–1120 (2013)
12. Leap Motion Controller. https://www.leapmotion.com
13. Fanini, B.: A 3D interface to explore and manipulate multi-scale virtual scenes using the leap motion controller. In: Proceedings of ACHI 2014, the Seventh International Conference on Advances in Computer-Human Interactions, pp. 258–263 (2014)
14. Apostolellis, P., Bortz, B., Peng, M., Polys, N., Hoegh, A.: Poster: exploring the integrality and separability of the leap motion controller for direct manipulation 3D interaction. In: IEEE Symposium on 3D User Interfaces (3DUI), pp. 153–154. IEEE (2014)
15. Coelho, J. C., Verbeek, F. J.: Pointing task evaluation of leap motion controller in 3d virtual environment. In: Creating the Difference: Proceedings of the Chisparks 2014 Conference, pp. 78–85. The Hague, The Netherlands (2014)

16. Hart, S.G., Staveland, L.E.: Development of a multi-dimensional workload rating scale: results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. Elsevier, Amsterdam (1988)
17. Laugwitz, B., Held, T., Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. Springer, Heidelberg (2008)