Prop-Chromeleon: Adaptive Haptic Props in Mixed Reality through Generative AI

ANONYMOUS AUTHOR(S)

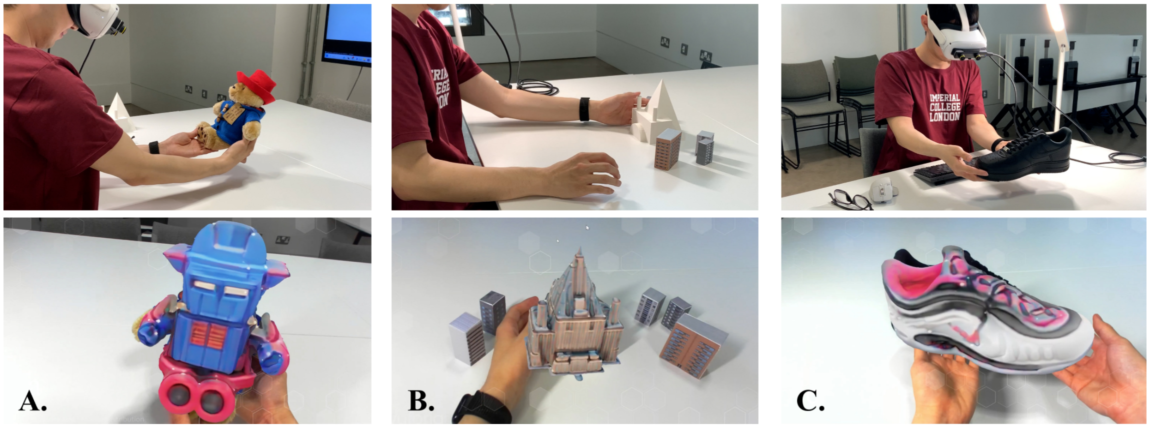

Fig. 1. Prop-Chromeleon transforms physical objects into adaptive passive haptic props: A) transforming a Paddington bear toy into “a cute transformer toy”; B) transforming a 3D printed white building model into “a Empire State Building Architecture”; and C) transforming a black Nike AF1 shoe into “a mix of Nike Air Foamposite One with Airmax 97 sneaker”.

Mixed Reality (MR) aims to seamlessly blend digital and physical worlds, but the absence of haptic feedback often disrupts the cohesion between visual and tactile experiences. We introduce Prop-Chromeleon, which dynamically repurposes everyday objects as haptic props for virtual assets using generative artificial intelligence (AI). Based on user prompts, Prop-Chromeleon generates virtual models and textures that are fit and overlaid on the physical prop shape. Prop-Chromeleon enables a single physical prop to act as haptic proxies for a wide range of virtual objects that through generative AI can change in real-time without the need for predefined proxies or virtual libraries. We demonstrate Prop-Chromeleon’s effectiveness in adaptively transforming physical objects to haptic virtual objects via a technical evaluation of diverse shapes and prompts. We also showcase Prop-Chromeleon’s ability to deliver believable passive haptic experiences in an MR user study. Prop-Chromeleon offers a scalable and adaptive approach to improving haptic MR experiences.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; **Virtual reality**; **User studies**.

Additional Key Words and Phrases: Mixed Reality, Generative AI, Passive Haptics

ACM Reference Format:

Anonymous Author(s). 2025. Prop-Chromeleon: Adaptive Haptic Props in Mixed Reality through Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI ’25), Apr 26th - May 1st, 2025, Yokohama, Kanagawa Prefecture, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Mixed reality (MR) removes the barrier between the physical and digital world, unlocking natural and intuitive 3D human, computer, and environmental interactions [29]. Consumer-level head-mounted displays like the Apple Vision Pro [21] have made significant strides in display and audio technologies. However, haptic feedback – a critical component for truly immersive experiences [20, 22] – is still notably absent in many MR applications today. This deficiency becomes evident when users attempt to physically interact with virtual objects but only to see themselves passing through the models, breaking the carefully crafted illusion of physical and virtual worlds blending together.

To bridge this gap, haptic devices have received significant attention in research from early force-feedback joysticks like the Phantom Device [28] to grounded [9, 46], hand-held [16, 18] and various wearable devices [6, 10, 11, 14, 27, 36, 41, 52]. Despite these advancements, current haptic solutions remain costly and bulky, restricted to laboratory use, or limited to simulate basic tactile sensations. Leveraging everyday objects as haptic props has been proposed as an alternative approach because it eliminates the need for external devices [3, 15, 19]. However, this method also faces challenges such as the mismatch between the varied shapes of virtual objects and the limited forms of physical ones. For optimal immersion, the physical props should closely match their virtual partner to avoid visual-haptic discrepancies that can disrupt user engagement [20, 42]. But in real-world environments, finding a universal set of physical props that accommodate most virtual content is unrealistic.

We believe that an ideal MR passive haptic system should overcome these challenges by handling various everyday objects across different user environments, creating case-by-case, on-demand haptic experiences. Generative artificial intelligence (AI) has seen significant progress in recent years with 2D text-to-image applications like Stable Diffusion [45] and Midjourney [30], ControlNet for precise control of generated images [51], and image-to-3D tools [47, 48, 53] have made it possible to reconstruct 3D meshes from a single image input. We propose to leverage these models to enable adaptive passive haptic props for MR.

We present Prop-Chromeleon, a MR system that leverages generative AI to adaptively transform daily objects into tangible proxies to be used in MR interactions. Prop-Chromeleon takes user prompts as input, captures the depth map of the physical object, and employs generative AI to create a corresponding 3D mesh. This mesh is then anchored to the physical object in real-time using AR authoring tools with 6-DoF (Degrees of Freedom) tracking, ensuring that the transformation remains accurate while the user is engaging with the object. Unlike conventional methods, Prop-Chromeleon does not require sourcing real-world proxies, carrying compatible objects, or following preset procedural steps. Instead, by leveraging generative AI, our system enables flexible transformations across various user settings.

We evaluated Prop-Chromeleon in two studies. First in a technical evaluation, we tested Prop-Chromeleon on a wide range of object shapes and user prompts. We measured the generated shape disparity using established Chamfer and Hausdorff distances, and assessed overall generation results through qualitative coding. The results showed a 90% successful generation rate, increasing with personalized prompts, and Hausdorff distances of 0.290 and Chamfer distances of 0.431 for generated shapes relative to the original object shape. The results demonstrated Prop-Chromeleon's ability to adaptively and accurately generate virtual objects according to the physical prop and the user prompt. Second in an MR user study, we explored users' perceptions of Prop-Chromeleon's performance and ability to provide passive haptic experiences with various physical objects and prompts. Our results showed that compared to an AI-generated baseline that does not consider the shape of the physical object, Prop-Chromeleon provides a significantly better perceived haptic experience and was preferred by the majority of study participants. In summary, our main contributions are as follows:

(1) Prop-Chromeleon, an adaptive MR system capable of transforming any physical object into specified virtual content while preserving its 3D structure.
(2) Results of a technical evaluation showing that Prop-Chromeleon is capable of handling a wide variety of transformations from a combination of diverse user prompts and object shapes, and an accompanied open-source dataset.
(3) Results of a user study showcasing Prop-Chromeleon's ability to provide immersive haptic experiences while also being preferred over a static baseline.

2 RELATED WORK

Our work builds on previous work on passive haptics, virtual and physical integration, and generative AI.

2.1 Virtual and Physical Alignment

Researchers in the XR domain have proposed multiple works that attempt to align virtual and physical objects and environments for various reasons. In virtual reality (VR), aligning virtual and physical objects enables passive haptics to provide additional immersion to the virtual experience. For example, *Haptic Retargeting* [3] is a technique that capitalizes on the dominance of visual cues to redirect users' hands when interacting with virtual objects, thus improving the sense of presence by aligning physical actions with virtual feedback. Similarly, *VR Haptics at Home* [15] uses common household objects, such as chairs and pillows, as passive haptic props for predefined game elements such as cannons and cats.

In the MR domain, several works attempt to align virtual and physical environments and objects to better integrate the worlds for increased immersion, enhancing physical objects with virtual components for additionally functionality, or simply replacing the texture of physical objects depending on the current need of applications. For example, Annexing Reality [19] identifies physical objects in the user's environment and overlays similarly shaped virtual objects for haptic feedback during interaction with virtual objects. *ARchitect* [25] shares a similar idea and allows users to actively reuse physical objects as interactables and obstacles with a preset of virtual models. Other approaches, such as *KinectFusion* [23], create live voxel-based reconstructions of the environment, enabling intimate interaction between virtual and physical realities. Projects like Remixed Reality [26] and *SceneCtrl* [50] expand on this concept, offering options such as selecting, moving, deleting, and copying objects to tangibly interact with and manipulate MR scenes. For outdoor applications, *DreamWalker* [49] allows users to walk in the real world while remaining fully immersed in large virtual environments using a headset.

Although these projects show the potential for integrating physical and virtual worlds and the importance of haptics for immersion, applications are commonly restricted to laboratory settings. As such, scalability issues arise due to their lack of adaptability across diverse user environments. For example, ARchitect, Annexing Reality, and VR Haptics at Home require a preset collection of haptic virtual objects, and when the system cannot find suitable real-world proxies, the experience suffers due to visual-haptic mismatch. In Prop-Chromeleon, we leverage generative AI to adaptively create virtual objects based on user prompts. This enables us to create virtual objects for a wide range of scenarios, furthermore, adapting the shape of the generated virtual objects provides believable haptic experiences and limits mismatches between the virtual and physical worlds.

2.2 Visual-Haptic Mismatch

To perceive the surrounding environment, our brain needs to integrate sensory signals from a common source but segregate those from other independent sources [39]. Thus, a conflict between visual and tactile perception can cause confusion and negatively affect the user experience. Rock and Victor [38] were among the first to investigate human experience when presented with conflicting visual and tactile information about an object's properties, who showed that although visual perception tends to dominate during such conflicts, a certain degree of mismatch is tolerable.

More recently, XR works have studied the extent to which virtual and physical environments can mismatch without being noticeable or causing a significant degradation to the user experience. In *Substitutional Reality*, Simeone et al. [42] explored the impact of these mismatches on the believability of VR experiences. In their study, participants interacted with a virtual environment paired with corresponding physical proxies and experienced different types of mismatches (e.g., shape, size, texture). They found that while some level of mismatch is tolerable, significant discrepancies largely reduce the authenticity of the haptic experience, and that object shape was the most critical factor affecting believability. Based on these findings, *Smart Substitutional Reality* [13] further explored this mismatch and found that enhanced tactile feedback not only increases immersion and spatial presence, but also increases user tolerance for higher levels of freedom in application design. These insights led us to develop Prop-Chromeleon which addresses visual-haptic mismatch by adjusting the shape of the virtual object to the physical object's shape. This enables Prop-Chromeleon to deliver believable passive haptic experiences.

2.3 3D Content Creation with Generative AI

Generative AI has recently seen significant advancements in delivering text, images, videos, and other media to a new era of AI for general creativity [5]. Text-to-image generation models that are pre-trained diffusion models on large text-image datasets such as Stable Diffusion [45], Midjourney [30] and DALL·E [32, 33, 35] have shown remarkable capabilities in producing high-quality images based on text-based user input. Furthermore, models such as ControlNet [51] have expanded the input space to enable spatial control of these image generation models by integrating features such as depth maps, normal maps, or soft edges which enable the creation of images based on specific 3D input models. Finally, advances in 3D reconstruction from 2D images have progressed through the evolution of generative models [24, 31, 47, 48, 53]. These methods can convert 2D images into 3D representations, such as meshes or point clouds, and enables the integration of generated content into 3D environments such as MR.

The use of generative AI tools in virtual environments has attracted widespread interest. Sra et al. [43] investigated using pre-scanned depth data and 3D reconstruction models to generate virtual reality scenes based on real-world templates. In another study, Sra et al. [44] uses music and generative AI to automatically create kaleidoscopic textures for virtual environments to elicit emotional responses in users. Recent work *LLMR* [12] presented a framework using Large Language Models (LLM) as an architect of C# Unity code to create interactive real-time MR experiences. In a related approach, Aghel Manesh et al. [1] users can prompt a LLM programming agent to create virtual environments without needing a programming or technical background. We build on these works and combine text-to-2D models with 3D reconstruction models and achieve prompt-guided 3D generation with spatial conditioning based on the user's prop and text-prompt to enable adaptive passive haptics in MR.

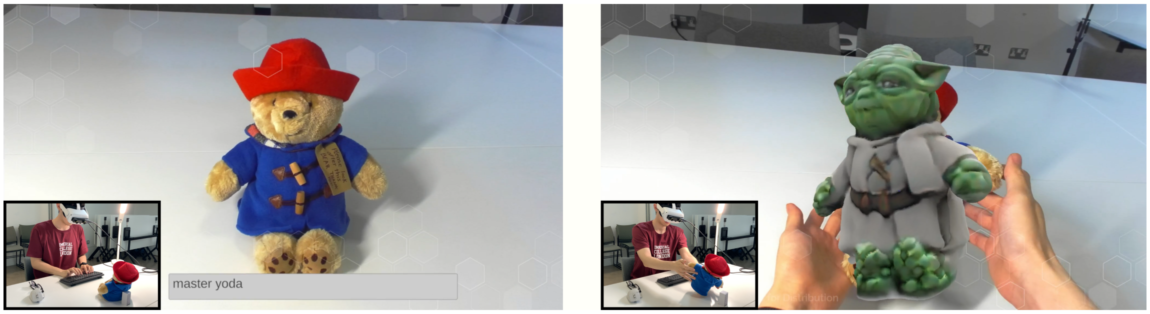

Fig. 2. Prop-Chromeleon from the user's point of view. Left: The user inputs a "Master Yoda" text prompt to transform the Paddington bear. Right: Prop-Chromeleon generates a 3D model that matches the prompt and is overlayed on the physical prop for haptic interaction with the 3D model.

3 PROP-CHROMELEON SYSTEM

Prop-Chromeleon visually transforms real-world physical objects into 3D models with matching shapes, based on user prompts for passive MR haptics. Designed for versatility and flexibility, Prop-Chromeleon can adapt to a wide range of object shapes, enabling users to engage their imagination and reshape their reality in a real-user environment, without the need for specialized equipment. In this section, we explain Prop-Chromeleon's user flow, design, and implementation. Prop-Chromeleon uses a VR headset and a depth camera to enable MR, a combination of generative AI models for the generation of 3D content, and AR authoring tools for spatial tracking.

3.1 User Journey

The user journey can be described as an interactive transformation process. After entering the designated scene, the user provides a prompt via a physical keyboard or speech (Figure 2). The user's prompt will then appear in an input field UI within the headset. For example, on Figure 2 left, the user types "Master Yoda" and clicks the space button on the keyboard. Once the prompt is confirmed, the system captures the scene's data and hide the input UI, sends the information to a cloud server for processing, and returns a corresponding 3D model, which will be anchored to the physical object in 6 Degrees of Freedom (DoF) in real-time. This process typically takes 20-30 seconds for the cloud server (with a single A10G) to generate and return the model. The user then sees the object being visually transformed into the generated 3D model in the headset, which remains aligned during movement and rotation. In our example (Figure 2 right), a generated Master Yoda appears aligned on the bear. The user can now interact directly with the real object while perceiving it as the virtual one in MR.

3.2 Prop-Chromeleon Pipeline

To generate 3D content based on user prompts and the shape of real-world objects, we combined a text-to-image AI model with a 3D reconstruction model, along with AR authoring tools and a capture system to create a unique pipeline (Figure 3). The processes of image generation and 3D transformation occur sequentially as a coroutine in a cloud server, preventing any heavy computational load from affecting the user.

3.2.1 Model Input. The input stage of the pipeline begins once the user completes their prompt (Figure 3A), and its primary goal is to capture precise information about the physical object to ensure accurate alignment of shape during

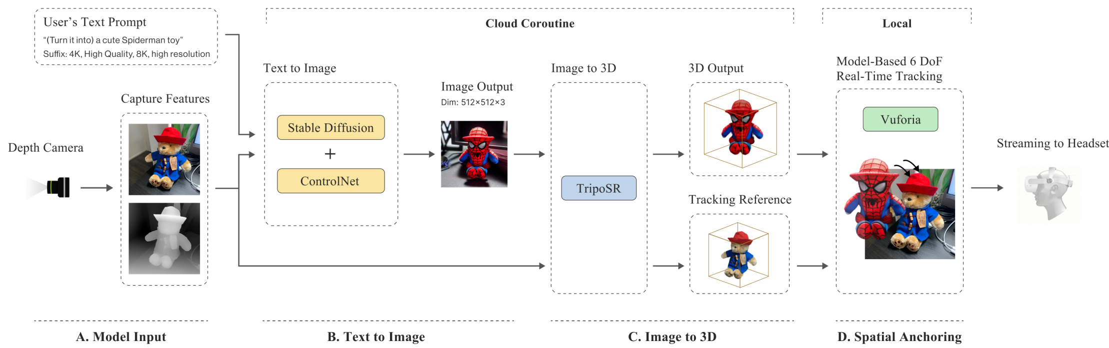

Fig. 3. The Prop-Chromeleon processing pipeline combines depth and object tracking, and integrates the user's prompt by combining Stable Diffusion and ControlNet to generate a 3D object that is overlaid and aligned with the physical prop.

generation. When the user finishes input and clicks the space button, the depth camera mounted in front of the HMD captures a depth map from the user's perspective which serves as Prop-Chromeleon input together with the user's prompt. If a depth camera is not available, an alternative way to add depth input is by taking a regular picture and using depth estimation algorithms. Although this eliminates the need for a depth camera, it adds extra time for depth inference and may yield less accurate results in low-light or blurry conditions.

3.2.2 Text to image. Our second component (Figure 3B) converts the model input into a 2D image that aligns with the user's prompt while adhering to the physical object's shape by leveraging generative AI models. The depth map from the Model Input component is used as the control frame in ControlNet [51], ensuring the image conforms to the physical object's shape. The user prompt guides the content creation, resulting in an image that aligns with the view, ready for the next stage (Image to 3D). We selected Stable Diffusion 2.1 [40] and ControlNet v1.1 [51] for this process due to their strong performance and open-source availability. ControlNet ensures precise shape control, while Stable Diffusion generates diverse and accurate images based on the prompt.

3.2.3 Image to 3D. To enable the user to view a generated 3D object rather than a flat image anchored to the physical object, this component (Figure 3C) reconstructs a 3D mesh based on the 2D image generated by the previous step. We used TripoSR [47] an open-source 3D reconstruction model for this purpose. Before initiating the mesh generation process, the background of the image is removed using the Rembg tool [17], which is natively implemented in the TripoSR to ensure that only the primary object in the scene is processed to a 3D object. The successful removal of the background is vital, as any residual elements can shift the center of the generated model, potentially causing errors in the tracking system and leading to a mismatch between the real and virtual objects. Maintaining a clean and simple background can help with the segmentation process.

3.2.4 Spatial anchoring. The final component of the pipeline uses AR authoring tools to anchor the generated 3D mesh to the real object (Figure 3D). We chose Vuforia [37] because of its robust tracking capabilities and general availability. In our system, we used Vuforia's Model Target function for rapid generation of tracking references from 3D models, supporting real-time 6 DoF tracking. This function allows us to use generated models as a tracking reference, scaling the system to accommodate various objects. Rather than directly using prompt-generated models as tracking references, we used TripoSR to transfer images of the original items for more stable tracking performance to items of various shapes (Figure 3). Once processing is complete, the model is returned to the user's view, resized to match the run-time tracking reference of the physical object, and attached to it in real time. This completes the Prop-Chromeleon process, and the user can now physically interact with the physical object that has been visually transformed to the virtual object (Figure 2b).

3.3 Implementation

We used Unity version 2022.3.20f1, Vuforia Engine version 10.22, ControlNet version 1.1, and Stable Diffusion version 2.1 with the version 1.4 checkpoint in our system. ControlNet is used as an extension for Stable Diffusion. We call their function through the webuiapi library using Python. We mounted a ZED Mini depth camera in front of a Meta Quest 2 headset as our hardware setup. It is essential for the depth camera to stream from the user's natural point of view to prevent confusion and ensure that it captures the depth map from the same perspective. This setup can be replicated using other depth-sensing devices, or even regular cameras by using ControlNet's native depth estimation.

4 STUDY 1: GENERATION

The Prop-Chromeleon system is designed to transform real-world objects into items prompted by the user while maintaining their original shape. Therefore, we conducted a technical analysis to assess its performance across a dataset of various object shapes and prompts. Our analysis consisted of two parts. First, we evaluated the system's ability to generate accurate 3D models which are overlaid on the physical object. Second, we qualitatively assessed the generated results of each combination of prompt and shape based on a set of heuristics relevant to the quality of Prop-Chromeleon output.

4.1 Preliminary Test

To better understand the types of prompts users might employ when interacting with Prop-Chromeleon, we conducted a preliminary test with 9 participants (6 Male, 3 Female, average age: 29.6, SD: 11.76) to explore user behavior patterns during prompt generation. All participants were recruited from a local university campus, with backgrounds in design, fine art and engineering. Each participant was first introduced to the system, after which they were provided with a Paddington bear (Figure 2) as a reference object and asked to generate at least 3 different prompts. We then collected all the prompts from the participants and listed them in Figure 4.

	Prompt 1	Prompt 2	Prompt 3
Participant 1	Casper *	a Casper ghost in smiling face ‡	Teddybear holding a rocket and spear weapon §
Participant 2	Spaceship *	silver spaceship with sharp head and blue tail ‡	Eiffel Tower *
Participant 3	Green Goblin from the original Spiderman movie †	Italian woman screaming bloody murder §	Tin Foil *
Participant 4	sport car †	a cute shark child toy †	a cute shark doctor child toy †
Participant 5	magician *	a witch wearing a hat §	a pink sheep †
Participant 6	a blue Smurfs wearing hats §	a boy wearing a suit §	Appleman *
Participant 7	a cowboy barbie in all pink †	an alien from Jupyter with four eyes ‡	a VC jelly gummy in orange flavour †
Participant 8	astronaut *	Alien bear †	Peppa pig *
Participant 9	deadpool *	SpongeBob *	kim jong un *

Fig. 4. User-generated prompts from preliminary study. Symbols indicate the belonging prop template specified in subsection 4.2: * = '[Noun]', † = '[Adjective] [Noun]', ‡ = '[Noun] with [Description]' and § = '[Noun] [Doing something]'.

4.2 Prop-Chromeleon Dataset

To comprehensively evaluate our system and ensure its relevance to real-user environments, our goal was to evaluate its performance in a wide variety of shapes and prompts that users are likely to encounter in everyday scenarios. As such, we built a generation dataset consisting of a diverse range of shapes and prompts. First, we randomly selected 50 3D models with distinct shapes from ShapeNetCore v2 [7], a large-scale 3D shape dataset, to represent a wide range of objects. This allowed us to evaluate a larger diversity of shapes instead of using real-world objects. We utilized Blender [4] to generate a depth map for each model from their Isometric View, positioned 35 degrees above the ground and rotated horizontally by 30 degrees, for application in Prop-Chromeleon.

For the prompts, we analyzed the patterns of the user-generated prompts from our preliminary tests, focusing on both syntax and content (Figure 4). This ensured that the prompts we use in our analysis represented the prompts we could expect users to input. Based on these prompts, we identified four template structures:

(1) '[Noun]'
(2) '[Adjective] [Noun]'
(3) '[Noun] with [Description]'
(4) '[Noun] [Doing something]'

We then generated two sets of prompts based on these templates using ChatGPT 4 [34]. The first set of general prompts was created without considering the shape of the objects and applied to all 50 models. We generated two prompts for each template, resulting in eight prompts. The second set of object-specific prompts was tailored to each object's shape, and we used ChatGPT's image capabilities to assist this process. Again, we generated two prompts per structure resulting in 8 prompts $\times$ 50 objects = 400 distinct prompts. We tested Prop-Chromeleon by applying general and object-specific prompts to the same set of 50 models, generating a total of 800 generations for evaluation. Please refer to our dataset repository [anonymized] for the selected models, prompts, and generated results.

4.3 Mesh Similarity Analysis

Shape similarity is a crucial metric to evaluate, as visual-haptic mismatches can lead to negative user experiences. We begin by capturing both the input geometry and the generated geometry from the pipeline as triangle meshes. Inspired by the evaluations in recent state-of-the-art neural surface reconstruction work in the computer graphics community [8], we utilize Hausdorff distance and Chamfer distance, both of which are well-established methods in visual computing to evaluate the geometry generation quality of our pipeline quantitatively. In datasets like ShapeNet, many objects have hollow interiors (e.g., thin volumes enclosing empty spaces, such as dressers or closets), rendering naive volumetric metrics like volume Intersection over Union infeasible. Therefore, we rely on surface-based measures. Hausdorff distance provides a lower bound measurement of geometric difference by identifying the maximum point-to-point distance between the input and generated shapes. Chamfer distance, as a more general point cloud-wise metric, offers a robust average-case analysis by summing the squared distances between corresponding points. Together, these metrics capture both outliers (extreme differences) and average similarity. Given that the initial poses of the input and generated geometries (derived from the depth map of the current camera view) often differ significantly, we normalize both shapes into a unit sphere. We then randomly sample rotations of the generated shape and perform rigid alignment for each permutation, taking the minimum squared sum distance before computing the final Hausdorff and Chamfer distances. This approach ensures that the results are not trapped in local minima, providing a more accurate and representative similarity assessment.

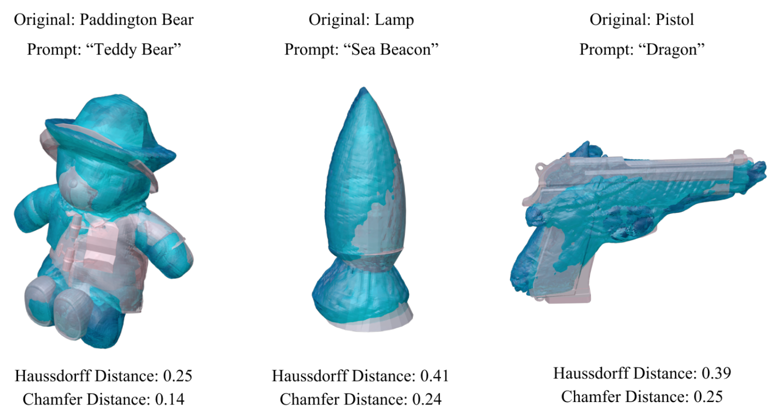

Fig. 5. Comparison between generated meshes (blue) and original object shapes (white) produced by our pipeline. The Hausdorff and Chamfer distances quantify the geometric differences between the original and generated shapes for each object.

4.3.1 Results. We computed the Chamfer and Hausdorff distances over 50 models, each evaluated with 16 unique prompts from two conditions, resulting in a total of 800 input-generated geometry pairs. We observe that excluding the top outlier groups significantly impacts the shape similarity metrics, with Chamfer and Hausdorff distances decreasing as more extreme entries are removed. After excluding the top 100 groups, the Chamfer distance drops to 0.270 and the Hausdorff distance to 0.410, while excluding the top 50 groups results in slightly higher values of 0.290 and 0.431, respectively. For consistent comparison, all shapes are normalized to a unit sphere with a radius of 1, where their size is scaled by the maximum displacement, and all objects are centered.

In addition to numerical results, visual comparison (Figure 5) shows that the input and generated 3D models are almost identical, with only slight differences in fine details. The low Chamfer distance confirms minimal average deviations between the models, while the low Hausdorff distance indicates that even the largest differences are small and localized. These visual similarities indicate that the calculated distances are indeed very low, reflecting the high accuracy of the generated shapes.

4.4 Prop-Chromeleon Generation Evaluation

Due to the subjective nature and variety of the generated results, it is challenging to evaluate prompt fidelity for each group with a single success-or-failure metric. In our study, we adopted a multi-dimensional approach and formulated three binary questions:

- **Q1:** *Does the model include the main elements from the prompt?*
- **Q2:** *Does the model accurately represent specific features from the prompt, like color or texture?*
- **Q3:** *Does the model reflect the intended theme of the prompt?*

We defined these three questions as we found that after reviewing multiple generations, the generated results of the prompts were expressed in three distinct ways. *Q1* incorporates the main element into the designated shape. This assesses how well the generated object's shape fits into the original one. For example, a successful generation in this category could be a spherical object transformed into a crouching dragon that fits the shape of the object. A failed generation would be generating a ball with dragon scales but lacking other essential dragon characteristics, such as the

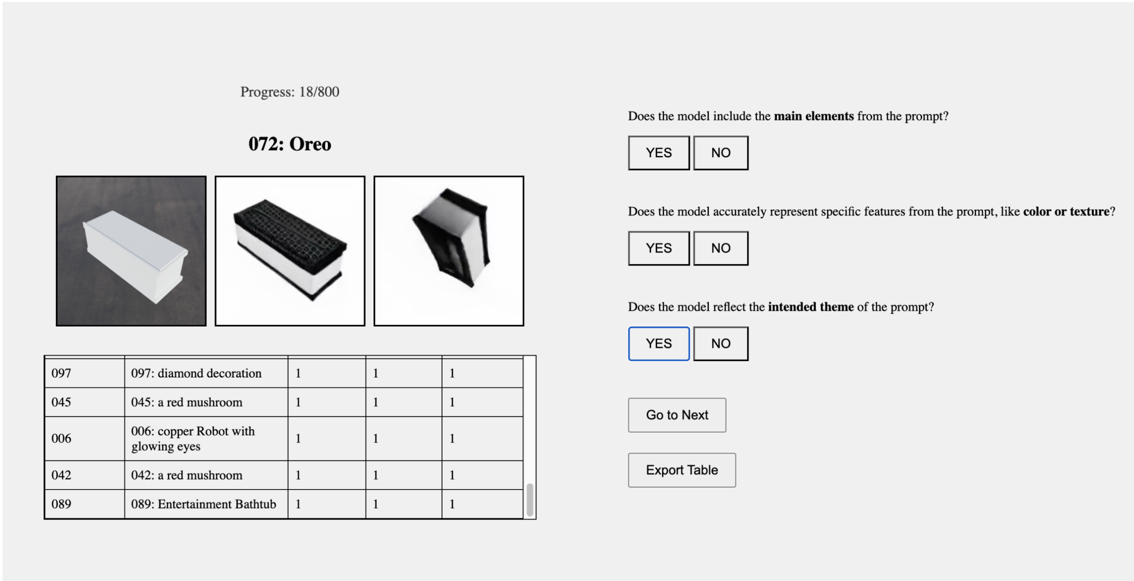

Fig. 6. The website interface used for evaluating generated results. It presents the prompt, original object, and generated object, along with a 360-degree rotating video of the generated object. Each generation was evaluated by answering three questions, with the order of presentation randomized to ensure unbiased evaluation.

head or eyes. *Q2* represents the color or texture. For instance, while the dragon-scaled ball may fail in terms of shape, it would be a successful result in terms of texture. Some prompts, such as "jewelry with gold and silver", express their features primarily through this category. *Q3* represents the intended theme, which refers to the general perception of the result. Prompts like 'A talking cactus with a sombrero' or 'futuristic vehicle' can be interpreted in various ways when considering the first two categories. Also, sometimes the "generated results" for these prompts can be more abstract. Therefore, we included the "intended theme" to evaluate this aspect from a broader perspective.

We created a website for the evaluation that presents all the necessary information for coding (Figure 6), including the prompt, the original object, the generated object, and a 360-degree rotating video of it. Each generation was coded by answering the three questions. The generation order was randomized. A randomly selected subset of 112 generations were coded by three authors to ensure alignment between raters. We then used Fleiss' kappa to evaluate the alignment between coders. Overall, there was a very good alignment between the raters. Q1 had a very good alignment, $\kappa = .848$ (95% CI, .741 to .955), $p < .001$, 105 of 112 aligned (93.75%). Q2 had a good alignment, $\kappa = .747$ (95% CI, .640 to .854), $p < .001$, 111 of 112 aligned (99.11%). Finally, Q3 also had a very good alignment $\kappa = .851$ (95% CI, .744 to .958), $p < .001$, 104 of 112 aligned (92.86%). After establishing a high alignment between coders, the rest of the dataset were rated by a single author.

4.4.1 Results. Overall, our results indicate that the Prop-Chromeleon system effectively generates 3D models that align with user prompts, particularly when customized prompts are tailored to specific requirements. For Q1, which assesses whether the main elements from the prompt were included, custom prompts achieved a 95.5% success rate, compared to 64.5% for general prompts, with an overall success rate of 80.0%. In Q2, evaluating color and texture accuracy, the system performed consistently well, with a 97.0% success rate for custom prompts and 90.3% for general prompts (93.4%

overall). Q3, focusing on thematic accuracy, showed a 94.3% success rate for custom prompts and 63.5% for general prompts, with an overall success rate of 78.8%.

These results indicate that customized prompts generate better results in terms of key elements and thematic accuracy compared to general prompts. This improvement is due to the alignment between the original object and the target output, as well as the flexibility allowed by the prompt. For example, transforming a bucket into a mug yields more accurate results than transforming it into an owl, which requires a more constrained bird-like shape, limiting the system's ability to capture the necessary features. Similarly, prompts like "copper robot" achieve higher accuracy in both elements and theme compared to the owl, as the robot allows for more flexible shapes. However, color and texture generation is less dependent on shape similarity, resulting in higher accuracy across both prompt types.

In conclusion, the Prop-Chromeleon system performs optimally when using custom prompts that consider the context of the original object, especially when there is flexibility in shape. Color and texture accuracy remain consistently high, while thematic alignment improves significantly with personalized input, highlighting the system's adaptability to user-specific needs.

4.5 Summary

In summary, our evaluations demonstrated Prop-Chromeleon's strong performance in capturing key prompt elements, color, texture and thematic accuracy. The results also showed that generation was most successful when prompts were customized to match the object's form. Additionally, mesh similarity analysis using Chamfer and Hausdorff distances confirmed low geometric deviations between the input and generated models, validating the pipeline's effectiveness in maintaining high shape accuracy. The dataset used for our analysis can be found at [ANONYMIZED], accompanied by detailed instructions to replicate the experimental setup.

5 STUDY 2: USER TESTS

To evaluate the effectiveness of Prop-Chromeleon in enhancing haptic experiences and achieving usable transformations, we conducted a user study focused on how shape alignment within Prop-Chromeleon improves haptic feedback, as well as collecting user feedback on the overall experience and transformation quality.

5.1 Participants

Twelve participants (7 males, 5 females) aged 23-54 (M = 30.33, SD = 10.44) took part in the study. Participants were asked to rate their experience with VR/AR and Generative AI on a scale of 1 (minimal experience) to 5 (extensive experience). The sample had moderate experience with VR/AR (M = 2.50, SD = 1.38) and Generative AI (M = 2.67, SD = 1.37).

5.2 Conditions

Two conditions were created to evaluate the impact of shape alignment on haptic feedback in Prop-Chromeleon. The key difference between the conditions was the alignment of the virtual model with the physical object.

Prop-Chromeleon Condition: In this condition, ControlNet was used to align the virtual model with the haptic prop, ensuring that the virtual and physical objects matched in shape, as shown in the middle of Figure 7.

Baseline Condition: In this condition, no shape alignment was provided, serving as a baseline for comparison, as shown in the right part of Figure 7.

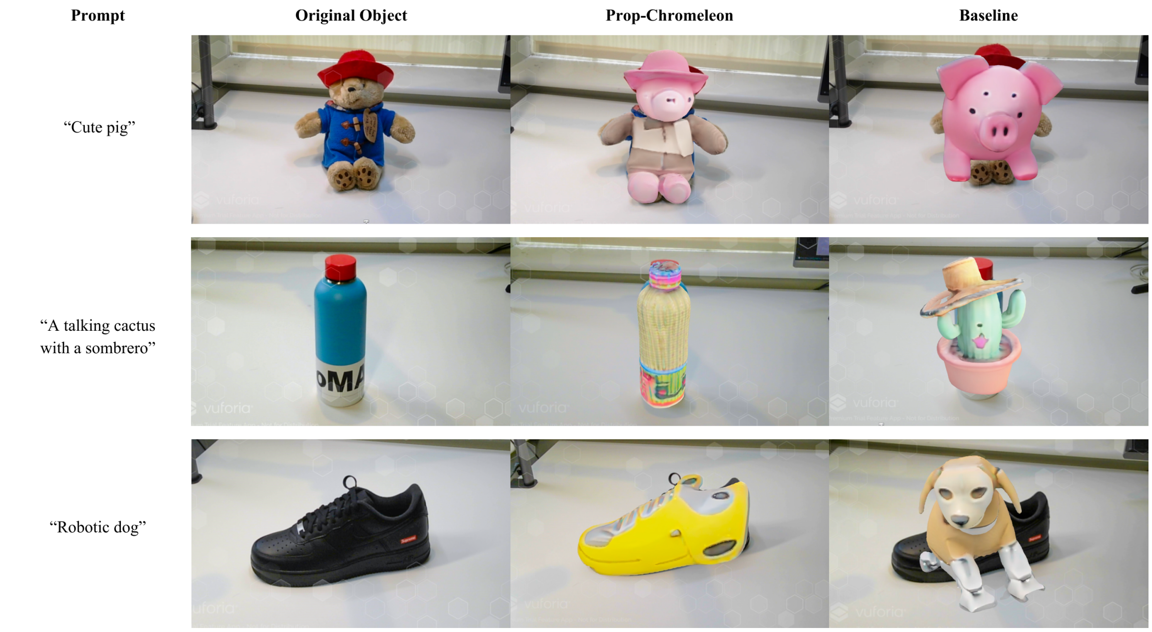

Fig. 7. Prompts, original physical objects, Prop-Chromeleon generations, and baseline generations used for Task 1 in our user study.

Both conditions were tested across three scenes, each with a different physical prop: a Paddington Bear toy, a Nike Air Force One shoe, and a stainless steel bottle(Figure 7). The same virtual models were generated in both conditions using the following prompts: 'cute pig' for the bear, 'yellow robotic dog' for the shoe, and 'talking cactus with a sombrero' for the bottle. The virtual models were anchored to their corresponding physical objects in Unity.

5.3 Tasks

5.3.1 Task 1: System Comparison. Task 1 was designed to compare the effectiveness of shape alignment in Prop-Chromeleon with a baseline condition. Participants interacted with three physical objects across both conditions: one where the virtual model was aligned with the physical object (using Prop-Chromeleon) and one where no alignment was provided (baseline). For each object, participants were asked to follow these instructions:

(1) Without touching it, observe the object from different angles.
(2) Slowly hold the object with both hands, and move it to the other side of the table.
(3) Touch and feel the shape and outline of the object.
(4) Imagine it's a game, interact with the model in any way you prefer.

5.3.2 Task 2: Prop-Chromeleon Exploration. In the second task, participants fully engaged with Prop-Chromeleon by transforming the same three physical objects from Task 1. Participants could choose any prompt for each object, and Prop-Chromeleon generated corresponding virtual models based on those prompts. After each transformation, they were asked to rate the generation result on "how well it met their expectations" on a scale of 1-5. Additionally, their behavior and prompt choices were observed and discussed to gather further insights into the user experience.

5.4 Procedure

Before starting the study, participants first signed a consent form and answered a demographic questionnaire. Each participant was then introduced to the concept of VR haptics, which included an explanation of the tactile sensations experienced when interacting with virtual objects in typical VR environments. Participants then put on the headset equipped with the attached depth camera system and were given adequate time to adjust it for comfort. Following this, they received a tutorial on using Prop-Chromeleon. Once participants were comfortable with the system, they were guided to proceed to the two tasks described below, which designed to evaluate Prop-Chromeleon's effectiveness in enhancing haptic experiences. All participants began with Task 1, as shown in Figure 9. To balance the study, half of the participants started with the Prop-Chromeleon condition, while the other half began with the baseline condition. After completing both conditions, participants engaged in a semi-structured interview where they discussed their ratings, system preferences, and how various factors influenced their experience. After completing all three scenes in one condition, participants filled out a questionnaire based on the multidimensional Haptic Experience (HX) scale. The HX scale is an 11-item questionnaire designed to measure four key dimensions of haptic interaction: Realism, Harmony, Involvement, and Expressivity [2]. For Task 2, participants engaged in another semi-structured interview following the rating process, providing further feedback on their experiences with Prop-Chromeleon. This discussion helped to identify patterns in how participants interacted with the transformed models and prompted a deeper understanding of user preferences and engagement.

5.5 Results

We evaluated Prop-Chromeleon's effectiveness and usability through two tasks. Overall, both tasks was conducted successfully. A combination of quantitative results and insights from post-study interviews led to several key findings.

5.5.1 Questionnaires and Preferences. When asked about their preferred system, eleven out of the twelve participants stated that they preferred Prop-Chromeleon, while only one participants mentioned the baseline as their favorite. These results were further highlighted by the results of the HX questionnaire (Figure 8) which were generally in favour of Prop-Chromeleon over the baseline. The participants perceived that the haptic feedback for Prop-Chromeleon was significantly more "realistic" (R1: z=2.59, p=.010, r=.75), "believable" (R2: z=3.10, p=.002, r=.89), and "convincing (R3: z=2.87, p=.004, r=.82) than the baseline. These results were further confirmed by the significant difference for the combined "Realism" factor (PA1: z=3.06, p=.002, r=.88).

The participants further indicated that the haptic feedback for the baseline was significantly more "disconnected from the rest of the experience" (H3: z=3.13, p=.002, r=.90), "felt out of place" (H5: z=2.98, p=.003, r=.86), and "distracting" (I1: z=2.70, p=.007, r=.77) than Prop-Chromeleon. These results were further confirmed by the significant difference in the "Harmony" factor (PA2: z=3.06, p=.002, r=.88). Meanwhile, participants also indicated that for Prop-Chromeleon, participants significantly more "liked having the haptic feedback part of the experience" (H2: z=2.88, p=.004, r=.83) and felt significantly more "engaged with the system due to the haptic feedback" (I2: z=2.76, p=.006, r=.80). These results were again confirmed by the combined "Involvement" factor (PA3: z=3.06, p=.002, r=.88).

"Engagement" was the only factor that we found no significant differences between Prop-Chromeleon and the baseline (PA4: z=1.68, p=.092, r=.48). Although the results showed that the participants found that "the haptic feedback changes based on changes in the system" (E4: z=2.07, p=.038, r=.60) significantly more for Prop-Chromeleon. We found no significant differences when asked whether each system's haptic feedback "all felt the same" (E1: z=0.96, p=.336, r=.28), or whether "the haptic feedback reflects varying inputs and events" (E5: z=1.55, p=.121, r=.45). These results

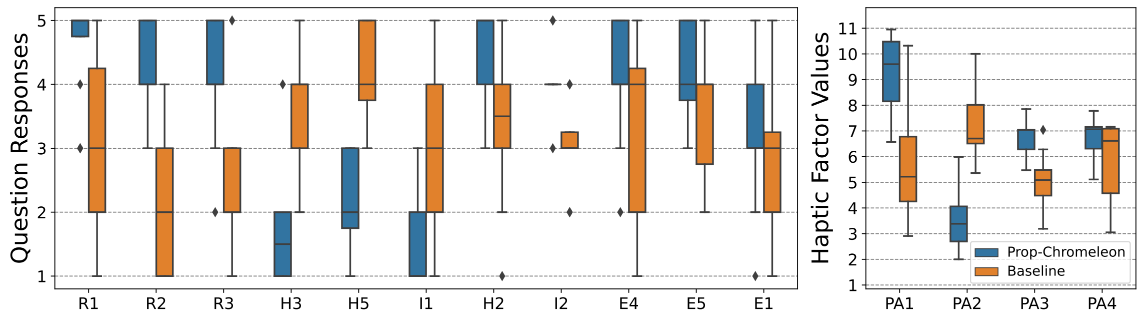

Fig. 8. Questionnaire results for individual questions, and haptic factors. We found a significant difference for all questions and factors except E1, E5, and PA4.

Fig. 9. Participants interacting with different condition groups in the user study, Left: Participant moving the prop in the shoe-Robotic Dog scene, baseline condition; Middle: Participant touching the prop's hat brim in the Paddington Bear-Cute Pig scene, Prop-Chromeleon condition; Right: Participant engaging with prop in the Paddington Bear-Cute Pig scene, Prop-Chromeleon condition;

are not surprising, since we only evaluated passive haptic systems where the haptic experience was static based on the shape of the physical object.

5.5.2 Interview Results.

Prop-Chromeleon improves user interaction and engagement. Prop-Chromeleon notably improved user interaction and the quality of the haptic experience. We noticed that participants demonstrated a smoother and more active interaction, with the ability to precisely locate and manipulate objects with Prop-Chromeleon (Figure 9). This contrasted with the baseline, where several participants missed their initial grip due to a misalignment between real and virtual objects.

When asked which condition they preferred, 11 of the 12 participants favored Prop-Chromeleon, describing it as more "realistic"(P1, P3, P9) and "convincing" (P2, P7, P8, P10), and the haptic experience felt more "natural" (P2). The participants also specifically commented on the difference between Prop-Chromeleon and the baseline. P8 described that moving objects with the baseline felt like "using a controller", while with Prop-Chromeleon, it felt like they were "really touching the object". Some participants mentioned that although the haptic feedback for the baseline was "helpful to some extent" (P12), they doubted "how much lift it brings to the experience" (P10).

Shape alignment enhances the haptic experience. Participants highlighted that Prop-Chromeleon provided a more immersive haptic experience. Several noted that the third task of touching and feeling the object's shape and outline felt the most different between the two conditions. While interacting with the Paddington Bear, P10 remarked: "For a moment, I totally forgot it was something else there". P9 and P2 said that touching the bear's hat brim in Prop-Chromeleon, while seeing themselves touching a virtual hat in the headset, was "the most impressive moment". P4 shared a similar experience when touching the hand and feet of the bear. For the baseline, such deep experiences regarding haptic feeling were not mentioned. These results indicate that Prop-Chromeleon's shape-alignment approach enhances the perceived haptic experience.

Prop-Chromeleon provided enjoyable and transformative experiences. Prop-Chromeleon not only improved the perceived haptic experience but also made the experience enjoyable and transformative. For task two, participants responded overall very positively to the generated transformations (M=3.94, SD=0.92). Many participants described the Prop-Chromeleon experience as "fun" and "interesting". Participants interaction during the second task was often joyful and filled with laughter during the generated transformations. Many expressed excitement while waiting for the results after inputting their prompts, and surprise upon seeing the outcomes. In post-study interviews, many said they would like to try it again.

When asked "If you could take this home, where would you apply it?" Participants offered creative answers beyond hand-held props, such as having a "Star Wars themed bedroom", a "cool painted bike", a "luxury villa", or turning a person into "James Bond". Others envisioned more practical uses for design and productivity, such as "helping generate design ideas", "rapid prototyping", or "making a pill bottle more child-friendly". Some participants also suggested adding a feature to recommend what to generate based on the real object.

Influence of Physical Props on Prompt Choice. The choice of prompts varied significantly among the participants and was at times influenced by the physical props themselves. The participants gave a wide range of prompts in task 2 (Figure 10). Some referenced pop culture, like Star Wars, superheroes, or cartoon characters for the Paddington Bear, while others chose real-world items such as "Palm tree with big leaves and a giant trunk" for the bottle or "Seafood risotto" for the shoe. Imaginative prompts included "Silver Grater" for the shoe, and real figures such as "Donald Trump" and "Kim Jong Un" for the Paddington Bear. Among the props, the Paddington Bear transformations received the highest rating (M=4.17, SD=0.58), followed by the Shoe transformations (M=4.00, SD=0.95), and the water bottle transformations (M=3.67, SD=1.15). We also observed that the participants' choice of prompts was sometimes influenced by the haptic prop. For example, when using Paddington Bear as the prop, many prompts related to animals, such as 'A blue lion' or 'Pig'. Similarly, with the bottle prop, prompts like "Palm tree" and "R2-D2," which are column-shaped, were common. The variation of prompts combined with the positive response from participants showcases the highly transformative capabilities of Prop-Chromeleon.

Texture mismatch can negatively affects realism and believability. Although 11 of the 12 participants favored Prop-Chromeleon over the baseline group, one participant preferred the baseline. They found that texture mismatches, such as a hairy sensation instead of the expected pigskin for the Paddington bear, were unrealistic. They also preferred to be able to see uncovered parts of the real object, which made them feel more "secure" and "safe" when reaching for it.

Similarly, when asked what factors stopped you from trusting Prop-Chromeleon's generated object, other participants also mentioned texture mismatches. For example, P7 expected a cactus prompt input to feel "prickly" or "soft" but the stainless steel bottle underneath felt "cold and metallic", with P8 noting that this kind of mismatch "reminded me this is not real". These comments show that in addition to shape, texture alignment also plays an important role in enhancing the believability of haptic experiences.

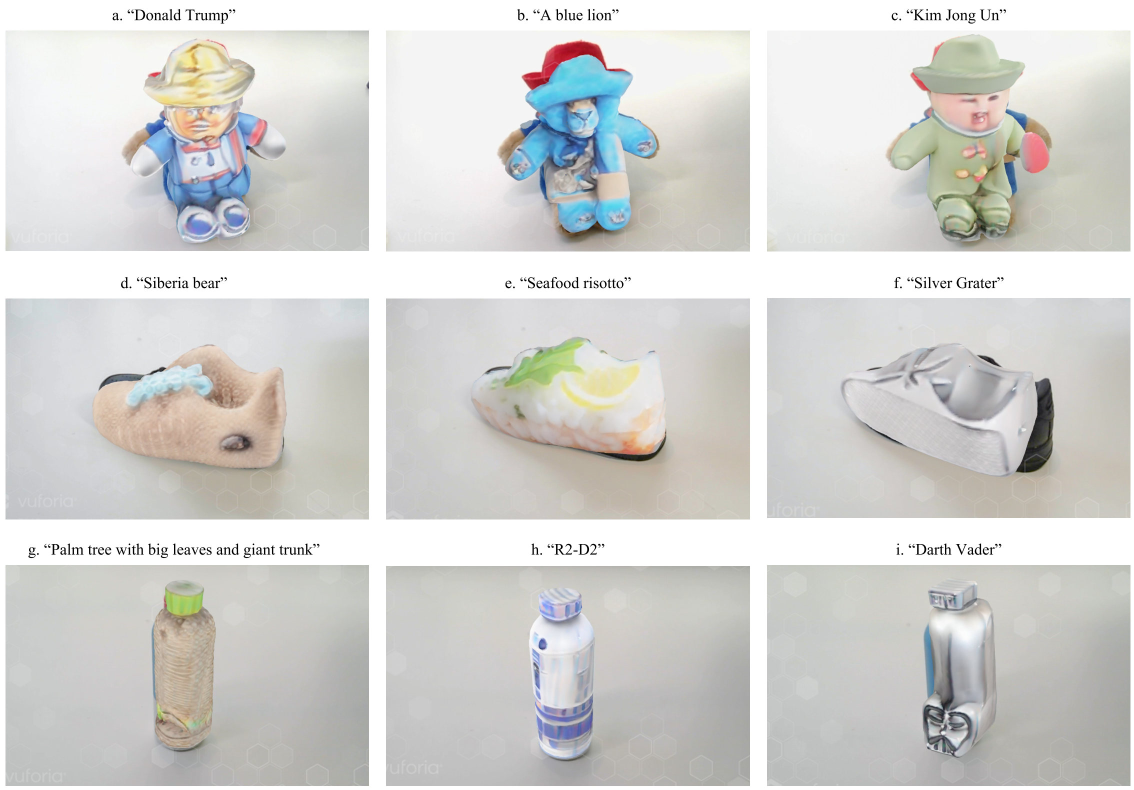

Fig. 10. Examples of user-generated prompts and their corresponding generated results from Task 2.

Occlusion conflicts can negatively affect immersion and spatial perception. Conflicting occlusion cues between hands and 3D object due to tracking limitations would occasionally negatively affect the haptic experience. When the generated model was first generated, some participants perceived it as a flat layer on the screen rather than a 3D object, as the model would seem to block the hands, making it look closer to their eyes. Most participants accordingly adjusted their hand interactions accordingly to avoid confusion. However, a participant (P6, with limited VR experience, 1/5) struggled with this in later steps.

6 DISCUSSION

We introduced Prop-Chromeleon, a novel approach that transforms everyday physical objects into haptic props for virtual models created through user prompts, seamlessly integrating generative AI into MR environments. Prop-Chromeleon enable highly adaptive passive haptic that do not rely on pre-existing libraries of virtual objects to match the shape of the physical objects, while at the same time supporting a wide variety of physical shapes. Our technical evaluations and user studies demonstrate the system's ability to generate and align 3D models with physical objects by considering the original shape of the physical object during generation, thus creating immersive and engaging haptic experiences, which underscore Prop-Chromeleon's potential to bridge the gap between physical interaction and virtual content.

Our generation study highlighted several key factors that contribute to a successful transformation. Most notably, customized prompts that align closely with the haptic prop's shape consistently led to superior results. Prompts that describe objects with flexible or adaptable forms, such as a "spaceship" or "jewelry with silver and gold", were particularly effective. This result is largely due to the way that generative models interpret and transform textual input into visual content. Ensuring harmony between the user's prompt and the shape of the original physical is essential to achieve visually convincing transformations. In addition, our user study revealed the importance of managing texture mismatches, as they can influence the user's perception of the transformation. Future work could explore ways to help users craft optimal prompts or using alternative modalities such as images as input for generation. One promising avenue is to integrate additional models to analyze the texture and shape of the physical object to provide intelligent, contextually relevant suggestions for more accurate and effective transformations.

The user study confirmed that Prop-Chromeleon significantly enhances the haptic experience, with shape alignment emerging as a critical factor in this improvement. By ensuring that the physical object closely mirrors the virtual model, users can engage in more immersive interactions, which enhances the overall sense of realism. Texture also plays a role, as noted in participant feedback. Users consistently highlighted the impact of texture congruence between the virtual model and the physical object on their haptic experience. Although Prop-Chromeleon already provides a highly engaging experience, future developments could focus on better aligning texture properties to further enhance realism.

In Prop-Chromeleon, we opted to generate 3D models using single-view depth maps rather than full 3D scans. Although full scanning could provide more detailed geometry, the trade-offs in terms of processing complexity and real-time responsiveness made single-view depth capture a more practical solution. Both our generation analysis and user studies demonstrated that this approach produces sufficiently accurate 3D models, enabling seamless alignment between virtual and physical objects. Future work may explore multi-angle depth capture, progressive updates to the depth capture, or other techniques to address edge cases where model fidelity could be further improved. However, the current implementation shows that the trade-off made in favor of speed and accessibility does not significantly compromise the quality of interaction or user satisfaction.

Participants responded enthusiastically to their experience with Prop-Chromeleon, describing it as "fun" and "interesting." The ability to personalize transformations and interact with physical objects in novel ways added a layer of creativity and engagement that was well received. This positive feedback highlights Prop-Chromeleon's potential to captivate users by combining tangible interactions with the flexibility of generative AI. The participants also mentioned multiple avenues for generation beyond hand-held objects such as entire rooms or people. We believe that integrating Prop-Chromeleon with systems such as *LLMR* [12] or *RemixedReality* [26] for room-scale generation introduces an exciting design space for future exploration and the general use of generative AI in MR.

7 CONCLUSION

We presented Prop-Chromeleon, a novel passive haptic system that transforms everyday physical items into haptic props for AI-generated virtual objects in MR that adapt to the shape of the physical object. We show in two user studies that Prop-Chromeleon is capable of handing a wide variety of object shapes and user prompts. Furthermore, an MR user study showed that Prop-Chromeleon provides believable haptic experiences and was preferred to a static baseline. In sum, Prop-Chromeleon show that Generative AI can be effectively used for passive haptics in MR. Furthermore, considering the shape of the physical object during generation enhances the haptic experience. Our work opens up further research on additional model input and beyond hand-held objects for object transformation to accurately shape haptic MR experiences in our daily lives.

REFERENCES

[1] Setareh Aghel Manesh, Tianyi Zhang, Yuki Onishi, Kotaro Hara, Scott Bateman, Jiannan Li, and Anthony Tang. 2024. How People Prompt Generative AI to Create Interactive VR Scenes. In *Designing Interactive Systems Conference (DIS '24, Vol. 7)*. ACM, 2319–2340. https://doi.org/10.1145/3643834.3661547

[2] Ahmed Anwar, Tianzheng Shi, and Oliver Schneider. 2023. Factors of Haptic Experience across Multiple Haptic Modalities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 260, 12 pages. https://doi.org/10.1145/3544548.3581514

[3] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[4] Blender Foundation. 2023. Blender - a 3D modelling and rendering package. https://www.blender.org Version 3.4.0, Accessed: 2023-09-12.

[5] Hanqun Cao, Cheng Tan, Zhangyang Gao, Yilun Xu, Guangyong Chen, Pheng-Ann Heng, and Stan Z. Li. 2022. A Survey on Generative Diffusion Models. *IEEE Transactions on Knowledge and Data Engineering* 36 (2022), 2814–2830. https://api.semanticscholar.org/CorpusID:265039918

[6] D. Chakarov, I. Veneva, M. Tsveov, P. Mitrouchev, and P. Venev. 2019. Design of a Two Arms Exoskeleton as Haptic Device for Virtual Reality Applications. In *Advances on Mechanics, Design Engineering and Manufacturing II*, Francisco Cavas-Martínez, Benoit Eynard, Francisco J. Fernández Cañavate, Daniel G. Fernández-Pacheco, Paz Morer, and Vincenzo Nigrelli (Eds.). Springer International Publishing, Cham, 252–262.

[7] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. 2015. *ShapeNet: An Information-Rich 3D Model Repository*. Technical Report arXiv:1512.03012 [cs.GR]. Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[8] Zhiqin Chen, Andrea Tagliasacchi, Thomas Funkhouser, and Hao Zhang. 2022. Neural dual contouring. *ACM Trans. Graph.* 41, 4, Article 104 (jul 2022), 13 pages. https://doi.org/10.1145/3528223.3530108

[9] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3718–3728. https://doi.org/10.1145/3025453.3025753

[10] Inrak Choi, Heather Culbertson, Mark R. Miller, Alex Olwal, and Sean Follmer. 2017. Grabity: A Wearable Haptic Interface for Simulating Weight and Grasping in Virtual Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 119–130. https://doi.org/10.1145/3126594.3126599

[11] Turhan Civelek and Arnulph Fuhrmann. 2022. Virtual Reality Learning Environment with Haptic Gloves. In *Proceedings of the 2022 3rd International Conference on Education Development and Studies* (Hilo, HI, USA) *(ICEDS '22)*. Association for Computing Machinery, New York, NY, USA, 32–36. https://doi.org/10.1145/3528137.3528142

[12] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time Prompting of Interactive Worlds using Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 600, 22 pages. https://doi.org/10.1145/3613904.3642579

[13] Benjamin Eckstein, Eva Krapp, Anne Elsässer, and Birgit Lugrin. 2019. Smart substitutional reality: Integrating the smart home into virtual reality. *Entertainment Computing* 31 (2019), 100306. https://doi.org/10.1016/j.entcom.2019.100306

[14] Cathy Fang, Yang Zhang, Matthew Dworman, and Chris Harrison. 2020. Wireality: Enabling Complex Tangible Geometries in Virtual Reality with Worn Multi-String Haptics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376470

[15] Cathy Mengying Fang, Ryo Suzuki, and Daniel Leithinger. 2023. VR Haptics at Home: Repurposing Everyday Objects and Environment for Casual and On-Demand VR Haptic Experiences. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 312, 7 pages. https://doi.org/10.1145/3544549.3585871

[16] Martin Feick, Cihan Biyikli, Kiran Gani, Anton Wittig, Anthony Tang, and Antonio Krüger. 2023. VoxelHap: A Toolkit for Constructing Proxies Providing Tactile and Kinesthetic Haptic Feedback in Virtual Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 104, 13 pages. https://doi.org/10.1145/3586183.3606722

[17] Daniel Gatis. 2023. rembg. https://github.com/danielgatis/rembg Accessed: 2023-09-12.

[18] Eric J Gonzalez, Eyal Ofek, Mar Gonzalez-Franco, and Mike Sinclair. 2021. X-Rings: A Hand-mounted 360° Shape Display for Grasping in Virtual Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 732–742. https://doi.org/10.1145/3472749.3474782

[19] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1957–1967. https://doi.org/10.1145/2858036.2858134

[20] H.G. Hoffman. 1998. Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*. 59–63. https://doi.org/10.1109/VRAIS.1998.658423

[21] Apple Inc. [n.d.]. Apple Vision Pro. https://www.apple.com/apple-vision-pro/.

[22] Brent Edward Insko. 2001. *Passive haptics significantly enhances virtual environments*. Ph.D. Dissertation. The University of North Carolina. Advisor(s) Brooks, Frederick P. AAI3007820.

[23] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 559–568. https://doi.org/10.1145/2047196.2047270

[24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 42, 4, Article 139 (jul 2023), 14 pages. https://doi.org/10.1145/3592433

[25] Chuan-en Lin, Ta Ying Cheng, and Xiaojuan Ma. 2020. ARchitect: Building Interactive Virtual Experiences from Physical Affordances by Bringing Human-in-the-Loop. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376614

[26] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173703

[27] Yuxin Ma, Tianze Xie, Peng Zhang, Hwan Kim, and Seungwoo Je. 2024. AirPush: A Pneumatic Wearable Haptic Device Providing Multi-Dimensional Force Feedback on a Fingertip. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 430, 13 pages. https://doi.org/10.1145/3613904.3642536

[28] Thomas H. Massie and Kenneth J. Salisbury. 1994. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*.

[29] Microsoft. 2023. Discover Mixed Reality. https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality. Accessed: 2023-07-30.

[30] Midjourney. [n.d.]. Midjourney: An independent research lab. https://www.midjourney.com/home. Accessed: 2024-08-06.

[31] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (dec 2021), 99–106. https://doi.org/10.1145/3503250

[32] OpenAI. 2021. DALL·E: A Neural Network that Creates Images from Text. https://openai.com/index/dall-e. Accessed: 2024-08-06.

[33] OpenAI. 2022. DALL·E 2: A New Generation of Image Synthesis with Improved Capabilities. https://openai.com/index/dall-e-2. Accessed: 2024-08-06.

[34] OpenAI. 2023. ChatGPT: Optimizing Language Models for Dialogue. https://www.openai.com/chatgpt

[35] OpenAI. 2023. DALL·E 3: Advanced Image Generation and Editing with Text. https://openai.com/index/dall-e-3. Accessed: 2024-08-06.

[36] J. Perret and E. Vander Poorten. 2018. Touching Virtual Reality: A Review of Haptic Gloves. In *ACTUATOR 2018; 16th International Conference on New Actuators*. 1–5.

[37] PTC. 2023. Vuforia Engine Product Page. https://www.ptc.com/en/products/vuforia/vuforia-engine. Accessed: 2023-07-30.

[38] Irvin Rock and Jack Victor. 1964. Vision and Touch: An Experimentally Created Conflict between the Two Senses. *Science* 143, 3606 (1964), 594–596. https://doi.org/10.1126/science.143.3606.594 arXiv:https://www.science.org/doi/pdf/10.1126/science.143.3606.594

[39] Tim Rohe and Uta Noppeney. 2015. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biology* 13, 2 (02 2015), 1–18. https://doi.org/10.1371/journal.pbio.1002073

[40] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[41] Vivian Shen, Tucker Rae-Grant, Joe Mullenbach, Chris Harrison, and Craig Shultz. 2023. Fluid Reality: High-Resolution, Untethered Haptic Gloves using Electroosmotic Pump Arrays. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 20 pages. https://doi.org/10.1145/3586183.3606771

[42] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3307–3316. https://doi.org/10.1145/2702123.2702389

[43] Misha Sra, Sergio Garrido-Jurado, Chris Schmandt, and Pattie Maes. 2016. Procedurally generated virtual reality from 3D reconstructed physical space. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 191–200. https://doi.org/10.1145/2993369.2993372

[44] Misha Sra, Prashanth Vijayaraghavan, Ognjen Rudovic, Pattie Maes, and Deb Roy. 2017. DeepSpace: Mood-Based Image Texture Generation for Virtual Reality from Music. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2289–2298. https://doi.org/10.1109/CVPRW.2017.283

[45] Stability AI. 2023. Stability AI Official Website. https://stability.ai. Accessed: 2023-07-30.

[46] Ryo Suzuki, Eyal Ofek, Mike Sinclair, Daniel Leithinger, and Mar Gonzalez-Franco. 2021. HapticBots: Distributed Encountered-type Haptics for VR with Multiple Shape-changing Mobile Robots. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1269–1281. https://doi.org/10.1145/3472749.3474821

[47] Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. 2024. TripoSR: Fast 3D Object Reconstruction from a Single Image. *ArXiv* abs/2403.02151 (2024). https://api.semanticscholar.org/CorpusID:

268248244

[48] Zhengyi Wang, Yikai Wang, Yifei Chen, Chendong Xiang, Shuo Chen, Dajiang Yu, Chongxuan Li, Hang Su, and Jun Zhu. 2024. CRM: Single Image to 3D Textured Mesh with Convolutional Reconstruction Model. *ArXiv* abs/2403.05034 (2024). https://api.semanticscholar.org/CorpusID:268297409

[49] Jackie (Junrui) Yang, Christian Holz, Eyal Ofek, and Andrew D. Wilson. 2019. DreamWalker: Substituting Real-World Walking Experiences with a Virtual Reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1093–1107. https://doi.org/10.1145/3332165.3347875

[50] Ya-Ting Yue, Yong-Liang Yang, Gang Ren, and Wenping Wang. 2017. SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 427–436. https://doi.org/10.1145/3126594.3126601

[51] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3813–3824. https://doi.org/10.1109/ICCV51070.2023.00355

[52] Mengjia Zhu, Amirhossein H. Memar, Aakar Gupta, Majed Samad, Priyanshu Agarwal, Yon Visell, Sean J. Keller, and Nicholas Colonnese. 2020. PneuSleeve: In-fabric Multimodal Actuation and Sensing in a Soft, Compact, and Expressive Haptic Sleeve. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376333

[53] Zi-Xin Zou, Zhipeng Yu, Yuan-Chen Guo, Yangguang Li, Ding Liang, Yan-Pei Cao, and Song-Hai Zhang. 2024. Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10324–10335.