MozArt: An Immersive Multimodal CAD System for 3D Modeling

Anirudh Sharma and Sriganesh Madhvanath
Hewlett-Packard Labs, Bangalore, India
{anirudh.sharma, srig} @hp.com

ABSTRACT

3D modeling has been revolutionized in recent years by the advent of computers. While computers have become much more affordable and accessible to the masses, computer modeling remains a complex task involving a steep learning curve and extensive training. In this paper we describe the MozArt Table, our effort to redefine the interface for computer modeling to make it more accessible to lay users. We have explored both the hardware and software aspects of the interface, specifically, the use of intuitive speech commands and multitouch gestures on an inclined interactive surface. The paper describes our approach, hardware setup and the technology used to make it work.

General Terms

Design, Human Factors

Keywords

Multimodal interfaces, multi-touch, computer aided design (CAD), 3D modeling, speech input, gestures, touch table

1. INTRODUCTION

Modeling of three-dimensional shapes and structures comes naturally to human beings, as can be evidenced from the ubiquity and antiquity of human activities such as clay modeling, pottery, carpentry etc. [3]. 3D modeling has been revolutionized in recent years by the advent of computers, so much so that it is difficult to come across a product today that was not first visualized on a computer. While computers have become much more affordable and accessible to the masses, computer modeling remains a complex task involving a steep learning curve and extensive training.

One of the key issues with these interfaces is the extensive use of complex GUIs operated using keyboard and mouse [4]. These WIMP (Windows, Icons, Menu and Pointer) based interfaces in use today make it difficult for the average user to participate in CAD modeling due to the complex button sets, toolbars and keyboard shortcuts. It takes a substantial effort to attain design proficiency with these tools; designers still prefer to use pen and paper for early stages of design where they prefer to sketch freely without getting into the complexities of the software, even though

the results are much superior with the latter.

The last few years have witnessed a rapid shift away from WIMP interfaces towards ones that use modalities such as touch, speech, and vision etc. to interact with digital content. In this paper we describe *MozArt*, a research effort to explore how multimodal interfaces employing touch and gesture may be applied to simply 3D modeling. We have explored both hardware form factors as well as the use of multiple modalities of interaction, and we describe our efforts in this paper.

This paper is organized as follows. We first cover in brief some of the recent literature related to human-computer interfaces for computer modeling. In Section 3, we describe our hardware prototype for 3D modeling - a tiltable drafting-table style interactive surface. In Section 4, we describe the user interfaces and touch and speech interactions we have enabled for 3D modeling. Section 5 describes the software design of the system, where we have leveraged open source extensively. We conclude the paper with a discussion of what we have learned so far and next steps.

2. RELATED WORK

Recently, there have been a great variety of multitouch table-top displays and research prototypes that have made it possible to directly manipulate user interface components using tactile gestures. Some of these efforts have looked at the problem of 3D modeling. For example, SpaceClaim recently demonstrated a prototype of a CAD tool with a multitouch screen [9]. However, the demonstration does not address ease of task-switching and continues to use the same toolbars and buttons that might be used with keyboard and mouse.

The work of Arangarasan & Gadh [1] makes use of multimodal input and output for working with 3D models, however, it requires the use of head-mounted displays. In addition to the obvious cost implications, this can make the system difficult for beginners due to ergonomic issues.

ILoveSketch [8] combines stylus input with automation to allow users to create accurate 3D models primarily by sketching using an electronic stylus. While well suited for the creation of curved and organic shapes, the interface may be more complex than necessary for simpler 3D forms.

CAD had traditionally been viewed as the domain of experts, and these systems are in general targeted towards their needs. The motivation for our work has been to enable even lay users to explore 3D modeling. In the following sections, we describe the MozArt prototype system that we have built, in two parts: the hardware, and the user interface and interactions.

3. MozArt HARDWARE PROTOTYPE

In this section, we describe the hardware prototype that we have built to support 3D modeling. The hardware is essentially a tilt-mounted multi-touch display, similar to an architect's drafting table.

A traditional drafting table can be considered a large-format display that affords idea sharing by virtue of its visibility, and provides a platform that enables collaboration and discussion about designs [2]. Our setup involves a fabricated tilt mount having a touch screen with which the user can interact using voice and touch (Figures 1 & 2).

Figure 1: MozArt tiltable multitouch table schematic

Figure 2: MozArt hardware prototype

The key feature of the prototype is the customizable orientation. The display can be titled from 0 to 45 degrees. The dimensions of the structure are 1m x 0.5m x 1m, and this makes it well-suited for use standing, or bending, and for collaborating with others.

The projection display is obtained by feeding the VGA output of a computer to a projector. The projector's output is reflected onto the surface using a mirror placed below the tabletop.

Multi-touch is implemented using the LLP (Laser Light Plane) technique, shown schematically in Figure 3. The four corners of the tabletop are fitted with 850nm Infrared lasers. The scattered light generated upon touching the surface is captured by a modified infrared camera below the table surface, and processed using Community Core Vision (CCV) libraries [10] to detect and track blobs (Figure 4).
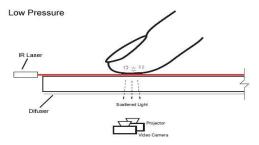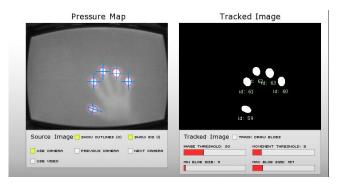

Figure 3: Laser Light Plane technique

Figure 4: Blobs generated upon multitouch

4. USER INTERFACE

We decided to use Google SketchUp [11] as the underlying modeling tool. Google SketchUp is predominantly used to create 3D models early in the design process where the designs are still in a fluid and tentative state. An informal online survey of users conducted on NUIGroup and SketchUp forums also suggested that Google SketchUp was the preferred tool for designing initial 3D models, as compared to AutoCAD, Blender and SolidWorks. Moreover, it comes with a Ruby API which enables writing extensions and adding new capabilities.

The default SketupUp UI is designed for keyboard and mouse interaction, and features extensive use of menus for selecting modes, views and shapes (Figures 5, 6). In our interface, we have replaced the use of toolbars and buttons for a subset of common tasks with touch and speech commands, enabling the user to focus on design rather than locating buttons for intended tasks (Figure 5).
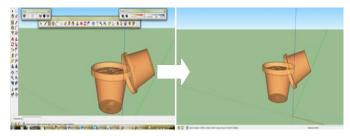

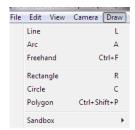**Figure 5: Default UI transformed into minimal UI**

Figure 6: Command list in one of the menus

Drawing shapes

For example, the speech command "*draw rectangle*" activates the 'rectangle' mode which allows a rectangle to be drawn or modified using touch gestures; the dimensions of the rectangle are specified by dragging of finger on screen (Figure 7). Speech when used in this manner allows the visual attention of the user to remain on the design rather than being diverted to locating buttons and toolbars to switch modes of operation.

Figure 7: Drawing a rectangle after speech input

The design of the multimodal commands exploits the key SketchUp feature of defining all shapes using edges (straight lines) and faces (2D shapes created when several edges form a flat loop). Further, the pushing and pulling feature easily allows 2D shapes to be extruded into 3D ones.

The speech command "*pull it up*" enables the extruding mode; the user can now touch the face of the 2-D shape and pull it up intuitively using fingers, transforming the 2-D shape into its volumetric extrusion (Figures 8, 9).
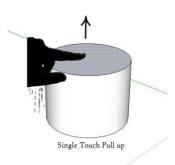

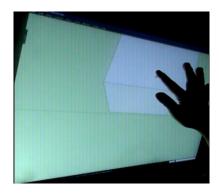**Figure 8: Extruding a circle using touch drag on face**

Figure 9: "Pulling up" a rectangle to form a cuboid

Orbiting and panning

Changing view port is a frequent activity during CAD modeling. We employ the speech command *"Show me around"* to activate the orbiting mode. The two finger pinch-to-zoom gesture may now be used to zoom-in and out or into the perspective (Figure 10).

The vocabulary of speech and touch commands implemented in our interface is specified in tables 1 and 2.
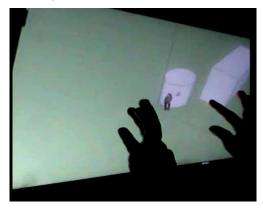

Figure 10: Zooming in and out of the 3D model

Table 1: Speech commands

Speech Command	Category	Objective
"Draw a Circle"	Activate(circle)	Modify(M)
"Draw a rectangle"	Activate(Rectangle)	Modify(M)
"Pull it up"	Activate(Push/Pull)	Operate(O)
"Show Isometric"	Show isometric model view	Change viewport(V)
"Top View"	Show top view	Change viewport(V)
"Show me around"	Activate 3D orbiting	Change viewport manually(V)
"New model"	Close and load new	File Tasks

Table 2: Touch gestures

Touch Gestures	Outcome
Two finger pinch	Zoom-in/Zoom-out
Single touch drag	Specify extent of M and O

5. SYSTEM DESIGN

We wrote a touch-based listener for Google SketchUp that interprets TUIO [6] events from the touch surface, and parses them into gestures such as drag, pan, zoom, touch down, touch up. The incoming TUIO data is parsed using a Touch Client written in C++ and loaded into Google SketchUp using a Ruby script. The multitouch gestures are then mapped to events such as dragging, zooming, and panning.

CMU Sphinx [12] is used to process incoming speech from the microphone. The commands are forwarded to SketchUp using WindowsHooks [13] to initiate keyboard shortcuts for switching between tasks. The overall system design is shown pictorially in Figure 11.
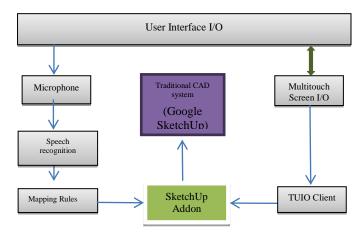

Figure 11: Overall system design

6. DISCUSSION AND FUTURE WORK

We have described the current status of our prototype implementation of MozArt, a tiltable multi-touch table with a multimodal interface that attempts to present an alternate and easier way to interact with 3D models. The form factor chosen provides a greater sense of directness because users put their fingers directly onto the item of interest and manipulate it. Also, the tilt enhances collaboration and prevents fatigue over prolonged working sessions [7]. Our next step is to undertake a formal user study comparing our prototype with the original desktop UI in both objective and subjective terms such as task completion times and user satisfaction.

We are also working on further expanding the multimodal vocabulary. For many modeling tasks, precision of extents becomes a key issue. To address this, we intend to support spoken commands that allow precise dimensions to be specified.

For example, if

$S=$ *"Create a circle with radius 15"*

then

$P=\{$*"create a circle", "with radius 25"*$\}$,
$M=\{$*"create", "circle")*,
$D=\{$*"radius", "15")*

P=Separated phrase

S=Sentence spoken

M=Modify task

D=Dimension specification

Also, speech input using an open microphone suffers from the problem of false positives (spurious input) due to ambient sound and more so in multi-user environments. We are working to add lip movement detection based on image processing to eliminate such errors.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Arangarasan R. , Gadh R., 2000 Geometric Modeling and Collaborative Design in a Multi-Modal Multi-Sensory Virtual Environment. In *DECT 2000*

[2] Buxton W. , Fitzmaurice G., Balakrishnan R., Kurtenbach G., 2000, Large Displays in Automotive Design, *IEEE Computer Society 20 (4), 68–7.*

[3] Cani M., Angelidis A. 2006. Towards Virtual Clay. *SIGGRAPH 2006 Courses*

[4] Darren A., Yaser G., Frank Maurer 2010 Adapting existing applications to support new interaction technologies: technical and usability issues. *Proceedings of the 2nd ACM SIGCHI symposium on Engineering Interactive Computing Systems.*

[5] Park, J. and Han T., 2010. LLP+: multi-touch sensing using cross plane infrared laser light for interactive based displays, *SIGGRAPH 2010 Posters*

[6] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza. TUIO: A protocol for table-top tangible user interfaces. *In Proceedings of Gesture Workshop 2005, 2005*.

[7] Piper AM, Campbell R., Hollan J. 2010. Exploring the accessibility and appeal of surface computing for older adult health care support, Technical Report. *Proceedings of the 28th international conference on Human Factors in Computing Systems*

[8] BAE, S.-H., BALAKRISHNAN, R., AND SINGH, K. 2008. *ILoveSketch: As-Natural-As-Possible Sketching System for Creating 3D Curve Models. In Proc. UIST, ACM, New York, 151–160.*

[9] SpaceClaim Multitouch CAD modeling. http://www.spaceclaim.com/en/

[10] Community Core Vision Library
http://nuicode.com/projects/tbeta

[11] Google SketchUp
http://sketchup.google.com/

[12] CMUSphinx Libray
http://cmusphinx.sourceforge.net/

[13] Windows Hooks
http://support.microsoft.com/kb/3188