

Collaborative 3D Manipulation using Mobile Phones

Jerônimo G. Grandi^{*1} Iago Berndt^{†1} Henrique G. Debarba^{‡2} Luciana Nedel^{§1} Anderson Maciel^{¶1}

¹Institute of Informatics, Universidade Federal do Rio Grande do Sul, Brazil

²Immersive Interaction Group, École Polytechnique Fédérale de Lausanne, Switzerland

ABSTRACT

We present a 3D user interface for collaborative manipulation of three-dimensional objects in virtual environments. It maps inertial sensors, touch screen and physical buttons of a mobile phone into well-known gestures to alter the position, rotation and scale of virtual objects. As these transformations require the control of multiple degrees of freedom (DOF), collaboration is proposed as a solution to coordinate the modification of each and all the available DOFs. Users are free to define their manipulation roles. All virtual elements are displayed in a single shared screen, which is handy to aggregate multiple users in the same physical space.

1 INTRODUCTION

Manipulating a three-dimensional object in a virtual environment is a complex task, as it requires the control of multiple degrees of freedom (DOFs) to translate, rotate and scale the object. One alternative toward efficiency is to share the task and solve it in a collaborative manner. Collaborating users may wish to split the different manipulation aspects among them. For instance, while one user translates an object, a second user could rotate this same object.

However, the choice of input device in such collaborative scenario is not straightforward. With the recent surge of virtual reality (VR) consumer products – such as head mounted displays and stereo televisions – displays seems to be converging to some few standards. On the other hand, input devices are still much more diverse. Conventional input devices (mouse and keyboard) are rather limited in DOFs one can easily and simultaneously control. Non-conventional input devices, in turn, tend to favor a natural movement mapping metaphor. However, these are usually expensive and time consuming to setup (e.g. require the fixation of multiple sensors or the use of a dedicated tracking system) or not precise (e.g. markerless systems such as the MS Kinect or the Leap Motion).

To overcome the limitations of previous techniques, we propose the use of mobile phones, which are ubiquitous to the contemporaneous life, for 3D manipulation. Current smartphones include inertial sensors and a touchscreen, from which one can retrieve the device orientation and multi-touch gestures.

Our proposal relies on the mobile phone orientation, touchscreen, and physical volume buttons to perform the three main transformations: rotation, scale and translation. The actions required to perform the transformations are already consolidated in the everyday tasks when interacting with mobile apps and games, such as: touch and slide to translate, device rotation to rotate, and pinch and spread to scale the virtual object.

Additionally, our approach accepts the simultaneous connection of multiple mobile phones, allowing users to collaborate by simul-

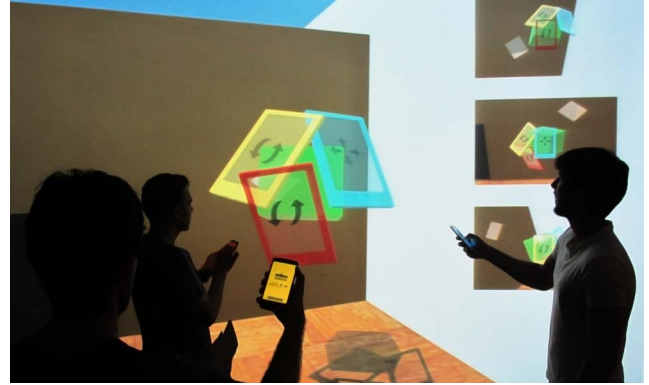


Figure 1: Three users simultaneously driving the object through the virtual environment. The colored rectangles indicate the position and orientation of each user in the VE. The three windows at the right-side show the three personal views.

taneously controlling any manipulation DOFs. This provides equal participation, individual responsibility and positive interdependence, which are mandatory features in a collaborative task.

2 RELATED WORK

The use of mobile phones as input devices for an external computing environment has already been studied to some extent. Notably, Boring et al. used the built in camera of a mobile phone to directly interact with a distributed computing environment, performing tasks such as transferring photos across different displays [1]. Moreover, Katzakis et al. also explored the usage of a mobile phone for 6DOF manipulation [4] with their plane-casting metaphor. Nonetheless, our group has also adopted mobile devices to interact with virtual environments, such as for the direct control of a docking object orientation [2], and to address the problem of 3D selection by coupling orientation and tactile sensors toward increased selection precision [3]. The later also included simple scale and translation of objects in a 2D scene. However, none of the previous works approached the problem of collaborative manipulation.

3 MATERIALS AND METHODS

3.1 Infrastructure

Our setup is composed of Android-based mobile devices with inertial sensors, touchscreen and WiFi connection, a server computer and a screen, preferentially large. An app on the phone communicates with the server that manages the clients data and the virtual environment. The screen is shared by the users.

The server application runs on a PC and was developed on Unity 3D. It allows the visualization of the VE, manages the connections with the mobile phones, and the manipulation transformations to be performed. Communication between server and phones uses TCP network protocol to ensure that the packets will always arrive in the correct order. This prevents motion flicker during transformations.

*e-mail: jggrandi@inf.ufrgs.br

†e-mail: iago.berndt@inf.ufrgs.br

‡e-mail: henrique.galvandebarba@epfl.ch

§e-mail: nedel@inf.ufrgs.br

¶e-mail: amaciel@inf.ufrgs.br

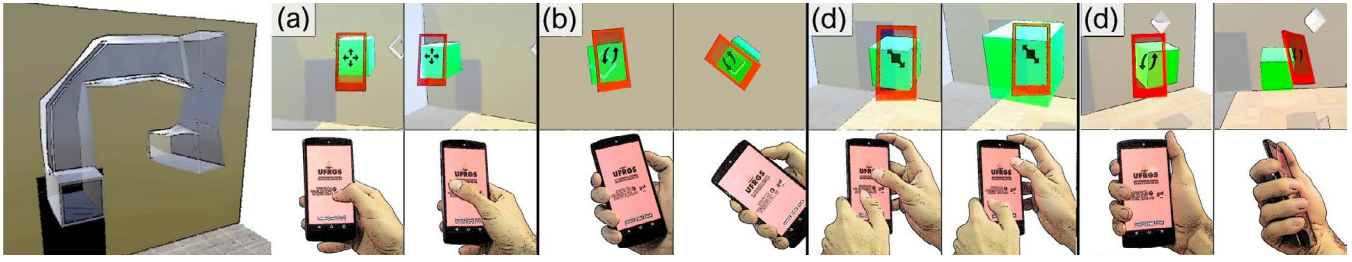


Figure 2: Overview of the circuit through which the selected object has to be taken and walkthrough of manipulations that a mobile phone cursor can perform over the selected object: (a) translation can be applied by touching the mobile phone screen and sliding the point of contact, the object translates on the plane defined by the mobile device orientation; (b) by holding the volume down button and rotating the phone one can rotate the object likewise; (c) scale is applied by touching the screen with two fingers and producing a pinch/spread gesture; (d) the camera orientation can be controlled by holding the volume up button and rotating the mobile phone.

3.2 Technique

The technique we propose for interaction in the 3D virtual environment (VE) uses the mobile phone physical orientation, the touchscreen, and the physical volume buttons to manipulate a total of 7 DOF of a selected 3D object. More specifically, there are 3-DOF for translation, 3-DOF for rotation and 1-DOF for uniform scale.

To translate the selected object, a plane is defined using the device orientation (1:1 mapping). This plane is aligned to the touchscreen and, as the thumb (or any other finger) starts to touch and slide on the screen, a similar translation is applied to the object (Fig. 2a). The selected object is rotated by pressing and holding the volume down button. Throughout the time the button is pressed, the virtual object will rotate following the orientation the phone (Fig. 2b). It is possible to use the clutch metaphor to reach a total rotation beyond wrist limits (perform object rotation, reposition the mobile phone, then perform a new object rotation). Uniform scale is performed by pinch and spread gestures with two fingers on the touchscreen (Fig. 2c). Finally, the main camera (shared among users) follows the selected object at a fixed distance. If a user wishes to change the camera orientation, this can be done in a similar manner as the virtual object orientation. Using the volume up button instead (Fig. 2d).

Since all the users actions are transmitted over WiFi and are applied directly to the manipulated virtual object, the object's transformations might be discontinuous. To minimize these undesired jumps, all movements are smoothed with a low-pass filter.

3.3 Implementation

In the implemented case study, the interactive task consists in manipulating a cube, which in turn has to be taken through a circuit of obstacles while avoiding collisions (Fig. 2left). The user interface includes a global camera, and the data respective to each mobile phone connected. This includes current orientation the activation of three basic transformations: rotation, translation and scale. The cube used in this task is referred on this paper as *selected object*, as our interface works in the same way for any 3D object.

The server application display and main camera are shared among all collaborating users, and all mobile phones are calibrated with respect to the screen position. The representation of each mobile phone in the shared display includes two elements: (i) a smartphone 3D model that orbits around the selected object following the phone orientation (Fig. 2a-d); (ii) a picture-in-picture (PIP) camera containing the point of view of the orbiting rectangle, positioned at the top right corner of the display (Fig. 1). The PIP camera is meant to provide additional depth cues, which are rather limited in the main camera as it is shared among multiple users.

We also rendered icons on the virtual phone representation indicating the task being performed by each user, so that collaborating users can be aware of each other current activity without the need for verbal communication. The icons can indicate that a mobile phone

is either performing a translation, rotation or scale on the selected object (Fig. 2a-c).

All actions are stored in a matrix representation and are sent through the network to the server. Every action performed by each individual user counts as a transformation step. Action matrices are multiplied by the transformation matrix of the virtual object. Therefore, every contribution from each user is summed up into the final object's transformation without restrictions or weights.

4 DISCUSSION AND FUTURE WORK

We presented a smartphone-based collaborative user interface for 3D object manipulation in virtual environments. Our solution simplifies complex 3D tasks by mapping gestures into transformation actions. Furthermore, all transformations can be performed by an unlimited number of participants at the same time. Smartphones are omnipresent and people are familiar with touch and orientation gestures. We assume that this will provide users of our technique with a very high affordance.

Regarding visualization, we chose to use a large screen as a shared visualization platform. It is a handy solution to aggregate multiple users in a single VE. Any person passing by the display is able to connect to the application and interact with it. This setup corroborates with the objectives of interactive public displays.

Depending on the application, our technique can be easily adapted to HMDs, CAVEs and other immersive displays. Individual displays allowing parallax and binocular stereo have the potential to increase depth perception and user performance.

Considering collaboration, our hypothesis is that allowing users to perform transformations freely will permit them to naturally understand each other participation and negotiate their role in the task. User experiments should be designed to test the hypothesis and measure how collaboration helps to perform better as a team than by oneself.

REFERENCES

- [1] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch projector: mobile interaction through video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2287–2296. ACM, 2010.
- [2] H. Debarba, J. Franz, V. Reus, A. Maciel, and L. Nedel. The cube of doom: A bimanual perceptual user experience. In *2011 IEEE Symposium on 3D User Interfaces*, pages 131–132. IEEE, 2011.
- [3] H. Debarba, L. Nedel, and A. Maciel. Lop-cursor: Fast and precise interaction with tiled displays using one hand and levels of precision. In *2012 IEEE Symposium on 3D User Interfaces*, pages 125–132. IEEE, 2012.
- [4] N. Katzakis, R. J. Teather, K. Kiyokawa, and H. Takemura. Inspect: Extending plane-casting for 6-dof control. *Human-centric Computing and Information Sciences*, 5(22), 2015.