# RealFusion: An Interactive Workflow for Repurposing Real-World Objects towards Early-stage Creative Ideation

Cecil Piya* Vinayak† Yunbo Zhang‡ Karthik Ramani §, ¶

School of Mechanical Engineering
Purdue University

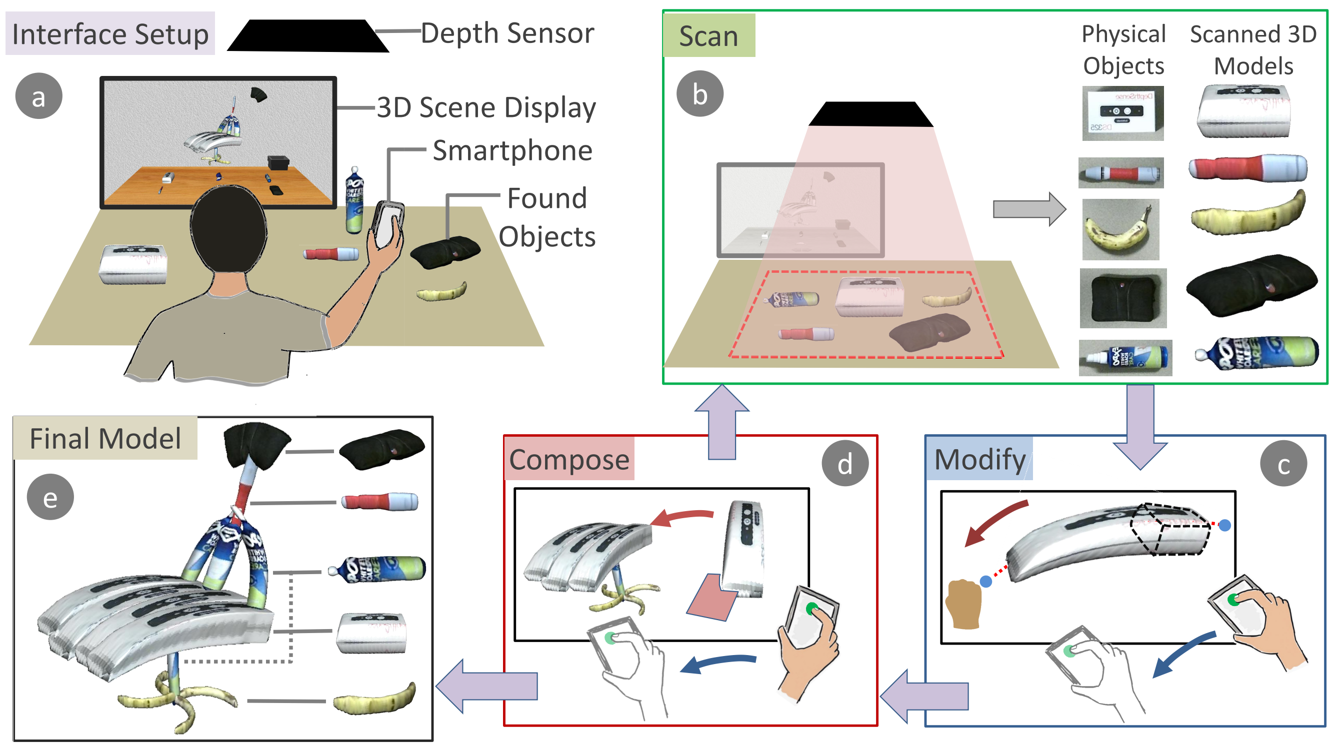


Figure 1: (a) Physical setup used in *RealFusion*, (b-d) three stages of the workflow: (b) digitizing physical objects using 3D scanning, (c-d) using mid-air gestures with a smartphone for interactive modification of scanned shapes and 3D composition of shapes, (e) 3D model representing an early-stage design idea.

**ABSTRACT**

We present *RealFusion*, an interactive workflow that supports early stage design ideation in a digital 3D medium. *RealFusion* is inspired by the practice of *found-object-art*, wherein new representations are created by composing existing objects. The key motivation behind our approach is direct creation of 3D artifacts during design ideation, in contrast to conventional practice of employing 2D sketching. *RealFusion* comprises of three creative states where users can (a) repurpose physical objects as modeling components, (b) modify the components to explore different forms, and (c) compose them into a meaningful 3D model. We demonstrate *RealFusion* using a simple interface that comprises of a depth sensor and a smartphone. To achieve direct and efficient manipulation of modeling elements, we also utilize mid-air interactions with the smartphone. We conduct a user study with novice designers to evaluate the creative outcomes that can be achieved using *RealFusion*.

**Index Terms:** Information Interfaces and Presentation (e.g. HCI)Miscellaneous;

*e-mail: cpiya@purdue.edu
†e-mail:fvinayak@purdue.edu
‡e-mail:ybzhang@purdue.edu
§e-mail:ramani@purdue.edu
¶School of Electrical Engineering (by courtesy)

## 1 INTRODUCTION

Early-stage creative ideation is an essential activity in art and design, as it allows designers to conceive and explore preliminary ideas while informing downstream processes. It's goal is not to generate full-fledged designs, but rather to aid visual observation and communication of coarse mental images [34]. Current design practices primarily utilize sketching and 3D modeling for early-stage ideation [13]. While sketching is an efficient means for expressing ideas, it is limited to a single viewpoint and also requires good drawing skills. Most 3D modeling software on the other hand are


Figure 2: Found object art: (a) Picasso's *Head of a Bull*, (b,c) Works by other found object artists: (b) *Kathryn Rodenbach*, (c) *David Calfo*

tailored towards detailed design, and are therefore unsuitable for creative tinkering and freeform explorations [6]. Our work aims to bridge such gap between low-fidelity tools and feature rich software to support creative ideation in a digital 3D medium.

In this paper, we present *RealFusion*, a creative ideation workflow that allows designers to express early-stage ideas by (a) quickly scanning physical objects, and (b) virtually modifying and composing them into meaningful design representations (Figure 1). Here, the underlying concept is inspired by *found object art*, where physical objects lend themselves as components of a 3D collage. This artform has been shown to effectively convey both abstract and concrete ideas [9], and is also reflected in the process of physical mock-up design. However, physical shape composition can be time consuming and also require mechanical skills. In contrast, *RealFusion* leverages a virtual environment where real-world constraints are non-existent, making shape modification and assembly tasks significantly easier to perform.

Prior works have shown the utility of augmenting digital 3D modeling with physical reality. For instance, Modeling in Context [20] enables design of new artifacts within a 2D image of their physical settings. Using motion sensing and 3D data acquisition hardware, other works have extended this notion into the 3D space by either utilizing physical objects as spatial references [35] or coarse-level scaffolds [26, 24] to guide new designs. The objective in *RealFusion* however is not to *build-around* or *build-over* existing objects, but rather to personalize the objects themselves as building blocks of a new design representation.

We anticipate several benefits from *RealFusion* in early-stage design. First, by using an instantaneous 3D scanning system, designers can avoid the time-intensive process of modeling each component from scratch, and instead acquire the geometry of suitable objects. Second, this process helps expand creative thinking by allowing designers to look past mundane identities of ordinary objects, and view them as potential elements of new ideas [8]. Third, while we enable digital modeling operations that cannot be easily performed in the physical world, we also provide interactions that encourage physical engagement via mid-air gestures with a smartphone. Finally, given that the non-invasive scanning system cannot damage or alter the scanned objects, any object in the physical world becomes amenable for use in a design.

We demonstrate *RealFusion* using a prototype interface, which leverages new technologies like a 3D scanner, depth sensor, and smartphone based mid-air interactions. Such known technologies allows us to ensure robustness in the system and facilitate fluid interactions. In this work, we mainly focus on evaluating the creative outcomes of *RealFusion* in the context of early-stage design. We also conduct our studies with novice designers given their amenability towards unstructured design approaches and quick design explorations through trial-and-error processes [1].

## 2 Background

In computer graphics, *digital shape composition* (DSC) serves as a powerful means for combining pre-existing shapes into new representations. Several works have proposed techniques to synthesize 3D collages [9] and product/multimedia families [15] from a finite set of shapes. However, in contrast to such generative modeling approaches which aim to automate design, creative ideation is an interactive process where a designer's involvement is critical for externalizing ideas and exploring the design space. Since the goals here are not always clearly defined and subject to unexpected changes, it is characterized by a sense of play where creative thinking is inspired from *reflection-on-action* [6]. Thus, our focus in *RealFusion* is to provide users with an engaging design experience conducive towards creativity.

Several works, such as SnapPaste [29] and MeshMixer[1], have explored interactive DSC systems, where GUI widgets are used to manipulate and position 3D shapes. While these works show precise 3D control that are suitable for detailed design, their use of 2D inputs for 3D operations can constrain free explorations during early-stages [32]. In contrast, mid-air gestures have been shown to provide an *"expressive nature that enables less constraining and more intuitive digital interactions"* [31]. For instance, Kim and Maher [18] demonstrated how such gestures can support more efficient 3D operations that lead to increased design form explorations and creative cognition. Thus, by using mid-air gestures in *RealFusion*, we expect to enable *easy personalization* and *juxtaposition* [3] of shapes in 3D space using suggestive actions like picking, placing, manipulating, pulling, and bending.

To enhance physical engagement, tangible shapes have been used as building blocks within DSC workflows. For example, Anderson et al. [2] utilized rectangular blocks with embedded sensors to define coarse-level architectural models. Other works [14, 4, 19] have made similar use of non-instrumented primitive shapes in mixed-reality design systems. While *RealFusion* adopts a similar approach, we do not pre-define the physical building blocks. Instead, we allow users to identify suitable objects from their surroundings to serve as design components. Given the diversity of shapes available in our everyday lives, we expect this approach to not only inspire creativity but also provide a unique aesthetic quality in the resulting models. In addition, we allow users to modify scanned shapes to better reflect their intent and also generate 3D forms that are unavailable.

Similar to *RealFusion*, systems like CopyCAD [7] and KidCAD also [8] enable users to compose new design forms by repurposing full of partial geometries from existing objects. However, given that such tools compose shapes by imprinting objects over a planar surface, the resulting designs are 2.5D in nature. In *RealFusion*, we provide mid-air interactions to enable 6 DOF inputs for configuring shapes into a compound design. This allows users to express the full 3-dimensionality of an idea, and also explore a wide variety of 3D forms.

## 3 System Design Rationale

We design the *RealFusion* workflow based on the generic process followed in *found object art*, i.e. collecting physical objects, modifying them according to need, and composing them into a 3D structure. It comprises of three modeling states (Figure 1):

**State 1 (Scan):** Physical objects of interest are collected and digitized into 3D models.

**State 2 (Modify):** The scanned shapes are modified to better match user intent, and also diversify a single object into multiple forms.

**State 3 (Compose):** Using mid-air interactions, the shapes are composed into 3D models that reflect preliminary design ideas.

We implemented a prototype system (Figure 1 (a)) using off-the-shelf components to minimize custom implementations and ensure system robustness. It comprises of a computer-monitor setup, an RGB-D sensor for motion tracking and 3D scanning, and a smartphone for 3D interactions. Here, we discuss two primary components of the system.

[1]http://www.meshmixer.com/

### 3.1 3D Scanning Interface

Given the emphasis on rapid expression of ideas in early-stage design, we find it essential to enable instant scanning of physical objects in *RealFusion*. For this, we implemented a *quick scanning system*, where an overhead RGB-D sensor streams a live-video of the physical space into the monitor (Figure 4(a)-top). To scan an object, it is placed over the desk such that it is contained within a scanning window shown in the video display. A scan button is then pressed to acquire 3D data inside this window and perform shape reconstruction. This system is analogous to image based reconstruction presented by Olsen et al [23]. However, by using depth data we avoid the need for user inputs to define geometric boundaries and features. Due to a fixed sensor view-point, the *quick scanning system* assumes planar symmetry of objects and also generates coarse-level shapes. However, given the *quick-and-dirty* nature of early-stage design, the precision of the shapes does not affect the creative outcomes in *RealFusion*.

To support more aesthetic contexts like art and industrial design, we also provide a *detailed scanning system*. Here, we consider three options: (a) depth sensor based SDKs like KinectFusion [11] or Skanect, (b) smartphone camera based shape reconstruction apps like Autodesk 123D Catch, and (b) dedicated 3D scanners (Figure 4(a)-bottom). While options (a) and (b) can be acquired without any changes to the system, option (c) requires additional cost. But, our tests indicated that dedicated scanners are more ergonomic and provide better scanning fidelity. Compared to quick scanning, detailed scanning is time consuming as it relies on multiview data acquisition and manual data clean up. Thus, users can decide which system to use based on time constraints, design contexts, and accuracy requirements.

### 3.2 Shape Composition Interface

To support 3D interactions during the *modify* and *compose* states, we considered hand gestures, digital controllers, and smartphones as viable media for mid-air inputs. We find that while hand gestures allow users to express spatial intent using natural modes of human communication, they lack a means for tactile feedback. In fact, they typically involve grabbing gestures to manipulate imaginary objects, which not only precludes kinesthetic control of virtual objects [10], but also leads to discomfort and physical fatigue [17]. Haptic gloves with rendered tactile feedback on the other hand require wearing of obtrusive hardware.

In contrast, a tangible medium provides tactile feedback that inherently comes from grasping real-world objects, and also uses our natural prehensile ability for manipulating virtual objects. Digital controllers are examples of such media, and their utility within 3D creativity tools has been demonstrated [12]. But, given the generic functionality of such devices (i.e. motion tracking and click buttons) we find their use to be limiting in terms of future extensions of *RealFusion*. Instead, we utilize smartphones as 3D controllers, as they provide a wider variety of interactive capabilities like multi-touch inputs, orientation sensing, GUI display and control, web browsing etc. In addition, the commonality of smartphones make them significantly more accessible.

While smartphones have been predominantly used as GUI systems, recent works [16, 33] have explored their spatial interactive capabilities with computer displays by combining tilt gestures with multi-touch inputs. Similarly, Mine et al. [22] showed that by augmenting smartphones with position tracking hardware, they can be used as 6 DOF controllers. In *RealFusion*, we adopt a similar approach. However, we repurpose the over-head depth sensor from the *quick scanning system* to track the phone's position using a non-intrusive vision based method.

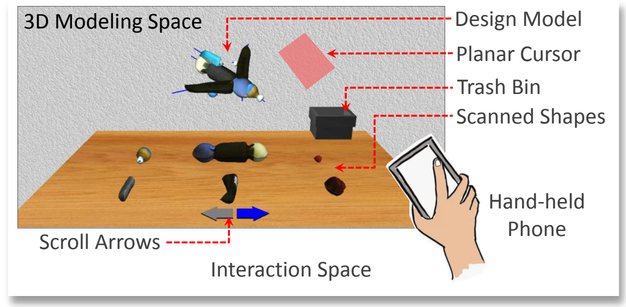


Figure 3: 3D modeling scene displayed in the computer monitor.

## 4 Modeling Interactions

The 3D composition scene (Figure 3) is designed to resemble the interaction space, and consists of a horizontal desk over which scanned shapes are laid out. To avoid clutter, only 6 shapes are displayed at a time, but the scroll arrows allow access to other shapes. Since we use a flat screen display, shadows are rendered on the desk surface to assist depth perception. Users can express *intent-to-interact* with the scene by holding the phone towards the sensor for a brief moment (1.5 sec) [27]. This gesture activates a planar cursor whose motion can be controlled by manipulating the phone in mid-air. To stop interactions, users can simply place the phone on the desk. Touch gestures on the phone allow users to indicate discrete events like clutch, release, scale, and replicate. The following operations are performed during model construction (Figure 4).

Shape Selection and Manipulation. To pick up a shape, the cursor is first hovered over it (Figure 4(c)-top). A bounding box around the shape indicates the cursor's proximity. A single tap gesture is then used to clutch (or release) the shape. Under a clutched state, the cursor is fixed into the shape, allowing users to control it by manipulating the cursor.

3D Model Composition. Multiple copies of the scanned shapes can be picked up and assembled into a 3D model. At any point, users can rotate or translate the assembly to change its viewpoint. For translation, the cursor is brought close to the assembly center. The assembly is then clutched with a touch-and-hold gesture and constrained to only translate with the cursor. The same interaction is used for assembly rotation, except here the cursor is placed away from the center, and its motion used to pivot the assembly. By grouping shapes at different proximal locations, users can also create sub-assemblies that can be separately manipulated.

Shape Modification. The *modify* state is activated by hovering the cursor over a shape and selecting it with a double tap gesture. In this state, a two finger pinch gesture (inward or outward) uniformly scales the selected shape. To deform the shape, users can *grab* either one of its axial endpoints using a touch-and-hold gesture, and suggestively move the cursor to elongate, compress, or bend the shape (Figure 4(b)). Here, the cursor is displayed as open or closed hand icons to indicate proximity and clutching of the end-points.

Other Operations. A copy of an assembly shape can be picked up by hovering the cursor over the shape and using a two finger single tap gesture (Figure 4(c) bottom-left). Similarly, a three finger single tap gesture invokes an undo command, allowing users to trace back upto their last five steps. For this, we maintain an undo stack, each storing a snapshot of a prior assembly configuration. Users can also discard an assembly shape by picking it up and releasing it over the trash-bin.

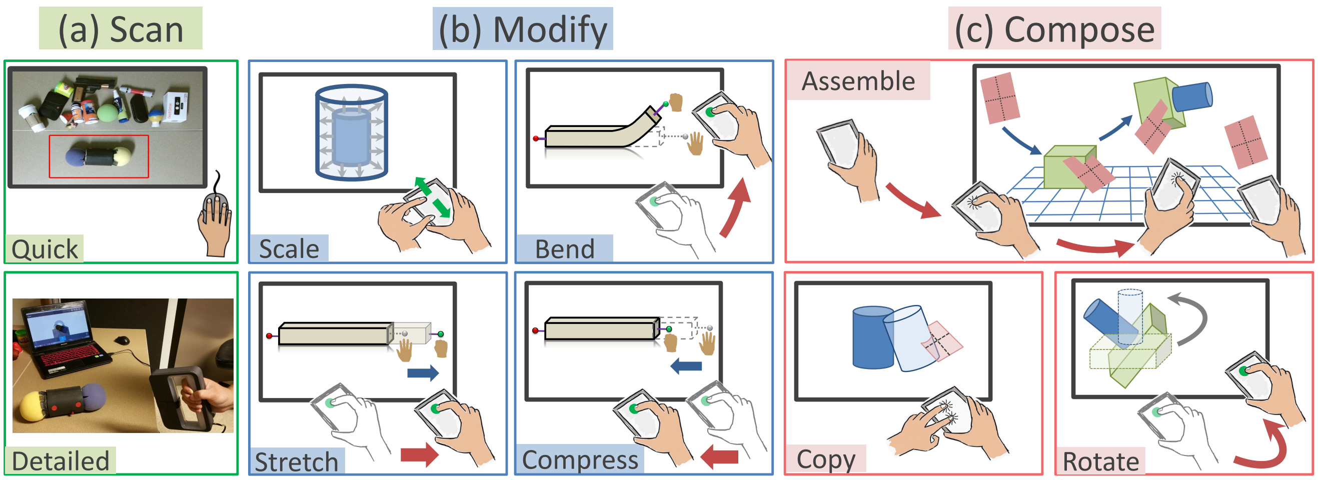


Figure 4: Modeling operations during each stage of the *RealFusion* workflow.

## 5 System Implementation

Our system used a PC (i7 2.40GHz, 8GB memory, NVIDIA GeForce GT 750M), a Samsung Note 4 device (Android 4.4, 3GB memory), and a Kinect v2 sensor. The following sections describe the implementation of its primary components.

### 5.1 Quick 3D Scanning

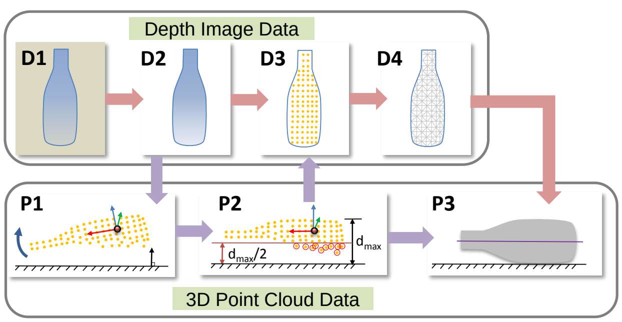


Figure 5: 3D reconstruction pipeline of physical objects.

In the *quick scanning system*, the depth sensor acquires data from the scanning window as an RGB-D image and a 3D point cloud (Figure 5). To reduce the effects of measurement errors, each image pixel and point cloud vertex is set as the mean value from 80 successive frames. Our method processes data in both the image space and the point cloud, depending on the information required. Since there is a direct correspondence between the two spaces, the operations applied in one space are analogously applied in the other. Each step of the reconstruction process is described below and referenced to Figure 5 with a specific label (e.g. D2, P2)

**Background Removal (D2):** To segment object data in the scanning window image, $\mathbf{I_1}$, we use a pre-defined empty scanning window image, $\mathbf{I_2}$, as the reference background. Here, the depth value at each pixel ($\mathbf{p_i}$) in $\mathbf{I_1}$ is compared with its counterpart in $\mathbf{I_2}$. If the difference $\delta = |\mathbf{I_2}(\mathbf{p_i}) - \mathbf{I_1}(\mathbf{p_i})|$ is less than 20 mm, the pixel is classified as background data, otherwise as object data.

**Object Data Smoothing:** The depth image and point cloud data are smoothened using three iterations of mean filtering. Here, each data point ($\mathbf{p_i}$), is set as the average between itself and its $n$ closest neighbors ($\mathbf{p_j}$), $\mathbf{p_i}' = [\mathbf{p_i} + \sum_j \mathbf{p_j}]/(n+1)$. The neighboring points correspond to the adjacent pixels of $\mathbf{p_i}$ in the image space.

**Tilt Adjustment (P1):** To ensure a symmetrical 3D model, this step adjusts the point cloud's orientation such that its symmetry plane is parallel with the desk surface. Here, PCA is used to compute the point cloud's principle normal direction. The 3D transformation between the normal and the global vertical direction (normal to desk surface) is then applied to each point cloud vertex.

**Plane of Symmetry Estimation (P2):** If $d_{max}$ is the maximum distance between the point cloud and the desk surface, the plane of symmetry is set at a distance of $d_{max}/2$ from the desk.

**Outlier Removal (P2):** All data points lying below the plane of symmetry are discarded from the object data. This prevents overlapping geometry within the 3D model.

**Boundary Extraction (D3):** The exterior boundary of the object data is computed in the image space by first converting the scanning window image into a binary image. Here, pixels with object data have a value of 1, while the rest have 0. An OpenCV based contour extraction algorithm [30] is then used to identify boundary pixels. The corresponding point cloud boundary vertices are constrained to lie on the plane of symmetry to prevent holes in the 3D model.

**3D Model Generation (D4,P3):** We use Constrained Delaunay Triangulation (from OpenCV) to generate a 2D mesh over the object data in the image space. The mesh connectivity is mapped to the point cloud vertices to obtain a 3D half mesh. This is then reflected across the plane of symmetry to form the final 3D model. Four iterations of the Laplacian filter are applied to further improve the model's surface quality. Finally, the RGB data from the image pixels are mapped onto the 3D mesh to generate a texture similar to its physical counterpart.

**Deformation Axis (P3):** Using PCA, we compute the 3D model's first principle direction. A line along this direction on the plane of symmetry is defined as the model's default deformation axis. It is represented as a sequence of 100 points, and has its endpoints extending 20% beyond the model's extremities.

### 5.2 3D Position Tracking

To track the phone's position during 3D interactions, depth data from the interaction space is represented as a stream of grayscale images. When users express an *intent-to-interact* (Section 4), 3D tracking of the phone is initialized. Here, the centroid of all pixels with 50-60 cm depth values represents the initial position of the phone in the image space. The motion of this position in ensuing frames is then tracked using Lucas-Kanade's Optical Flow algorithm (OpenCV) [5], characterized by equation 1. Here, $v_x$ and $v_y$ represent a pixel's velocity components between successive frames, while $\mathbf{I}_{x_i}$, $\mathbf{I}_{y_i}$, and $\mathbf{I}_{t_i}$ are partial derivatives of neighboring pixel $i$ with respect to position $x$, $y$ and time $t$. To speed up computation, we use a 2nd level pyramid representation of the input image, and also constrain the algorithm to search a local neighborhood of $5 \times 5$ pixels. To refine the tracked position at each frame and to prevent it from drifting away, it is mean shifted to the centroid of all pixels contained within a radius of 7.5 cm.

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \sum_i I_{x_i}^2 & \sum_i I_{x_i} I_{y_i} \\ \sum_i I_{x_i} I_{y_i} & \sum_i I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_{x_i} I_{t_i} \\ -\sum_i I_{y_i} I_{t_i} \end{bmatrix} \quad (1)$$

### 5.3 Mapping Phone Parameters to Virtual Cursor

We set the interaction space dimensions as $600 \times 440 \times 400$ mm, such that users of varying arm lengths can easily access different regions. The modeling space dimensions were set as $14 \times 7 \times 8$ units to ensure visibility of scene elements and minimize visual clutter. Given the spatial correspondence between the two spaces, the phone's 3D position can be linearly mapped onto the cursor with a scaling factor of 0.023 (5mm phone motion maps into 0.12 unit cursor displacement).

The Android SDK provides the phone's orientation in terms of roll, pitch, and azimuth (Figure 6 (a)). By having the depth sensor face downwards, we constrain the interaction space's X-Y directions to be equivalent to a horizontal plane. This allows us to directly use the roll and pitch readings, given that they are measured with respect to the horizontal. The azimuth reading however is measured with respect to magnetic north and thus requires a map between the phone's global frame and the interaction space. For this, we apply a calibration step before each session. Here, the phone is placed on the desk with its top face up and its major axis roughly aligned with the Y-direction (Figure 6 (b)). The azimuth value in this orientation (averaged across 30 successive readings) is then set as the constant offset angle, $\theta$. This angle is subtracted from each azimuth reading in order to read it in terms of the interaction space.

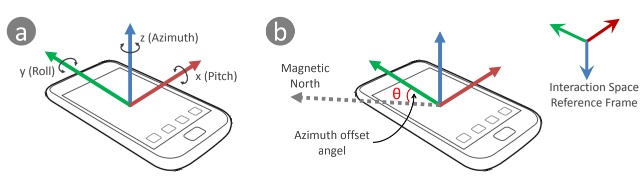


Figure 6: (a) Phone orientation with respect to global frame, (b) Calibrating azimuth offset angle.

The cursor in the 3D scene has its default location at the origin and lies parallel to the scene's horizontal desk. In each frame, we first map the phone's orientation onto the cursor, using the rotation matrix $\mathbf{M} = \mathbf{A} * \mathbf{P} * \mathbf{R}$, where $\mathbf{R}$, $\mathbf{P}$, and $\mathbf{A}$ represent rotation matrices about the interaction space's Y, X, and Z directions by the roll, pitch, and adjusted azimuth angles respectively. The phone's position ($\mathbf{v}_{phone}$) is then mapped as the cursor's position ($\mathbf{v}_{cursor}$) using, $\mathbf{v}_{cursor} = 0.023 * \mathbf{v}_{phone}$

### 5.4 Axial Deformation of Shapes

For shape modifications, we use an axis based mesh deformation technique [21]. This approach is suitable for our context given its computational efficiency and robustness against 3D mesh artifacts. Here, each discrete point $\mathbf{p}_k$ on the axis is assigned a local Frenet frame, comprising of a tangent ($\mathbf{t}_k$), normal ($\mathbf{n}_k$), and binormal ($\mathbf{b}_k$).

$$\mathbf{t}_k = \frac{\mathbf{p}_{k+1} - \mathbf{p}_k}{|\mathbf{p}_{k+1} - \mathbf{p}_k|} \qquad \mathbf{b}_k = \frac{\mathbf{t}_{k-1} \times \mathbf{t}_k}{|\mathbf{t}_{k-1} \times \mathbf{t}_k|} \qquad \mathbf{n}_k = \mathbf{b}_k \times \mathbf{t}_k$$

During shape deformation (Figure 4 (b)), users can clutch and freely manipulate one of the axial endpoints ($\mathbf{p}_0$), while the other ($\mathbf{p}_n$) is kept fixed. The displacement of point $\mathbf{p}_i$ between the endpoints is then computed using a cubic function $\mathbf{d}_i = \mathbf{D}.(1 - u_i^3)$, where $\mathbf{D}$ is $\mathbf{p}_0$'s spatial displacement and $u_i$ is $\mathbf{p}_i$'s parametric distance from $\mathbf{p}_0$. At each incremental displacement, the Frenet frames of the axis points are updated.

A vertex on the 3D mesh is displaced based on the transformation of its closest axis point $\mathbf{p}_i$. Here, if $\mathbf{v}_j$ is the position vector of the mesh vertex with respect to $\mathbf{p}_i$'s Frenet frame, $\mathbf{v}_j$ has three direction cosines ($\alpha$, $\beta$, $\gamma$) defined within that frame. When $\mathbf{p}_i$ gets displaced, the mesh vertex $\mathbf{v}_j$ is also transformed such that its new position vector $\mathbf{v}'_j$ maintains the same direction cosines and magnitude in the updated Frenet frame.

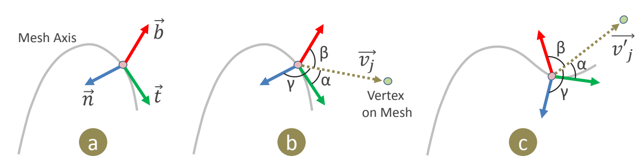


Figure 7: (a) Frenet frame of axis point, (b) Position vector of a mesh vertex, (c) Displacement of mesh vertex based on axial deformation.

## 6 User Evaluation and Results

We conducted a user study with 11 participants (7 male, 4 female) to evaluate *RealFusion's* support towards early-stage creative ideation. Here, we recruited undergrads from industrial design and engineering, as they represent novice designers with basic exposure to 3D modeling. None of them were familiar with mid-air interactions. Each user was first trained on the interface (15 min), and then assigned two tasks. In Task 1, they both conceptualized and constructed a design idea of their own choosing, while browsing the web for creative inspiration. Task 2 involved a 3D brainstorming activity (21 min), with a focus on quickly externalizing multiple forms of a specific design context. We provided users with 17 objects, but they could bring along personal items or look for other objects in the lab. Due to time constraints, we utilized the *quick scanning system*. We document user experiences with video recordings and surveys.

Drawing from principles in creativity support [25], shape composition workflows [3, 8], and design ideation [28], we frame our observations based on the following factors.

### 6.1 Creative Expressiveness

Most users were receptive to the notion of repurposing everyday objects within design mock-ups, and could construct identifiable models to express different ideas (Figures 8, 12). They commented that *RealFusion* offered *"encouragement of imagination through the use of existing parts"* and enabled them to *"come up with cool ideas just by putting simple forms together"*. We also observed the workflow to stimulate user engagement: *"I found myself designing things I wouldn't think of because it helped spark ideas when I was putting things together and rotating them in 3-D space."*

While some users expressed that *"the final outcomes were rougher than expected"*, they *"would not consider it (lack of precision) to be harmful"*, given *RealFusion's* role as *"an initial concept design tool which does not involve high precision modeling."* In

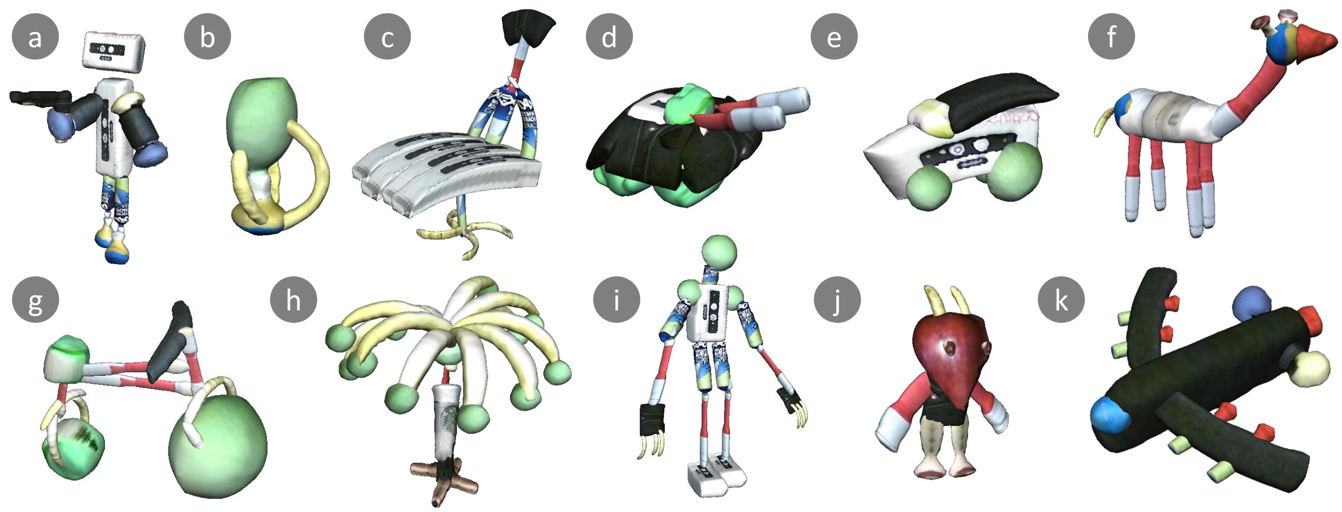


Figure 8: Design Models created by novice users in the user study.

| Model | Completion Time (minutes) | | | | Number of Objects Used | Number of Components |
|---|---|---|---|---|---|---|
| | Scan | Modify | Compose | Total | | |
| a | 2.9 | 1.0 | 6.8 | 10.7 | 5 | 9 |
| b | 4.4 | 2.4 | 8.5 | 15.3 | 3 | 4 |
| c | 2.5 | 2.2 | 12.4 | 17.1 | 5 | 14 |
| d | 1.6 | 1.35 | 7.3 | 10.3 | 4 | 12 |
| e | 2.2 | 2.4 | 16.5 | 21.1 | 4 | 7 |
| f | 2.1 | 1.9 | 12.6 | 16.6 | 6 | 12 |
| g | 2.0 | 0.8 | 7.1 | 9.9 | 5 | 14 |
| h | 2.3 | 08 | 11.9 | 22.2 | 5 | 29 |
| i | 2.2 | 2.1 | 15.5 | 19.8 | 6 | 23 |
| j | 2.0 | 1.3 | 10.0 | 13.3 | 5 | 10 |
| k | 2.7 | 2.1 | 21.1 | 25.9 | 5 | 14 |
| Average | 2.44 | 2.1 | 11.8 | 16.56 | | |

Figure 9: Task performance during model construction.

fact, most users found their results to reasonably match their mental image. When selecting objects in their design, users primarily focused on object geometry, but some also incorporated object texture and functionality to enhance their design (robot face and gun in Figure 8 (a)). They also found that textures *"added color to (their) designs."*, and helped in identifying scanned objects.

### 6.1.1 Multiple Styles for Shape Creation

In addition to the *scan-modify-compose* workflow, we found users exploring their own creative styles to (a) achieve flexibility in modeling, (b) better express design intent, and (c) produce complex geometries. These styles can be classified as follows.

**Collective Scanning** of different objects to generate shapes not available in the surroundings (Figure 10 (c)) or to compose a subassembly in the physical space before digitization.

**Creating Sweep Geometries** by arranging identical shapes along a straight or a curved path to generate extrusions (Figure 12) or organic surface designs (Figure 8 (c)) within a model.

**Patterning** identical shapes around a fixed reference for symmetricity (Figure 8 (i), (k)) and regularity (Figure 8 (h)) of components within the design.

**Scanning a Single Object in Different Views** to generate shapes with diverse forms based on scanning direction (Figure 10 (a-b)).

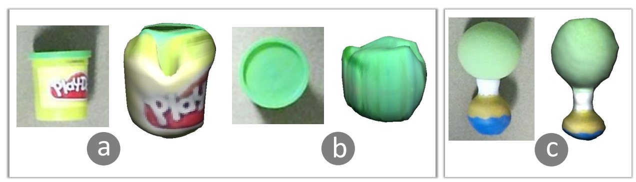


Figure 10: (a-b) Scanning a common object along two different orientations to get different shapes. (c) Scanning two objects together as a composite shape.

## 6.2 Design Exploration

Users found shape modification invaluable for better expressing their design intent: *"(it was) needed to make the (virtual) object look more like the object in mind."* As one user expressed *"I was able to find different ways to distort the objects which allowed for a large amount of design possibilities from a simple set of objects."* Users also indicated that real-time view manipulations (Figure 4(c) bottom-right) allowed them to visually inspect and reflect upon ideas for ensuing iterations: *"(it) helped me see what I have and fill in any holes that might be needed to create better ideas."* We found that the ability to observe ideas from multiple perspectives also stimulated users' creative cognition: *"new shape forms and ideas came to mind as I rotated the models,"*; *" I was able to add on changes that I had not previously thought of."*

### 6.2.1 Multiple Paths for Exploration

Using color-coded plots to represent trends in user behavior, we observed three distinct patterns in how users approached our workflow. In Figure 11, we show one example of each pattern.

**Structured Approach (SA)** Some users rigidly followed the *scan-modify-compose* sequence, by first conceptualizing an idea in their mind and proceeding to build it. They devoted chunks of time towards specific tasks and rarely transitioned across different states.

**Planned Exploration (PE)** Most users first scanned random objects that looked interesting, and then explored multiple ideas in the 3D scene by examining different forms and configurations. They also frequently switched between *compose* and *modify* states.

**Free Exploration (FE)** Several users followed completely unstructured paths by freely transitioning between different states and showing a willingness to test impulsive ideas. We found them to *reflect-on-action* by frequently combining results and insights from different ideas into a convergent form.

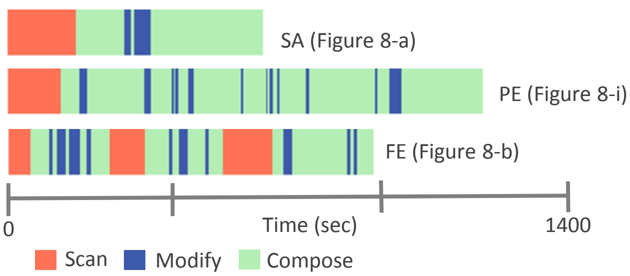


Figure 11: Timeplots showing user activity during design.

#### 6.2.2 Serendipitous Discoveries

Based on user feedback, it was evident that serendipity played a significant role in the final outcome for most designs. As one user noted, *"I went into each design with an open mind. Each outcome was unexpected and I got good things from them."* The system usage patterns showed that most users did not have a clear image of what they wanted at the beginning of the tasks. However, after tinkering with coarse ideas, they progressively developed and converged to a final solution: *"I usually started out with a rough picture in my imagination and when I got to each piece, I was able to just wing it and adjust the shape I chose as needed."* We find that serendipitous findings is largely attributable to *RealFusion's* support towards free explorations.

#### 6.2.3 Design Ideation & Creative Diversity

7 users completed the brainstorming task (Task 2) in the 21 minute time limit, while 4 needed an additional 5 minutes. Given the time constraint, users minimized scanning time by using fewer objects, and instead focused on *modify* and *compose* operations to explore different ideas. They also frequently used *Patterning* and *Sweep Geometries* to efficiently define complex shapes and structural regularity. Most users could explore divergent forms within a particular design context, and found themselves leveraging insights from earlier iterations into latter designs: *"when I was working on one design, I got new ideas for using other objects to create something different."* Figure 12 shows three sets of diverse forms within three design contexts, each set generated by a distinct user.

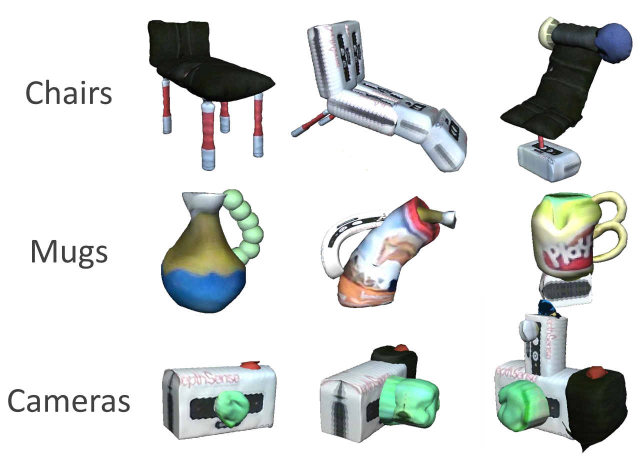


Figure 12: Exploring different forms of a design context.

### 6.3 Supporting Ideation in Found Art

While *RealFusion* was inspired by *found-object-art*, we were also interested in evaluating its implications towards the artform itself. For this, we tested our interface with 2 artists, experienced in constructing physical 3D collages. We found that while they were unsure about a digital system's utility in traditional artistic expression, they saw its potential in helping artists visually explore early ideas and better plan their artwork. The artists mentioned: *"this would save me going through my mind three-four times to make sure I have everything"*, and *"(this) can be helpful in laying out basic ideas to save time"*. One artist alluded that the system could also be useful for exploring solutions that are difficult to discover using the physical process: "Sometimes it (artwork) is missing something, and if I had something like this, it could greatly help me find out what it is rather than scratching my head". In addition, they also felt that an interactive digital tool could assist in art education settings, where time and resources are constrained.

### 6.4 Limitations

While mid-air gestures enabled direct and efficient 3D placement of shapes, they also cause arm fatigue particularly during fine level manipulations. Thus, it makes sense to seek a middle ground in our interactions, where users can quickly configure a shape's placement using mid-air gestures and then fine tune its orientation using multi-touch gestures only. Even though most users could easily sense the 3-dimensionality of the modeling space through the shadows and perspective view, a few users experienced difficulty with depth perception on the flat screen. In enhance the 3D immersive experience, we could also use a head-mounted AR display that renders shapes directly on the interaction space. Due to time constraints, we were also unable to fabricate user created models and assess their utility.

## 7 Discussions

**Low Thresholds and Wide Walls.** While *RealFusion* was primarily intended as a *scan-modify-compose* workflow, our study revealed that it could also support a variety of other creative mechanisms for constructing and exploring 3D designs. We believe that its simplicity allowed users to customize their approach for better expressing design intent and to achieve unique outcomes. Users also found such simplicity to stimulate creativity: *"compared to CAD programs, it forced me to think of different ways to make interesting things"*. As a result, multiple creative pathways emerged that were previously not apparent. Given that there is no single fixed pattern for the way designers generate ideas [6], such flexibility was revealed as an essential component in *RealFusion*. Here, its low thresholds made *RealFusion* easy to use, while its "wide walls" enabled users to explore interesting possibilities [25].

**Perceived Utility.** Both designers and artists found value in *RealFusion* towards early-stage creative ideation and brainstorming activities. Given their background in design, users saw its potential in enabling quick construction of design mock-ups for visual observation, inspection, and assessment of mental ideas. Some users mentioned that in contrast to traditional sketch-based ideation, *RealFusion* served as a more effective means for externalizing ideas without the need for specialized skills: *"it allows me to get across design concepts, even with my limited drawing skills"*; *"it helps avoid mis-communications that may come with flat drawings."* Given the variety of concepts produced both between and within different contexts and users, we find *RealFusion* conducive to divergent thinking, which is fundamental to early-stage design [28]. Similarly, its support of serendipitous findings allow users to freely explore the design space to uncover unexpected yet valuable results [3].

**Broader Implications.** While we used the smartphone as a purely mid-air input modality, we find a broader context for its use within *RealFusion*. For example, by leveraging the phone's GUI interface, we could enable advanced operations like precise shape manipulation, more complex shape modifications, and web based 3D shape retrieval. This could improve the quality of the resulting models and the creative outcomes, while pushing *RealFusion* towards design of functional artifacts that can be 3D printed. With advancements in 3D tracking techniques, there is also a scope for directly controlling scanned shapes in the modeling space by holding and moving their physical counterparts. This can significantly enhance the sense of realism and physical engagement for users. Given our results, we also find it promising to study *RealFusion* in other contexts such as education tools, multimedia, architecture, and engineering. This will allow us to evaluate *RealFusion's* utility beyond novice designers and artists, and with other demographics like youths, professional designers, animators etc.

## 8 Conclusion

We presented *RealFusion* as an interactive workflow for visually exploring early-stage design ideas in a digital 3D medium. To this end, we demonstrated a wide range of creative possibilities supported by *RealFusion*, particularly in context of design ideation. We hope that it leads to further exploration of computer supported and human-centric creativity tools that merge physical artifacts within digital settings in new ways.


## Acknowledgements

We thank the reviewers for their valuable feedback. This work was supported by NSF CMMI-ESD (#1538868) and NSF CPS:Synergy (#1329979). Any opinions, findings, and conclusions, or recommendations in this material are those of the author(s) and not necessarily the views of the National Science Foundation.


## References


[1] S. Ahmed, K. M. Wallace, and L. T. Blessing. Understanding the differences between how novice and experienced designers approach design tasks. *Research in engineering design*, 14(1):1–11, 2003.

[2] D. Anderson, J. L. Frankel, J. Marks, A. Agarwala, P. Beardsley, J. Hodgins, D. Leigh, K. Ryall, E. Sullivan, and J. S. Yedidia. Tangible interaction+ graphical interpretation: a new approach to 3d modeling. In *Proc. of CGIT*, pages 393–402. ACM, 2000.

[3] W. Benjamin, S. Chandrasegaran, D. Ramanujan, N. Elmqvist, S. Vishwanathan, and K. Ramani. Juxtapoze: supporting serendipity and creative expression in clipart compositions. In *Proc. of SIGCHI*, pages 341–350. ACM, 2014.

[4] H. Benko, R. Jota, and A. Wilson. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proc. of CHI 2012*, pages 199–208. ACM, 2012.

[5] J. Bouguet. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, 5:1–10, 2001.

[6] L. Candy and E. Edmonds. Creative design of the lotus bicycle: implications for knowledge support systems research. *Design Studies*, 17(1):71–90, 1996.

[7] S. Follmer, D. Carr, E. Lovell, and H. Ishii. Copycad: remixing physical objects with copy and paste from the real world. In *Proc. of UIST*, pages 381–382. ACM, 2010.

[8] S. Follmer and H. Ishii. Kidcad: digitally remixing toys through tangible tools. In *Proc. of SIGCHI*, pages 2401–2410. ACM, 2012.

[9] R. Gal, O. Sorkine, T. Popa, A. Sheffer, and D. Cohen-Or. 3d collage: expressive non-realistic modeling. In *Proc. of NPAR*, pages 7–14. ACM, 2007.

[10] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proc. SIGCHI 1994*, pages 452–458. ACM, 1994.

[11] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proc. of UIST*, pages 559–568. ACM, 2011.

[12] J. Jerald, P. Mlyniec, A. Yoganandan, A. Rubin, D. Paullus, and S. Solotko. Makevr: A 3d world-building interface. In *Proc. of 3DUI*, pages 197–198. IEEE, 2013.

[13] B. Jonson. Design ideation: the conceptual sketch in the digital age. *Design studies*, 26(6):613–624, 2005.

[14] R. Jota and H. Benko. Constructing virtual 3d models with physical building blocks. In *Proc. of SIGCHI*, pages 2173–2178. ACM, 2011.

[15] E. Kalogerakis, S. Chaudhuri, D. Koller, and V. Koltun. A probabilistic model for component-based shape synthesis. *TOG*, 31(4):55, 2012.

[16] N. Katzakis, R. J. Teather, K. Kiyokawa, and H. Takemura. Inspect: extending plane-casting for 6-dof control. *Human-centric Computing and Information Sciences*, 5(1):1–22, 2015.

[17] K. Kim, J. Kim, J. Choi, J. Kim, and S. Lee. Depth camera-based 3d hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions. *Sensors*, 15(1):1022–1046, 2015.

[18] M. J. Kim and M. L. Maher. The impact of tangible user interfaces on designers' spatial cognition. *Human-Computer Interaction*, 23(2):101–137, 2008.

[19] M. Lau, M. Hirose, A. Ohgawara, J. Mitani, and T. Igarashi. Situated modeling: a shape-stamping interface with tangible primitives. In *Proc. of TEI*, pages 275–282. ACM, 2012.

[20] M. Lau, G. Saul, J. Mitani, and T. Igarashi. Modeling-in-context: user design of complementary objects with a single photo. In *Proc. of SBIM*, pages 17–24. Eurographics Association, 2010.

[21] F. Lazarus, S. Coquillart, and P. Jancne. Axial deformations: an intuitive deformation technique. *Computer-Aided Design*, 26(8):607 – 613, 1994.

[22] M. Mine, A. Yoganandan, and D. Coffey. Making vr work: building a real-world immersive modeling application in the virtual world. In *Proc. of SUI*, pages 80–89. ACM, 2014.

[23] L. Olsen and F. Samavati. Image-assisted modeling from sketches. In *Proc. of GI*, pages 225–232. CIPS, 2010.

[24] R. Ramakers, F. Anderson, T. Grossman, and G. Fitzmaurice. Retrofab: A design tool for retrofitting physical interfaces using actuators, sensors and 3d printing. In *Proc. of SIGCHI*. ACM, 2016.

[25] M. Resnick, B. Myers, K. Nakakoji, B. Shneiderman, R. Pausch, T. Selker, and M. Eisenberg. Design principles for tools to support creative thinking. *Creativity Support Tools: A workshop sponsored by the NSF*, 2005.

[26] V. Savage, S. Follmer, J. Li, and B. Hartmann. Makers' marks: Physical markup for designing and fabricating functional objects. In *Proc. of UIST*, 2015.

[27] J. Schwarz, C. C. Marais, T. Leyvand, S. E. Hudson, and J. Mankoff. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proc. of SIGCHI*, pages 3443–3452. ACM, 2014.

[28] J. J. Shah, S. M. Smith, and N. Vargas-Hernandez. Metrics for measuring ideation effectiveness. *Design studies*, 24(2):111–134, 2003.

[29] A. Sharf, M. Blumenkrants, A. Shamir, and D. Cohen-Or. Snappaste: an interactive technique for easy mesh composition. *The Visual Computer*, 22(9-11):835–844, 2006.

[30] S. Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

[31] E. van den Hoven and A. Mazalek. Grasping gestures: Gesturing with physical artifacts. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 25(03):255–271, 2011.

[32] E. Varga. *Using hand motions in conceptual shape design: theories, methods and tools*. TU Delft, Delft University of Technology, 2007.

[33] D. Vinayak, Ramanujan, C. Piya, K. Ramani, et al. Mobisweep: Exploring spatial design ideation using a smartphone as a hand-held reference plane. In *Proc. of TEI*, pages 12–20. ACM, 2016.

[34] G. Voland. *Engineering by design*. Pearson Education India, 1999.

[35] C. Weichel, M. Lau, D. Kim, N. Villar, and H. W. Gellersen. Mixfab: a mixed-reality environment for personal fabrication. In *Proc. of SIGCHI*, pages 3855–3864. ACM, 2014.