DOI: 10.1109/CW.2016.14 Preprint Version. Manuscript submitted to and accepted by Cyberworlds 2016

Exploration of Natural Free-Hand Interaction for Shape Modeling using Leap Motion Controller

Jian Cui
School of Computer Science and Engineering
Nanyang Technological University, Singapore
e-mail: CUIJ0006@e.ntu.edu.sg

Arjan Kuijper
Fraunhofer IGD & Darmstadt Technical University
Darmstadt, Germany
e-mail: arjan.kuijper@igd.fraunhofer.de

Alexei Sourin
School of Computer Science and Engineering
Nanyang Technological University, Singapore
e-mail: assourin@ntu.edu.sg

Abstract— **In this paper, we propose a web-enabled shape modeling system with natural free-hand interaction, which can be easily learned by users while imposing least mental load on them. The deformation interface allows for performing various deformations, including stretching, compressing, squeezing, enlarging, twisting and tapering, on shapes interactively mimicking how they are done in real life. The manipulation interface allows an object to be directly grabbed and manipulated with either one or two hands, while also smoothly switching between them. Constrained methods are also provided for precise manipulation. An intuitive metaphor is designed to help the users to discover the interaction techniques by themselves without any manuals or instructions. A rendering pipeline, based on function-based extension of VRML/X3D, is designed with hidden complexity to support the proposed functionalities of the system. Hands motions are captured by Leap Motion controller. The user study proves the naturalness of the modeling system, and its easiness to be learned and remembered.**

Keywords—hand interaction; mid-air interaction; shape modeling; function-based representation; gestural interface; Leap Motion controller; natural interaction; real-life interaction behaviors; least mental load

I. INTRODUCTION

Most 3D shape modeling systems still rely on 2D interfaces, where 3D design and manipulation tasks have to be decomposed or mapped into 2D or 1D operations, e.g., using text editors, curve editors, or trial-and-error slider bars. They fail to map to the way of how people conceive and manipulate 3D shapes, imposing a large amount of time to be spent on learning how to use the tools rather than providing flexibility and fluidity during the process of shape construction. The efficient and natural user interaction then becomes a bottleneck for 3D shape design.

An alternative way is to use hands to create, deform and manipulate shapes directly in the 3D space. Compared to the traditional mouse- and keyboard-based 2D graphical user interfaces, free-hand interaction can provide additional degrees of freedom and valuable geometrical information. However, traditional shape modeling systems deal with a large amount of operations not linked with real-life interaction scenarios, such as point editing and surface extrusion. These complex operations inevitably result in a set of artificial gestures or unrealistic interaction to be designed in hand-based 3D shape modeling systems, which fails to match realistic interaction behaviors in real life. The users have to be trained for a long time to acquire the interaction skills, and spend much effort to remember the interaction ways. This observation motivates us to explore natural hand interaction and a new shape modeling paradigm mimicking realistic interaction behaviors.

We hypothesized that a natural interaction interface, allowing the users to interact in a way similar to real life, would enable them to learn, acquire and master shape modeling quickly with least mental load and training. Exploring this idea requires us to solve three research problems. *The first challenge* was to investigate various deformation and manipulation operations–how hands are used to stretch, compress, twist, taper, squeeze and manipulate objects uni-manually and bi-manually in real life–and to design interaction techniques mimicking these behaviors to be completed a hand skeleton model. However, it is difficult to use hands for precise modeling and manipulation operations without force feedbacks, which cannot be provided by optical tracking devices. Therefore, *our second challenge* was to come up with intuitive visual metaphors to effectively harness free hand interaction in a controllable manner for precise deformation and manipulation. Finally, a mathematical model had to be selected to support shape modeling and deformation in web-enabled virtual modeling spaces. Hence, *the third challenge* for us was to design an interaction paradigm accommodating the proposed natural interaction interface, which leaves users to concentrate on the interaction without knowing the complexity beneath the interface.

II. RELATED WORKS – SHAPE MODELING WITH FREE HAND INTERACTION

Free-hand interactions have been used in 3D shape modeling systems to facilitate modeling process. We classify them into three categories: free surface modeling systems, sculpting and claying systems, and CAD systems.

Free-hand interaction was used to assist in *constructing and modifying free-form surfaces.* In a surface modeling system Surface Drawing [1], hands were treated as brushes to directly draw surface strokes. In another system proposed in [2], generalized cylinders were created using three variables extracted from the hands geometry: the distance between the hands, the orientation, and the midpoint of the line joining the two hands. A sketch-based modeling system called Mockup Builder [3] allowed hands to be used for sketching the contours on a multi-touch table, and extruding the shape above the table with pinch gesture. Fuge et al. [4] proposed a modeling system with multiple shape representations. Hands were used to create and shape a point cloud, from which a smooth surface was finally rendered.

In *sculpting or claying systems*, where shapes are treated as plastic or clay-like materials, hands were used as a simple tool to add or remove materials or deform the shapes. Sato et al. [5] presented a claying system, where Compactly Supported Radial Basis Functions were centered on 10 fingertips to deform shapes through warping the space around fingertips. In another system presented in [6], mesh deformation was achieved through hand interaction with a 3D voxel space. Each voxel stored an integer number, which would be increased when the user's hands passed through it. The integer was used for calculating a force to pull the vertex within its effective range. Application Sculpting [7], published in the Leap Motion's app store, allowed the users to interact with the clay-like shape using two spheres controlled by two index fingertips. Chang et al. [8] presented a virtual clay system AiRSculpt based on a voxel sculpting model. It was accompanied by a gestural interface to implement system functionalities. A geometric interaction technique for bare-hand mid-air virtual pottery was presented in [9] where pottery shaping was modeled as a gradual convergence of pot's profile to the user's hand shape. This method was further investigated in [10], where the deformation process in a broad class of operations, such as pulling, pushing and fairing, could be well organized into effectively harnessed shape modeling without the need for a fixed set of gestures.

There were attempts to incorporate free-hand interaction into traditional *CAD modeling systems* for rapid creation of 3D conceptual shapes. Kang et al. developed a gesture language named HG3D for the CAD solid modeling system [11]. The gestural language controlled both 3D object operations and workspace operations. Another solid modeling system AIR-MODELLING introduced gestural interface combined with augmented reality technologies into a traditional CAD system [12]. Here, hand interface resulted in a significant reduction of shape modeling time, and also allowed the users to quickly conceptualize potential solutions.

With reference to the previous research, most systems relied on using hands to achieve various unrealistic shape modeling operations, resulting unnatural interaction interfaces which were hard to learn and use. The sculpting and claying systems are intrinsically natural and comparatively easy to be used, however they simply treated hands as 3D cursors, neglecting various dexterous uses of hand for deformation and modeling intentions reflected from hand motion. Although various natural mid-air gestures for virtual interaction and shape modeling were classified and implemented in [13]. It did not give proper hints and visual feedbacks to users eliciting their behaviors, which introduced a supervised training period for using the system. This gap motivated us to investigate people's natural real-life interaction modes, in order to make shape modeling more intuitive and easier to learn.

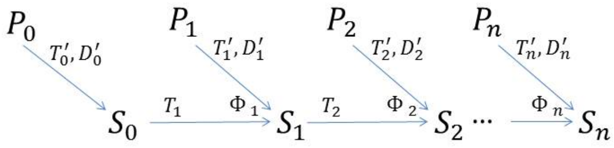


Figure 1. The shape modeling paradigm. The initial shape S_0, created from a primitive shape P_0, is procedurally modified by a newly incorporated primitive though operation Φ_i, until it evolves into the final shape S_i. T'_i and D'_i are the rigid transformation and deformation applied to primitive P_i, and T_i is the rigid transformation performed on the intermedia shape S_i.

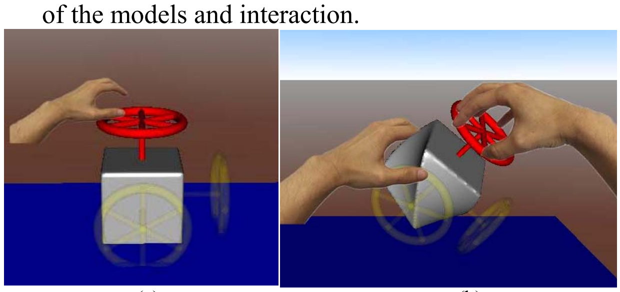
III. SHAPE MODELING FRAMEWORK

Many complex shape models can be seen as being built up from a large number of single simple components. Making such a model involves manufacturing each component and assembling them together procedurally. For example, architecture modeling deals with making isolated building/structure elements, made of cardboard, wooden blocks, foam or other materials, and by combining them into the final building model. Industrial assembly involves designing and manufacturing of every single mechanical component and then puts them together manually or through an assembly line.

We model this process through a paradigm, where the final shape S_n is constructed from an initial primitive P_0 through a sequence of modification operations. In each operation, a new primitive P_i is generated, which can be viewed as a modification tool. In order to make variations of the shapes available, we allow a tool to be deformed after its generation. In the next step, both the tool and the intermediate shape (called a main shape or a work piece) are transformed into the desired position and pose. Finally, the intermediate shape is modified by the tool through union (to simulate deposit of material), or subtraction (to simulate removal of material). The process continues until a final shape is created. The modeling process is illustrated in Fig. 1.

In the next sections, we consider designing:

- a natural deformation interface to deform shapes using hands;
- a manipulation interface allowing hands motor skills to be adopted for manipulation and assembling shapes;
- a rendering pipeline for construction and visualization of the models and interaction.



(a) (b)

Figure 2. The Different interaction modes with a handle: (a) when either hand approaches, the handle is in its highlighted mode; (b) The handle is in the movement mode, when the hand is pinched in its proximity. The second hand can be used as well to manipulate the object at any time.

IV. HAND INTERACTION INTERFACE

A. The steering wheel metaphor

We use a “steering wheel” metaphor as a visual clue giving the users hints on how to interact with a shape. The metaphor consists of three handles, which are fixed along the three principle axes of the local coordinate system of the shape. When the shape is being manipulated, the handles will move along with it. Interaction with the metaphor is done through interacting with the three wheels set at the end of each handle. The wheels can be stretched/pushed, rotated, and squeezed separately in the deformation modes of scaling, twisting and tapering. Deformation of the wheel will result in the corresponding deformation of the shape along the respective axis. Axis-based rotation and translation are also implemented by rotating and moving the corresponding wheel. This metaphor allows the users to explore and discover the corresponding deformation by themselves simply by touching and moving the handles with their hands. No specific gestures have to be then memorized.

There are three modes of each handle: deactivation mode, highlighted mode, and movement mode. A handle is semi-transparent in the deactivation mode, when no hands are located in its proximity. It becomes visible and opaque in the highlighted mode, when either hand approaches it, as it is show in Fig. 2(a). After the hand in the proximity of the handle is pinching, the handle bar goes to its movement mode and moves along with the hand. The pinching hand becomes invisible in this mode in order to eliminate the disparities between the hand motion and the handle motion. The flowchart diagram in Fig. 3 illustrates how the handle along X-axis reacts with hand interaction.

Pinching defines the start and the end of an interaction period. A natural way to hold the round wheel is by moving an open hand over the wheel while wrapping the fingers around it to grab it firmly. This action is detected by measuring the closeness of five fingertips. Specifically, we calculate the geometric center of the five fingertips and the average distances between the center and the tips. When the average distance is smaller than a threshold d_t, the pinching event is triggered, as calculated in (1):

$$d = (\sum_{i=1}^{5} dis(P_i, \bar{P}))/5, \quad (1)$$

where $\bar{P} = (\sum_{i=1}^{5} P_i)/5$ is the geometric center of the fingertips, and P_i, $i = 1, 2,\ldots, 5$ corresponds to the tip positions of the thumb, index, middle, ring and pinky fingers. This method can recognize the users’ pinching intentions, and it is robust on hand pose. Regardless of the hand orientation or its slight motion away from the handle, this pinching action can be tracked at any time.

Interaction with the handle is done by using one hand. The other hand can be used alternatively or for moving the object into a comfortable position, as it can be seen in Fig. 2(b), just as we hold the object using one hand while exerting force on it using another.

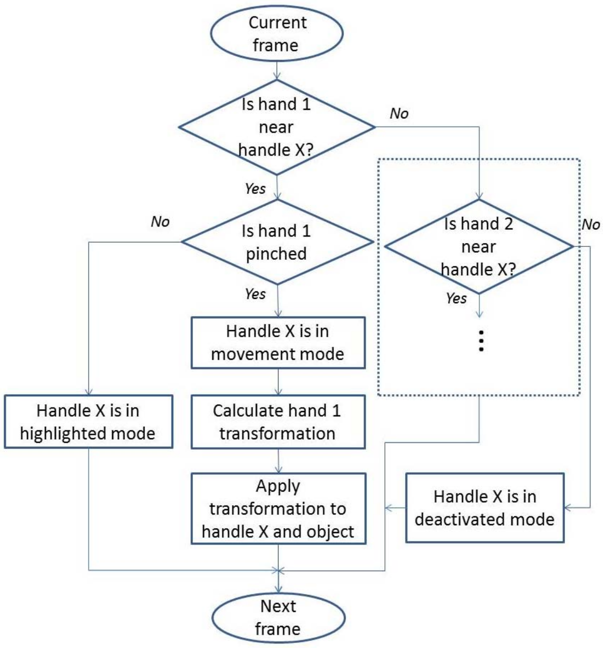


Figure 3. Flowchart of hand interaction with the X-axis handle.

B. Deformation interface

Here, we concentrate on designing a natural interaction interface for deforming objects. The operations include directional scaling (stretching and compressing), uniform scaling (expanding and squeezing), twisting and tapering. Mathematic definitions of these operations can be found in [14]. The challenge is to model and define these operations from complex motion of hand skeleton model, mimicking natural interaction behaviors. Considering that different individuals have different preferred “natural” ways of interaction, we propose interaction techniques based on the observation and understanding of common behaviors and mental patterns, which were investigated in our previous works in [15].

1) Directional scaling

This operation scales an object in a specific direction. In real life, it is decomposed into two operations and described as stretching (scaling up) and compressing (scaling down). Two hands are used to grab two opposite ends of the object while moving apart or together to exert a directional force.

Mimicking it, scaling along a principle axis requires the user to grab the corresponding interaction wheel using one hand (imagine the object is fixed at its center), and to move the hand towards or away from the object center. Translation amount of the wheel is decided by the translation of the hand projected onto the principle axis. Considering for example Y-axis, suppose the translation of the wheel along Y-axis is Δd, and the original shape is in the range of $[y_{min}, y_{max}]$. The scaling factor α is then

$$\alpha = (y_{max} + \Delta d)/\, y_{max}, \quad (2)$$

where $\alpha > 0$. The corresponding geometrical variables are labeled in Fig. 4(a). The deformation is then defined mathematically as

$$\begin{aligned} X &= x \\ Y &= \alpha y \\ Z &= z, \end{aligned} \tag{3}$$

where (*x, y, z*) are coordinates of a point on the shape, and (X, Y, Z) are coordinates of a new point after the deformation.

2) Uniform scaling

The respective deformations in real life are squeezing (scaling down) and enlarging (scaling up). Squeezing involves using hands to exert forces from different directions to make the object shrink uniformly. Although enlarging is not commonly done, unless the object is able to inflate, it can be generalized as the process inverse to squeezing. Similarly to directional scaling, uniform scaling is performed in the same way: any of the three handles can be grabbed, and used to deform the object, as it is shown in Fig. 4(b). This deformation is defined as

$$\begin{aligned} X &= \alpha x \\ Y &= \alpha y \\ Z &= \alpha z, \end{aligned} \tag{4}$$

where α is calculated in the same way as in (3).

3) Twisting

Twisting involves exerting a torque force through rotating fingers, wrist or the whole arm, depending on the force required. We achieve twisting along an axis through rotating the corresponding wheel. Taking *Y*-axis, for example, twisting angle $\bar{\theta}$ is determined by the average rotation angle of the fingers projected onto the *XZ* plane, as it is shown in Fig. 4(c). The method is robust to extract hand rotation features by projecting kinds of hand rotation into two directions. The deformation is then defined by

$$\begin{aligned} \theta &= \frac{y - y_{min}}{y_{max} - y_{min}} \bar{\theta} \\ X &= x\cos(\theta) - z\sin(\theta) \\ Y &= y \\ Z &= x\sin(\theta) + z\cos(\theta), \end{aligned} \tag{5}$$

where y_{max} and y_{min} are the extreme *y* values of the shape.

4) Tapering

Tapering involves using a hand to exert a shrinking force to one end of the object, while leaving unchanged the other end. Tapering is performed by further squeezing the wheel after pinching. After the hand is open, the wheel will restore back to its initial size, but the tapered shape cannot restore again. The decreased value of radius Δr is recorded (as it is shown in Fig. 4(d)), and accumulated to the scaling factor $\tilde{\beta} \in [\tilde{\beta}_{min}, 1]$. Taking *Y*-axis, for example, the deformation of tapering the upper end is then defined by

$$\begin{aligned} \beta &= \frac{y - y_{min}}{y_{max} - y_{min}} \tilde{\beta} \\ X &= \beta x \\ Y &= y \\ Z &= \beta z \end{aligned} \tag{6}$$

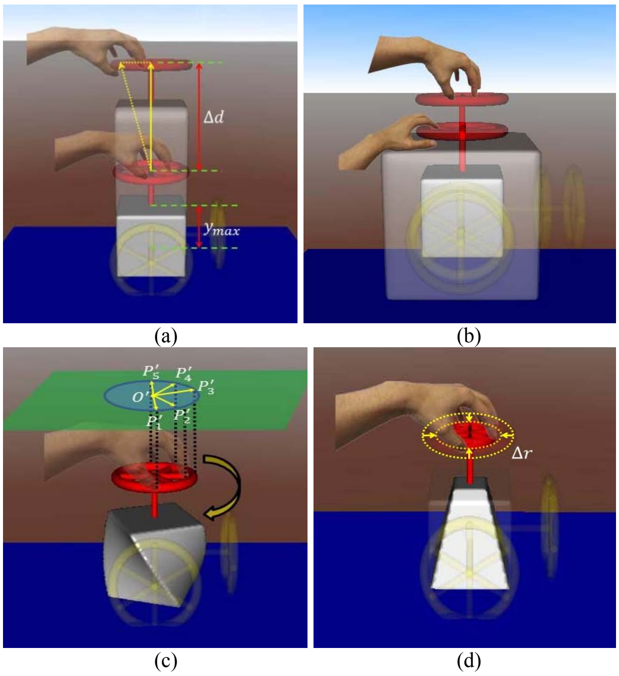


Figure 4. Illustration of different unimanual deformation operations: (a) directional scaling; (b) uniform scaling; (c) twisting; (d) tapering.

5) Combination, undo and reset

The proposed deformation operations have to be performed in a sequence, so that the final shape is resulted from combined deformations of the initial shape. Suppose P_0 is an arbitrary point on the initial shape, and D_i defines the i^{th} deformation operation, which can be any one from (3)-(6). After *n* times of transformation, P_0 would become

$$P_n = D_n D_{n-1} \ldots D_2 D_1 P_0 \tag{7}$$

where $D_1 \ldots D_n$ are stored in a stack. Starting a new deformation operation involves pushing the current one into the stack and putting the new operation on top of it. UNDO is done by discarding the operation on top and popping the stack. RESET is done through clearing the stack. The implementation of this method is illustration in Section V.

C. Manipulation interface

After deformation, the tool needs to be set to the desired position and orientation to be finally assembled into the main shape. In real-life scenarios, this process consists of a coarse but fast manipulation done first for approximate placing of the object, followed by a fine adjustment done by tiny hand or finger movements. Here, we incorporate into the interface these observed natural manipulation paradigms.

1) Grabbing

Grabbing indicates the beginning of manipulation. Different shapes, weights, sizes, and manipulation intentions will result in different grabbing gestures, which were classified into 17 types in [16]. Natural grabbing relies on closing fingers towards the thumb with the object in-between to exert a pair of opposite forces. One or more fingers could be used, depending on the size and weight of the object. We model this gesture by putting five sensing points on the five fingertips of each hand, and the grabbing status is decided by the status of collision of the sensing points with the object. The object is grabbed by one hand when the thumb tip and any of the other four fingertips collide with the object. It is released when either thumb or all the other four fingers move away from the object. The collision algorithm between a moving point and the object is discussed in Section V.

2) Free Manipulation

There are always one (the main shape) object or two (the main shape and the tool) present in the scene. In the free manipulation mode, either of the two objects can be grabbed by either hand. However, two objects cannot be simultaneously grabbed by one hand, in which case the tool has the priority to be grabbed first. The two objects can be grabbed by two hands, both separately and simultaneously, to be manipulated and put together, just as how we assemble components bimanually in real life. When the object is grabbed by two hands, bimanual interaction mode is activated, which is the case when the object is so big or heavy that we need two hands to take and manipulate it in real life.

When the object is grabbed by one hand, it is fixed in the hand local coordinate system and then it moves along with hand motion. Suppose the transformation of the object is W_{object} and the transformation of the hand is W_{hand} . After the object is grabbed, its relative transformation to hand T_{rel} is fixed. The transformation of object W_{object} is decided by transformation of the hand $\mathrm{W_{hand}}$ and the fixed relative transformation T_{rel}:

$$W_{object} = T_{rel} W_{hand}, \tag{8}$$

where

$$T_{rel} = W^0_{object} (W^0_{hand})^{-1}. \tag{9}$$

W^0_{object} and W^0_{hand} are the transformation of the object and shape on grabbing, as it is illustrated in Fig. 5(a).

When the object is grabbed by both hands simultaneously, bimanual manipulation mode is activated, where transformation is decided by the centers of two palms. Suppose the center positions of the two hands in current frame are $\overrightarrow{C_1}$ and $\overrightarrow{C_2}$, and in the previous frame they are $\overrightarrow{C_1'}$ and $\overrightarrow{C_2'}$. Then, the transformation of the object is:

$$\vec{T} = (\overrightarrow{C_1} + \overrightarrow{C_2} - \overrightarrow{C_1'} - \overrightarrow{C_2'})/2 \tag{10}$$

and the rotation is

$$R = Rotation(\vec{n},\ \theta), \tag{11}$$

where $\vec{n}$ is the rotation axis and θ is the rotation angle.

$$\vec{n} = \overrightarrow{C_1C_2} \times \overrightarrow{C_1'C_2'} \,/\, \left\| \overrightarrow{C_1C_2} \times \overrightarrow{C_1'C_2'} \right\|$$
$$\theta = \langle \overrightarrow{C_1C_2}, \overrightarrow{C_1'C_2'} \rangle \tag{12}$$

The involved vectors are labeled in Fig. 5(b).

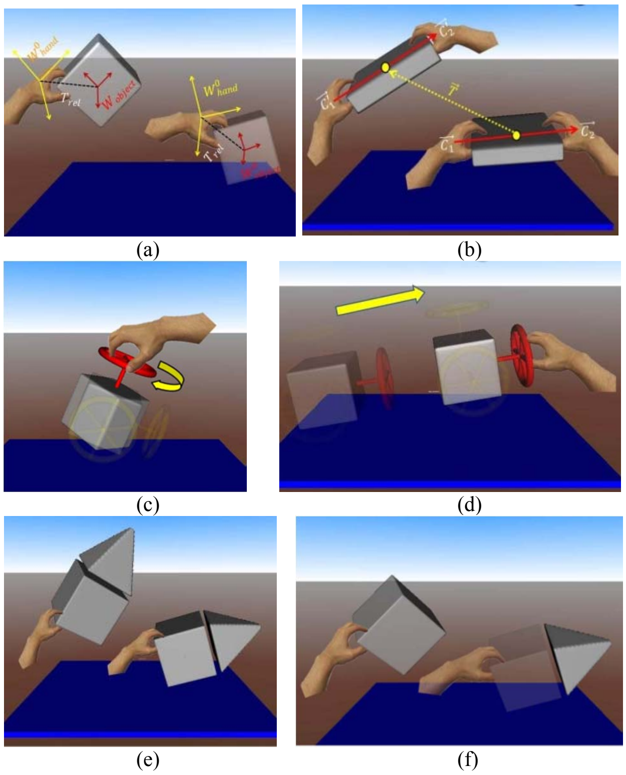


Figure 5. Different interaction modes: (a) unimanual manipulation; (b) bimanual symmetric manipulation; (c) axis-fixed rotation; (d) axis-fixed translation; (e),(f): free mode vs. joint movement mode.

3) Constrained manipulation

In the free manipulation mode, too many degrees of freedom of the object must be controlled simultaneously (6 DOF for uni-manual and 5 DOF for bi-manual). They are hard to be used for manipulating objects precisely in the mid-air virtual environment without constraints imposed and force feedbacks generated. Therefore, we provide another four manipulation modes: rotation (3 DOF); movement (3 DOF); rotation along an axis (1 DOF) and translation along an axis (1 DOF) for fine-tuning of the object.

In the rotation mode, the object can only be rotated but not translated, while in the movement mode its manipulation is constrained into translation merely. The “steering wheel” appears in the axis-constrained mode. Rotation along an axis requires rotation of the corresponding handle (in the same way as twisting), while moving the object along an axis requires moving the handle (in the same way as scaling), as described in Fig. 5(c) and (d).

4) Joint motion mode

Aligning objects involves constant switching between adjusting the pose of the objects and manipulating them to see whether they are aligned from another direction. Since two objects move separately, moving the main shape will result in the already assembled tool falling off, as it is shown in Fig. 5(f). Therefore, we provide a joint motion mode, which is very helpful for alignment. This mode is the similar to free manipulation mode, except when the main shape is grabbed and manipulated while the tool is not grabbed, the tool will move along and be fixed in the hand local coordinate system. The comparison between the two modes is shown in Figs. 5(e), (f).

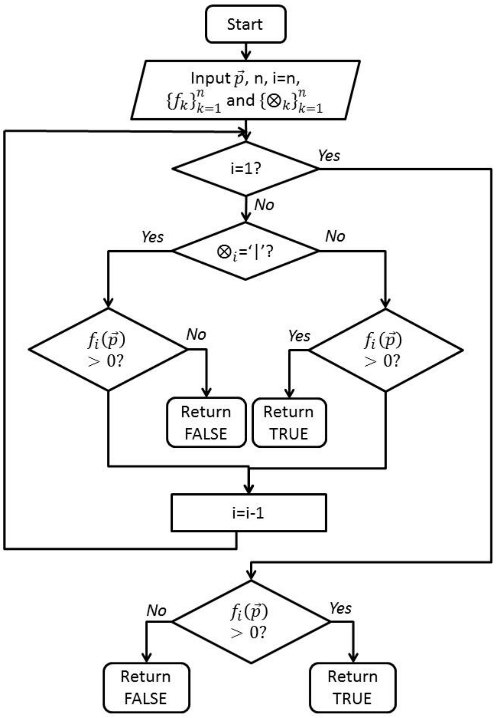


Figure 6. Flow chart of collision algorithm between a point and the main shape.

V. PROCEDURAL MODELLING DETAILS

We define shapes using function-based representations, where a geometric object is defined as a closed subset of the 3D space $f(x,y,z) \geq 0$. Applying a set of deformation operations $D_1 \dots D_n$, as defined in (7), on the shape $f_0(x,y,z)$ will result in a new shape defined by

$$f_d(x,y,z) = f_0(D_1^{-1}D_2^{-1} \dots D_n^{-1}(x,y,z)) \geq 0 \quad (13)$$

The scene deals with two objects – the main shape $f_{shape}(x,y,z) \geq 0$ and the tool $f_{tool}(x,y,z) \geq 0$. The tool is generated from a primitive after a sequence of deformations, and the main shape is a result of applications of the tools. Union '|' and subtraction '&-' are defined as

$$f_{union} = f_{shape} | f_{tool} = max(f_{shape}, f_{tool}) \geq 0 \quad (14)$$

$$f_{subt} = f_{shape} \& - f_{tool} = min(f_{shape}, -f_{tool}) \geq 0 \quad (15)$$

Suppose a primitive is defined as $f_0(x,y,z) \geq 0$. After deformation D (one or a combination of deformation), rotation R and translation T, the shape becomes $f_1(x,y,z) \geq 0$. Collision between a point $\vec{p} = (x_0, y_0, z_0)$ and $f_1(x,y,z) \geq 0$ is detected only when

$$f_1(x_0,y_0,z_0) = f_0\left(D^{-1}R^{-1}T^{-1}(x_0,y_0,z_0)\right) \geq 0 \quad (16)$$

Suppose f_{shape} is defined as

$$f_{shape} = f_1 \otimes_1 f_2 \otimes_2 f_3 \otimes_3 \dots \otimes_{n-1} f_n, \quad (17)$$

where $\otimes_i$ is a binary operator, which can be either union '|' or subtraction '&-'. Detecting collision between $\vec{p}$ and the main shape f_{shape} requires detecting the collision between $\vec{p}$ and all components consisting of f_{shape}. The algorithm is illustrated in Fig. 6.

VI. IMPLEMENTATION DETAILS

A. *Implementation platform*

The codes were written in VRML and java-script (vrml-scripts), and visualized using MS Internet Explorer with the BS Contact plugin. For shape modeling we used function-based extension FVRML/FX3D [17, 18] which allows the function definitions $f(x,y,z) \geq 0$ to be written as VRML scripts and embedded directly into the VRML/X3D codes. It also allows for an efficient model exchange in shared virtual worlds.

B. *Hand tracking and visualization*

We used Leap Motion controller [19] for hand tracking. Hand skeleton data, e.g., hand joint position, palm position, normal and direction, was obtained from the Leap Motion SDK version 2.3.1. The data was made available in VRML using a plugin which we previously developed [20]. The virtual hands were constructed from the hand skeleton data using balls (joints) and cylinders (bones).

C. *Menu and procedure*

The modeling menu consists of two levels. There are five major functionalities from the first level, and each item consists of 7 sub-items in the second level, as shown in Fig. 7.

At the beginning of a modeling session, a single primitive has to be selected from the PRIMITIVE menu. After deformation and manipulation, it will be deposited as the main shape. Next, a primitive is generated each time and combined with the main shape using union or subtraction. The menu of COLORS is used to assign a color to the primitive. Various kinds of deformation operations can be selected from the DEFORMATION menu (as described in Section IV A), where d-scale is short for directional-scaling, and u-scale is for uniform-scaling. The manipulation operations, as discussed in Section IV B, can be selected from the MANIPULATION menu. The functionalities of Union, subtraction, increasing or decreasing resolution, and adjusting bounding box are available in the OPERATION menu. Refer to the video illustrating the functionalities which can also be found at *https://youtu.be/dKoVmuPjXH0*.

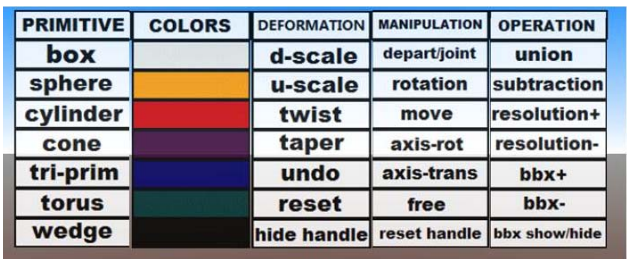
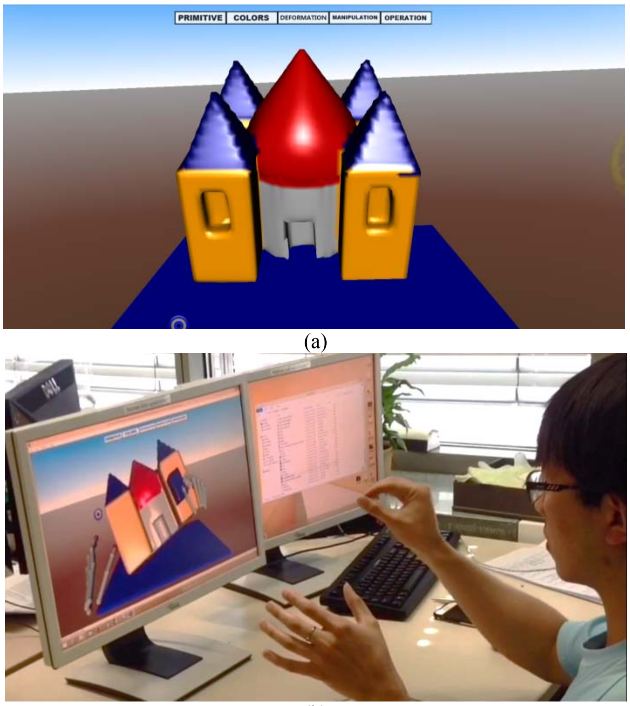


Figure 7. The menu consists of two levels. After clicking at a first-level menu item, the corresponding second level of menu will roll down.



(a)

(b)

Figure 8. User study task and set up: (a) the shape model for participants to build; (b) the apparatus set-up.

TABLE I. USER STUDY EVALUATION DATA

Question	Q1	Q2	Q3	Q4	Q5
Median	5.5	5.5	6.0	2.0	5.0
Deviation(MAD)	0.5	1.0	0.0	1.0	1

VII. USER STUDY AND ANALYSIS

We hypothesized that if our system had a natural and intuitive interface the users can learn how to interact by themselves easily with least mental loads. We also hypothesized that the users can properly adopt constrained techniques for precise deformation and manipulation tasks. Here we describe the user study tests conducted to prove our hypothesis.

A. Procedure

At the beginning of the experiment, a participant was shown the final model he/she was required to made, as it is shown in Fig. 8 (a). The model consisted of 13 primitives of three kinds: sphere, cube and cylinder. The process involved scaling and tapering (for the rooftops) of the primitives and assembling them one by one together. However, no instructions or manuals were given to them on how to do it–they had to explore the modeling functionalities by themselves. The help could only be provided when a participant needed explanation of terminology like 'bbx+' (Fig. 7). The participants were also allowed to make the major parts of the model, if they felt too tedious and tiring the build the whole one. After the model was made and the modeling functionalities were fully explored, the participants were required to evaluate their attitude towards the following five statements:

Q1: The interaction methods are easy to learn and understand.
Q2: It is natural to interact in these ways for deforming and manipulating objects.
Q3: It is easy for me to remember the interaction
Q4: The system is easy to control.
Q5: I like the system.

For each statement, the participants were required to give a score based on a 7 point Likert scale reflecting to which extent they agreed with the statements, where 1 corresponded to strongly disagree and 7 corresponded to strongly agree. They were also required to explain the reasons for their scores. A discussion followed at the end of the experiment. The discussion was centered on but not limited to advantages and disadvantages of the system; difficulties to use the system; comparison of the system with the current shape modeling tools and suggestions on how to improve the system.

8 participants were invited to do the user study. Four of them used commercial shape modeling systems before. None of them had much experience with mid-air interaction. Average age of the participants is 25.6 (*sd*=2.8). During the experiment, the participants were seated in a chair in front of the computer screen, and the controller was placed on the desktop facing up, as shown in Fig. 8 (b). The experimenter sat next to them and took notes of their interaction behavior. The interaction part lasted 50-60 minutes, including the time spent on getting familiar with the device, exploration of the functionalities, designing their modeling strategies and building the model. The discussion part took 20 minutes on average.

B. Results

The system has received good feedbacks on its learnability, naturalness and mental loads, as reflected in Table 1. The observations are described in the following four aspects:

Learnability: the system is relatively simple and easy to be learned. All participants could discover interaction techniques by themselves. For various deformation operations, they tried their hands on the handles and found out the interaction ways very quickly. The free and constrained manipulation methods are straightforward, so that the participants did not spend much time to understand them. However, the function of "depart/joint" was merely explored, because its usage is rather case-depended.

Naturalness: most participants admitted they could interact in a similar way as how they did it in real life, e.g., as commented "*these interaction ways seem what they are meant to be and feels pretty natural and intuitive to me.*" Negative feedbacks were received from 3 participants centering on the handles, e.g., "*the virtual handles are intuitive but not very natural, since there are no handles in realistic cases. But it seems to be the best way I can come up with to implement these operations*."

Mental Load: we received a unanimous positive feedback on the aspect of mental load. The designed free hand interaction techniques exerted a least mental load on users, as commented by one of the participants "*it took me some time to figure out the functions. But after I learned how to interact with hands, I can remember them for a long time and be confident to make shape models*".

Controllability: controllability was largely affected by the insufficient tracking accuracy using Leap Motion controller – participants relied on small finger motion during interaction, which could not be sensitively detected by the device. The lack of disparity estimation added to the difficulty of predicting the collision between the hand and the objects. An interesting phenomenon was observed that the users were very reluctant to use the provided constrained manipulation methods to increase precision and stability. Two participants explained: "*it is interesting to try on this free manipulation way with my hands just like how I do it in real life. It is hard to control but I could improve during the process of practicing. However I did not consider much about the constrained ways, since it limited my interaction freedom largely*"; "*I did not try the constrained ways, because the direct grabbing way could fulfill how I wanted to interact and manipulate. I was also trying to find hints helping me align the objects precisely*." The fact that the users had no prior experience–a prerequisite for Q1-Q3–has therefore influenced the evaluation of Q4. They were indulging more on finding out tracking features of the device, rather than planning an effective interaction strategy. It would be interesting to see how more experienced users would tend to use the provided constrained manipulation methods in order to perform the tasks faster.

VIII. CONCLUSION

We designed a web-enable shape modeling system with natural free hand interaction, to be learned and acquired quickly by the users, while imposing least mental loads. We explored natural ways of how people stretch, compress, squeeze, enlarge, twist and taper shapes, as well as how they grab and manipulate objects with one hand or two in real life, and modeled those behaviors to be used in a shape modeling system. Intuitive constrained interaction techniques were provided to increase the controllability of interaction. A shape modeling paradigm, accommodating the proposed natural interaction interface, was designed. It leaves users to concentrate on the interaction without knowing the complexity beneath the interface. The user study proved the naturalness of the modeling system, and its easiness to be learned and remembered.

IX. FUTURE WORK

We will come up with more intuitive virtual metaphors and interaction techniques for people to learn and understand shape modeling, based on the understanding of their mental models of how they plan to interact with shape. In order to check the controllability, we intend to perform a follow-up experiment with the same and new users performing the same task, encouraging the same users to use the provided constrained manipulation methods to increase precision and stability and measure the time difference and their impression of the controllability.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centers in Singapore Funding Initiative, joint PhD Degree Program NTU-TU Darmstadt, and MOE Singapore Funding RG17/15 "Haptic Interaction With Images And Videos".

REFERENCES

[1] S. Schkolne, M. Pruett, and P. Schröder. "Surface drawing: creating organic 3D shapes with the hand and tangible tools" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01). ACM, New York, NY, USA, pp. 261-268, 2001.

[2] S. Murugappan, H. Liu, and K. Ramani, " Shape-It-Up: Hand gesture based creative expression of 3D shapes using intelligent generalized cylinders," *Computer Aided Design.* 45, pp. 277-287, February 2013.

[3] R. Bruno, A. De, C. Géry, and J. A. Joaquim, "Mockup builder: direct 3D modeling on and above the surface in a continuous interaction space," In Proceedings of Graphics Interface (GI '12), Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 173-180. 2012

[4] M. Fuge, E.M. Yumer, G. Orbay, and L.B. Kara, "Conceptual design and modification of freeform surfaces using dual shape representations in augmented reality environments," Computer Aided Design. 44, 10, pp. 1020-1032, October 2012.

[5] M. Sato, V. Savchenko and R. Ohbuchi, "3D freeform design: interactive shape deformations by the use of CyberGlove," Cyberworlds, 2004 International Conference on, pp. 147-154, 2004.

[6] K.G. Kobayashi, et al., "3D Shape Modification based on Gesture Input," International Journal of CAD/CAM, 13(2), 2013

[7] *Sculpting*. Available from: https://apps.leapmotion.com/apps/sculpting/windows.

[8] S.-A Jang, et al., "AiRSculpt: A Wearable Augmented Reality 3D Sculpting System," in Distributed, Ambient, and Pervasive Interactions. Springer. p. 130-141, 2014.

[9] K. Ramani, "A gesture-free geometric approach for mid-air expression of design intent in 3D virtual pottery," Computer-Aided Design, 69: pp. 11-24, 2015.

[10] R.K. Vinayak, "Extracting hand grasp and motion for intent expression in mid-air shape deformation: A concrete and iterative exploration through a virtual pottery application," Computers & Graphics, Volume 55, pp.143-156, April 2016.

[11] J. Kang, et al., "Instant 3D design concept generation and visualization by real-time hand gesture recognition," Computers in Industry64(7): pp. 785-797, 2013.

[12] S. Arroyave-Tobón, G. Osorio-Gómez, and J.F. Cardona-McCormick, "Air-modelling: a tool for gesture-based solid modelling in context during early design stages in AR environments," Computers in Industry, 66, pp. 73-81. 2015.

[13] J. Cui, D.W. Fellner, A. Kuijper, and A. Sourin, "Mid-Air Gestures for Virtual Modeling with Leap Motion," in Distributed, Ambient and Pervasive Interactions: 4th International Conference, DAPI 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings, Springer International Publishing: Cham. pp. 221-230, 2016

[14] H. Barr. Alan, "Global and local deformations of solid primitives," SIGGRAPH Comput. Graph. 18(3): pp. 21-30, 1984.

[15] J. Cui, A. Kuijper, D.W. Fellner, and A. Sourin, "Understanding People's Mental Models of Mid-Air Interaction for Virtual Assembly and Shape Modeling," In Proceedings of the 29th International Conference on Computer Animation and Social Agents (CASA '16), ACM, New York, NY, USA, pp. 139-146, 2016.

[16] T. Feix, et al, "A comprehensive grasp taxonomy. in Robotics, Science and Systems", Workshop on Understanding the Human Hand for Advancing Robotic Manipulation, 2009.

[17] Q. Liu, and A. Sourin, "Function-based shape modelling extension of the Virtual Reality Modelling Language," Computers & Graphics, **30**(4): pp. 629-645, 2006.

[18] *FVRML/FX3D*. Available from: http://www3.ntu.edu.sg/home/assourin/FVRML.htm.

[19] Leap Motion. Available from: https://www.leapmotion.com/.

[20] J. Cui and A. Sourin, "Feasibility Study on Free Hand Geometric Modelling Using Leap Motion in VRML/X3D," Cyberworlds (CW), 2014 International Conference on, Santander, pp. 389-392, 2014.