

Ananya Ipsita*¹

School of Mechanical Engineering,
Purdue University, IN, USA
email: aipsita@purdue.edu

Runlin Duan*

School of Mechanical Engineering,
Purdue University, IN, USA
email: duan92@purdue.edu

Hao Li*

Department of Mechanical Engineering,
Stanford University, CA, USA
email: li2053@stanford.edu

**Subramanian
Chidambaram**

School of Mechanical Engineering,
Purdue University, IN, USA
email: subbu10123@gmail.com

Yuanzhi Cao

School of Mechanical Engineering,
Purdue University, IN, USA
email: yuca@ptc.com

Min Liu

School of Mechanical Engineering,
Purdue University, IN, USA
email: liu66@purdue.edu

Alex Quinn

School of Electrical & Computer
Engineering,
Purdue University, IN, USA
email: aq@purdue.edu

Karthik Ramani

School of Mechanical Engineering,
Purdue University, IN, USA
email: ramani@purdue.edu

The Design of a Virtual Prototyping System for Authoring Interactive VR Environments from Real World Scans

Domain users (DUs) with a knowledge base in specialized fields are frequently excluded from authoring Virtual Reality (VR)-based applications in corresponding fields. This is largely due to the requirement of VR programming expertise needed to author these applications. To address this concern, we developed VRFromX, a system workflow design to make the virtual content creation process accessible to DUs irrespective of their programming skills and experience. VRFromX provides an in-situ process of content creation in VR that (a) allows users to select regions of interest in scanned point clouds or sketch in mid-air using a brush tool to retrieve virtual models, and (b) then attach behavioral properties to those objects. Using a welding use case, we performed a usability evaluation of VRFromX with 20 DUs from which 12 were novices in VR programming. Study results indicated positive user ratings for the system features with no significant differences across users with or without VR programming expertise. Based on the qualitative feedback, we also implemented two other use cases to demonstrate potential applications. We envision that the solution can facilitate the adoption of the immersive technology to create meaningful virtual environments.

Keywords: Virtual Reality, Embodied interaction, Point cloud interaction, Behavioral modelling, Scene reconstruction, Scene manipulation, Human-centered AI, Graphical user interface

1 Introduction

We are living in an era where substantial efforts are being made to create seamless pathways between the physical reality and its digital counterpart. Due to its tremendous potential to capture the reality in authentic self, the digital counterpart can be made useful in a variety of applications, such as remote collaboration, hologram teleconferencing, and simulated training [1–3]. Consequently, the past decade has witnessed the emergence of Virtual Reality (VR) technologies and applications in various domains, e.g. manufacturing, construction, sports, etc. that provide the domain users (DUs) with realistic and immersive platforms to explore and interact with the digital counterparts of the realities [4]. Despite the spatial affordances provided by the immersive VR applications in allowing situated and hands-on experiences for the DUs, there is high barrier for adoption of these technologies in domain specific fields [5]. This is partly due to lack of accessible methods for the DUs to author these applications for their respective domains.

Conventional methods of digital content creation for VR applications are tedious and fragmented, involving stacks of various technologies e.g., VR programming software and technical modelling [6,7]. The creation process is usually limited to the hands of experts in VR programming (EVRPs), and thereby alienates novices in VR programming (NVRPs) from the authoring process [5]. Most importantly, the requirement of VR programming expertise creates an entry barrier for the DUs, who are oftentimes also found to be NVRPs, to author field-specific virtual applications. Considering the above problems that the DUs face to utilize the technology to their benefit, our work is focused on designing an accessible system workflow that can assist DUs in authoring interactive VR environments from real world scans without needing technical expertise.

To make the scene reconstruction more accessible to NVRPs, prior works have introduced numerous authoring systems for VR applications [8–10]. The idea is to provide users with a modelling tool that effectively uses the reality scene as a modelling reference for derivative scene reconstruction with added interactive functionalities. However, the amount of user interaction and

¹Corresponding Author.

*These authors contributed equally to this work



Fig. 1 Workflow of VRFromX (from left to right). (1) User scans a real environment to get the raw point cloud of the scene using an iPadPro. (2) User creates virtual content by interacting with the point cloud through AI assistance. (3) User attaches real functions to the virtual objects using an Affordance Recommender. (4) User interacts with the reconstructed scene in a virtual environment.

technical modelling knowledge required is still substantial. On the other hand, recent works in the artificial intelligence (AI) area have been using deep learning to automatically reconstruct the digital scene from 3D scan of the real-world [11–15]. However, these automated approaches usually lack the ability to provide user-in-the-loop customization. Inspired by the previous works, we choose an approach that balances user interaction with AI automation.

Because of the state-of-the-art sensing hardware and computer vision algorithms, real-world digitization can be easily done by commercially available 3D scanners and RGB-D cameras [16]. The results from these devices are usually the point cloud or the mesh model reconstruction of the reality which provides a unique benefit of laying out the objects reasonably with realistic dimensions and unaltered spatial relationships. However, they can only serve as a spatially situated visual reference for the user without any interactive functionalities. In order to preserve the interactivity of the scanned real-world scene for a desired VR experience, it requires professional developers to manually recreate the virtual scene while using the scanned results as a background modelling reference [17]. Accompanied by the affordances provided by our system, users can spatially interact in these visually-immersive virtual worlds in an embodied fashion [18,19] for authoring applications.

There are a few different interface platforms for users to interact with the digital version of the real world, including 2D (as in mobile, tablet and desktop-based interfaces) and 3D (as in VR-based interfaces). In this paper, we focus on VR due to its capability to provide immersive experiences while supporting in-situ spatial and embodied interactions from the users [18]. The choice of 3D platform affords better 3D perception and direct testing of the authored content within the execution environment, without needing to switch between 2D and 3D interfaces. Moreover, 2D interfaces bring interaction issues while adding the extra dimension in unnatural and unconventional ways, and restrict an effective “flow” of using hands to work in virtual environments [20]. Considering these advantages of VR, our work is focused on creating an accessible and immersive system for users to transform physical reality into meaningful virtual environments.

We present VRFromX, an end-to-end system design that supports the authoring of interactive VR scenes from real-world scans. Our workflow is illustrated in Figure 1: the user starts from scanning a real-world scene (a welding workshop in this case) into point cloud using a handheld 3D scanner (Figure 1-1); then replaces the point cloud objects with their corresponding virtual models via embodied interaction with AI-assisted object retrieval (Figure 1-2); the user then attaches functionality onto the virtual objects and defines logical connections between them (Figure 1-3); finally the virtual scene is ready for VR interactions (this particular scene can be used for virtual welding training) (Figure 1-4). The contributions of this paper are as follows:

- The system workflow design of VRFromX that enables the creation of interactive VR scenes, by scanning real-world environments into point cloud scans, then transforming the point

cloud objects into virtual 3D models, and establishing functionalities and logical connections among virtual contents.

- The design of user interaction that supports point cloud segmentation and editing via embodied interactions, the AI assistant that guides the object retrieval and alignment, and the spatial and visual interface for functionality and logic authoring.
- A detailed system development of a welding use case scenario, evaluation results obtained from a three-session user study, in addition to two example use case demonstrations to show the breadth and nature of potential applications of VRFromX.

In order to evaluate the usability of the system, we conducted a user study using a welding use-case with 20 DUs, comprising of 12 NVPRs and 8 EVRPs. Through the findings of the user study, we wanted to answer the following research questions:

- To what extent does the users’ expertise in VR programming affect their performance and experiences in using VRFromX to author virtual environments?
- How does the domain knowledge of users affect their perception and experience in using VRFromX to author virtual environments?

Our study results indicated high user ratings about the accessibility, engaging and intuitive aspects of the system to author virtual experiences. A comparative evaluation across the EVRPs vs. NVPRs indicated similar ratings towards perception and experiences in using the system, across the two groups with no significant differences. Two other use cases in the (1) 3D printing and (2) Robot-IoT domains were further explored to envisage the potential applications of VRFromX.

2 Related Work

2.1 Authoring tools for Scene Reconstruction and Manipulation. Traditional 3D reconstruction techniques that targeted transfer of real world scenes to virtual environments have mostly lived in desktop-based environments. Computer vision-based methods used in these environments have explored reconstructing virtual scenes using RGB/RGB-D images [21–28], point clouds [29–32], and 360° images [33]. However, these computer vision techniques are usually not adaptive to users’ interaction and the 2D interfaces also limits immersion from the users. On the other hand, prior work have highlighted the immersive affordance provided by the 3D platforms [34] that has been used in VRFromX to facilitate intuitive interaction from users while leveraging their spatial reasoning skills during the scene reconstruction.

Over the past few years, research works have deployed 3D reconstruction techniques to automatically reconstruct virtual scenes within VR using scanned RGB-D information [35–39]. Reality Skins [35] uses consumer-level RGB-D cameras to pre-scan and process the user’s physical surrounding to translate it directly into

a virtual environment with the best object layout. Oasis [36] procedurally generates an interactive VR experience using pre-scanned data from the physical environment as a template. By allowing the seamless transition between the physical and virtual realities, RealityCheck [37] selectively shows real-world objects as part of a situated VR experience. Dreamwalker [38] leverages RGB-D sensors, Windows Mixed Reality tracking, and GPS to substitute real-world walking experience with a dynamic virtual environment. VRoamer [39] scans the physical environment in real-time and dynamically instantiates pre-authored virtual experiences that fit the physical world. These systems mainly focus on employing RGB-D information to provide realistic immersion within virtual spaces, ensure safe walking in the physical world and construct pre-authored virtual environments. However, the above construction techniques do not provide interaction capabilities for the user to create or modify the authored environments, thereby restricting their abilities for prototyping purposes. Moreover, these techniques result in the creation of virtual environments without any semantic meaning associated with the virtual objects, and thus do not facilitate any interaction from the end-users. On the other hand, VRFromX considers the RGB-D information as a medium where users can interact with to transform static point clouds into virtual experiences, which allows user involvement to create meaningful experiences.

In the context of object manipulation in the reconstructed scenes, authoring tools have been developed that allows users to capture and manipulate scene scans of physical environments [40–47]. Remixed Reality [40] performs a live 3D reconstruction with multiple external RGB-D cameras and provides direct manipulation of the reconstructed virtual scene. AffordIt! [41] offers a solution to assign affordances and constraints to virtual objects which do not have intrinsic behaviors after 3D reconstruction and segmentation. Prouzeau et. al. [42] developed an authoring tool for building management using digital twins in VR where users can design situated and embedded AR visualizations to 3D models, and later, view those visualizations in AR in the physical space. ARchitect [10] reconstructs an interactive virtual scene by scanning the physical environment and aligning virtual proxies to physical objects with similar affordances. DistanciAR provides an authoring interface that allows the user to scan an environment at a remote site, add and manipulate virtual objects and create AR experiences for the remote environment [43]. PhotoShopAR allows users to quickly prototype design ideas in a spatial context using point cloud editing operations and later experience the design via AR [46]. ScalAR allows users to create AR experiences and define semantic behaviors of AR contents with the help of a VR authoring interface, while using existing scenes as semantic references [44]. However, the manipulation techniques offered by the above authoring tools are specifically restricted to the spatial affordances provided by the virtual objects by virtue of their position and location in the environment and thus limits the semantic representation of the reconstructed scene. Although VRFromX bears some similarities with the above tools in the mode of reconstructing the virtual scene, the nature of scene manipulation techniques offered by the proposed system is strikingly different. In contrast to the above methods, VRFromX allows users to manipulate and interact with the reconstructed scenes to facilitate the prototyping of context-aware and interactive virtual experiences. A logic authoring interface is employed using which users can link different virtual objects in the scene and embed contextually relevant functionalities to associate semantic meaning to the virtual objects.

In the context of embedding interactive functionalities in virtual scenes, prior work have also studied spatially situated authoring tools to attach logic constructs onto virtual objects [8,9,48]. Ens et al. [8] developed an application to enable end-users with limited programming skills to author Internet of Things (IoT) devices in virtual environments with multiple logic nodes. Ng et al. [9] provided a game authoring system that enables users to add virtual contents to the physical world and attach game-logics to virtual objects. Flowmatic provides an immersive authoring tool that allows

programmers to specify reactive behaviors that react to discrete events such as user actions, system timers, or collisions [48]. The tool introduces primitives for programmatically creating and destroying new objects, for abstracting and re-using functionality, and for importing three-dimensional (3D) models, while allowing their direct representation in VR. However, the above authoring tools disconnect the users from the physical reality by utilizing pre-authored virtual scenes [8] and/or still require technical experience from the users to be able to program the interactive behaviors [48]. Although the logic-authoring interface developed in VR-FromX draws insights from the above works, the system workflow differs from the prior work by providing DUs with an end-to-end workflow to create interactive and meaningful virtual experiences from real world scans.

2.2 AI-assisted User Interaction and Sketch-based Object Retrieval.

In the last decade, research works have merged automated modelling and reconstruction techniques with human interaction workflows to explore the different degrees and benefits of human and AI involvement in authoring virtual content [49–53]. While AI can aid the user by reducing the requirement of skills, time and effort during the content transformation and creation process, bringing the user in-the-loop adds more control and customization over the content design, and drives the transformation in a more goal-oriented fashion. Inspired by the prior works in this area, VRFromX combines the benefits of human-in-the-loop systems and AI techniques to create context-aware VR environments. The AI-assisted human-in-the-loop interactive capability in VRFromX is specifically utilized in the scene reconstruction where user performs sketching operations in the point cloud scans. And accordingly, the AI algorithms return the corresponding virtual objects that closely resemble the inputs from the user.

Since early times, humans have used sketching as a natural and free-form input modality to convey visual information [54,55]. Because of the easy to use and intuitive aspects of the sketch-based interactions coupled with the inherently fine-grained visual descriptions [56], sketch-based retrieval methods have witnessed a substantial interest within the computer vision community for virtual content creation. SemanticPaint [51] offers a classical 3D reconstruction pipeline in which the user can physically touch any desired object in the real world to enable real-time segmentation and labelling of scanned environments. Shao et. al. [57] proposed an interactive approach for semantic modelling of indoor scenes where the user can draw strokes on the RGB-D image of the scene to alter results from an automatic segmentation. Recent research have also explored immersive sketching in mid-air to retrieve 3D models for virtual content creation. Giunchi et al. [49] proposed a multi-view Convolutional Neural Networks (CNN) based retrieval method where free-form sketches within a virtual environment can be used as queries for searching virtual models. Zhu et al. [58] proposed a skeleton-based approach to draw 3D models in Head Mounted Displays (HMD) based virtual environments using convolution surface-based brushes. Li et. al. [59] implemented a CNN-based 3D sketch-based shape retrieval system, which allows users to freely draw 3D sketches in the air to retrieve relevant 3D models. On the same note, Luo et al. [60] studied the retrieval problem from sparse human VR sketches by training on human and synthetic sketches. BuildingSketch [61] allows users with limited modelling knowledge to draw freehand strokes in mid-air which subsequently gets processed by deep neural networks to generate detailed building models. Similar to prior work in the context of sketch-based object retrieval, VRFromX leverages a PointNet-based architecture [62] to facilitate interactive object retrieval for virtual content creation using free-form mid-air sketches and point cloud selections as query inputs inside a virtual environment.

Our current work directly builds upon our prior work published in CHI extended abstracts [63] which presented the system workflow design of VRFromX. In addition to providing a clearer description concerning the design details, justifying the design choices and carefully situating the work with respect to relevant

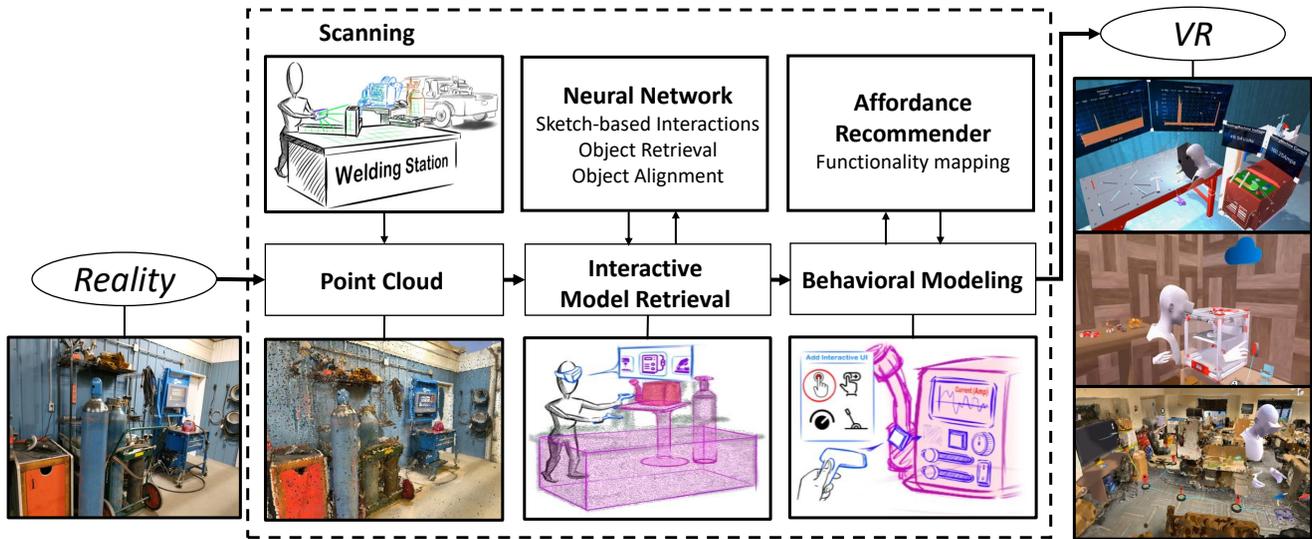


Fig. 2 Overview of VRFromX System

literature, the current version conducts a revised evaluation of the system to justify the research objectives. The study results provides novel insights on how the domain and technical expertise of the users affect their perception and performances in utilizing VRFromX to author VR experiences.

3 Design Goals

Considering the difficulties faced during the authoring process for transforming physical reality to VR, and insights gathered from relevant prior work, the following design goals of VRFromX were identified to make the system accessible to DUs while providing them with an enjoyable experience.

- **Accessibility:** System features of VRFromX should be easy to understand and use, and eliminate the need of any modelling and programming expertise. This can enable users with varied technical backgrounds and expertise levels to easily develop familiarity with the system features.

First, hardware complexity is kept low involving the use of OculusQuest™, which is chosen due to its easier configuration compared to other VR devices [64]. Next, easy adoption of the authoring process is facilitated by the use of freehand sketching on the point clouds as input modalities from the users [65], and easy-to-use drag-and-drop user interfaces for attaching affordances [66] to virtual objects. While the point cloud sketching can provide users with a visual and interactive means of interacting with the spatial data [67,68], the drag-and-drop based interfaces allow users to simply drag elements and drop them into specific areas or targets by adopting simple motor skills [69]. Furthermore, such interactions and interfaces have shown to reduce learning curve by reducing cognitive load, enhance understanding of the UI, and facilitate easy user interaction. Furthermore, the sketching and behavioral modelling tools developed in VRFromX have been designed with user friendly interfaces and features that are easy to find, use and navigate to further lower the floors in accommodating different abilities of the users during the authoring process.

- **Engaging:** System features of VRFromX should provide an immersive experience and provide opportunities for active user engagement during the authoring process. This is achieved by providing an in-situ authoring environment with gamified and dynamic interactions, and options for user customization.

Prior work have shown enhanced sense of immersion, engagement and motivation in users during authoring using VR as compared to 2D-based platforms such as desktop-based environments and tablets [70]. The 3D authoring environment in VRFromX enables free and embodied interactions from users while allowing them to directly experience the resulting application, a concept coined by Lee et. al. as WYXIWYG: What You eXperience Is What You Get [70]. The system provides a prompt interface for users to embrace natural and free-form creation methodologies such as 3D sketching in VR [49,65]. Active user engagement in the point cloud environment is supported with the AI-assistance during point cloud sketching. The sketching interface in VRFromX easily adapts to the user's choices and modifications, and works on the changes to provide desired outputs in an iterative manner. Prior work has shown the impact of such AI-assisted interactions to induce a sense of motivation and immersiveness in users by enhancing the scope of creativity and support, and promoting sustained interest during the authoring process [71].

- **Effective:** The interaction procedures offered by VRFromX should be intuitive, goal-oriented and make sense to the DUs about what they are authoring.

By bringing in the digital twin of the real world into VR, users can more closely relate to the real world tools and interactions while authoring using the system. As evident in the research conducted by prior literature [67], the point cloud data can provide a visual representation of the spatial reality, and thus make it easier for DUs to understand and analyze the complex 3D information contained within it. Using the sketching tools and AI-assisted object retrieval in VRFromX, users can (1) manually segment or edit specific points or regions on the point cloud scan to retrieve the desired virtual counterpart, and (2) personalize and/or iterate on their sketches or selections. Such interaction flows with human-in-the-loop and AI-assistance have shown to improve decision-making in users during the authoring process [71,72].

Being able to understand and relate the tasks in the virtual environment to real world activities can add more meaning to the creation process. In order to achieve this, the behavioral modelling interface employed in VRFromX allows users to attach contextually relevant properties to the virtual objects. Such context-aware authoring can enhance user motivation and confidence during the authoring process [73]. Finally, the outcomes of the authoring process i.e., the resulting VR

application should behave as expected by the DUs. After completion of authoring, the system allows the users to test and verify the virtual environments that they have created.

4 System Design and Development

We designed VRFromX to enable users to transfer a scanned point cloud scene into an interactive virtual experience (Figure 2). The major components of our system include: (1) an interactive interface for embodied interaction on point cloud for AI assisted object retrieval, and (2) behavioral modelling for virtual models. Point cloud acquisition of a real environment is performed with the help of an inbuilt LIDAR in iPadPro, using an application called *3d Scanner App*TM [74]. This raw point cloud serves as a medium for the user to interact within a virtual environment. Spatial and colour information of the point cloud helps the user in perceiving various objects present inside the scene. The user then employs a brush tool to segment regions of interest in the point cloud scene. Using the segmented point cloud as the input query, AI algorithms implemented in the back-end assist the user in retrieving corresponding 3D models. Finally, the user shifts to the behavioral modelling mode to assign functions to the virtual models with the aid of an affordance recommender.

4.1 Point Cloud Interaction and Interface. In VRFromX, we designed a system for users to interact with a scanned point cloud in an embodied manner inside a virtual environment. A brush tool which enables users to select regions on the point cloud or sketch in mid-air in a free-form manner was introduced. The various features of the brush tool (Figure 3-1) are explained as follows.

Select, *Delete*, and *Add* (Figure 3d, c, b respectively) are used for selecting, deleting and adding points in the point cloud scene respectively. *Clear All* (Figure 3e) is used to clear all selections or sketches in the scene. *Query* (Figure 3f) is used to send the point cloud selection or sketch as query input to the back-end neural networks. *Scale* (Figure 3g) is used to scale or orient the point cloud scene as per convenience. *Tip Size* (Figure 3a) and *Tip Shape* (Figure 3h) are used to change the size and shape of the *Brush Tip* (Figure 3i) respectively. To prevent confusion, the *Brush Tip* changes color on selection of different features, e.g., blue for selection, white for adding points, and red for deleting points.

After forming a clear perception of the context of the scanned scene and the objects present inside it, the users can exploit the functionalities offered by the brush tool for content creation inside the virtual environment. They can manually segment regions of interest inside the point cloud or modify them (by adding or deleting points) to get corresponding virtual models from the database. The tool also allows the users to sketch the object shape in mid-air if they do not find the objects in the partial point cloud. Keeping in consideration the accessibility of regions inside the scene, the *Scale* feature enables users to manually scale the environment or orient it to a desired setting to suit requirements using simple hand gestures.

4.2 Object retrieval and alignment using AI assistance.

Neural networks are implemented to assist the user in object retrieval and model alignment in an interactive manner. Object retrieval in VRFromX is performed as a two-step process, object classification followed by similarity check. After user selects a region of interest in the point cloud (Figure 4-2) or sketches the object shape, the point cloud selection or sketch can be fed to the back-end system as an input query. A PointNet [75] based classification network deployed in the back-end provides the user with the top five possibilities for object classes present in the database (Figure 4-3). When the user selects the desired class, the network performs a L2-distance based similarity check and returns the top five plausible results of the object models that closely resemble the input query (Figure 4-4). The L2-distance is calculated using the

feature vector of each object extracted by the classification network. The retrieval process is also adaptive to the user's selections, i.e. if the user decides to change the shapes of the input query using the brush tool features (Figure 4-5, 4-6), the AI algorithms adapts to the input and modifies the search results accordingly (Figure 4-7, 4-8). After the desired model is selected, an alignment network based on PointNet-LK [76] assists the user by automatically aligning the 3D virtual model with the region of interest according to scale, position and orientation (Figure 4-8). Users can also perform simple hand gestures to grab and alter the position, orientation and scale of the virtual models based on their requirements. Further details on the back-end implementation are provided in Section 5.

It is worth mentioning that the interaction capabilities provided by VRFromX can also enable users to re-purpose the point cloud to retrieve models of interest. In Figure 5, a user re-purposes the point cloud of a table to retrieve a 3D model of a chair. As explored in prior literature [46,77], such re-purposing can be particularly helpful in redesigning physical spaces, creativity exploration, etc.

4.3 Behavioral modelling Design and Interface. After setting the environment with the virtual models retrieved from the database, users can attach affordances to those objects to make them interactable. According to Gibson et al. [78], the term affordance is defined as the property of an object that defines its uses, or an action that can be performed over the object by an agent in an environment. In our work, we defined such behaviors of virtual objects to be categorized under three categories: *Animations*, *Displays*, and *Controls*. *Animations* and *Displays* are used to understand and visualize the different states of the virtual object(s). *Animations* are used to symbolize the various states of a virtual object in terms of motion or effects (sound, visual). As shown in Figure 6-2, the light or glow animation is used to determine the states (ON, OFF) of the virtual lamp. *Displays* are panels used to visualize data for various attributes of virtual objects. For example, the display panel in Figure 6-3 is used to display the intensity of the virtual lamp. *Controls* are used to perform actions that enable transition between the different states of the virtual object(s). These affordances can exist in the form of other virtual objects or widgets, and can trigger actions upon collisions between objects, or user manipulation of the widgets. [For more details on selection of widgets for the Controls affordances, see Section 4.3.1.] For example, the user manipulates the Switch (Figure 6-4) and Slider (Figure 6-5) *Controls* to switch ON the lamp and change its intensity respectively. Finally, *Links* are used to represent the logic connections between the virtual objects and their affordances, and fall under two categories: *Trigger Link* and *Data Flow Link*. The *Trigger Link* (Figure 6-4) represents the input connections from the *Controls* affordances to the virtual objects, whereas the *Data Flow Link* (Figure 6-3) represent the output connections from the virtual objects to the *Displays*.

We provide an Affordance Recommender for the user to choose the behavioural attributes of a virtual object. The interface design of the affordance recommender is shown in Figure 6-1. Upon grabbing the virtual model of interest, users can enable the affordance recommendation by pressing a virtual button that appears on the left hand. Object properties for specific classes are pre-defined and stored in JSON format along with the 3D models in the database. The recommender extracts these properties for the target object and provides the users with possible options. Users can grab one or more functions from the list of choices displayed by the recommender and attach them to the corresponding virtual object.

4.3.1 Controls Widgets for Behavioral modelling. Buttons, levers, and switches are ubiquitous to control physical machines, thus it would be valuable to include them as digital metaphors for *Controls* widgets in an immersive interface. The mapping affordances of the real buttons, levers, and switches can be relied upon for interactivity, and most users would already be familiar with

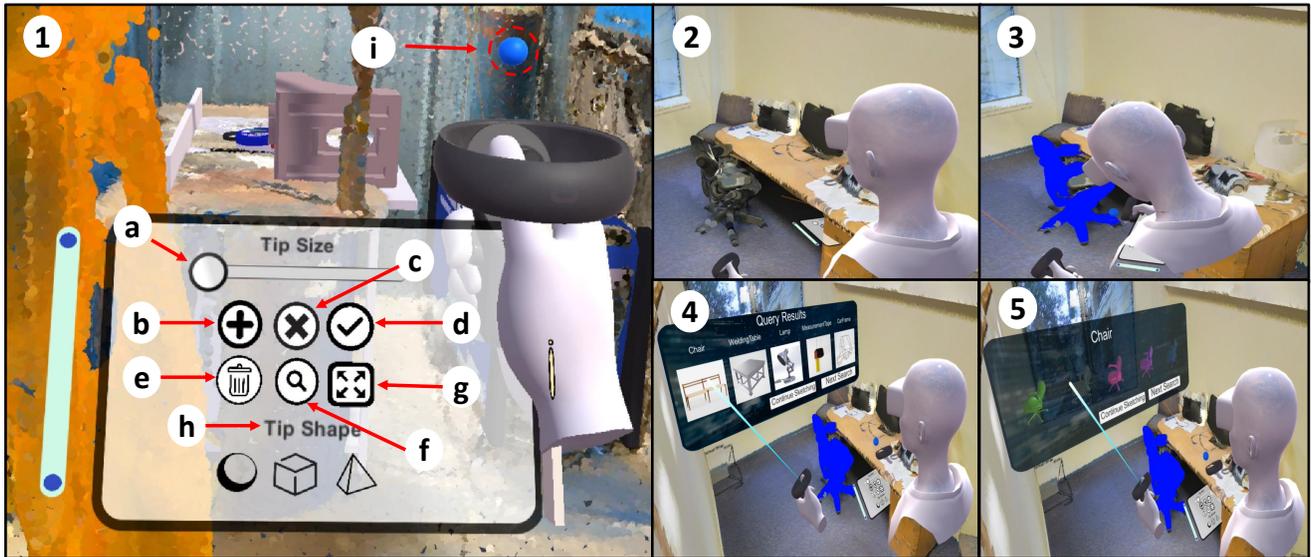


Fig. 3 Interface of the Brush Tool: (a) Tip Size, (b) Add, (c) Delete, (d) Select, (e) Clear All, (f) Query, (g) Scale, (h) Tip Shape, (i) Brush Tip. User employs the Brush Tool to retrieve a chair (2-5).

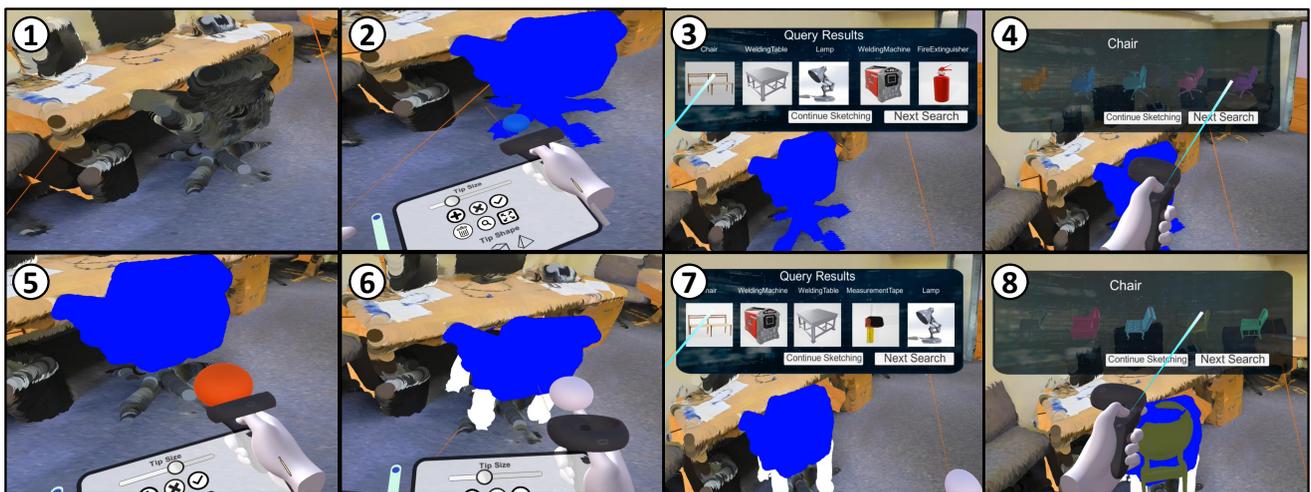


Fig. 4 Point cloud interaction for Object Retrieval (1) User identifies the region of interest and (2) selects the desired point cloud using the Brush tool. (3) AI shows top 5 object classes and (4) top 5 similar models for the input query. (5) User changes the selection by removing and (6) adding points. (7) AI provides a new set of class results as well as (8) top 5 similar models.

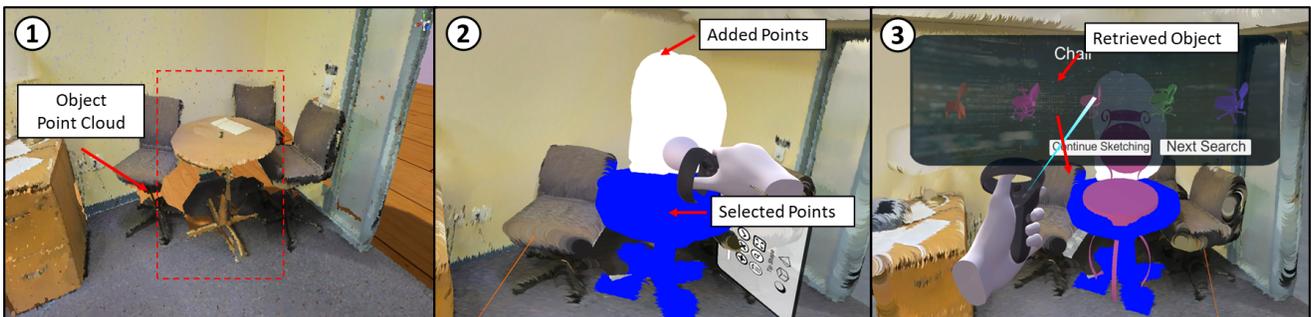


Fig. 5 (1) User defines a region of interest and (2) re-purposes the point cloud of a table (3) to retrieve 3D model of a chair.

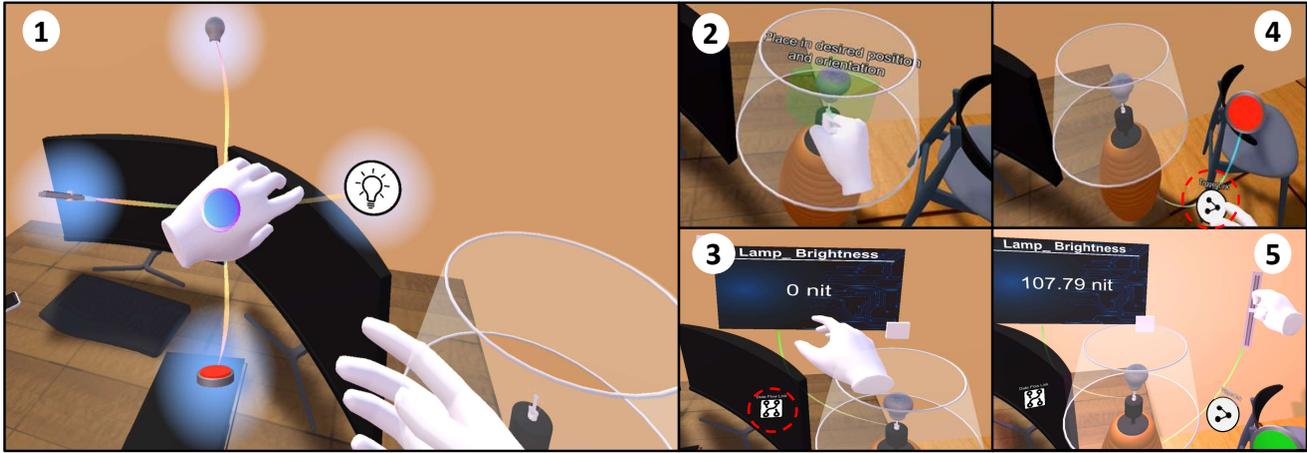


Fig. 6 User assigns functions to a virtual model of a lamp with the aid of (1) an Affordance Recommender and attaches (2) Animations - Lights, (3) Displays - Intensity Panel with Data Flow Link, (4) Controls - Switch button with Trigger Link and (5) Controls - Slider.

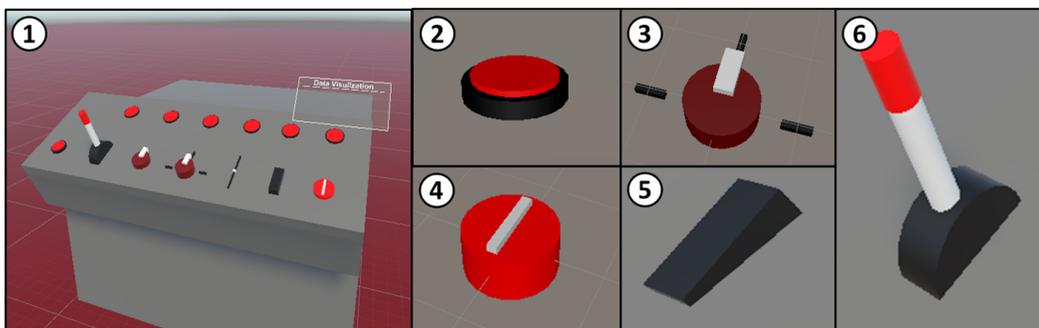


Fig. 7 Controls widgets used for Behavioral modelling in VRFromX

them. Thus, this mapping reduces the effort of learnability for the UI. To identify the most frequently used objects to model the prefabs for *Controls* widgets, we performed a small-scale study of a hundred different commonly used machines. Machines such as 3D printers, LASER cutter, lathe, household printers, CNC mills, water-jet cutter, TIG welder, miter saw, table saw, smart lights, etc. were chosen for the study, and the same machines from different companies were also identified and grouped as part of the study.

The list of physical interfaces that are frequently used such as buttons, levers, and switches were identified and categorized based on their functions. All items identified were part of one of the following three categories: *Toggle*, *Linear*, *Discrete*. A *Toggle* is a device that provides only two states of affordances (Figure 7-2, 7-5, 7-6), while a *Linear* device can provide a range of states within a continuous function with limits (Figure 7-4). A *Discrete* device provides output with more than one state but is not continuous (Figure 7-3). Each one of these categories serves a different purpose, at least one device was chosen from each category. Figure 7-1 shows the list of seven chosen button prefabs that were identified, and digital metaphors of which were developed.

5 Use case implementation

The implementation of the system workflow design was based on a welding use case scenario. Welding is a hard skill which involves a set of tools and interactions under controlled settings. We chose the welding use case to test the capability of VRFromX to author a virtual Metal Inert Gas (MIG) welding simulator. We adjusted the system requirements in terms of object models and their functions in database to enable an experienced welder for authoring the welding simulator. We created a welding dataset with object classes that can be typically found in an actual MIG-

welding setting (See Figure 8) and trained the back-end neural networks on these object classes. Object properties were defined such that the authoring process could involve replicating machine set-up (amperage/voltage), creating a realistic welding puddle with accurate sounds and movements, simulation of sparks and slag, and tracking key weld parameters, such as work angle and contact-to-tip distance [79]. In order to obtain the point cloud for the implementation, the welding station at a local welding training company was scanned using an iPadPro. A researcher walked around the scene holding the iPadPro scanning every possible detail using the *3d Scanner App*TM [74]. The process was repeated 5 times each of which took around 10 minutes and the best scan was used for the implementation. The best scan was identified by the same researcher upon subjective consideration of the visual details captured in the scans.

We used the Oculus Quest HMD with Oculus Link as well as Oculus Touch controllers to perceive the virtual experience. The system ran on a laptop PC with a Processor Intel® Core™ i7-8700K, NVIDIA GeForce RTX 2070 super graphics card and 64 GB of RAM. The raw point cloud obtained was directly rendered and used in UnityTM, and no data augmentation techniques were utilized during this process. Front-end in VRFromX system was developed within Unity EngineTM in C# and was connected with the pre-loaded back-end neural networks through a separate service thread in UnityTM.

The back-end implementation of the neural networks employed in VRFromX is as follows. The implementation of the PointNet [75] based classification network was based on a PyTorch version of the original PointNet [62]. The classification network was trained 250 epochs on the ModelNet40 dataset [80], which includes 40 classes of objects. The network was fine-tuned with 200 epochs

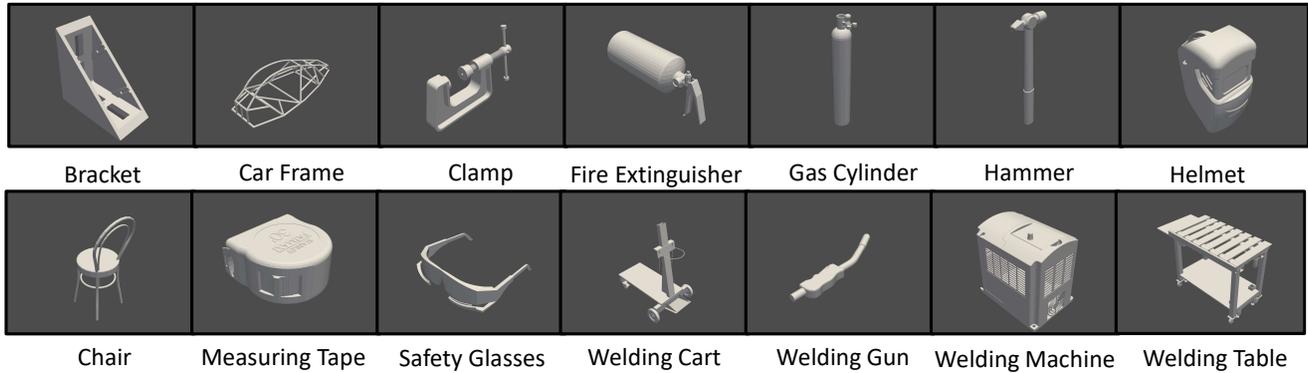


Fig. 8 14 classes in welding dataset used to train the classification network for object retrieval in VRFromX

Demographics	DUs (N=20)		NVRPs (N=12)		EVRPs (N=8)		Mann Whitney U Analysis	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	U	p<0.05
Prior welding hands-on experience [Yes = 1, No = 0]	1.00	0.00	1.00	0.00	1.00	0.00	48.00	1.00
DIY approach to your technical tasks [Yes = 1, No = 0]	0.80	0.41	0.83	0.39	0.75	0.46	44.00	0.76
Prior technical modelling experience [Yes = 1, No = 0]	0.95	0.22	0.92	0.29	1.00	0.00	44.00	0.76
Familiarity with using VR applications [1 = Much Lower, 7 = Much Higher]	3.45	1.96	2.42	1.08	5.00	2.00	14.50	0.01 (*)
Sketching/designing skills [1 = Much Lower, 7 = Much Higher]	4.65	1.23	4.42	1.38	5.00	0.93	35.00	0.32
Prior authoring experience [1 = Much Lower, 7 = Much Higher]	4.00	1.59	3.58	1.68	4.63	1.30	30.00	0.16
Worked with point-cloud data [1 = Far Too Little, 7 = Far Too Much]	1.75	0.97	1.67	0.98	1.88	0.99	40.00	0.54
Expertise in welding [1 = Strongly Disagree, 7 = Strongly Agree]	4.55	1.43	4.75	1.36	4.25	1.58	37.00	0.40
Expertise in VR development [1 = Strongly Disagree, 7 = Strongly Agree]	3.00	2.03	1.67	0.49	5.00	1.77	20.50	0.02 (*)

Table 1 Table showing Demographics of Users who participated during the system evaluation of VRFromX

on the welding dataset, including 14 classes of welding equipment, each consisting of 50 models each (See Figure 8). The fine-tuning with the welding dataset was performed to prepare the system for the welding use case and improve the classification accuracy. The network's accuracy for top 5 results was over 98%, with the average testing time of 1.8 seconds. Accuracy was calculated by running predictions on the testing dataset which contained a subsample of the training dataset. The alignment network based on PointNet-LK [76] was implemented using a PyTorch version of the original PointNet-LK [81] on the ModelNet40 dataset with a data augmentation method that applies to random rotation and variation. The average testing time of the network for aligning an input with 2000 points was 1.3 seconds.

6 User Study

6.1 Study Setup. In order to evaluate the ability of the system to author interactive virtual experiences from real-world scans, a user study was conducted using the welding use case. The study was conducted with 20 participants (3 female, 17 male) with ages between 18-35 years, from which 12 users (3 female, 9 male) were NVRPs, and the rest 8 users were EVRPs (0 female, 8 male). All users had prior hands-on experience with welding, and thus can also be considered as DUs. Furthermore, 11 users were experts in welding. For more details on the user demographics, please see Table 1. Participants were recruited using word of mouth and the study was approved under IRB protocols. During this study, we observed the usability of VRFromX while examining how the users interacted with the system and what features they expected to achieve using the system. During the study, we also tried to evaluate the potential of the virtual welding simulator to be used for training purposes based on the comments of the welding experts.

6.2 Study Procedure. At first, users were provided with an iPadPro and were asked to experience the scanning process by scanning any surrounding region using the application *3d Scanner*

*App*TM [74]. This step was included in the user study to give the user an idea about the point cloud and the ease of it's acquisition using a consumer-level product. This was followed by giving the user a virtual tour of the welding station in VR by placing virtual markers at appropriate positions in the scan showing the pictorial representation of the objects in the actual scene. After the users had developed a proper understanding of the scene and objects present in the scan, the actual study process started. A researcher explained the instructions verbally about the functions of the physical buttons on the Oculus controllers (OC) and virtual buttons in the system. Using a video demonstration, the objectives of each task were explained to the user. The entire session was divided into three tasks; (1) object retrieval using the brush tool, (2) behavioral modelling of virtual objects, and (3) interaction with the final scene. During the first task, users were asked to either select the objects from the scanned point cloud or sketch in mid-air using the brush tool to retrieve models from the database. After building the virtual environment with the models, the next task was to assign functions to the objects. The final task was to interact with the objects inside the welding scene to perform virtual welding. After each task, users recorded their experience about the system using a 7-point Likert survey (1=Strongly Disagree, 7=Strongly Agree). After completion of the three tasks, we interviewed the users by asking about their experience with the system, comments on improvement, and applications where the system could be used. Data was collected for each user in terms of: (i) time required to finish each task, (ii) 7-point Likert Scale based questionnaire to test the usability of the system and (iii) qualitative feedback. Each user study session was recorded from three camera views, first person camera view in VR, third person camera view in VR and third person camera view in real environment. We segmented the first person camera view in VR to record the objective results in terms of time needed to perform the tasks. We audio-recorded all the subjective comments and suggestions from the users for post-study analysis and summary. The subjective feedback is later used in the paper to explain the study results and inspire for future design in-

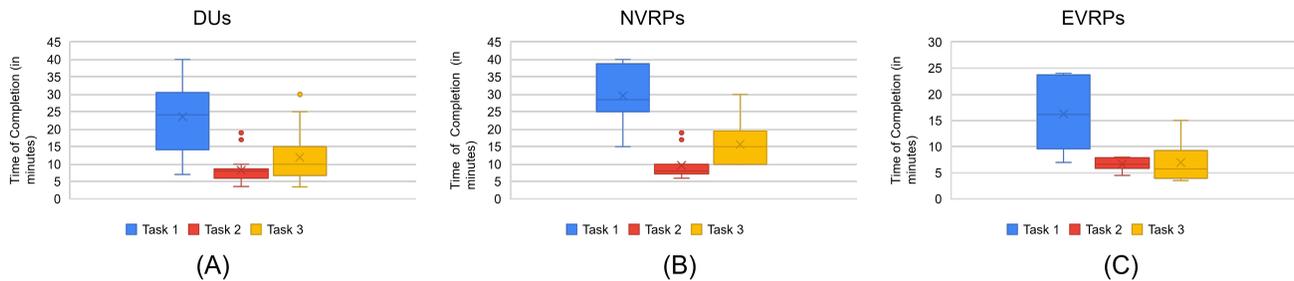


Fig. 9 Box Plots showing the data distribution for the time of completion of the three tasks across (A) DUs, (B) NVRPs and (C) EVRPs

sights. The study lasted for approximately 60 minutes. Participants were given \$20 compensation for taking part in the study.

6.3 Results. All participants were able to complete every task. Results from questionnaire surveys were analysed to identify users' scores on each individual aspect of the system, how usable were the interaction techniques, how relatable were the actions as compared to the real world entities and how immersive was the experience. All Likert Scale questions are reported with mean (M) and standard deviation (SD) in Table 2. We also provide user's reactions that were gathered during the conversational interview in the Discussion section, depicting their experience while performing the tasks.

Task1 : Point Cloud Interaction for Object Retrieval: During this task, users used the brush tool for setting up the virtual welding environment. The time of completion for the task was first extracted from the video analysis, and then checked for any presence of outliers. No outliers were observed for the time of completion of task 1 across the DUs. See Figure 9A for the Box plot for the data distribution of the time of completion. The average time (in minutes) needed to finish this task was found to be $M=24.24$, $SD=10.16$. User ratings from the questionnaire survey following this task were grouped into three categories to check the accessibility (A1-A6), engaging (A12-A17) and effectiveness (A7-A11) aspects of the point cloud interaction using VRFromX. See Figure 10 for more details.

Task2: Behavioral modelling for virtual objects. With the help of the affordance recommender, users attached various functions to the virtual objects to make the virtual scene interactable. The time of completion for the task was first extracted from the video analysis, and then checked for any presence of outliers. Two outliers were observed for the time of completion of task 2 across the DUs, with values of 19 and 17. See Figure 9A for the Box plot for the data distribution of the time of completion. The presence of the outliers was thoroughly examined to ensure they were not data collection errors. It was observed that the high values of the outlier data points accounted from (A) the corresponding users having low familiarity using VR applications, (B) the corresponding users being novices in VR programming, and (3) the system's ability to allow users to iterate upon the authored prototypes, e.g., users had the flexibility to revisit the point cloud interaction during Task 2 to retrieve object models if needed. Thus, the outliers were retained in the analysis due to their potential relevance to the research, the use of clear instructions and standardized procedures for data collection across all users, and the use of non-parametric analysis methods [Mann-Whitney U analysis] which are less sensitive to outliers. The average time (in minutes) taken to complete this task was found to be $M=8.46$, $SD=3.56$ (including outliers), and $M=7.40$, $SD=1.46$ (excluding outliers). User ratings from the questionnaire survey following this task were grouped into three categories to check the accessibility (B1-B3), engaging (B11-B16)

and effectiveness (B4-B10) aspects of the behavioral modeling using VRFromX. See Figure 11 for more details.

Task 3: Virtual welding. During this task, users interacted with the virtual scene that they created during the previous tasks. The time of completion for the task was first extracted from the video analysis, and then checked for any presence of outliers. An outlier was observed for the time of completion of task 3 across the DUs, with value of 30. See Figure 9A for the Box plot for the data distribution of the time of completion. The presence of the outlier was thoroughly examined to ensure they were not data collection errors. It was observed that the high values of the outlier data point accounted from (A) the corresponding user having low familiarity using VR applications, (B) the corresponding user being novice in VR programming, and (3) the system's ability to allow users to interact with the authored scenes as desired. In the case of the outlier, the corresponding user performed virtual welding 8 times to achieve a desired weld, while keeping track of the work parameters. Thus, the outlier was retained in the analysis due to the user's potential relevance to the research, the use of clear instructions and standardized procedures for data collection across all users, and the use of non-parametric analysis methods [Mann-Whitney U analysis] which are less sensitive to outliers. The average time (in minutes) taken to complete this task was found to be $M=12.18$, $SD=7.00$ (including outlier), and $M=11.25$, $SD=5.76$ (excluding outlier). User ratings from the questionnaire survey following this task were grouped into two categories to check the engaging (C8-C14) and effectiveness (C1-C7) aspects of the virtual welding using VRFromX. See Figure 12 for more details.

6.3.1 Comparative Analysis of EVRPs vs. NVRPs: The results from survey questionnaires and video analysis were divided into two groups based on VR programming expertise of the users. This analysis was performed to examine the effect of prior VR programming experience on the user perception and performance during authoring using VRFromX. It was verified that the normal distribution assumption was not met by conducting Shapiro-Wilk normality test on the data ($p<0.05$). Therefore, Mann-Whitney U test was conducted on the survey results and the quantitative data for time of completion for the three tasks, to compare the performance of the two user groups. The mean (M), standard deviation (SD), U-scores and p-values from the comparative analysis are reported in Table 2 with statistical significance indicated by $p<0.05$. From the analysis results, the user experience and perception of EVRPs vs. NVRPs were almost similar for the three tasks during the study with no significant differences. An on-the-border significance was found for the user rating if the retrieval results were obtained as expected during the Task 1 [$U=22.00$, $p=0.04$], where the average rating obtained for EVRPs and NVRPs were 5.75 ($SD=1.28$) and 4.50 ($SD=0.90$), respectively. The time of completion for the three tasks, extracted from the video analysis, was checked for any presence of outliers across the two user groups. For task 2, two outliers were observed for the time of completion

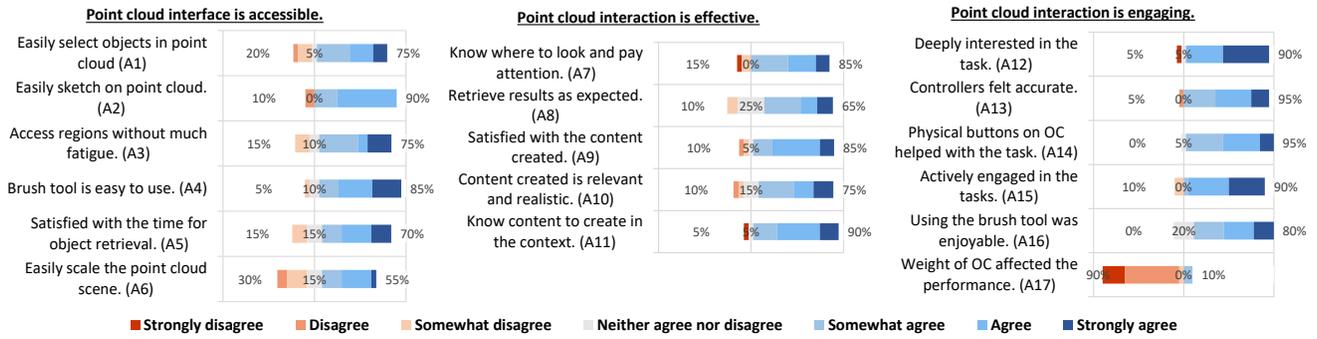


Fig. 10 Questionnaire survey results after Task 1: Point Cloud Interaction for Object Retrieval, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree)

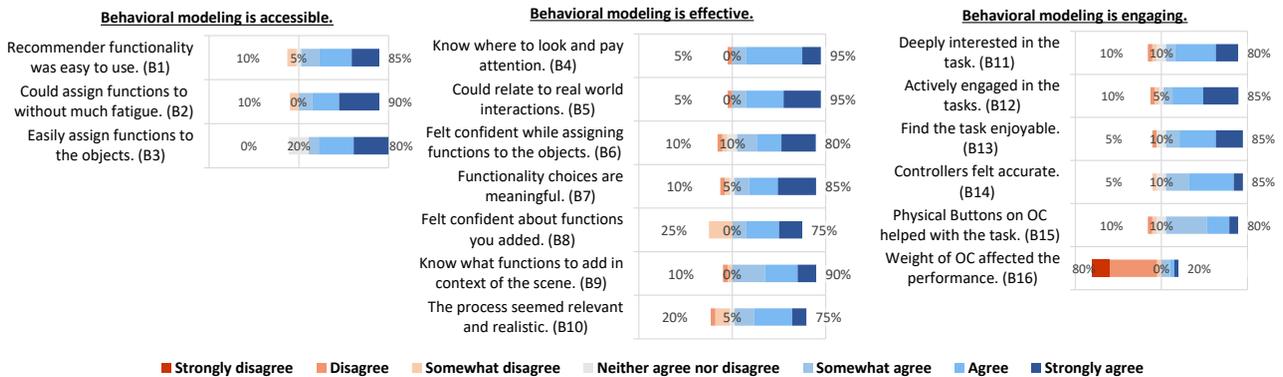


Fig. 11 Questionnaire survey results after Task 2: Behavioral modelling for virtual objects, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree)

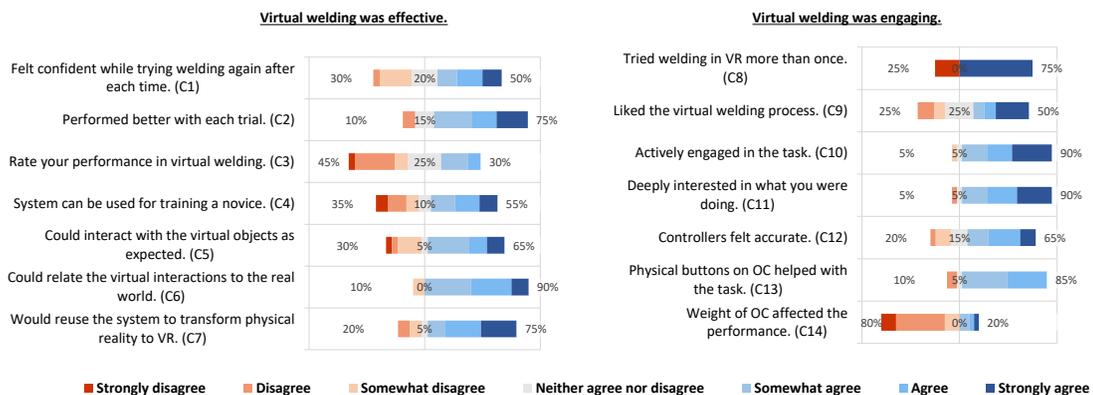


Fig. 12 Questionnaire survey results after Task 3: Virtual welding, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree)

Quantitative Data from User Study Tasks (Survey Data [1 = Strongly Disagree, 7 = Strongly Agree]/ Time of Completion)	DUs (N=20)		NVRPs (N=12)		EVRPs (N=8)		Mann Whitney U Analysis	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	U	p<0.05
Task 1: Point Cloud Interaction for Object Retrieval								
Easily select objects in point cloud. (A1)	5.05	1.43	4.75	1.29	5.50	1.60	30.00	0.16
Easily sketch on point cloud. (A2)	5.35	1.23	5.67	0.49	4.88	1.81	42.00	0.64
Access regions without much fatigue. (A3)	5.20	1.36	5.25	1.29	5.13	1.55	47.00	0.94
Brush tool is easy to use. (A4)	5.75	1.16	5.58	1.00	6.00	1.41	33.50	0.26
Satisfied with the time for object retrieval. (A5)	5.25	1.37	5.17	1.27	5.38	1.60	42.50	0.67
Easily scale the point cloud scene. (A6)	4.55	1.50	4.75	1.42	4.25	1.67	38.50	0.46
Know where to look and pay attention. (A7)	5.20	1.47	5.67	0.89	4.50	1.93	35.50	0.33
Retrieve results as expected. (A8)	5.00	1.21	4.50	0.90	5.75	1.28	22.00	0.04 (*)
Satisfied with the content created. (A9)	5.50	1.28	5.75	0.87	5.13	1.73	41.50	0.62
Content created is relevant and realistic. (A10)	5.20	1.36	5.50	1.17	4.75	1.58	36.00	0.35
Know content to create in the context. (A11)	5.60	1.35	5.83	0.94	5.25	1.83	41.50	0.62
Deeply interested in the task. (A12)	6.15	1.42	6.50	0.52	5.63	2.13	42.00	0.64
Controllers felt accurate. (A13)	5.65	1.14	5.50	1.24	5.88	0.99	44.00	0.76
Physical buttons on OC helped with the task. (A14)	5.65	0.81	5.75	0.87	5.50	0.76	43.50	0.73
Actively engaged in the tasks. (A15)	6.05	1.19	6.33	0.65	5.63	1.69	40.50	0.56
Using the brush tool was enjoyable. (A16)	5.50	1.05	5.50	0.80	5.50	1.41	48.00	1.00
Weight of OC affected the performance. (A17)	2.10	1.12	2.33	1.30	1.75	0.71	37.00	0.40
Time of Completion (in minutes)	24.24	10.16	29.58	8.13	16.22	7.36	8.00	0.00 (*)
Task 2: Behavioral modelling for virtual objects								
Recommender functionality was easy to use. (B1)	5.70	1.26	5.75	0.87	5.63	1.77	42.50	0.67
Could assign functions to without much fatigue. (B2)	6.00	1.26	5.75	1.48	6.38	0.74	38.50	0.46
Easily assign functions to the objects. (B3)	5.85	1.14	6.08	0.90	5.50	1.41	38.00	0.44
Know where to look and pay attention. (B4)	5.85	1.09	5.75	1.29	6.00	0.76	47.00	0.94
Could relate to real world interactions. (B5)	6.05	1.19	6.08	0.79	6.00	1.69	39.50	0.51
Felt confident while assigning functions to the objects. (B6)	5.60	1.47	5.67	1.23	5.50	1.85	46.50	0.91
Functionality choices are meaningful. (B7)	5.80	1.44	5.92	1.16	5.63	1.85	47.50	0.97
Felt confident about functions you added. (B8)	5.35	1.53	5.83	1.11	4.63	1.85	31.00	0.19
Know what functions to add in context of the scene. (B9)	5.50	1.28	5.75	0.75	5.13	1.81	42.00	0.64
The process seemed relevant and realistic. (B10)	5.20	1.47	5.42	1.38	4.88	1.64	36.50	0.37
Deeply interested in the task. (B11)	5.60	1.39	6.00	0.85	5.00	1.85	34.00	0.28
Actively engaged in the tasks. (B12)	5.85	1.42	6.17	1.19	5.38	1.69	32.50	0.23
Find the task enjoyable. (B13)	5.75	1.29	5.83	0.94	5.63	1.77	44.50	0.79
Controllers felt accurate. (B14)	5.50	1.00	5.33	0.98	5.75	1.04	38.00	0.44
Physical Buttons on OC helped with the task. (B15)	5.10	1.21	5.25	1.06	4.88	1.46	40.00	0.54
Weight of OC affected the performance. (B16)	2.60	1.73	2.67	1.67	2.50	1.93	45.00	0.82
Time of Completion (in minutes)	8.46	3.56	9.67	4.12	6.65	1.21	18.50	0.02 (*)
Task 3: Virtual welding								
Felt confident while trying welding again after each time. (C1)	4.65	1.57	4.67	1.67	4.63	1.51	47.50	0.97
Performed better with each trial. (C2)	5.25	1.52	5.50	1.51	4.88	1.55	35.50	0.33
Rate your performance in virtual welding. (C3)	3.55	1.50	3.58	1.38	3.50	1.77	47.00	0.94
System can be used for training a novice. (C4)	4.35	2.01	4.00	2.13	4.88	1.81	36.50	0.37
Could interact with the virtual objects as expected. (C5)	4.65	1.69	4.50	1.83	4.88	1.55	45.00	0.82
Could relate the virtual interactions to the real world. (C6)	5.45	1.10	5.25	1.22	5.75	0.89	40.00	0.54
Would reuse the system to transform physical reality to VR. (C7)	5.35	1.69	5.25	1.82	5.50	1.60	47.00	0.94
Tried welding in VR more than once. (C8)	0.75	0.44	0.75	0.45	0.75	0.46	48.00	1.00
Liked the virtual welding process. (C9)	4.80	1.85	4.83	2.12	4.75	1.49	47.00	0.94
Actively engaged in the task. (C10)	5.90	1.17	6.25	0.87	5.38	1.41	30.00	0.16
Deeply interested in what you were doing. (C11)	5.80	1.28	6.17	1.03	5.25	1.49	28.50	0.13
Controllers felt accurate. (C12)	5.00	1.49	4.92	1.68	5.13	1.25	46.50	0.91
Physical buttons on OC helped with the task. (C13)	5.05	1.19	5.17	1.11	4.88	1.36	42.00	0.64
Weight of OC affected the performance. (C14)	2.75	1.68	3.08	1.88	2.25	1.28	34.50	0.30
Time of Completion (in minutes)	12.18	7.00	15.67	6.50	6.96	3.85	9.00	0.00 (*)

Table 2 Table showing Comparative Analysis of EVRPs vs. NVRPs from survey questionnaire ratings and Time of Completion during the User Study Tasks



Fig. 13 Screenshots showing participants interacting with the different virtual welding simulators authored by them during the study (from third person view in Unity™)

of task 2 across the NVRPs, with values of 19 and 17. See Figure 9B,C for the Box plots for the data distribution of the time of completion across the two user groups. EVRPs took an average of 16.22 (SD=7.36), 6.65 (SD=1.21), and 6.96 (SD=3.85) minutes to complete Tasks 1, 2, and 3, respectively. NVRPs took an average of 29.58 (SD=8.13), 9.67 (SD=4.12), 8 (SD=1.41), and 15.67 (SD=6.50) minutes to complete Tasks 1, 2 (including outliers), 2 (excluding outliers), and 3, respectively. There was a significant difference observed in the time of completion of the three tasks in both groups: (i) Task 1 [$U=8.00$, $p^*=0.00$], (ii) Task 2 [$U=18.50$, $p^*=0.02$] (including outliers), and (iii) Task 3 [$U=9.00$, $p^*=0.00$]. Given the potential relevance of the outliers, the use of clear instructions and standardized procedures for data collection across all users, and the utilization of non-parametric analysis methods [Mann-Whitney U analysis], the outliers were retained in the analysis. The results are further analysed in the Discussion section in greater detail.

6.3.2 Summary: We now summarize the main outcomes of the user study. The results obtained from the questionnaire survey indicates the feasibility of VRFromX as an authoring platform to transform physical environment of a welding station into a virtual welding simulator. Figure 13 shows the interactions of participants with the virtual welding simulators that they created during the user study (from the third person view in Unity™). User ratings from the survey questionnaire proved that the interaction procedures offered by the system are simple, and easy to understand and use. Most importantly, the execution of the different interactions do not necessarily demand skill expertise in VR programming, and thus NVRPs could use the system for creating meaningful virtual experiences. Moreover, the entire process can be completed in the same environment without needing the users to switch between platforms. The free and embodied interactions inside the 3D environment did not cause fatigue to the users, rather they rated the authoring experience to be enjoyable and immersive. Being already familiar with the real welding settings and processes from before, the welding DUs were more attached to the authoring process since they could connect their activities in the virtual world to the corresponding real world tools and interactions. The perceived response from the users regarding the weight of the OC affecting the overall performance during the tasks was low. Additionally, the perceived accuracy of the system and OC during the authoring process was rated highly. This further implies that the users enjoyed the authoring process in VR and were satisfied with the system response to their actions.

7 Discussion and future work

From the findings of the study and the qualitative feedback from the conversational interview, we now try to answer the research

questions that were posed before. We also discuss future extensions and directions for evaluating VRFromX, and some limitations therein.

(1) To what extent does the users' expertise in VR programming affect their performance and experiences in using VR-FromX to author virtual environments? From the comparative analysis shown in Table 2, it can be observed that NVRPs and EVRPs provided positive ratings of their experiences with VR-FromX during all three tasks. This indicates that prior experience in VR programming is not a requisite for interacting with the system. This is in accord with the system interface of VRFromX being designed with easy-to-use and immersive features that are accessible across users irrespective of their programming skills. It was observed that both groups of users were highly enthusiastic about trying the different features and enjoyed their experience with the system. Both groups felt that the system features were easy to use and provided them with an immersive experience. There was a significant difference found in the times of completion for the three tasks for EVRPs vs. NVRPs. This can be reasoned due to the lack of familiarity with using VR devices and applications in the case of NVRPs. This resulted in NVRPs taking more time to get acclimated to the immersive scenes and the button controls, as compared to the EVRPs. Despite the high completion times, the positive ratings in the survey indicate that the longer duration did not quite affect the user experience that NVRPs gained using the system. Next, we attempt to relate each of the system components to their underlying causes of providing users with accessible and immersive experiences.

Point cloud interaction and 3D sketching in VR: VRFromX enables users to use hands freely in their surrounding space to create virtual scenes. The free sketching affordance provided by the system enables the users to portray their thoughts in the 3D space, which in its entirety acts as an unrestricted canvas to the curious minds. From the study results, it was observed that users were enthusiastic to use the 3D canvas during their interactions with the system. As pointed out by a user, "I did not know we could do point cloud sketching to retrieve [3D] models. 3D sketching is different from 2D sketching, [it seems] very easy for beginners, [there is] no need to learn any new software, [and] we don't need experts to do that." This can be reasoned by stating that sketching is very natural to human beings, and is something one learns from childhood [82]. A user remarked, "I liked the freeform nature of the drawing, drawing was usually smooth. Matching the drawing (may be the shape of the object) to get objects was a cool feature." So, it would not take much time to get accustomed to the 3D sketching as an input modality for content creation. Some users also commented on the cognitive benefits that sketching would provide when used as the input modality during the content creation, "[It can be useful,] Especially if you know what an object looks like. Maybe people would remember things better if they would sketch

to get the objects”.

The easy and free interactions enabled by the sketching modalities along with the simple features of the brush tool further helped users to select point clouds or sketch in mid-air without much effort. The brush tool interface, being attached to the user’s hands, is easy to find and reach, whenever necessary. The features of the tool span across the basic operations needed to select point clouds or sketch in mid-air to retrieve 3D models from the database. Users could get hold of the features easily, and were able to successfully construct the virtual welding scene using the brush tool. The point cloud environment also simplified the task for the user by providing a reference to build upon. Being able to move freely inside the virtual environment, and using the system features to manipulate the position and scale of the virtual scene, the user could easily reach desired locations in the scene.

Users felt fascinated by the idea of turning their sketch into a virtual model, and the interactive feature provided by the system. An user pointed out, “[I] Like the 3D sketching aspect, drawing things in mid-air and the object detection, trying to turn the sketch into the object that I was trying to convey”. Through the experiment, it was shown that the object retrieval performed using point cloud selection and 3D sketch as the query input provided acceptable results in terms of the accuracy of 3D models retrieved from the database and time of retrieval using our system. The action-reaction scheme facilitated by the AI assistance was perceived well by the users and accounted further towards enhancing the accessibility and user experience of our system. Interactions with the point cloud are also interactive, meaning the system responds to the manipulations made by the user. As pointed out by another user, “[The system could] Easily detect the sketch, simple sketching detects the objects - [The system interface is] user friendly”. Although both EVRPs and NVRPs highly rated the efficiency of the retrieval system to fetch the objects as expected, there was an on-the-border significance result achieved between the two groups. This could be possibly explained by the observation from the demographics survey which indicated that the EVRPs had higher sketching skills as compared to the NVRPs (Table 1). This calls for potential improvements to the point cloud interface component of the system to further enhance its accessibility.

During our user study, some users pointed out augmenting the capabilities of the brush tool to provide some geometrical features to create simple shapes to assist during sketching. “Use some features to draw straight lines, circles, rectangles; [We need] not draw them ourselves. Some features which are used in tools like paint, put simple shapes together to build complex shapes”. This can be achieved by adding features in the brush tool to enable creation of constraint-based geometric shapes in 3D [83,84]. According to some users, similar-looking shapes were a bit harder to retrieve. “But sometimes it (the object retrieval system) got confused. Like I was trying to get a gas cylinder, the network was giving me fire extinguisher. Maybe color and the finer details could be used to compute the differences between tall gas cylinder and short fire extinguisher”. Currently, the PointNet network used for retrieving the object models from database only use (x, y, z) coordinates as the point’s channels [62]. To improve the accuracy of the classification network, extra feature channels such as color, normal, etc. can be used. Furthermore, using additional modalities such as voice input [50] can be explored as viable solutions to improve the system’s accuracy in such cases.

Another direction to look into is to make the human and AI interaction in VRFromX more effective. AI algorithms can be used to assist the users in performing the semantic segmentation of the point cloud scans [51]. While selecting point cloud using the brush tool, the network can provide some suggestions for areas to select to the users based on their intent. On the same note, the network can recommend features for sketch completion when the users are performing 3D sketching in VR [72]. Further collaboration between humans and AI could be targeted towards improving the accuracy and scalability of the retrieval networks through providing scope for active learning during point cloud interaction in

VRFromX [85]. Using active learning techniques would make the Human-AI collaboration symbiotic, meaning the AI-based networks could learn better using human interaction. For example, these techniques can especially be useful in case of unseen object classes, where user can segment and label point cloud scans for the AI-based networks to learn, and in return user interaction can be enhanced by the improved outputs from the system. We also plan to explore the potential of VRFromX in reconstructing shapes by making use of shape primitives [86] or geometric features in form of planes, surfaces, etc. present in the point cloud [87] for object reconstruction in case of unseen classes.

Affordance assignment in VR: Attaching affordances to virtual objects in VRFromX was facilitated by simple drag and drop interactions. The interaction methods also created a direct connection between the user and the authoring process, by inducing a sense of realism and presence inside the immersive environment. Users could spatially attach the various functions to the virtual objects in terms of animations, displays and controls while perceiving the spatial relationship between the object(s), their parts and functions with respect to their virtual self. The three dimensional authoring provides freedom to the user to perform free transformations of self or objects with respect to each other and other objects, thereby the sense of spatial awareness experienced by the users during the authoring process can be very closely related to the similar settings in real environments. For example, while attaching the functions to the virtual welding gun during the user study, the users could correctly position the *Spark* animation to the tip of the welding gun.

Interaction with the virtual environment: User interactions with their authored content added more significance to the value of the authoring process, resulting in enhanced user motivation and confidence in using the system for transforming physical reality to virtual experiences. The system enables DUs to directly test the authored content in the virtual environment without needing to switch between multiple platforms. During the testing phase, if the user figures out any scope for modifications to the existing content, the user can go back to the authoring stage effortlessly by pressing buttons on the OC and make necessary changes to the content.

(2) How does the domain knowledge of users affect their perception and experience in using VRFromX to author virtual environments?

Point cloud interaction and 3D sketching in VR: During the point cloud interaction for object retrieval, prior experience in welding enabled the users to get familiarized with the point cloud scan of the welding station quickly and in no time. Users could make use of the spatial and color information offered by the scan to recall and recognize the welding objects found in a typical workplace setting. e.g., Tools like clamp, hammer and wirebrush were found above the welding table; The welding gun was found wired to the welding machine; the red fire extinguisher attached to the wall was an easy find.

Affordance assignment in VR: The nature of the functions displayed by the Affordance Recommender, being very similar to the real world processes, draw the DUs more towards the functionality authoring process, by exercising their motor and cognitive faculties to transfer the knowledge and skills gained from their real-world experience. As pointed out by a user, “[I] Liked dragging and attaching functions. [It] was easy. Which components [were] meant for each tool was obvious. Graphs were nice. Overall the attaching components part was smooth.” DUs can directly make use of and transfer their understanding that they have developed during the real world interactions and processes easily in the virtual environment. For example, during the user study, the users could figure out the nature of the functions needed to be attached to the welding gun, e.g. *Spark* animation to represent the graphical effects during virtual welding, *Angle* and *Distance* displays to visualize the welding parameters, and *Welding Plate* and *Trigger Button* controls to trigger virtual welding. Therefore, the transfer of understanding and perception can reduce the overall cognitive effort required during the process. A user pointed out, “It was very interesting. [It]

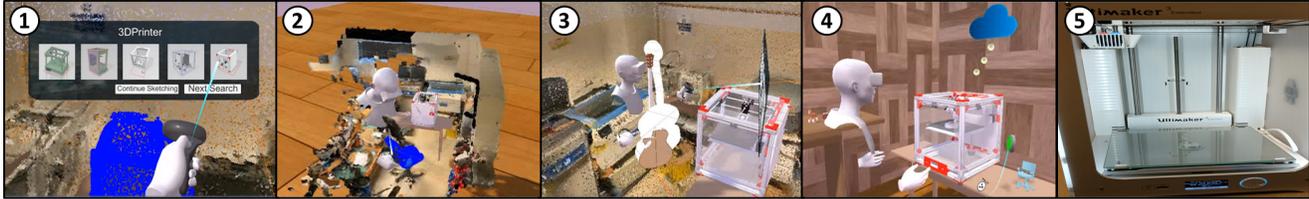


Fig. 14 Remote 3D-Printing Use Case: (1) After retrieving a 3D Printer, (2) user creates 3D-models by point cloud selections and (3) sketching in mid-air, and finally (4) sends the model for remote printing. (5) IoT-enabled 3D printer prints the model.

Works very smoothly. [It is] Easy to use, and very easy to figure out”.

Interaction with the virtual environment: When performing virtual welding, the realistic content and high fidelity of the affordances associated with the virtual objects enhanced the user perception about the virtual welding simulator. As pointed out by a user, “*Virtual welding behavior [was] somewhat realistic, that part is somewhat reasonable*”. Being aware of the real-world processes creates a sense of satisfaction when the user can test that the authored content in VR behave as the real world version of the object would behave in the real world. A user pointed out, “[*The processes are] Relatable to the real world interactions, because I have done it before*”. Few users even highlighted that the virtual welding simulator could be used for basic MIG welding training purposes, “*When you are holding the gun, you can see the angle and improve the performance. [It is] Cool how you can track that and get some feedback. That’s the part which can make it good for training. It was cool to see the feedback. You can try it in VR first before doing it in real.*” Some welding experts also mentioned about the use of the virtual welding simulator to provide beginner level experiences to welding novices, e.g., develop knowledge regarding welding related objects, setting up the welding space, etc. A welding expert pointed out, “[*The virtual welding simulator can be] Good first space [to teach] how to set up a welding space, what the various parts are, [and] for the beginners, to get used to the system.*”

Insights were gained from the study to improve the efficacy of the virtual welding trainer. Based on the comments of the DUs, the 3D models in the virtual welding environment looked cleaner and visually more attractive as compared to real-world settings. The issue could be fixed by adding textures to the list of affordances for the virtual objects. As pointed out by a welding expert, “[*There might be some] Trouble translating from real to VR, [especially, the] Way to hold the welding gun. [It feels] Sensitive not having the actual way of holding the welding gun in your hand*”. One way to solve address this issue can be achieved by replacing the OC from user’s hands with an IoT-enabled real welding gun. The effect of haptic perception on the user experience inside the virtual welding environment can be explored as a scope of future work [88]. Furthermore, future research can explore the effectiveness of such authored applications for training purposes and evaluate the transfer of skills from VR to the real environment [88,89].

A limitation of the the current work is related to the absence of statistical power analysis to estimate the sample size for the user study, which was due to constraints in finding users with welding experience for the study. Despite this limitation, we believe the evaluation results from the study still provides valuable insights into evaluating the effect of technical expertise and domain-specific experience towards interacting with the system components of VR-FromX. Although the sample size was constrained, we made efforts to mitigate potential biases by keeping welding experience as the only criteria for the user selection, utilizing clear instructions and standardized procedures for data collection across all users, and utilizing non-parametric statistical analysis methods. We believe that the insights gained from this study contribute to the existing body of knowledge in the field of providing accessible methods for adoption of the immersive VR technology for domain-specific

applications, particularly in the context of accessible authoring of VR applications [7]. Nevertheless, we acknowledge the benefits of future studies with a proper sample size estimation using statistical power analysis to further validate and extend the current findings, and hope that our work provides design insights to explore this area in greater detail.

8 Potential applications and Use case demonstrations

We further analyse insights from prior work and our own observations during the user study to derive potential applications of our system.

Potential applications: The accessibility of authoring of the VR-based applications should keep in pace with the increasing trend of use of the technology in various sectors such as training and education, gaming and entertainment, and industry. By enabling easy authoring of virtual content, VRFromX facilitates easy accessibility of the technology to create useful applications in VR. During the conversational interview, users commented about the use of VRFromX to author applications for (1) training in industry, e.g., “*Training purposes - drill mills, saw, CNCs.*”, “*In my old factory, I could use this for setting parameters, training for setting parameters. Training how to do set up parameters for machines. Useful when the task might be dangerous. Especially when they are doing some dangerous tasks, VR takes the danger out.*”, “*A lot of potential in medical usage. Training people to perform surgery.*”, (2) education, e.g., “*Comfortable way to teach a soft skill, before you actually try on a physical set-up.*”, “*Lab scenario - students can do it at home, can go to lab for data collection, rather than having problems with the set-up.*”, (3) entertainment, e.g., “*Escape room-It’s a game so that people can escape from room to any mysterious world.*”, (4) safety, e.g., “*Trial and error, process is expensive to perform in real. Scan the process, bring in VR, do in VR for user experience, save the cost. Process is hazardous, VR would help there.*”, (5) assembly, e.g., “*Assemblies and stuff like that where you don’t have multiple choices.*”, (6) human-robot interactions, e.g., “*Stories about up and coming robotic devices, [where] surgeon [is] in remote location and patient [is] with the robot. Surgeons with basically the joystick and perform the procedure remotely.*”, etc.

Motivated by the insights collected from the user study, we implemented two different use case scenarios with an aim to justify the capabilities of our system to perform activities in Mixed Reality (MR) environments, such as (1) Remote 3D Printing, and (2) Robot-Iot Task Planning.

Remote 3D Printing. The second use case demonstrates an example integration of VRFromX in maker education to enhance students’ learning experiences. Prior work has shown that maker education, which emphasizes hands-on based experiential learning and student creativity [90], can be enriched using the MR technology by offering a more interactive and engaging learning experience [91,92]. First, MR learning environments can allow students to visualize and manipulate objects in three-dimensional space. Secondly, the seamless integration of real-world elements into the virtual space allows students to bridge the gap between



Fig. 15 Robot IoT Use Case: (1) User specifies the path and authors actions in VR and (b) the physical robot performs the tasks as authored by human in real world.

abstract ideas and tangible creations, making their learning more tangible and meaningful.

In this context, the second use case demonstrates an example of MR fabrication using IoT-enabled devices (Figure 14). User interacts with the scanned point cloud of a 3D-printing area to retrieve a virtual 3D-Printer and assigns functions to facilitate interaction with the IoT-enabled 3D-Printer in the real world. Using the brush tool, virtual models are retrieved using point cloud selections or sketching operations. When the user puts the selected virtual models on the plate of the virtual printer, the models are sent to the physical 3D-printer for remote printing. Such applications can specifically be used in maker-based educational settings to encourage high school students with little knowledge in technical modelling. Using sketching features, they can retrieve 3D designs for maker-based prototyping and print models remotely without the need of going into the complexity of constraint-based 3D modelling softwares [93].

Robot-IoT Task Planning. Addressing the challenge of effective communication of intent between humans and robots in smart environments has been a longstanding issue [94–96]. Previous studies have presented promising solutions for Human-Robot Interaction (HRI) by leveraging the surrounding environment as a shared canvas for visual and spatial cues [97], and enabling humans to convey instructions to robots for performing tasks in smart environments that incorporate multiple IoT devices [98]. In this context, the third use case aims to leverage the advantages of VR-FromX in HRI specifically for authoring robot-IoT tasks, fostering seamless interactions among humans, robots, and the smart environment.

The third use case is demonstrated in Figure 15. Using the scanned point cloud overlaid in the virtual environment, user specifies the path that the robot has to follow by selecting waypoints and authors the robot tasks at desired waypoints. Using the brush tool features, user retrieves the models of IoT-enabled devices (TV, 3D-Printer) that the physical robot has to interact with and assigns functions to them (Figure 15-1). Finally the physical robot navigates on the path and interacts with the IoT-enabled devices in the real world as authored by the human (Figure 15-2). With intuitive and user-friendly interfaces, VRFromX can empower users to scan any physical environment for task authoring and remotely control the robot-IoT tasks at their convenience. By directly programming the robot through simple spatial interactions, humans assume the role of the robot’s “brain”, guiding it to perform the intended tasks within the smart environment [98].

9 Conclusion

We have presented the system design and workflow of VR-FromX that transforms physical reality into interactive virtual experiences inside a VR environment. We have demonstrated how the system design facilitates an embodied user interaction on scanned point cloud to create virtual content using AI assistance. We also demonstrated the usability of the functionality and logic authoring tool to attach affordances to the virtual objects for a meaningful end-user experience. The design concepts used in VRFromX were successfully tested using a welding use case on 20 domain users. The study results confirmed accessibility, effectiveness and engaging aspects of the system in authoring virtual experiences irrespective of the users’ VR programming expertise. We further implemented two more use cases to demonstrate the diversity of potential applications of VRFromX in Mixed Reality scenarios. By leveraging direct knowledge transfer from DUs to create VR applications, VRFromX provides an accessible platform for easy and greater adoption of the authoring process by allowing DUs to become the makers of VR applications themselves. By leveraging cloud-based architectures inside VRFromX, the system’s potential can be further extended to foster collaborative work where multiple participants from interdisciplinary fields can author virtual content together. Such collaboration, while facilitating the scalability of the authoring process across inter-related fields, would also promote creativity by supporting the joint exploration of design spaces in the virtual world.

Acknowledgment

This work was partially supported by the National Science Foundation under grant FW-HTF 1839971. We also acknowledge the Feddersen ChairFunds. We thank Hyung-gun Chi and Asim Unmesh for their help regarding this work. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

References

- [1] Lehner, V. D. and DeFanti, T. A., 1997, “Distributed virtual reality: Supporting remote collaboration in vehicle design,” *IEEE Computer Graphics and Applications*, **17**(2), pp. 13–17.
- [2] Kato, H. and Billinghurst, M., 1999, “Marker tracking and hmd calibration for a video-based augmented reality conferencing system,” *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99)*, IEEE, pp. 85–94.
- [3] Barsom, E. Z., Graafland, M., and Schijven, M. P., 2016, “Systematic review on the effectiveness of augmented reality applications in medical training,” *Surgical endoscopy*, **30**(10), pp. 4174–4183.
- [4] Zhang, Y., Liu, H., Kang, S.-C., and Al-Hussein, M., 2020, “Virtual reality applications for the built environment: Research trends and opportunities,” *Automation in Construction*, **118**, p. 103311.

- [5] Ashtari, N., Bunt, A., McGrenere, J., Nebeling, M., and Chilana, P. K., 2020, "Creating augmented and virtual reality applications: Current practices, challenges, and opportunities," *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13.
- [6] Gaspar, H., Morgado, L., Mamede, H., Oliveira, T., Manjón, B., and Gütl, C., 2020, "Research priorities in immersive learning technology: the perspectives of the iLRN community," *Virtual Reality*, **24**, pp. 319–341.
- [7] Coelho, H., Monteiro, P., Gonçalves, G., Melo, M., and Bessa, M., 2022, "Authoring tools for virtual reality experiences: a systematic review," *Multimedia Tools and Applications*, **81**(19), pp. 28037–28060.
- [8] Ens, B., Anderson, F., Grossman, T., Annett, M., Irani, P., and Fitzmaurice, G., 2017, "Ivy: Exploring spatially situated visual programming for authoring and understanding intelligent environments," *Proceedings of the 43rd Graphics Interface Conference*, pp. 156–162.
- [9] Ng, G., Shin, J. G., Plopski, A., Sandor, C., and Saakes, D., 2018, "Situated game level editing in augmented reality," *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 409–418.
- [10] Lin, C.-e., Cheng, T. Y., and Ma, X., 2020, "ARchitect: Building Interactive Virtual Experiences from Physical Affordances by Bringing Human-in-the-Loop," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
- [11] Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M., 2017, "ScanNet: Richly-annotated 3d reconstructions of indoor scenes," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5828–5839.
- [12] Tchapmi, L., Choy, C., Armeni, I., Gwak, J., and Savarese, S., 2017, "Segcloud: Semantic segmentation of 3d point clouds," *2017 international conference on 3D vision (3DV)*, IEEE, pp. 537–547.
- [13] Naseer, M., Khan, S., and Porikli, F., 2018, "Indoor scene understanding in 2.5/3d for autonomous agents: A survey," *IEEE Access*, **7**, pp. 1859–1887.
- [14] Avetisyan, A., Dahnert, M., Dai, A., Savva, M., Chang, A. X., and Nießner, M., 2019, "Scan2CAD: Learning CAD Model Alignment in RGB-D Scans," *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE.
- [15] Han, M., Zhang, Z., Jiao, Z., Xie, X., Zhu, Y., Zhu, S.-C., and Liu, H., 2021, "Reconstructing interactive 3d scenes by panoptic mapping and cad model alignments," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 12199–12206.
- [16] Chen, K., Lai, Y.-K., and Hu, S.-M., 2015, "3D indoor scene modeling from RGB-D data: a survey," *Computational Visual Media*, **1**(4), pp. 267–278.
- [17] Cowan, B. and Kapralos, B., 2014, "A survey of frameworks and game engines for serious game development," *2014 IEEE 14th International Conference on Advanced Learning Technologies*, IEEE, pp. 662–664.
- [18] Kiltner, K., Groten, R., and Slater, M., 2012, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, **21**(4), pp. 373–387.
- [19] Giunchi, D., James, S., Degraen, D., and Steed, A., 2019, "Mixing realities for sketch retrieval in Virtual Reality," *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–2.
- [20] Schultheis, U., Jerald, J., Toledo, F., Yoganandan, A., and Mlyniec, P., 2012, "Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3D tasks," *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, IEEE, pp. 117–124.
- [21] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R., 2006, "A comparison and evaluation of multi-view stereo reconstruction algorithms," *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, Vol. 1, IEEE, pp. 519–528.
- [22] Snavely, N., Seitz, S. M., and Szeliski, R., 2008, "Modeling the world from internet photo collections," *International journal of computer vision*, **80**(2), pp. 189–210.
- [23] Furukawa, Y., Curless, B., Seitz, S. M., and Szeliski, R., 2010, "Towards internet-scale multi-view stereo," *2010 IEEE computer society conference on computer vision and pattern recognition*, IEEE, pp. 1434–1441.
- [24] Lun, Z., Gadelha, M., Kalogerakis, E., Maji, S., and Wang, R., 2017, "3d shape reconstruction from sketches via multi-view convolutional networks," *2017 International Conference on 3D Vision (3DV)*, IEEE, pp. 67–77.
- [25] Ebner, T., Feldmann, I., Renault, S., Schreer, O., and Eisert, P., 2017, "Multi-view reconstruction of dynamic real-world objects and their integration in augmented and virtual reality applications," *Journal of the Society for Information Display*, **25**(3), pp. 151–157.
- [26] Shin, D., Ren, Z., Sudderth, E. B., and Fowlkes, C. C., 2019, "Multi-layer depth and epipolar feature transformers for 3d scene reconstruction," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 39–43.
- [27] Song, S., Yu, F., Zeng, A., Chang, A. X., Savva, M., and Funkhouser, T., 2017, "Semantic scene completion from a single depth image," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1746–1754.
- [28] Dourado, A., de Campos, T. E., Kim, H., and Hilton, A., 2019, "Edgenet: Semantic scene completion from rgb-d images," arXiv preprint arXiv:1908.02893, 1.
- [29] Mandikal, P. and Radhakrishnan, V. B., 2019, "Dense 3d point cloud reconstruction using a deep pyramid network," *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 1052–1060.
- [30] Ochmann, S., Vock, R., and Klein, R., 2019, "Automatic reconstruction of fully volumetric 3D building models from oriented point clouds," *ISPRS journal of photogrammetry and remote sensing*, **151**, pp. 251–262.
- [31] Yi, L., Shao, L., Savva, M., Huang, H., Zhou, Y., Wang, Q., Graham, B., Engelcke, M., Klokov, R., Lempitsky, V., et al., 2017, "Large-scale 3d shape reconstruction and segmentation from shapenet core55," arXiv preprint arXiv:1710.06104.
- [32] Izadi, S., Kim, D., Hilliges, O., Molyneux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., et al., 2011, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera," *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 559–568.
- [33] Kim, H., Remaggi, L., Dourado, A., Campos, T. d., Jackson, P. J., and Hilton, A., 2021, "Immersive audio-visual scene reproduction using semantic scene reconstruction from 360 cameras," *Virtual Reality*, pp. 1–16.
- [34] Jackson, B. and Keefe, D. F., 2016, "Lift-off: Using reference imagery and freehand sketching to create 3d models in vr," *IEEE transactions on visualization and computer graphics*, **22**(4), pp. 1442–1451.
- [35] Shapira, L. and Freedman, D., 2016, "Reality Skins: Creating Immersive and Tactile Virtual Environments," *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp. 115–124.
- [36] Sra, M., Garrido-Jurado, S., and Maes, P., 2017, "Oasis: Procedurally Generated Social Virtual Spaces from 3D Scanned Real Spaces," *IEEE transactions on visualization and computer graphics*, **24**(12), pp. 3174–3187.
- [37] Hartmann, J., Holz, C., Ofek, E., and Wilson, A. D., 2019, "Realitycheck: Blending virtual environments with situated physical reality," *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.
- [38] Yang, J., Holz, C., Ofek, E., and Wilson, A. D., 2019, "Dreamwalker: Substituting real-world walking experiences with a virtual reality," *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1093–1107.
- [39] Cheng, L.-P., Ofek, E., Holz, C., and Wilson, A. D., 2019, "VRoamer: Generating On-The-Fly VR Experiences While Walking inside Large, Unknown Real-World Building Environments," *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, IEEE, pp. 359–366.
- [40] Lindlbauer, D. and Wilson, A. D., 2018, "Remixed reality: manipulating space and time in augmented reality," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
- [41] Masnadi, S., González, A. N. V., Williamson, B., and LaViola, J. J., 2020, "AffordIt!: A Tool for Authoring Object Component Behavior in VR," *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, pp. 741–742.
- [42] Prouzeau, A., Wang, Y., Ens, B., Willett, W., and Dwyer, T., 2020, "Corsican twin: Authoring in situ augmented reality visualisations in virtual reality," *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9.
- [43] Wang, Z., Nguyen, C., Asente, P., and Dorsey, J., 2021, "Distanciar: Authoring site-specific augmented reality experiences for remote environments," *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.
- [44] Qian, X., He, F., Hu, X., Wang, T., Ipsita, A., and Ramani, K., 2022, "Scalar: Authoring semantically adaptive augmented reality experiences in virtual reality," *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–18.
- [45] Wang, Z., Nguyen, C., Asente, P., and Dorsey, J., 2022, "Point Cloud Capture and Editing for AR Environmental Design," *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–3.
- [46] Wang, Z., Nguyen, C., Asente, P., and Dorsey, J., 2023, "PointShopAR: Supporting Environmental Design Prototyping Using Point Cloud in Augmented Reality," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.
- [47] Kumar, H. G., Khargonkar, N. A., and Prabhakaran, B., 2018, "ScanToVR: An RGB-D to VR Reconstruction Framework," .
- [48] Zhang, L. and Oney, S., 2020, "Flowmatic: An immersive authoring tool for creating interactive scenes in virtual reality," *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 342–353.
- [49] Giunchi, D., James, S., and Steed, A., 2018, "3d sketching for interactive model retrieval in virtual reality," *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, pp. 1–12.
- [50] Giunchi, D., Sztrajman, A., James, S., and Steed, A., 2021, "Mixing Modalities of 3D Sketching and Speech for Interactive Model Retrieval in Virtual Reality," *ACM International Conference on Interactive Media Experiences*, pp. 144–155.
- [51] Valentin, J., Vineet, V., Cheng, M.-M., Kim, D., Shotton, J., Kohli, P., Nießner, M., Criminisi, A., Izadi, S., and Torr, P., 2015, "Semanticpaint: Interactive 3d labeling and learning at your fingertips," *ACM Transactions on Graphics (TOG)*, **34**(5), pp. 1–17.
- [52] Xu, K., Chen, K., Fu, H., Sun, W.-L., and Hu, S.-M., 2013, "Sketch2Scene: Sketch-based co-retrieval and co-placement of 3D models," *ACM Transactions on Graphics (TOG)*, **32**(4), pp. 1–15.
- [53] Yu, L.-F., Yeung, S.-K., and Terzopoulos, D., 2015, "The clutterpalette: An interactive tool for detailing indoor scenes," *IEEE transactions on visualization and computer graphics*, **22**(2), pp. 1138–1148.
- [54] Eitz, M., Hays, J., and Alexa, M., 2012, "How do humans sketch objects?" *ACM Transactions on graphics (TOG)*, **31**(4), pp. 1–10.
- [55] Henschilwood, C. S., d'Errico, F., van Niekerk, K. L., Dayet, L., Queffelec, A., and Pollarolo, L., 2018, "An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa," *Nature*, **562**(7725), pp. 115–118.

- [56] Yu, Q., Liu, F., Song, Y.-Z., Xiang, T., Hospedales, T. M., and Loy, C.-C., 2016, "Sketch me that shoe," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 799–807.
- [57] Shao, T., Xu, W., Zhou, K., Wang, J., Li, D., and Guo, B., 2012, "An interactive approach to semantic modeling of indoor scenes with an rgbd camera," *ACM Transactions on Graphics (TOG)*, **31**(6), pp. 1–11.
- [58] Zhu, X., Song, L., You, L., Zhu, M., Wang, X., and Jin, X., 2017, "Brush2Model: Convolution surface-based brushes for 3D modelling in head-mounted display-based virtual environments," *Computer Animation and Virtual Worlds*, **28**(3–4), p. e1764.
- [59] Li, B., Yuan, J., Ye, Y., Lu, Y., Zhang, C., and Tian, Q., 2021, "3D sketching for 3D object retrieval," *Multimedia Tools and Applications*, **80**(6), pp. 9569–9595.
- [60] Luo, L., Gryaditskaya, Y., Yang, Y., Xiang, T., and Song, Y.-Z., 2020, "Towards 3D VR-Sketch to 3D Shape Retrieval," *2020 International Conference on 3D Vision (3DV)*, IEEE, pp. 81–90.
- [61] Liu, Z., Zhang, F., and Cheng, Z., 2021, "BuildingSketch: Freehand Mid-Air Sketching for Building Modeling," *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp. 329–338.
- [62] fxia22, 2020, "A pytorch implementation for pointnet," <https://github.com/fxia22/pointnet.pytorch>.
- [63] Ipsita, A., Li, H., Duan, R., Cao, Y., Chidambaram, S., Liu, M., and Ramani, K., 2021, "VRFromX: from scanned reality to interactive virtual experience with human-in-the-loop," *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–7.
- [64] Checa, D. and Bustillo, A., 2020, "A review of immersive virtual reality serious games to enhance learning and training," *Multimedia Tools and Applications*, **79**(9), pp. 5501–5527.
- [65] Bhattacharjee, S. and Chaudhuri, P., 2020, "A Survey on Sketch Based Content Creation: from the Desktop to Virtual and Augmented Reality," *Computer Graphics Forum*, Vol. 39, Wiley Online Library, Paper No. 2, pp. 757–780.
- [66] Annett, M. K. and Bischof, W. F., 2010, "Investigating the application of virtual reality systems to psychology and cognitive neuroscience research," *Presence: Teleoperators and Virtual Environments*, **19**(2), pp. 131–141.
- [67] Garrido, D., Rodrigues, R., Augusto Sousa, A., Jacob, J., and Castro Silva, D., 2021, "Point cloud interaction and manipulation in virtual reality," *2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 15–20.
- [68] Yang, E. K. and Lee, J. H., 2020, "Cognitive impact of virtual reality sketching on designers' concept generation," *Digital Creativity*, **31**(2), pp. 82–97.
- [69] Steinicke, F. and Hinrichs, K., 2006, "Grab-and-throw metaphor: Adapting desktop-based interaction paradigms to Virtual Reality," *3D User Interfaces (3DUI'06)*, IEEE, pp. 83–86.
- [70] Lee, G. A., Kim, G. J., and Billinghurst, M., 2005, "Immersive authoring: What you experience is what you get (wxyiwyg)," *Communications of the ACM*, **48**(7), pp. 76–81.
- [71] Chaudhuri, S. and Koltun, V., 2010, "Data-driven suggestions for creativity support in 3D modeling," *ACM SIGGRAPH Asia 2010 papers*, pp. 1–10.
- [72] Oh, C., Song, J., Choi, J., Kim, S., Lee, S., and Suh, B., 2018, "I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
- [73] Wang, T., Qian, X., He, F., Hu, X., Huo, K., Cao, Y., and Ramani, K., 2020, "CAPtURAR: An augmented reality tool for authoring human-involved context-aware applications," *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 328–341.
- [74] Labs, L., 2021, "3d Scanner App™," <https://apps.apple.com/us/app/id1419913995#?platform=ipad>.
- [75] Qi, C. R., Su, H., Mo, K., and Guibas, L. J., 2017, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660.
- [76] Aoki, Y., Goforth, H., Srivatsan, R. A., and Lucey, S., 2019, "Pointnetlk: Robust & efficient point cloud registration using pointnet," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7163–7172.
- [77] Cecil Piya, V., 2016, "RealFusion: An interactive workflow for repurposing real-world objects towards early-stage creative ideation," *Graphics interface*.
- [78] Gibson, J. J., 1977, "The theory of affordances," Hilldale, USA, **1**(2).
- [79] Wang, Y., Chen, Y., Nan, Z., and Hu, Y., 2006, "Study on welder training by means of haptic guidance and virtual reality for arc welding," *2006 IEEE international conference on robotics and biomimetics*, IEEE, pp. 954–958.
- [80] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J., 2015, "3d shapenets: A deep representation for volumetric shapes," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920.
- [81] Aoki, Y., 2020, "A pytorch implementation for PointNetLK," <https://github.com/hmgoforth/PointNetLK>.
- [82] Yang, M. C. and Cham, J. G., 2007, "An analysis of sketching skill and its role in early stage engineering design," .
- [83] Conesa-Pastor, J. and Contero, M., 2022, "EVM: An Educational Virtual Reality Modeling Tool; Evaluation Study with Freshman Engineering Students," *Applied Sciences*, **12**(1), p. 390.
- [84] Machuca, M. D. B., Asente, P., Stuerzlinger, W., Lu, J., and Kim, B., 2018, "Multiplanes: Assisted freehand vr sketching," *Proceedings of the Symposium on Spatial User Interaction*, pp. 36–47.
- [85] Luo, H., Wang, C., Wen, C., Chen, Z., Zai, D., Yu, Y., and Li, J., 2018, "Semantic labeling of mobile LiDAR point clouds via active learning and higher order MRF," *IEEE Transactions on Geoscience and Remote Sensing*, **56**(7), pp. 3631–3644.
- [86] Li, L., Sung, M., Dubrovina, A., Yi, L., and Guibas, L. J., 2019, "Supervised fitting of geometric primitives to 3d point clouds," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2652–2660.
- [87] Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B., 2018, "Pointcnn: Convolution on x-transformed points," *Advances in neural information processing systems*, pp. 820–830.
- [88] Ipsita, A., Erickson, L., Dong, Y., Huang, J., Bushinski, A. K., Saradhi, S., Villanueva, A. M., Pepler, K. A., Redick, T. S., and Ramani, K., 2022, "Towards modeling of virtual reality welding simulators to promote accessible and scalable training," *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–21.
- [89] Bossard, C., Kermarrec, G., Buche, C., and Tisseau, J., 2008, "Transfer of learning in virtual environments: a new challenge?" *Virtual Reality*, **12**(3), pp. 151–161.
- [90] Blikstein, P., 2018, "Maker movement in education: History and prospects," *Handbook of technology education*, **419**, p. 437.
- [91] Radu, I., Joy, T., and Schneider, B., 2021, "Virtual makerspaces: merging AR/VR/MR to enable remote collaborations in physical maker activities," *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–5.
- [92] Pellas, N., Kazanidis, I., and Palaigeorgiou, G., 2020, "A systematic literature review of mixed reality environments in K-12 education," *Education and Information Technologies*, **25**(4), pp. 2481–2520.
- [93] Birt, J. and Cowling, M., 2017, "Toward future 'mixed reality' learning spaces for STEAM education," *International Journal of Innovation in Science and Mathematics Education*, **25**(4).
- [94] Kaminka, G. A., 2013, "Curing robot autism: A challenge," *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, Citeseer, pp. 801–804.
- [95] Szafrir, D., Mutlu, B., and Fong, T., 2014, "Communication of intent in assistive free flyers. 358–365. ACM," *IEEE International Conference on Human-Robot Interaction (HRI)*.
- [96] Zhang, Y., Sreedharan, S., Kulkarni, A., Chakraborti, T., Zhuo, H. H., and Kambhampati, S., 2017, "Plan explicability and predictability for robot task planning," *2017 IEEE international conference on robotics and automation (ICRA)*, IEEE, pp. 1313–1320.
- [97] Williams, T., Szafrir, D., Chakraborti, T., and Ben Amor, H., 2018, "Virtual, augmented, and mixed reality for human-robot interaction," *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 403–404.
- [98] Cao, Y., Xu, Z., Li, F., Zhong, W., Huo, K., and Ramani, K., 2019, "V. ra: An in-situ visual authoring system for robot-iot task planning with augmented reality," *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 1059–1070.

List of Figures

1 Workflow of VRFromX (from left to right). (1) User scans a real environment to get the raw point cloud of the scene using an iPadPro. (2) User creates virtual content by interacting with the point cloud through AI assistance. (3) User attaches functions to the virtual objects using an Affordance Recommender. (4) User interacts with the reconstructed scene in a virtual environment. 2

2 Overview of VRFromX System 4

3 Interface of the Brush Tool: (a) Tip Size, (b) Add, (c) Delete, (d) Select, (e) Clear All, (f) Query, (g) Scale, (h) Tip Shape, (i) Brush Tip. User employs the Brush Tool to retrieve a chair (2-5). 6

4 Point cloud interaction for Object Retrieval (1) User identifies the region of interest and (2) selects the desired point cloud using the Brush tool. (3) AI shows top 5 object classes and (4) top 5 similar models for the input query. (5) User changes the selection by removing and (6) adding points. (7) AI provides a new set of class results as well as (8) top 5 similar models. 6

5 (1) User defines a region of interest and (2) re-purposes the point cloud of a table (3) to retrieve 3D model of a chair. 6

6 User assigns functions to a virtual model of a lamp with the aid of (1) an Affordance Recommender and attaches (2) Animations - Lights, (3) Displays - Intensity Panel with Data Flow Link, (4) Controls - Switch button with Trigger Link and (5) Controls - Slider. 7

7 Controls widgets used for Behavioral modelling in VRFromX 7

8 14 classes in welding dataset used to train the classification network for object retrieval in VRFromX 8

9 Box Plots showing the data distribution for the time of completion of the three tasks across (A) DUs, (B) NVRPS and (C) EVRPs 9

10 Questionnaire survey results after Task 1: Point Cloud Interaction for Object Retrieval, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree) 10

11 Questionnaire survey results after Task 2: Behavioral modelling for virtual objects, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree) 10

12 Questionnaire survey results after Task 3: Virtual welding, % denotes the number of participants from left to right who responded negatively (Strongly Disagree/Disagree/Somewhat Disagree), Neutral or positively (Somewhat Agree/Agree/Strongly Agree) 10

13 Screenshots showing participants interacting with the different virtual welding simulators authored by them during the study (from third person view in Unity™) 12

14 Remote 3D-Printing Use Case: (1) After retrieving a 3D Printer, (2) user creates 3D-models by point cloud selections and (3) sketching in mid-air, and finally (4) sends the model for remote printing. (5) IoT-enabled 3D printer prints the model. 14

15 Robot IoT Use Case: (1) User specifies the path and authors actions in VR and (b) the physical robot performs the tasks as authored by human in real world. 15

List of Tables

1 Table showing Demographics of Users who participated during the system evaluation of VRFromX 8

2 Table showing Comparative Analysis of EVRPs vs. NVRPs from survey questionnaire ratings and Time of Completion during the User Study Tasks 11