Hand Grasp and Motion for Intent Expression in Mid-Air Virtual Pottery

Vinayak *   Karthik Ramani †‡

School of Mechanical Engineering
Purdue University

ABSTRACT

We describe the design and evaluation of a geometric interaction technique for bare-hand mid-air virtual pottery. We model the shaping of a pot as a gradual and progressive convergence of the pot-profile to the shape of the user's hand represented as a point-cloud (PCL). Our pottery-inspired application served as a platform for systematically revealing how users use their hands to express the intent of deformation during a pot shaping process. Through our approach, we address two specific problems: (a) determining start and end of deformation without explicit clutching and declutching, and (b) identifying user's intent by characterizing grasp and motion of the hand on the pot. We evaluated our approach's performance in terms of intent classification, users' behavior, and users' perception of controllability. We found that the expressive capability of hand articulation can be effectively harnessed for controllable shaping by organizing the deformation process in broad classes of intended operations such as pulling, pushing and fairing. After minimal practice with the pottery application, users could *figure out* their own strategy for reaching, grasping and deforming the pot. Further, the use of PCL as mid-air input allows for using common physical objects as tools for pot deformation. Users particularly enjoyed this aspect of our method for shaping pots.

Index Terms: H.5.2. [Information Interfaces and Presentation]: User Interfaces—; I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1 INTRODUCTION

Mid-air gestures have been widely used to provide a *symbolic* way to express user's intent for 3D shape modeling [25, 32, 10, 29, 27]. Gesture-based interactions enable the user to focus on the design task rather than dedicating significant time towards learning the usage of the tool itself [13]. With the recent commercialization of depth cameras, gesture-based interactions have become accessible to the common user; creative applications for free-form shape modeling [1] in mid-air have gained significant popularity. The user input in these applications is represented as a combination of some special hand posture (such as pointing with a finger), and the motion of a representative point (such as the palm or finger-tip) on the hand.

Hand and finger movements in real-world shaping processes (such as pottery or clay sculpting) are complex, iterative, and gradual. Such processes are essentially governed by the physics and geometry of contact between the hand and clay. Thus, the true expressive potential of finger movements remains under-utilized despite advances in hand pose and skeletal estimation [14, 2]. In this paper, we seek a method that can determine user's intent directly from the shape the of the user's hand. What distinguishes our approach from existing works, is the representation of the hand as a point cloud (PCL). This representation essentially converts the problem of *interaction with geometry* to that of understanding the *geometry of interaction*.

In this paper, our goal is to enable the expression of design intent for shape deformation by determining the user's hand grasp and motion on a given shape. We achieve this goal by geometric characterization of contact made by the hand's PCL on the surface model of a shape. To this end, we design and evaluate an interaction technique that integrates the geometric information in user's actions with shaping operations for bare-hand mid-air virtual pottery. Our focus here is not to build a comprehensive and feature-rich 3D modeling system. Instead, we intend to investigate spatial interactions for 3D shape deformation with a raw representation of the hand. To this end, our pottery-inspired application serves as a platform for systematically revealing how users use their hands to express the intent of deformation during a pot shaping process.

1.1 Contributions

We make two contributions. First we demonstrate, with a practical implementation, that it is possible to achieve controllability in mid-air shape deformation using raw PCL data of the user's hand. We present a method that does require to compute any finite set of gestures or hand skeleton. Instead, it implicitly extracts the grasp and motion from the hand PCL for deforming the shape of a pot in 3D space. This feature of our method directly allows a user to shape pots by using physical artifacts as tools. Secondly, we evaluate our proposed method in terms of user performance, behavior and perception in pottery design. Our evaluations help reveal two core aspects of mid-air interactions for shape deformation, namely, intent & controllability. We demonstrate the engagement, utility, and ease of learning provided by our approach.

2 RELATED WORK

2.1 Mid-air Gestures

Gestures can be designed effectively for pointing, selection [23, 30], and navigation, since they define an unambiguous mapping between actions and response. Such tasks are implemented using deictic gestures [16] and can usually be segmented into discrete phases, with each phase triggering an *event* or a *command* [4]. Pointing in the direction of a virtual object creates the association between the user and the object. A recent study [31] shows dwell-time to be an effective method of pointing and selecting objects without hint to the users. In manipulative tasks such as ours, a direct spatial mapping is required between the user's input and the virtual object [21, 16]. Particularly in our case, such an association would be in terms of the proximity of the user's virtual hand to the shape being deformed.

2.2 Gestures for 3D Modeling

Consider a mid-air interaction scenario of selecting and displacing a mesh vertex for deforming the mesh. Since the user's hands are interacting in the air, there is no physical or natural mechanism for

*e-mail: fvinayak@purdue.edu
†e-mail: ramani@purdue.edu
‡School of Electrical and Computer Engineering (by courtesy)

triggering events. Here, gestures could serve two fundamental purposes. First, they help define a beginning (e.g. reaching and clutching some region of interest) and end (e.g. de-clutching the region after required deformation) of an interaction [4, 22]. Secondly, they help define the exact operation from a set of operations defined in the context of an application. For example, the type of deformation could be selected by using different gestures (e.g. fist to pull, point to push, open palm to flatten).

On these lines, most existing bare-hand interaction techniques for 3D shape conceptualization, use gestures combined with arm and full-body motions. Segen and Kumar [25] showed examples of computer-aided design (CAD) with their *Gesture VR* system, using computer vision for general virtual reality (VR) applications. Wang et al. [32] presented *6D Hands* to demonstrate CAD using marker-less hand tracking. Modeling of sweep surfaces using hand gestures and body motion was demonstrated by Vinayak et al. [29]. Han and Han [10], demonstrated an interesting surface-based approach with particular focus on audiovisual interfaces for creating 3D sound sculptures. Holz and Wilson proposed *Data miming* [13] as an approach towards descriptive shape modeling wherein voxel representation of a user's hand motion is used to deduce the shape which the user is describing. This approach uses hands without explicit determination of gestures for recognizing the user's description of an existing shape.

2.3 Hand Grasp

Prehension is a common phenomenon in real-world interactions. Jeannerod [15] notes two functional requirements of finger grip during the action of grasping, (a) adaptation of the grip to the size, shape and use of the object to be grasped, and (b) the coordination between the relative timing of the finger movements with hand transportation (i.e. whole hand movements). Intended actions strongly influence motion planning of hand and finger movements [3]. This suggests that the intent for deformation can be recognized before the user makes contact with the surface being deformed. Grasp classification [8] and patterns of usage and frequency [34] have been integral to robotics research. It is worth noting that the primary methodology for investigating grasp taxonomies is mostly derived from the geometry of the hand in relation to a physical object held or manipulated by the hand. What we aim to do is to understand what is the minimal and sufficient characterization of the user's hand and finger movements, that could be used for mid-air deformation. Our goal is not to explicitly detect the hand grasp, but to design a deformation approach where the grasp is automatically and implicitly taken into consideration during shape deformation.

2.4 Grasp Emulation Through Hardware

Literature in virtual reality [19, 5] has studied and implemented grasping in the context of object manipulation (mainly pick-and-place). Works by Sheng et al. [26], Kry et al. [17], and Pihuit et al. [20] have leveraged finer finger level movements and grasping for 3D shape deformation by designing ingenious hardware systems. In these systems, users can actually grasp and deform virtual objects. However, such systems are not accessible to the common user outside a lab environment. Further, wearing or holding can be intrusive to the user during a focused modeling task. In our work, we aim to determine user's intent from fine finger level movements while retaining the non-intrusiveness and accessibility of depth cameras.

3 Design Principles

3.1 Intent & Controllability

The general term *intent* is literally defined as "*the thing that you plan to do or achieve : an aim or purpose*". In our case, intent (*what one wants to achieve*) can be described in terms of the context of shape deformation (*what operations one can perform on the shape*). Based on Leyton's perceptual theory of shapes [18], Delamé et al. [9] proposed a process grammar for deformation by introducing structuring and posturing operators. Here, structuring operators involve adding/removing material to the shape, while posturing operators allow for modifications such as bending or twisting some portion of the shape. Since our context is that of deformation, we define the intent in terms of three basic operations: pulling, pushing, and smoothing. Here, pushing and pulling can be seen as counterparts to the structuring operators. Our experience with users strongly indicated that users frequently seek an explicit method that could allow them to intermittently refine a shape after reaching a certain desired shape. We added the smoothing operation as a means for the user to fair the shape after achieving rough version of a desired shape.

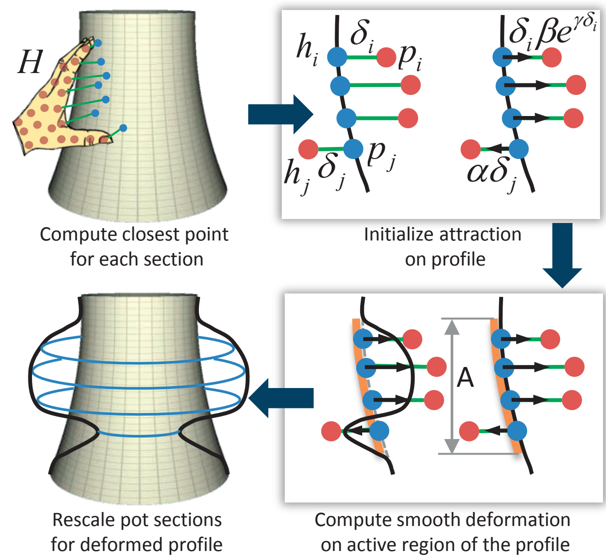

Figure 1: Algorithm for pot deformation is illustrated for proximal attraction. The profile is deformed based on the proximity of the points on a given hand PCL. Subsequently, each section is re-scaled to obtain the deformed pot surface.

We see controllability as the quality of intent recognition and disambiguation *as perceived by the user*. Specifically, in our context, controllability is defined as a function of two factors: (a) the disparity between what a user intends for the shape to be and what the shape actually becomes after the deformation, and (b) the responsiveness of the deformation. The goal is to minimize the disparity and optimize the responsiveness.

3.2 Rationale for Pottery

We have two goals in this paper. First, we seek a concrete geometric method that takes a general representation of the user's hand (PCL) and allows the user to deform 3D geometry. Secondly, we want to investigate this method in light of intent and controllability.

In a general shape deformation scenario, an arbitrary triangle mesh is the ideal and generic shape representation. However, a controlled study is prohibitively challenging in such a case, for two reasons. First, the hand PCL data obtained from a single depth sensor is partial and noisy. Secondly, dynamic and complex finger motions further add complexity to the occlusions and noise. Subsequently, designing interaction tasks for a quantitative evaluation is difficult, particularly for users with no prior experience with mid-air interactions for free-form 3D modeling. Hence, it is essential to constrain the geometric representation of the object being modified.

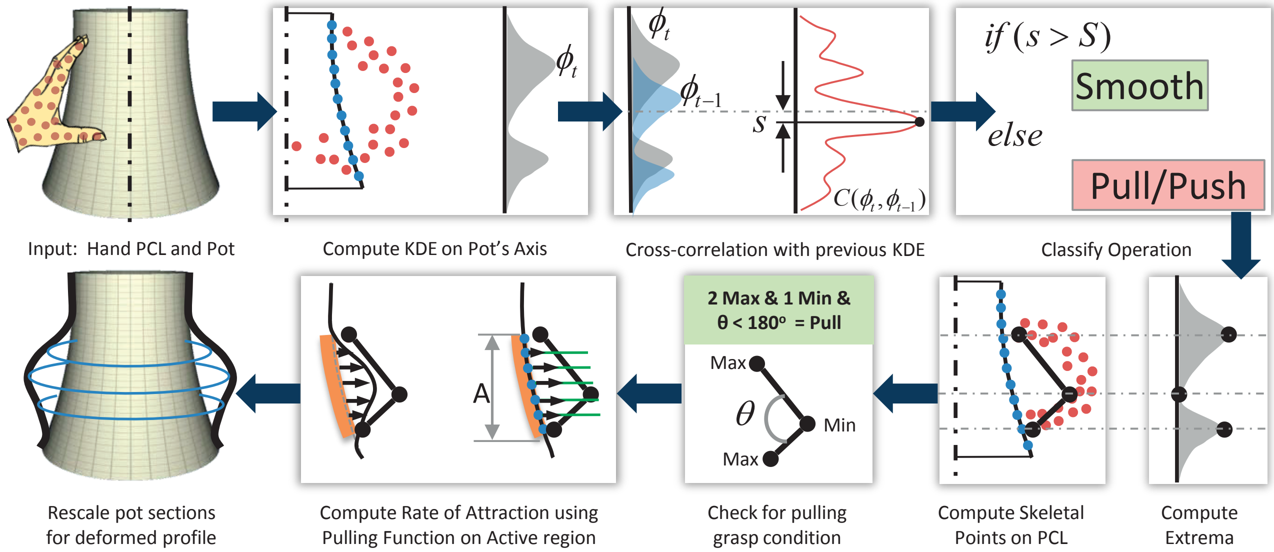

Figure 2: Algorithm for intent classification is illustrated. The main steps involve computation of axial KDE for hand PCL, detection of intent for smoothing, differentiation between pulling and pushing, and deformation of the pot. In this example, we show the details of the pulling deformation (row 2).

We use pottery as our application context for two reasons. First, it offers a well-defined and intuitive relationship between the use of hands and the shaping of pots to a user. This allows us to concretely construct a geometric relationship between the shape of the hand PCL and the corresponding user intent. Secondly, the simplicity of the geometric representation and deformation lends itself to quantitative measurement of the user's response to our system.

4 Algorithm

We represent a pot as a simple homogeneous generalized cylinder. The surface of the pot is defined as a vertical stack of circular sections. Each section is a polygonal approximation of a circle, i.e. a closed regular polygon. Note that a sequenced list of pairs (radius, height) is the *profile curve* of the pot. Given the profile, the surface mesh of the pot can be generated using a simple quad-mesh topology. The deformation of a pot is achieved by deforming the profile curve, i.e by modifying the radii of each section. For a 3D pot, this essentially corresponds re-scaling each section by the corresponding amount of deformation.

4.1 Problem & Approach

Given a 3D mesh of a pot and a PCL of the hand, our algorithm is required to address two problems. The first problem is that of determining when to begin and end a deformation process, i.e. to determine when the user wants to clutch and release the shape of the pot. The second problem is to determine the kind of deformation operation (e.g. pushing, pulling, smoothing) that the user wants to perform. This problem also implicitly involves the determination of the extent (e.g. local or global) of deformation.

For the first problem, our basic idea is to progressively conform the pot's shape to that of the user's hand. This idea, dubbed *proximal attraction* (Figure 1), is inspired by the notion of *dwell-time* used in 3D object selection [31]. Here, each point in the hand's PCL *attracts* a local region on the pot, hence deforming the pot without explicitly clutching or declutching the pot.

To address the second problem of intent classification, we first conducted a preliminary study to find out how users commonly reached, grasped and deform the pot. Here, proximal attraction

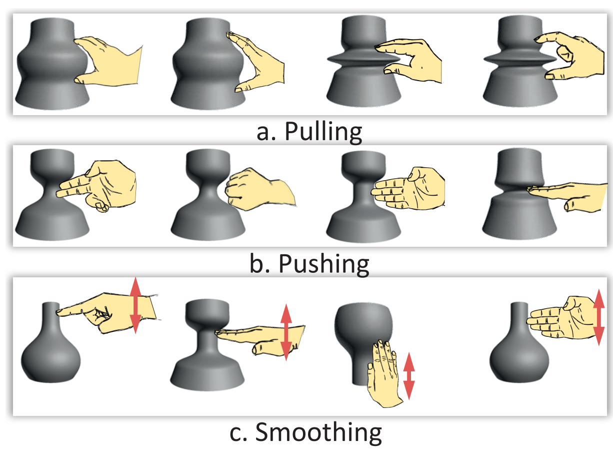

Figure 3: Common user patterns are shown in terms of grasp and motion performed by users for each target shape (in decreasing order of occurrence along columns). The hand images represent the grasp and the arrows (red) show the motion of the hand.

served as a naive approach for studying users' strategies and preferences. Subsequently, we implemented a method using kernel-density estimation to characterize the contact between the hand and the pot (Figure 2). This allowed us to classify the users' intent to push, pull or fair the surface of the pot depending on the hand grasp, finger movements, and motion of the hand on the pot's surface.

4.2 Deformation Without Clutching: Proximal Attraction

Consider the hand H as a set of points $\{h_i\}$ in 3D space and let p be the point on the pot that is closest to h. The main idea of proximal attraction is to deform the pot gradually by attracting p towards h in the *horizontal* plane. The condition of proximity is simply given as $\|h-p\|_2 < \varepsilon$, ε being a pre-defined threshold (say ε). A point $\{h\}$ in H locally deforms a small region on the pot if using the proximal

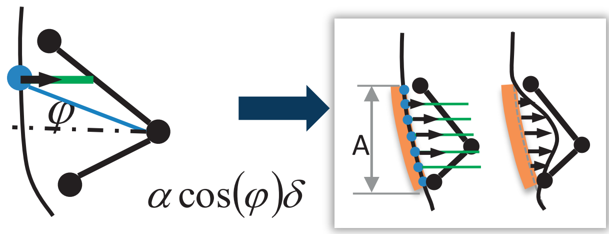

Figure 4: Computation of attraction rate using the angle of grasp.

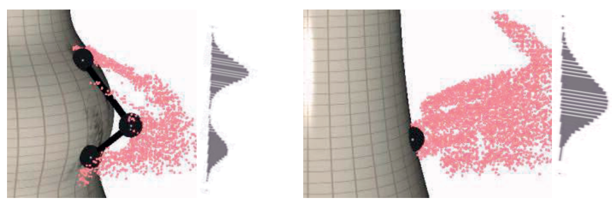

Figure 5: KDE functions are shown for a pulling (left) and pushing (right) intents.

attraction approach. On the whole this amounts to a gradual and progressive convergence of the pot-profile to the shape of the user's hands (Figure 1).

Pushing vs. Pulling: A push is characterized by an inward displacement ($\delta < 0$). This is the simplest case wherein a user would typically approach the pot and subsequently recede away once the desired deformation has occurred. A pull is characterized by an outward displacement ($\delta > 0$). This is a non-trivial intent to recognize since a user would invariably approach the surface first and then recede to pull. The overall motion of the hand is similar to that of a push. In order to distinguish pulling and pushing, we used two different rates of attraction. For pulling, we defined the attraction rate as a smooth function of the distance between the hand point and pot. The function is given by $\beta e^{\gamma \delta_i}$. For pushing, we defined the rate of attraction as α [1]. This essentially allows the user to first approach the pot without deforming it during the process of approach. The algorithm is as follows:

1. For each section i
   Compute j such that $\|h_j - p_i\| < \varepsilon$ is minimum.
   Set δ_i to horizontal distance between h_j and p_i
2. Set δ_r to $\delta_{max} - \delta_{min}$
3. Set γ to $\frac{1}{\delta_r}$
4. For each i on profile
   if($\delta_i < 0$): Set attraction at p_i to $\alpha\delta_i$
5. Compute Active region A
6. Smooth deformation ($\nabla^2\delta = 0$ for all points in A)
7. Compute deformed profile
8. Rescale pot sections

4.3 User Grasp & Motion Strategies

We conducted a lab experiment with 15 participants who were asked to create convex, concave and flat features on an initial seed pot. We used proximal attraction as a naive approach for this experiment. Generally, preferences towards grasping varied across users based on their expectation of the system and subsequent trial and error. However, we observed some common patterns for reaching, grasping and deforming each target shape. Users generally preferred small finger level movements for thin features. For fat and flat features, we observed that the users first formed a grasp according to the amount of deformation required and then moved the whole hand to achieve the feature as expected [3]. Most users spent time smoothing and refining the surface of the pot after the general shape had been obtained. The motion of the hand was performed vertically along the surface of the pot (Figure 3). This led to frustration due to the lack of an explicit way for the users to smooth or straighten a region of the pot.

There were two main issues with the proximal attraction approach. First, pulling was clearly more difficult since the rate of attraction was *designed* to be lower than that of pushing. Secondly, the users clearly distinguished between several operations of fairing, straightening, carving, pulling and pushing. However, the proximal attraction approach, was not designed to explicitly identify or classify the type of operation the user intended to perform. Our main goal was to resolve these two issues as described below.

4.4 Intent Classification: Grasp+Motion

The basic idea of the *grasp+motion* approach is to *summarize* the grasp of the hand *in relation to* the surface of the pot and subsequently classify the user's action (Figure 2). We achieve this by using kernel-density estimation of the point cloud on the axis of the pot. In our context, this kernel-density estimate (KDE) is essentially a smoothed histogram of the distribution of the hand's PCL on the pot's. We use the exponential function to determine the KDE. For a given section i, the KDE is given by:

$$\phi_{i,j} = \sum_{j=1}^{j=|H|} e^{a\|\delta_{i,j}^2\|} \tag{1}$$

There were three main observations that helped us use the KDE to classify the user's intent. First, users moved their hands in a fixed pose along the surface of the pot to express their intent for smoothing. This corresponds to detecting the vertical shift of the KDE. We used normalized cross-correlation [33] between the two consecutive KDE signals to determine the shift. Secondly, for pulling the pot, we observed that users used specific grasps (Figure 3). In this case, we note that the KDE has two maxima and one minima (Figure 5). Here, each maxima corresponds to the fingers making contact with the pot and the minima corresponds to the center of the grasp. This essentially allows us to track an abstract skeletal representation of the hand. We then define the attraction rate using a based on the angle of grasp (ϕ) (Figure 4) [2]. Finally, all actions that do not correspond to either smoothing or pulling, are assigned as pushing. For pushing, we use the proximal attraction approach for deformation. The steps of the intent classification algorithm are:

1. Compute the KDE ϕ_t at time t
2. Compute normalized cross-correlation $C(\phi_t, \phi_{t-1})$
3. Compute Active region A
4. Set s to the shift of correlation
5. if($s < S$): Smooth pot profile in A
6. else:
   Compute extrema
   Detect skeleton
   Compute θ
   if(#maxima = 2 & $\theta < 2\pi$): Apply pulling in A
   else: Apply proximal attraction in A
   Smooth deformation ($\nabla^2\delta = 0$ in A)
7. Compute deformed profile
8. Rescale pot sections

[1]See supplementary material for parameter values

[2]See supplementary material for parameter values

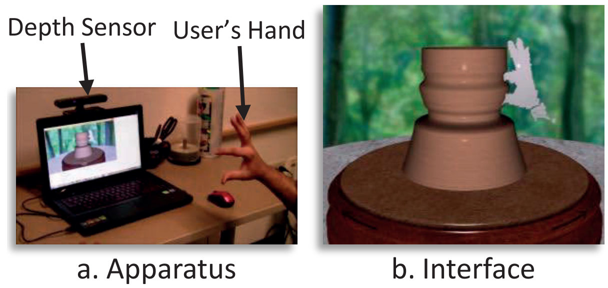

Figure 6: The apparatus (a) consists of the user, a computer and a depth camera. The user sees a PCL of their hand deforming a rotating pot (b).

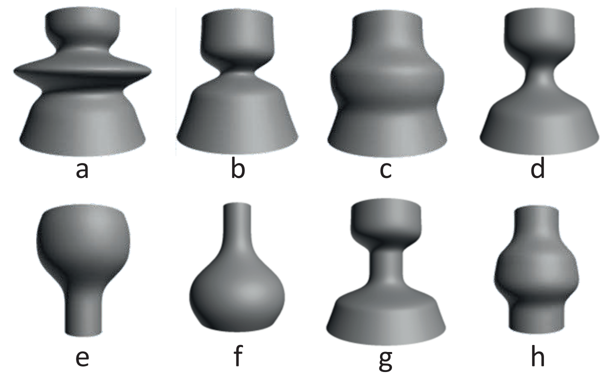

Figure 7: Eight pre-defined pots were shown to participants in the quiz. These are: (a, b) thin convex and thin concave, (c,d) fat convex and concave, (e, f) round and flat, and (g, h) flat at center and ends.

5 Evaluation

5.1 Apparatus

Our setup consisted of a Lenovo ideaPad Y500 laptop computer with an intel i7 processor and 8GB RAM, running 64-bit Windows 8 operating system with a NVIDIA GeForce GT 750M graphics card, and the SoftKinetic DS325 depth sensor (Figure 6(a)). SoftKinetic DS325 is a close range (0.1m-1.5m) time-of-flight depth sensor that provides a live video stream of the color and depth image of the scene. Every pixel on a given depth image can be converted to a 3D point using the camera parameters [3].

5.2 Implementation & Interface

Our first step was to segment the hand from the scene. The SoftKinetic iisu API provides robust segmentation and tracking of hands and additionally provides a smooth PCL of the hand. However, their tracking method does not work with hand-held objects - a feature that we required in order to allow users to utilize physical objects for deformation. We used a pre-defined a volumetric workspace as the active region in front of the computer screen. We designed our interface based on the guidelines provided by Stuerzlinger and Wingrave [28]. Our interface comprises of a 3D scene with a rotating pottery wheel on natural outdoor background (Figure 6(b)). The user sees the potter's wheel and the PCL of their hands, or the tools held in their hands. We also rendered a shadow of the hand PCL on the surface of the pot. From our pilot experiments we found this to be particularly helpful to users in estimating their proximity to the surface. We also observed that while trying to reach the surface from the side, several users unintentionally made contact with the back facing regions of the pot by moving their hands too close to the depth camera. Thus, we clamped the hand PCL so as not to allow points on the hand to reach behind the surface of the pot.

5.3 Participants

We recruited 15 (11 male, 4 female) science and engineering graduate students within the age range of $19-30$ years. 10 participants had familiarity with mid-air gestures and full body interactions, and 11 participants reported familiarity with 3D modeling and computer-aided design. 5 participants had amateur experience with ceramics and pottery.

5.4 Procedure

The length of the study varied between 45 to 90 minutes. In the beginning of the study, each participant was given a verbal description of the setup, the purpose of the study and functionality of the pottery application. This was followed by a practical demonstration of our pottery application. The participants were then asked to perform the following tasks:

- **P** *Practice:* Each participant used our application for three minutes to get an overall familiarity with the interaction of their hands with the pot surface. During this phase, the participants were allowed to ask questions and were provided guidance when required.
- **T1** *Quiz:* The participant was shown a pre-defined target shape and asked to shape a "blank" pot so as to roughly match the most noticeable feature of the pre-defined shape. A total of eight target shapes were shown in a random order (Figure 7). The participants were allowed to undo and redo a particular deformation at any time using keyboard shortcuts. The participant could also reset the current shape to the blank pot. Once the participant was satisfied with the result, they would move to the next pre-defined pot shape.
- **Q1** *Questionnaire 1:* The participants were asked a series of questions regarding the intuitiveness, quality of intent recognition, responsiveness of the deformation and consistency of pushing and pulling during the quiz.
- **T2** *Composition with Hands:* The participants were given a duration of five minutes during which they were asked to think of certain specific pot shapes and shape them using their hands. They were asked to describe what they intended to make before beginning their creation. Although the duration of time was fixed, the users were allowed to complete their last composition that was started before the end of the specified duration.
- **T3** *Composition with Tools:* The participants were given a duration of five minutes to create pots using a set of physical artifacts as tools. Our "tools" comprised of day-to-day objects (e.g. white-board marker, pair of scissors, ruler) and also some special objects such a Shapescapes™[4].
- **Q2** *Questionnaire 2:* Each participant was asked a series of questions regarding enjoyability, ease of use and learning. The participants were also asked regarding the utility, ease of use and preference over hands. User comments were requested about what they liked and disliked about the application, interface and interaction.

[3] See supplementary material for details

[4] www.shapescapes.com

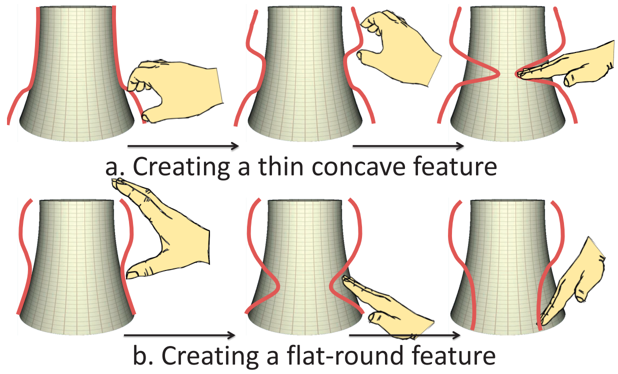

Figure 8: Two examples are shown of common deformation strategies are shown through which users created (a) thin concave and (b) flat-round features.

6 Results

6.1 Quality of User Response:

The main aspect that we sought from the *Quiz* was the quality of the final outcome across participants for a given quiz problem. We used curvature cross-correlation (*CCC*) as a measure to compare the quality of user created profiles. We first compute the curvature signature of an observed profile. Each point on the signature is the curvature of a point in the profile. This allows for the target-to-response comparison to be sensitive towards local dissimilarities across users. Subsequently, we compute the normalized cross-correlation [33] between the curvature signatures of the observation with that of the ground truth. This allows the comparison to be invariant to the shift between the features of the profiles along vertical direction. The quality is then defined as the maximum value of the correlation . The value of *CCC* lies in the range $[0,1]$. Here, higher values represent better quality (1 corresponding to perfect match and 0 no match).

6.2 User Performance & Behavior (T1)

We observed significant variations in how each user perceived and approached a given target shape. There was no evident correlation between the time taken by each user and the quality (*CCC*) of the final pot created by the user for any of the target shape. Hence, we represent the user performance as a bivariate dataset given by the ordered pair of the response quality and completion time. Consequently we visualize performance as a bag-plot [24] (Figure 9). Here, the spread of the data (i.e. variations in user responses) is given by the area of the *bag*.

Users clearly performed best for thin-concave targets with Tukey median value of $(0.94, 1.46)$. Performance was most consistent for fat-concave feature (Figure 9(d)). Users also performed consistently for round-and-flat features (Figures 9(e) and (f)). Variations were significant for central flat feature (Figure 9(g)). Further, the quality was very low for central-flat and top-bottom-flat features (Figures 9(g) and (h)). This was mainly because users typically spent considerable time pulling and smoothing the top and bottom regions after performing an initial push. Consequently, the median completion times were also higher for round-flat and central flat features (Figure 9(f) and (g) respectively).

6.3 User Behavior - Hands (T1)

The tendency to approach the pot from the sides was common. Upon asking, the users typically answered" "*my own hand blocks the view of the pot*". Users frequently tried using two hands, particularly for round-flat combinations. Some users also changed their manipulating hand from dominant to non-dominant due to arm fatigue as typically expected in focused mid-air interactions. This, however, was problem due to: (a) the limited volume of the workspace, and (b) difficulty of avoiding unintended deformation due to asynchronous motions of two hands.

6.4 User Behavior - Hands (T2)

On an average, users created 5 pots on an average (max: 12, min: 2) within 5.80 minutes (std: 0.66 min). We made two interesting observations in **T2**. First, we found that users were able to repeat the process of getting from an initial shape to the same final shape across multiple trials. Similarly the users could also deform a current shape back to some previous shape, akin to the *undo* operation, but with the hands. In fact, most participants preferred using their hands to *undo* a pot deformation instead of the keyboard-shortcut. One user stated: "*I thought it was easier to learn the software when I was trying to make my own pot not a model one*". This was expected because of the learning and practice that the users had during the quiz (**T1**). However, during **T1**, the users mentioned that their attention was divided due to the need to intermittently look at the target shape during the shaping process. Thus, they generally perceived **T1** to be more demanding than **T2**.

Typically, while shaping thin-convex features, we observed most users achieved a general convexity followed by pushing the top and bottom portions inward. We had assumed that users will create concave features in a single inward action. Interestingly, they first pulled the top and the bottom portions of the pot and subsequently pushed the central region of the pot (Figures 8(a)). Similar was the case with flat-round features (Figures 8(b)). Many users first pulled out the round feature followed by straightening the flat regions of the pot.

6.5 User Behavior - Tools (T3)

Most users showed immediate enthusiasm during the use of tools. However, users created 4 pots on an average (max: 8, min: 2) within 6.0 minutes (std: 0.8 min). Almost all users first inspected the objects provided to them and planned how to use them for shaping the pots. In contrast to the use of hands, we observed exploratory behavior in users while using tools. Rather than creating pots, most users were more interested in finding out the effect of each of the objects provided to them. This explained the decrease in the average number of pots in the composition task. One of the difficulties with the use of hands was the inability to create thin concavities. With the use of tools (Figure 10(a),(d)), users could achieve this easily. The most interesting behavior that was observed was the tendency to create convex deformations, which the users achieved by combining two different objects, so as to simulate a grasping hand. This was evident from the users' fascination with scissors (Figure 10(b)).

Another important aspect that we observed was the direct association the users made between the shape of the tool and the purpose it could be used for. The motion of the hand was affected by this association. For instance, while using a white-board eraser (Figure 10(c)), the most common motion was that of smoothing the pot. Similarly, for objects with grasp-like geometries, users invariably tried convex deformations by pulling (Figure 10(e)). One user fashioned a new tool by combining different Shapescape parts. This provided the convenience of holding the tool at the "handle" and deforming the pot using fine hand movements (Figure 10(f)).

6.6 User Feedback - Intent & Controllability

In general, users agreed that the pot behaved according to the way the users shaped their hands (Figure 11). About 50% of the users perceived the deformation response to be slow while the remaining considered it balanced. In general, we also found a common agreement on the initialization time and robustness to accidental

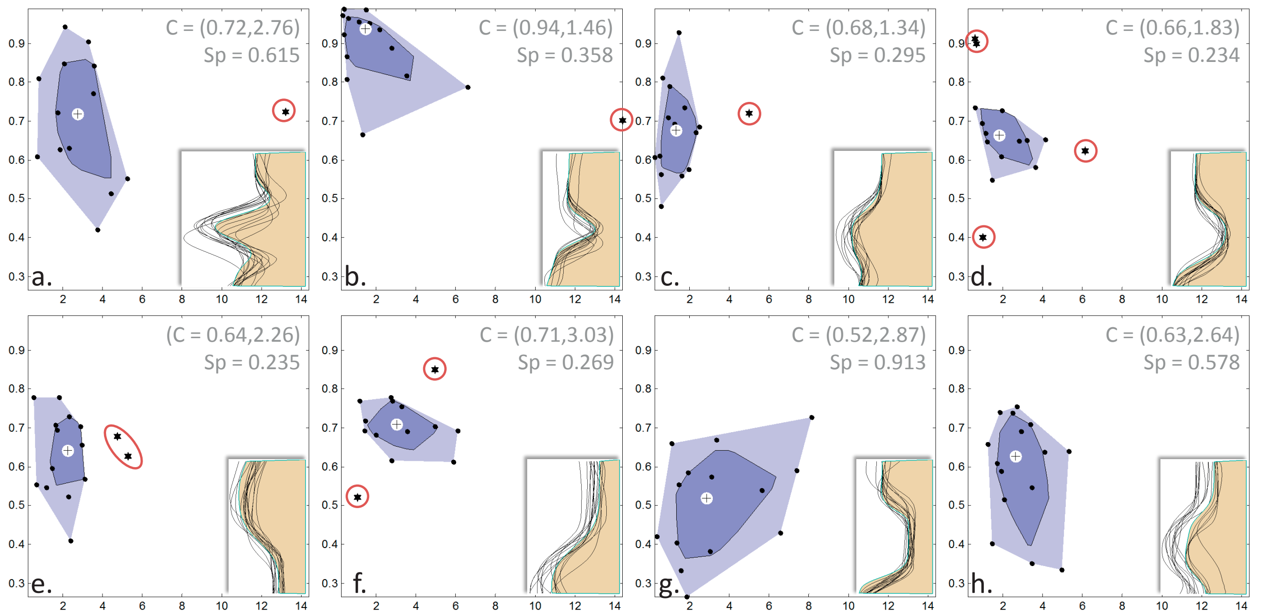

Figure 9: User performance is shown for the each quiz problem as a bag-plot. The x-axis is time in the range $[0, 14]$ minutes and the y-axis is the curvature cross-correlation in the range $[0, 1]$. The dark and light blue regions show the bag and fence regions, respectively. The white circle is the Tukey depth median and the points marked with red circles are the outliers. The insets show the actual pot profiles (black lines) created by the users in comparison to the target shapes (beige region) of the Quiz. The coordinates of the the depth median (C) and the spread (Sp) are provided for each target shape.

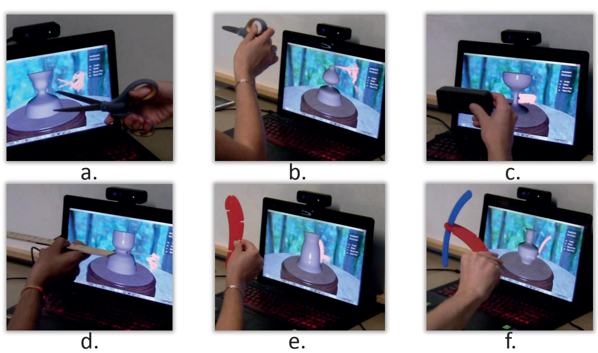

Figure 10: Examples of tool usage are shown.

deformation. The most common and expected difficulty that users faced was that of pulling specific regions of the pot. The reason for this perception was primarily related to the visual and tactile feedback rather than the algorithm for pulling itself. This was evident from the user's comments such as: "*I think the reason pushing and pulling were different were because the pulling you had to [make] 2 contacts with the pot and pushing you only needed one. I had a hard time understanding the depth of the pot making it hard to get two contacts on the pot*". One user also suggested: "*I think it would be better if I get some feeling when I touch the pottery. It [would] make me feel more real and easier to control my hand. Then it would be better to have some sounds when I touch the pottery*".

6.7 User Feedback - Experience

We found the user experience to be mostly positive (Figure 11(b)). The main aspects that the users liked were (a) realism of pottery, (b) ease of learning, and (c) the freedom of choosing how to deform the pot. In particular, users liked the use of tools and the smoothing operation the most. One user commented: "*The freeform design with tools was the most fun, as I could spend most of my time focusing on the design aspect as opposed to focusing on minimizing errors.*". According to another user: "*The pottery changing according to my hand shape is so real. While smoothing, I could shape it as well, I like to do it this way a little bit.*".

7 Discussion

7.1 Implications

Spectrum of Expressiveness: One aspect that is both advantageous and disadvantageous in our approach is that different users can achieve the same target shape using different strategies for grasping, reaching and deforming a shape. While this provides flexibility and intuitiveness to the user, it also results in increasing the time taken by the user to reach to a desired shape. The evaluation of proximal attraction evidently indicated that there needs to be a balance between completely free-form interaction and symbolic approaches. This is what we attempted through the KDE based approach. The main advantage that our process provided was the discovery of relevant grasp information that is useful to design continuous operations such as shape deformation. Our grasp based approach can serve as a starting point for designing grasp-based interactions using cleaner data such as hand-skeleton [2].

Definition of Intent: We began with a simple classification of intent through the analogy of structuring operators inspired by Delame's [9] work. However, users' description of actions and expectation strongly indicates towards a richer and more complex mental model for deformation processes. To this effect, we had to include a third class of operation, namely "smoothing" which evidently improved the performance of the user. Though this aspect is not new in 3D modeling in general, this aspect of refinement is certainly worth

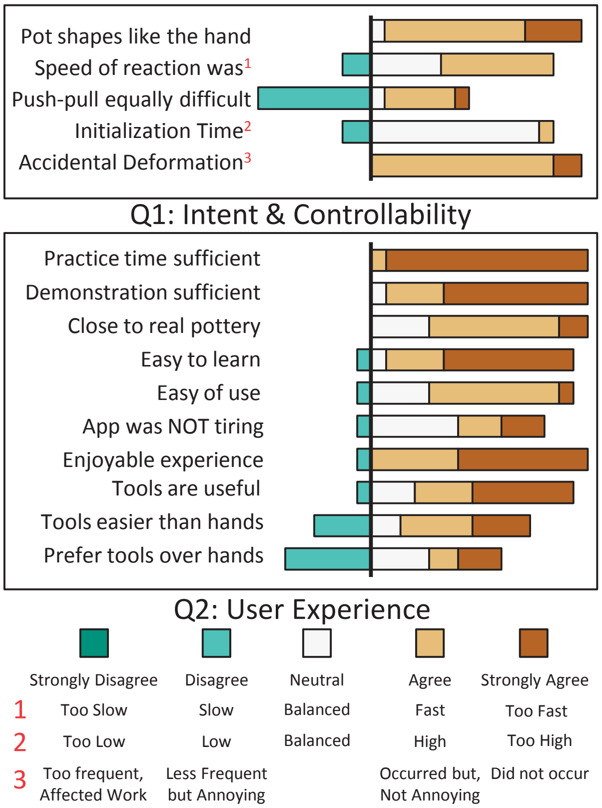

Figure 11: User response for our approach showed that while the robustness to accidental deformations was perceived to be negligible (a), many users still perceived pulling to be difficult. Users agreed regarding the usefulness of tools but were not in general agreement about preferring them over hands.

investigating from a perceptual point of view. Although we demonstrated intent classification for rotationally symmetric shapes, we believe that our results can provide insights into how hand grasp can be used for free-form modeling of arbitrary shapes. In particular, the combination of using bare hands and hand-held physical tools can be effectively employed for coarse and fine shape deformation respectively.

Precise & Selective Reachability: One user aptly commented: "*Sometimes it is hard to use the palm because it may deform the surface too much. The context of barely touching does not seem too well implemented. However, if you do this very carefully you can do the barely touching but may make your arm tired a little.*". This is the problem of *precise and selective reachability* wherein one is required to reach and manipulate a local region of an object without affecting neighboring regions. There is extensive volume of work that investigates *distal* selection, manipulation and navigation [11, 6, 12] of objects. We believe that precision and selectivity are problems worth investigating for close-range, i.e. *proximal* 3D manipulations in mid-air.

7.2 Limitations

Depth perception was a major cause of difficulty, particularly considering the unstructured and complex representation of the hand as a PCL. One user noted: "*Pushing seems easier than pulling. Part of the reason I suspect is the visual feedback. It is easier to determine if my hand starts to touch the pot, while it's not as easy to determine if my hand is still attached with the pottery or leaving it.*". Our method is currently implemented for pottery, which is essentially a one dimensional deformation. Further, we observed that the use of 2D displays is a factor due to which users tend to use side configurations. We believe that 3D visual feedback will encourage users to access the front and back faces.

Severe occlusion resulting from camera position and hand orientation is an issue particularly for skeletal based gesture recognition. We partly addressed this challenge using our PCL-based approach which can make use of partial data even when the full hand skeleton is intractable. However, occlusion is an inherent problem in any camera-based method. Investigation of optimal camera position and use of multiple cameras at strategic locations is an important future work.

In our current implementation, the definition of active regions is in terms of 2D profile topology rather than actual distances in real space. Thus, our implementation is dependent on PCL sampling relative to the mesh resolution of the pot. Independence from the sampling resolution may be addressed with an adaptive approach wherein new sections could be added according to manipulators or old ones removed based on geometric properties of the pot profile such as curvature.

7.3 Future Directions

Our first goal is to extend the KDE based approach for grasp and motion characterization to arbitrary meshes. Secondly, we intend to study how user perception ad performance is affected by adding 3D visual feedback and also tactile feedback. Finally, with our approach, it is possible to perform deformation using existing hand skeletal tracking approaches. We intend to investigate this in comparison to the PCL based hand representation. One key advantage of using tracked skeletons is that there is a direct correspondence between the fingers and palm which can give useful movement information for better intent detection. This would help segmenting users intentional and unintentional movements [7].

8 Conclusions

We presented a spatial interaction technique that uses hand grasp and motion for intent expression in virtual pottery. This approach enables a paradigm shift from existing gesture-based procedural events towards non-procedural and temporally continuous processes in the context of shape deformation. In other words, our work enables users to achieve what they intend in the way they see fit. To the best of our knowledge, no existing hand-based spatial modeling scheme offers such diverse contexts of user input, for instance the use of everyday real objects as tools for virtual shaping, with controllable outcomes. The idea creates new pathways for further research exploring creative design contexts in a "what you do is what you get" framework.

Acknowledgements

We thank the reviewers for their valuable feedback and comments. This work was supported by the NSF-AIR Award No. 1312167, the NSF Award No. 1235232 from CMMI-EDI (A-DRIVE), and the Donald W. Feddersen Chair Professorship and the School of Mechanical Engineering. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Dr. Raja Jasti (zeroUI) for providing valuable suggestions and feedback on our work. We also thank Kevin Lee Jr. (zeroUI) for the artwork of our pottery interface.

References

[1] Leap motion sculpting. `https://airspace.leapmotion.com/apps/sculpting/windows`.
[2] Leap motion skeletal tracking. `https://developer.leapmotion.com/features#hand-model`.

[3] C. Armbruster and W. Spijkers. Movement planning in prehension: do intended actions influence the initial reach and grasp movement? *Motor Control*, 10(4):311, 2006.

[4] T. Baudel and M. Beaudouin-Lafon. Charade: Remote control of objects using free-hand gestures. *Communications to the ACM*, 36(7):28–35, 1993.

[5] R. Boulic, S. Rezzonico, and D. Thalmann. Multi-finger manipulation of virtual objects. In *In Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST '96*, pages 67–74, 1996.

[6] D. Bowman and L. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. pages 53–58, 1997.

[7] A. Choumane, G. Casiez, and L. Grisoni. Buttonless clicking: Intuitive select and pick-release through gesture analysis. In *IEEE Virtual Reality Conference*, pages 67–70, March 2010.

[8] M. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *Robotics and Automation, IEEE Transactions on*, 5(3):269–279, Jun 1989.

[9] T. Delamé, J. C. Léon, M. P. Cani, and R. Blanch. Gesture-based design of 2d contours: An alternative to sketching? In *Proceedings of the ACM Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pages 63–70, 2011.

[10] Y. Han and B.-j. Han. Virtual pottery: a virtual 3d audiovisual interface using natural hand motions. *Multimedia Tools and Applications*, pages 1–17, 2013.

[11] C. Hand. A survey of 3d interaction techniques. *Computer Graphics Forum*, 16(5):269–281, 1997.

[12] K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell. Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, 5(3):260–302, 1998.

[13] C. Holz and A. Wilson. Data miming: inferring spatial object descriptions from human gesture. In *Proceedings of the ACM conference on Human factors in computing systems*, pages 811–820, 2011.

[14] N. K. Iason Oikonomidis and A. Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *Proceedings of the British Machine Vision Conference*, pages 101.1–101.11, 2011.

[15] M. Jeannerod. The formation of finger grip during prehension. a cortically mediated visuomotor pattern. *Behavioural Brain Research*, 19(2):99 – 116, 1986.

[16] M. Karam and M. C. Schraefel. A taxonomy of gestures in human computer interactions. Technical report, University of Southampton, 2005.

[17] P. G. Kry, A. Pihuit, A. Bernhardt, and M.-P. Cani. Handnavigator: Hands-on interaction for desktop virtual reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 53–60, 2008.

[18] M. Leyton. A process-grammar for shape. *Artif. Intell.*, 34(2):213–247, 1988.

[19] D. P. Mapes and J. M. Moshell. A two handed interface for object manipulation in virtual environments. *Presence*, 4(4):403–416, 1995.

[20] A. Pihuit, P. Kry, and M.-P. Cani. Hands on virtual clay. In *In Proceedings of IEEE International Conference on Shape Modeling and Applications*, pages 267–268, 2008.

[21] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari. Multimodal human discourse: Gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9(3):171–193, 2002.

[22] H. Rateau, L. Grisoni, and B. De Araujo. Sub-space gestures. Elements of design for mid-air interaction with distant displays. Research Report RR-8342, INRIA, August 2013.

[23] G. Ren and E. O'Neill. 3d selection with freehand gesture. *Computers & Graphics*, 37(3):101 – 120, 2013.

[24] P. J. Rousseeuw, I. Ruts, and J. W. Tukey. The bagplot: A bivariate boxplot. *The American Statistician*, 53(4):382–387, 1999.

[25] J. Segen and S. Kumar. Gesture vr: vision-based 3d hand interace for spatial interaction. In *Proceedings of the ACM international conference on Multimedia*, pages 455–464, 1998.

[26] J. Sheng, R. Balakrishnan, and K. Singh. An interface for virtual 3d sculpting via physical proxy. In *Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, pages 213–220, 2006.

[27] J. Song, S. Cho, S.-Y. Baek, K. Lee, and H. Bang. Gafinc: Gaze and finger control interface for 3d model manipulation in cad application. *Computer-Aided Design*, 46(0):239 – 245, 2014.

[28] W. Stuerzlinger and C. A. Wingrave. The value of constraints for 3d user interfaces. In G. Brunnett, S. Coquillart, and G. Welch, editors, *Virtual Realities*, pages 203–223. Springer Vienna, 2011.

[29] Vinayak, 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(2):277–287, 2013.

[30] R. Walter, G. Bailly, and J. Müller. Strikeapose: Revealing mid-air gestures on public displays. In *Proceedings of the ACM International Conference on Human Factors in Computing Systems*, pages 841–850, 2013.

[31] R. Walter, G. Bailly, N. Valkanova, and J. Müller. Cuenesics: Using mid-air gestures to select items on interactive public displays. In *Proceedings of the ACM International Conference on Human-computer Interaction with Mobile Devices*, pages 299–308, 2014.

[32] R. Wang, S. Paris, and J. Popović. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the ACM symposium on User interface software and technology*, pages 549–558, 2011.

[33] J.-C. Yoo and T. Han. Fast normalized cross-correlation. *Circuits, Systems and Signal Processing*, 28(6):819–843, 2009.

[34] J. Zheng, S. De La Rosa, and A. Dollar. An investigation of grasp type and frequency in daily household and machine shop tasks. In *IEEE International Conference on Robotics and Automation*, pages 4169–4175, May 2011.