



Gesture Interface for 3D CAD Modeling using Kinect

Dipen Dave, Ashirwad Chowriappa and Thenkurussi Kesavadas

Virtual Reality Lab., University at Buffalo, kesh@buffalo.edu

ABSTRACT

Modeling in general can be viewed as a two-step process, the first being concept visualization followed by the final design. Most designers who work with CAD systems go through the concept visualization stage and follow the same steps required in the final design stage. A standard approach to concept visualization is traditionally performed using clay modeling, yet there is a lack of computer aided design tools that can faithfully mimic the process of traditional clay modeling both in terms of representation of virtual clay as well as an immersive user interface capable of capturing the user's intent in an intuitive sense. In this paper we describe a voxel based representation of virtual clay and introduce a hierarchical gestural interface using Kinect which together brings the intuitiveness of traditional clay modeling to a CAD environment. The framework was validated using a 3D clay modeling software (n=13). Results showed that most of the participants were able to adapt to the new interface quickly.

Keywords: free form modeling, virtual clay, kinect gesture interface.

DOI: 10.3722/cadaps.2013.663-669

1 INTRODUCTION

The goal of developing a free form modeling system is to maintain the paradigm of real life object interaction, while harnessing the benefits that come with the use of computer simulation. The objective of such a system, in context of clay or free form modeling, is to generate schemes for simulating composite behavior that includes operations of adding, deforming, squeezing, rolling and shearing making this a powerful approach to model 3D objects. In this paper we describe our approach of simulating elasto-plastic behavior of clay with joining, slicing and deforming capabilities. To provide an intuitive input paradigm to control the modeling process, we have used a tool paradigm, where tools are initial shapes that form the prime creator, modifier and destroyer of objects. These tool manipulators are controlled by a hierarchical gestural interface using Kinect to make the clay modeling process more intuitive.

2 PRIOR WORK

3D clay based concept is not new. Galyean and Hughes [3] introduced an all-in-one virtual sculpting process. They achieved volumetric operations on clay using a volumetric representation of clay data which they called Voxmap. Using Polhemus Isotrack[®] device to control the position and orientation of the sculpting tool they facilitated the process of carrying out local deformations on the volume in 3D space. Their method was extended by Dewaele and Cani[1]. Mizuno et al. [7] used a similar method to carry out chiseling like operations. Complex sculpting tools were described in Ferley et al. [2] in which they added, removed, painted, smoothed and stamped on virtual surface. Perng et al [8] used 3D Tracker with 6DOF to handle the sculpting tool. Their system provided capabilities to perform carving, stuffing, pottery making and painting operations. They achieved this by changing the state of voxels near the boundary of cutting tool. Kameyama [5] developed a haptic CAD system to give the sculptor a realistic sculpting experience of holding and touching the material through a haptic device. Kesavadas et. Al [6], Kamerkar and Kesavadas [4] used a pressure sensitive input device called ModelGlove to edit NURBS surfaces. In commercial domain two software products exist; Z-brush [13] and Autodesk Mudbox [14]. Mudbox uses a 3D brush based sculpting paradigm which is tuned to advanced painting and texturing applications commonly uses in digital art and animations. Z-brush is also a 3D sculpting package with excellent painting capabilities. However, all of the above mentioned tools, provide for modeling capabilities neglecting interactive user interface design.

With respect to the user interface, efforts have been made to improve the intuitiveness. MoZart[10] employs multi-modal interface of touch and sound for modeling. Kang et al. [7] designed an interface which uses hand gestures for where tracking was performed using reflective markers and motion capture cameras. KinectTM as an input device for CAD applications is new and no references except a one online video [12] were found during literature search. In our work, we have utilized some of the best features of above mentioned work but we have also overcome many inherent weaknesses of the previous systems. Protean[®] uses discrete elements (voxels) to model solids rather than surface polygons. Unlike conventional ways to handle voxels, a compressed format to store voxel data is used. In this paper we discuss the features of Protean software [15] and then demonstrate how a KinectTM input device can be modified to perform modeling in place of mouse and keyboard.

3 CAD MODELING

Inspiration for digital sculpting originated from familiar two dimensional paint software, such as Adobe Photoshop, Corel Painter to name a few. These applications are designed specifically to handle complex operations on bitmaps and achieve desired effects or shapes, where most of the operations are applied via brushes. Protean Clay (Fig.1), a commercially available software brings this paradigm into the three dimensional domain and emulates intuitive, physical, sculpting processes essentially through painting in three dimensions.

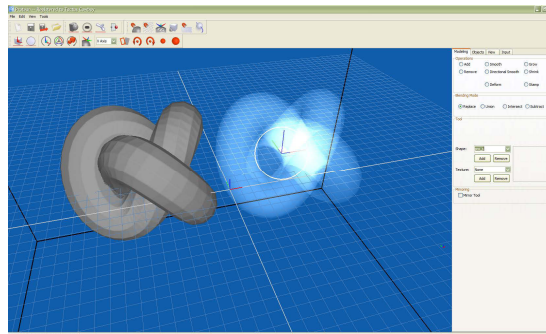


Fig. 1: Example of *Add* operation with custom tool shape. On the right side is the tool pallet.

Modeling with Protean Clay is performed using three major types of operations, constructive solid geometry operations, smoothing, and deformations, the last of which comprises the majority of the operations available in Clay. This makes Protean Clay a viable candidate to test the proposed gesture based interface.

4 INTERACTIONS

Object modeling in most free form CAD software requires that tool be moved continuously in 3D space. The use of standard 2D I/O devices such as keyboard and mouse for modeling makes this process tedious and unintuitive. In order to address these issues, use of interface with 3D input is one solution. Studies [11] have shown that gestural interfaces are highly intuitive and promote the level of involvement in the virtual 3D space unlike mouse and keyboard. We propose a 3D gesture based interface using Kinect, which is a sensor array with a depth sensor, RGB camera and four microphones. We make use of the skeletal tracker, part of Kinect SDK which can track 20 joints on the user and provides access to the spatial positional data of these joints in real time. Since there is no support for gesture interface currently on the Windows platform, we implement the same by emulating mouse and keyboard input using WinAPI.

The skeletal tracker algorithm implementation on Microsoft SDK, uses the depth data from Kinect to track up to six people in the scene. Once a person has been registered and tracked, it returns the spatial positional 3D data for 20 joints on each user. We use this tracked data for gesture recognition to identify user intent (in manipulating the 3D environment). In our current implementation we test the system by providing gesture recognition capabilities by a single user. However, we foresee its use in a collaborative 3D environment for multiple users.

The gesture recognition schema is implemented as a state machine (Fig. 2)

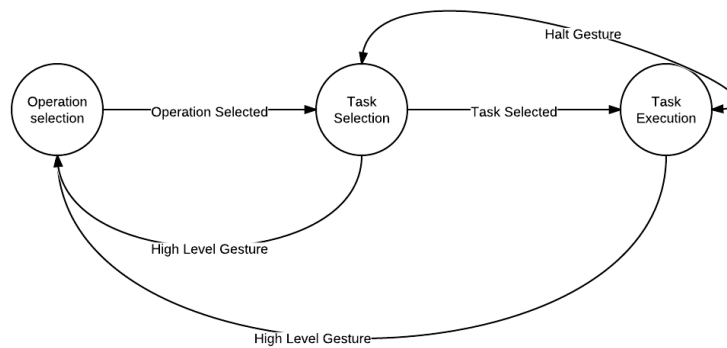


Fig. 2: State Transition Diagram.

For defining the gesture semantics, we identified gesture atomic primitives such that a unique combination of these primitives defines a specific gesture (Fig. 4). Right hand to the right of the body, raised left hand is some examples of primitive gestures. Right and left hand raised above the head defines the gesture to select an operation. Under this scheme of defining gestures as a combination of primitives, we classified them based on the ease of performing them into high level and low level gestures. Low level gestures are the ones that are to be performed more frequently and hence they have to be simple combination of primitives. On the other hand, high level gestures are to be performed less frequently and hence some complexity is permissible in their combination.

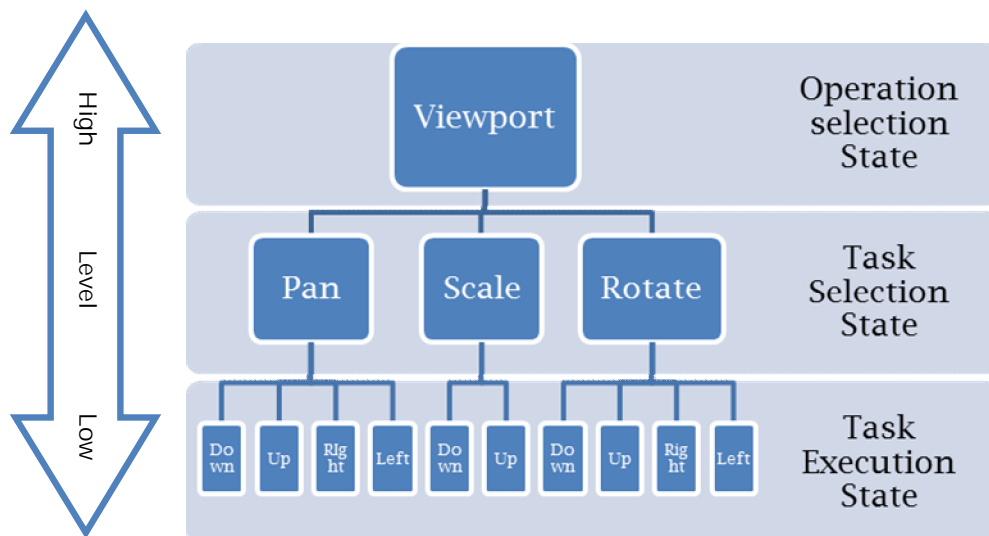


Fig. 3: Example of states for Viewport Operation.

A gesture continuum shown in Fig. 4 was structured such that high frequency gestures and low frequency gestures were modeled to best convey the user's intent. When the user executes a high level

gesture, the system halts in the operation selection state until an operation selection gesture is performed. On selection of an operation, the state machine switches to the task selection state and waits for the user to select a task. Finally after selecting a task, the system takes the user's actions as intent to perform low level action such as panning, scaling, rotation etc.

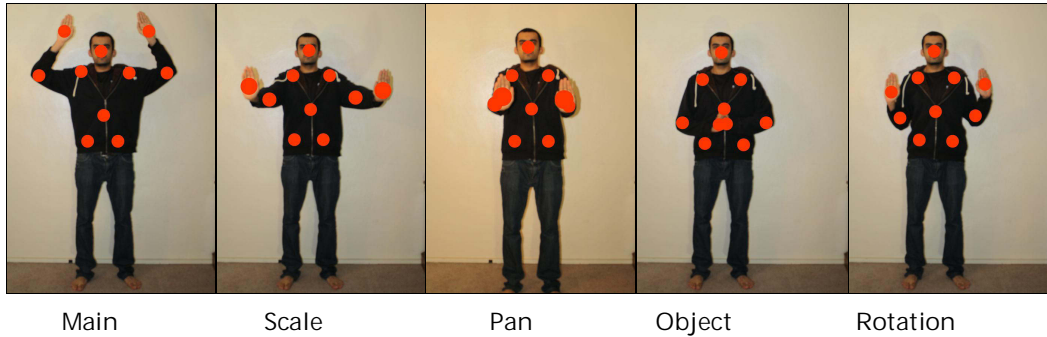


Fig. 4: Skeletal points overlaid on the RGB image.

5 VALIDATION STUDY

In order to validate our approach, a preliminary study was conducted. The task required the participants to manipulate a 3D environment using the approach mentioned above (section 4).

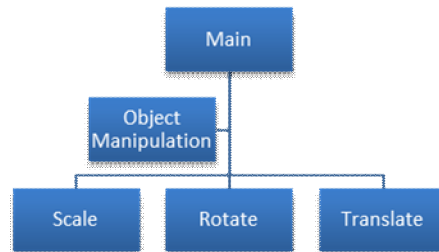


Fig. 5: Hierarchy of subtasks.

The tasks were further decomposed into sub-tasks give in Tab. 1, where the user could switch between tasks using the proposed hierarchy.

| Task # | Type | Tasks/Sub-tasks |
|--------|---------------------|--------------------------------------|
| 1 | Main | Go to Main |
| 2 | Object selection | Select Object Manipulation |
| 3 | Object manipulation | Select Scale |
| 4 | Scale | Scale the object to a specific value |
| 5 | Main | Go to Main |
| 6 | Object selection | Select Object Manipulation |
| 7 | Object manipulation | Select Translate |

| | | |
|----|---------------------|---|
| 8 | Translate | Translate the object in a specific trajectory |
| 9 | Main | Go to Main |
| 10 | Object selection | Select Object Manipulation |
| 11 | Object manipulation | Select Rotate |
| 12 | Main | Go to Main |

Tab. 1: Task the user was expected to perform.

The performance metrics for the task were time (t) and number of gestures (g) required to reach the goal. The subjects were divided into two groups. Group A (n= 10) were subjects who had prior experience in CAD modeling (experts) and Group B (n=3) had no prior experience using CAD systems (novices). Each participant was given a total of three trials. From the experimental study the average time taken to complete the task was 41.3 seconds for the expert group and 64.5 for the novice group. The average number of task switches was 14.5 for experts and 20.56 for novices. Results varied significantly across participants and across the trials. The maximum time taken was 112.7 seconds and the minimum was 15.3 seconds for the expert group; 116.7 and 21.7 for the novices. One important observation was that for both the groups participants improved the time to complete the tasks after each trial demonstrating that significant learning was possible with the new interface (Figure 7).

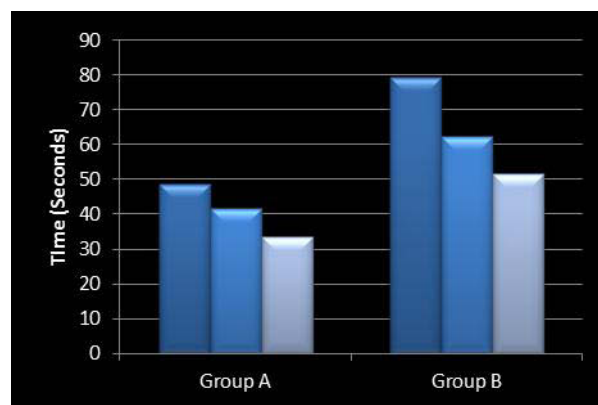


Fig. 6: Plots showing the average mean time taken across the three trials showing that participants improved their performance within three attempts.

After completion of the task, participants were presented with a questionnaire for a qualitative analysis. Questionnaire was followed by a brief discussion about the effectiveness of the system. The questionnaire required a scaled rating for Intuitiveness, and Fatigue. Majority of the participants found the interface to be intuitive however a few participants experienced fatigue while using the interface.

5.1 Discussion

The results look promising for a gesture based modeling environment as most of the participants improved their performance by the third trial. Our results indicate that the learning curve was short for both experts and novices which demonstrate the effectiveness of the gesture based approach for

modeling. The system is fairly intuitive and with careful task specific selection of the gestures, fatigue issue can be reduced.

6 CONCLUSION AND FUTURE WORK

In this paper a new method of carrying out sculpting of digital clay was proposed and successfully implemented. The preliminary tests have been successfully carried out demonstrating that the proposed system can be used to navigate through tasks required for 3D CAD modeling. An innovative new interface in the form of hierarchical gestures was implemented using Kinect. In the future, we plan to implement a native gestural interface in Protean to carry out more complex modeling of as opposed to the current iteration which uses mouse and keyboard emulation, which will give the user much more control in modeling.

REFERENCES

- [1] Dewaele G.; Cani M.P.: Interactive global and local deformations for virtual clay, Pacific Graphics, 2003.
- [2] Ferley, E.; Cani, M.P.; Gascuel, J.D.: Practical Volumetric Sculpting, The Visual Computer, 16(8):469-480, 2000. <http://dx.doi.org/10.1007/PL00007216>
- [3] Galyean, T.; Hughe, J.: Sculpting: An Interactive Volumetric Modeling Techniques, Proceedings of SIGGRAPH 1991: 267-274, 1991. <http://dx.doi.org/10.1145/127719.122747>
- [4] Kamerkar, A.; Kesavadas, T.: Touch Based Interactive NURBS Modeler using a Force/Position Input Glove, ASME 2004, Design and Technology Conferences, Salt Lake City, Utah, 2004.
- [5] Kameyama, K.: Virtual clay modeling system, Symposium on Virtual Reality Software and Technology (VRST 1997), New York, ACM: 197-200, 1997.
- [6] Zhong, K.; Kang, J.; Qin, S.; Wang, H.: Rapid 3D Conceptual Design Based on Hand Gesture 3rd International Conference on Advanced Computer Control, 2011.
- [7] Kesavadas, T.; Kamerkar, A.; Anand, A.: Touch-Based Interactive NURBS Modeler Using a Force/Position Input Glove, Virtual Engineering, Chap 2, Momentum Press, 2010.
- [8] Mizuno, S.; Okada, M.; Toriwaki, J.: Virtual Sculpting and Virtual Woodcut Printing, The Visual Computer 14(2): 39-51, 1998. <http://dx.doi.org/10.1007/s003710050122>
- [9] Perng, K-L; Wang, W-T; Flanagan, M; Ouhyoung, M.: A real-time 3D virtual sculpting tool based on modified marching cubes, SIGGRAPH, 2001.
- [10] Sharma, A.; Madhvanath, S.; Shekhawat, A.; Billinghamurst M.: MozArt: A Multimodal Interface for Conceptual 3D Modeling, 13th ACM International Conference on Multimodal Interaction, 2011.
- [11] Van Beurdena, M.H.P.H.; IJsselsteijna, W.A.; Hopf, K.: User centered design of gesture-based interaction technology, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3DVideo(3DTV-CON), 2011, vol., no., pp.1-4, 16-18 May2011 doi: 10.1109/3DTV.2011.5877225
- [12] <http://www.youtube.com/watch?v=IYaSf6wkI70> (accessed on April 7, 2012)
- [13] <http://www.zbrush.com> (Accessed on April 7, 2012)
- [14] <http://usa.autodesk.com/adsk/servlet/pc/index?id=13565063&siteID=123112> (April 7, 2012)
- [15] <http://www.proteanclay.com> (Accessed on April 7, 2012)