Mid-Air Modeling with Boolean Operations in VR

Daniel Mendes*, Daniel Medeiros
Maurício Sousa, Ricardo Ferreira
INESC-ID / IST, Universidade de Lisboa, Portugal

Alberto Raposo
Tecgraf / PUC-Rio, Brazil

Alfredo Ferreira, Joaquim Jorge
INESC-ID / IST, Universidade de Lisboa, Portugal

Abstract

Virtual Reality (VR) is again in the spotlight. However, interactions and modeling operations are still major hurdles to its complete success. To make VR Interaction viable, many have proposed mid-air techniques because of their naturalness and resemblance to physical world operations. Still, natural mid-air metaphors for Constructive Solid Geometry (CSG) are still elusive. This is unfortunate, because CSG is a powerful enabler for more complex modeling tasks, allowing to create complex objects from simple ones via Boolean operations. Moreover, Head-Mounted Displays occlude the real self, and make it difficult for users to be aware of their relationship to the virtual environment. In this paper we propose two new techniques to achieve Boolean operations between two objects in VR. One is based on direct-manipulation via gestures while the other uses menus. We conducted a preliminary evaluation of these techniques. Due to tracking limitations, results allowed no significant conclusions to be drawn. To account for self-representation, we compared full-body avatar against an iconic cursor depiction of users' hands. In this matter, the simplified hands-only representation improved efficiency in CSG modelling tasks.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles, Graphical User Interfaces

1 Introduction

We live in an age where there is ever more purely digital content. 3D virtual models are no exception. This kind of content is present in several fields: virtual mock-ups in architecture, human models in medicine, virtual worlds in cinema and video-games, and so forth. Traditionally, the creation of 3D models is made resorting to 2D input and output devices, but it can be improved using 3D input methods and additional depth cues such as stereoscopy.

Recent technological advances in hardware originated a regained interest in Virtual Reality, with the appearance of off-the-shelf solutions such as the Oculus Rift and the HTC Vive, among others. With improved displays and sensors, head-mounted displays (HMD) are now more accurate, comfortable and affordable, and can greatly enhance the visualization of 3D virtual content. Following other advances in tracking solutions, mid-air interactions in Virtual Reality have been proposed [4]. These allow users to operate with the same degrees of freedom as they would in the physical world, which has the power to accelerate 3D manipulation and modelling tasks.

To create 3D virtual models of objects and environments, several approaches can be followed: sketching, instantiation of primitives, manipulation of meshes' vertexes, or by combining existing objects. In order to combine objects, Constructive Solid Geometry is a powerful tool that uses Boolean operations to create more complex ones. Although Boolean operations are common in many commercial applications, natural metaphors to use them in mid-air are still elusive.

*e-mail: danielmendes@ist.utl.pt

2017 IEEE Symposium on 3D User Interfaces (3DUI)
March 18-19, 2017, Los Angeles, CA, USA
978-1-5090-6715-2/17/$31.00

Additionally, HMDs occlude the real self and as such users' representation is required to relate their actions to the virtual environment. Different representations have been followed, with full-body avatars increasing the sense of embodiment [10]. However, it is not yet related to efficiency on modelling scenarios.

In this paper, we address challenges of both natural metaphors to perform Boolean operations between two objects in mid-air and user representation in Virtual Reality settings. We propose two new techniques to perform Boolean operations: the first is follows a direct manipulation approach; the second implements a menu interface that provides immediate feedback. We also assess how the realism of users self-representation affects user performance in 3D modelling operations.

2 Related Work

There is considerable amount of previous research in both 3d modelling and user representation in immersive virtual environments. We will cover the most relevant works, and discuss how ours improves upon current state-of-the-art.

2.1 3D Modelling

Currently, there is a panoply of applications to create 3D virtual objects and environments. Commercial solutions, such as Blender, SketchUp, and so forth, offer an immensity of tools to create and edit 3D models, including CSG operations, resorting to mouse and traditional displays. Research on 3D modelling, however, has been primarily focused on moving away from WIMP interfaces and proposing more natural ways of creating such content.

Works such as Teddy [7] and Shapeshop [20] use 2D sketches to generate 3D content. In these, a created object can be modified by adding or removing content with additional sketches. Shapeshop was later extended by Lopes et al. [12] combining bimanual touch manipulation to the pen based input. They used pen input for precise operations such as sketching and touch input to secondary operations, like toolbar selection and camera manipulation.

Another usual approach is to obviate the mapping between 2D input and 3D actions, through 3D input. In the early days of VR, 3DM [5] used a handheld device with 6DOF tracking to manipulate a spatial cursor and create 3D models. More recently, Mockup Builder [1] introduced a direct modeling approach that mixes 2D touch input with 3D mid-air gestures above the interactive surface. It uses a stereo projection to co-locate imagery and users' hands. Users can create and edit 3D content by sketching in the surface or in the air, and then extrude resulting polygons of a mesh object. Also resorting to stereoscopic imagery and mid-air interactions, Takala et al. [23] extended the Blender software with 6DOF handheld controllers. Users control a 3D cursor to paint with meta balls in 3D, for modelling organic shapes, and to change the location and rotation of objects, which is used when placing objects or extrude polygons of meshes.

DIY World Builder [24] uses a magic wand metaphor to create 3D models and environments, by pointing in mid-air at objects. Although users perceive the virtual world through an HMD, tools and properties are chosen in a 2D interface in a smartphone placed on the non-dominant wrist. The Wonderland Builder [3] offers a multi-modal interface where users can interact using either a similar magic wand metaphor or through voice commands. These two works, however, do not allow any kind of object's mesh modification. They instead instantiate objects and then perform translation, rotation and scale operations on them.

MakeVR [9] follows a two-handed interface with hand-held controllers, allowing two objects to be manipulated simultaneously and independently. This work is the first that offers Boolean operations between objects in mid-air: union, intersection and difference. Boolean operations can be performed by making two objects intersecting each other and then pressing a button in the controllers. The operation's result is applied to the first selected object.

Focusing on transposing physical world modelling interactions to the virtual world, Cho et al. [6] implemented the metaphor of the potter's wheel. While the non-dominant hand spins the wheel, the dominant hand generates and modifies the model through 3D drawing poses, which are converted to virtual brushstrokes. Unlike conventional CAD systems, this does not rely on exact dimensions and geometries, offering instead greater flexibility for tasks such as conceptual design or idea visualization. Mine et al. [16] converted SketchUp desktop application into a VR application. They built a hybrid setup that collocates a touch display and physical buttons. 3D spatial input was used for coarse starting steps, and 2D touch for precision input, while perfoming modeling operations. Jackson et al. [8] presented Lift-Off, an immersive modeling system with a bimanual 3D user interface. It enables users to freely draw 3D model with fine control using both hands to define curves and surfaces.

Besides traditional WIMP interfaces, Boolean operations between two objects are often disregarded for 3D modelling. Virtual objects are usually created through primitive instantiation or sketching, and edited with extrusion operations or vertex manipulation. Approaches that try to bring natural interactions to the virtual world tend to move away from exact shapes, using brush like metaphors. Since Boolean operations between two objects can not be treated as natural tasks, due to physical constraints, they are usually applied through menu navigation or selected using physical buttons on handheld-controllers. In our work, we developed an innovative approach that employs natural manipulation metaphors to perform such operations with mid-air interactions.

2.2 Self Representation

Many are the factors that affects the VR experience, being presence the most important of them. Presence [22] relates to the feeling of "being there" on the virtual environment and is important for a good experience in immersive settings. As opposed to CAVE-like systems, the use of Head-Mounted Displays occludes the real self of the user, compromising the overall virtual-reality session. A way of overcoming this is by using a fully-embodied representation of the user within the virtual environment, which impacts the feeling of presence [21] and task performance [14] in such setups. The sense of embodiment into an avatar is constitutive of the sense of presence in VR and affects the way one interacts with virtual elements [10].

The level of realism of the avatar also plays an important part on the VR experience and how it relates to the sense of embodiment of an user. Another common problem on this matter is the uncanny valley [17]. To this matter, Piwek et al. [19] state that the effect of realism in the deepest part of the valley become more acceptable when it is animated. Additionally, previous work by Lugrin et al. [13] relate the uncanny valley effect to presence and embodiment of avatars when viewed through a HMD. Recently, Argelaguet et al. [2] found that the use of realistic self-representation can negatively impact object positioning tasks.

Even though the impact of avatar's graphical realism in user's presence is widely studied, none of the works mentioned relates to the specific scenario of 3D modelling. Previous work [11, 18] employed human-like representation of hands of the user to locate

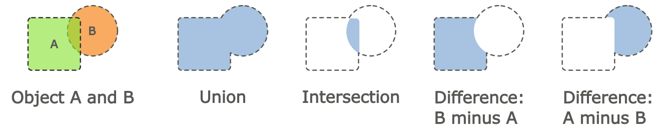


Figure 1: Boolean operations.

user's body movements in large-scale display setups in manipulation tasks. However, they do not relate manipulation efficiency with the representation used. A similar problem is found on MakeVR [9] which performs Boolean operation in VR using simplified visual indicators to map users actions in the VE.

3 Interaction Techniques for CSG Modeling in VR

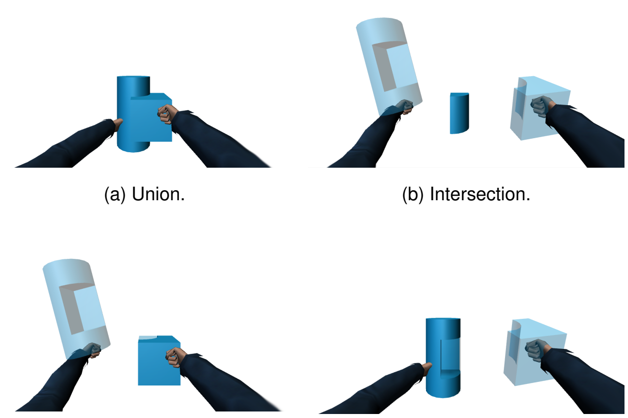
CSG through Boolean operations is widely used to create and edit 3D virtual models. These operations can either be union, intersection or difference (Figure 1). We propose two new techniques to perform Boolean operations in mid-air. One is based on mid-air manipulation gestures and the other follows a menu approach.

3.1 Gesture-based

For this approach, we followed natural manipulation gestures and applied them to Boolean operations. Firstly, users need to grab simultaneously the two objects A and B which the operation should be applied to, and make them intersect each other in the desired position. Then, after standing still for half a second, users can select the operation. For this, we used a metaphor of moving away parts that should be removed and releasing (Figure 2):

- Union (A or B): release both objects while in the intersection region (nothing is removed);
- Intersection (A and B): move both objects away;
- Difference (B not A): object A should be removed;
- Difference (A not B): object B should be removed.

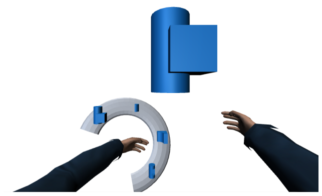
To decide if an object should be removed, we defined the intersection region as a sphere with 25 cm radius. While in the decision stage, a preview of the current resulting object is shown, which is confirmed by releasing the object.



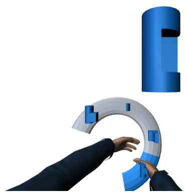
(a) Union. (b) Intersection.

(c) Difference: Left object minus the right object. (d) Difference: Right object minus the left object.

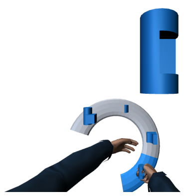
Figure 2: Boolean operations with mid-air gestures.



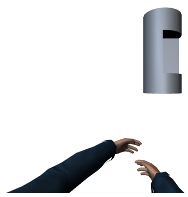
(a) Menu appears when releasing an object while colliding with another.



(b) Hovering to preview difference.



(c) Close hand to pick operation.



(d) Operation is applied and menu disappears.

Figure 3: Boolean operations with menu based approach.

3.2 Menu-based

The menu-based approach is inspired by traditional 2D menus present in traditional WIMP interfaces, with some improvements. To start Boolean operations, users can grab one object with the Dominant Hand (DH) and drag it so it intercepts another. Releasing the object while intercepting the other will reveal the menu.

The menu is represented in the VE through a semi-annulus divided into four sections, one for each selectable operation (Figure 3). The menu follows users' non-dominant hand (NDH) so that decisions can be made using the dominant hand. Operations in the menu are illustrated with previews of their resulting object. When users intersect one of the menu's sections with the DH, the preview is also shown in objects' position. Finally, a grab gesture with the DH confirms the operation.

4 Prototype

We built a prototype where we implemented our techniques to perform Boolean operations in mid-air to test and compare them.

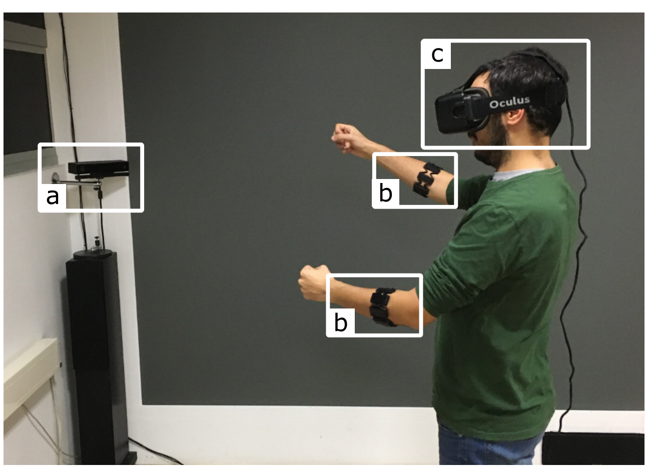


Figure 4: Setup of our prototype: (a) Oculus Rift DK2; (b) Myo armbands; (c) Microsoft Kinect v2.

(a) Simplified Hands (b) Avatar (c) User

Figure 5: Self-representations in relation to a real user.

4.1 Setup

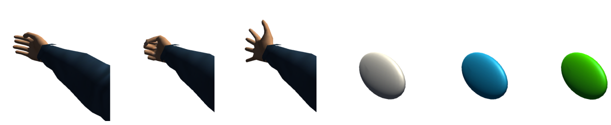
Our setup, illustrated in Figure 4, is composed by affordable off-the-shelf components that are able to detect both hand gestures and body movement, while providing an immersive experience. To display the VE we used an Oculus Rift DK2. To track user's full-body movement we used a Microsoft Kinect v2 depth camera. We combined Kinect's data with the orientation sensors embedded on the HMD to calculate users' point of view. Due to Kinect's limitations to properly recognize hands' gestures and orientation, we used additional Myo armbands. Myo armband is able to detect five poses, from which we use fist and spread fingers gestures.

4.2 Object Instantiation and Manipulation

To perform operations objects need to be created and placed in the desired position within the VE. To create objects, we implemented a palette metaphor, which is shown with a spread fingers gesture performed with the NDH facing up. Using the palette, users can choose a primitive object to be instantiated (cube, cylinder or sphere). The fist gesture is used for grabbing objects. We used an approach similar to the 6DOF-Hand [15]. It consists of a Simple Virtual Hand [4], directly mapping user's hand motion, with added scaling capabilities. The object will follow hand's position and orientation until a spread gesture is performed. Scale operations can also be done by grabbing an object with one hand and then closing the other. Increasing the distance between hands will uniformly enlarge the object. Analogously, moving hands closer will make the object smaller.

4.3 User Representation

There are two user representations in our prototype to test the influence of a self-avatar on the performance of a modelling task. We used two different representations, one with a full-body avatar, and another with a simplified representation of the user's hands (Figure 5). The avatar is scaled to match user's height and animated accordingly to the skeleton given by the depth camera. Since the Kinect depth camera does not provide accurate hand rotation values, we resort to Myo's rotations to animate avatar's hands, as well as the hands representation of the prototype's version without avatar. Hand poses are also captured with Myo armbands and that



(a) (b)

Figure 6: Avatar hand poses on both representations. From left to right: Idle, Closed, Spread.

information is used to animate the fingers of the avatar according to the current hand pose (Idle, Closed and Spread). On the simplified representation, we use different colors for each pose performed, as shown in Figure 6.

5 Preliminary Evaluation

We compared our proposed techniques against each other and against a baseline through a user evaluation with 24 participants. The baseline technique resorted to physical buttons on a handheld controller similarly to MakeVR [9]. Participants were asked to replicate a 3D model using the three techniques, which needed the usage of all Boolean operations. To test self-avatar efficiency, half the tests were performed with the full-body avatar, while the other half had the simplified hands' representation.

Although we found statistically significant differences between approaches in completion time, these were not related to Boolean operations' execution. Instead, they were noticeable in object manipulation, as this was mainly caused by the Myo armbands' inaccuracy to properly recognize hand gestures. It penalized the gesture based approach the most, as it relied on simultaneous grabbing with both hands. Qualitatively, participants reported preferring the menu based approach. Despite the interaction design of the gestures approach being identified as simple and natural, the menu was the easiest regarding recall. It showed the result of all possible operations beforehand, and participants could choose the desired outcome without really knowing which operation they were applying. Regarding user representation, we found that having a full-body avatar negatively influenced total completion time with all techniques ($t(22)=-3.003$, $p=0.007$).

6 Conclusions

Techniques that enable Boolean operations for 3D modelling in immersive virtual environments are scarce, with existing solutions following unnatural metaphors. In this work, we developed two novel mid-air techniques to perform Boolean operations between two 3D objects in Virtual Reality. The first is based on the gestures naturally used do to manipulate objects. The second follows a menu-based approach, which provides instantaneous feedback on all possible operations. We compared these techniques against each other and against a baseline based on the literature, which relies on handheld controllers' buttons. However, the Myo armbands used performed poorly, which, combined with far from perfect depth camera's positional tracking, lead to inconclusive results.

Although the use of fully-embodied avatars improves the sense of presence within the Virtual Environment, its impact on modelling tasks' performance had yet to be identified. To address this matter, participants in our evaluation experienced one of two different self representations. Having a hands-only simplified representation, participants performed significantly faster in comparison to when having full-body representation. This corroborates the findings from Argelaguet et al. [2] for object positioning, since the more realistic self-representation can occlude objects of interest.

As future work, we would like to assess if an improved user tracking leads to most significative results. Additionally, we believe that our techniques can be successfully combined with precision enhancing approaches for object manipulation, in order to create a fully capable CSG solution in VR.

Acknowledgements

This work was partially supported by FCT through grants UID/CEC/50021/2013, IT-MEDEX PTDC/EEISII/6038/2014, and SFRH/BD/91372/2012, and by CAPES through grant 9040/13-7.

References

[1] B. Araújo, G. Casiez, and J. Jorge. Mockup builder: direct 3d modeling on and above the surface in a continuous interaction space. In *Graphics Interface*, 2012.

[2] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *IEEE VR*, 2016.

[3] C. Barot, K. Carpentier, M. Collet, A. Cuella-Martin, V. Lanquepin, M. Muller, E. Pasquier, L. Picavet, A. Van Ceulen, and K. Wagrez. The wonderland builder: Using storytelling to guide dream-like interaction. In *IEEE 3DUI*, 2013.

[4] D. A. Bowman, E. Kruijff, J. LaViola, and I. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley, 2004.

[5] J. Butterworth, A. Davidson, S. Hench, and M. Olano. 3dm: A three dimensional modeler using a head-mounted display. In *Symposium on Interactive 3D Graphics*, 1992.

[6] S. Cho, D. Baek, S.-Y. Baek, K. Lee, and H. Bang. 3d volume drawing on a potter's wheel. *IEEE CGA*, 34(3), 2014.

[7] T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: a sketching interface for 3d freeform design. In *ACM SIGGRAPH*, 1999.

[8] B. Jackson and D. Keefe. Lift-off: Using reference imagery and freehand sketching to create 3d models in vr. *IEEE TVCG*, 22(4), 2016.

[9] J. Jerald, P. Mlyniec, A. Yoganandan, A. Rubin, D. Paullus, and S. Solotko. Makevr: A 3d world-building interface. In *IEEE 3DUI*, 2013.

[10] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence*, 21(4), 2012.

[11] K. Kiyokawa, H. Takemura, and N. Yokoya. Manipulation aid for two-handed 3-d designing within a shared virtual environment. In *HCI (2)*, 1997.

[12] P. Lopes, D. Mendes, B. Araújo, and J. A. Jorge. Combining bimanual manipulation and pen-based input for 3d modelling. In *SBIM*, 2011.

[13] J. Lugrin, J. Latt, and M. Latoschik. Avatar anthropomorphism and illusion of body ownership in vr. In *IEEE VR*, 2015.

[14] E. McManus, B. Bodenheimer, S. Streuber, S. De La Rosa, H. Bülthoff, and B. Mohler. The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments. In *ACM APGV*, 2011.

[15] D. Mendes, F. Fonseca, B. Araujo, A. Ferreira, and J. Jorge. Mid-air interactions above stereoscopic interactive tables. In *IEEE 3DUI*, 2014.

[16] M. Mine, A. Yoganandan, and D. Coffey. Principles, interactions and devices for real-world immersive modeling. *Computers & Graphics*, 48, 2015.

[17] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE RAM*, 19(2), 2012.

[18] N. Osawa. Two-handed and one-handed techniques for precise and efficient manipulation in immersive virtual environments. In *ISVC*, 2008.

[19] L. Piwek, L. McKay, and F. Pollick. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, 130(3), 2014.

[20] R. Schmidt, B. Wyvill, M. C. Sousa, and J. A. Jorge. Shapeshop: Sketch-based solid modeling with blobtrees. In *SBIM*, 2005.

[21] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM TOG*, 29(4), 2010.

[22] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM TOCHI*, 2(3), 1995.

[23] T. Takala, M. Mäkäräinen, and P. Hämäläinen. Immersive 3d modeling with blender and off-the-shelf hardware. In *IEEE 3DUI*, 2013.

[24] J. Wang, O. Leach, and R. Lindeman. Diy world builder: An immersive level-editing system. In *IEEE 3DUI*, 2013.