

RealityEffects: Augmenting 3D Volumetric Videos with Object-Centric Annotation and Dynamic Visual Effects

Jian Liao

University of Calgary
Calgary, Canada

jian.liao1@alumni.ucalgary.ca

Kevin Van

University of Calgary
Calgary, Canada

kevin.van@ucalgary.ca

Zhijie Xia

University of Calgary
Calgary, Canada

zhijie.xia@ucalgary.ca

Ryo Suzuki

University of Calgary
Calgary, Canada

ryo.suzuki@ucalgary.ca



Figure 1: RealityEffects is a desktop authoring interface to augment 3D volumetric videos with object-centric annotation and visual effects.

ABSTRACT

This paper introduces RealityEffects, a desktop authoring interface designed for editing and augmenting 3D volumetric videos with object-centric annotations and visual effects. RealityEffects enhances volumetric capture by introducing a novel method for augmenting captured physical motion with embedded, responsive visual effects, referred to as *object-centric augmentation*. In RealityEffects, users can interactively attach various visual effects to physical objects within the captured 3D scene, enabling these effects to dynamically move and animate in sync with the corresponding physical motion and body movements. The primary contribution of this paper is the development of a taxonomy for such object-centric augmentations, which includes annotated labels, highlighted objects, ghost effects, and trajectory visualization. This taxonomy is informed by an analysis of 120 edited videos featuring object-centric visual effects. The findings from our user study confirm that our direct manipulation techniques lower the barriers to editing and annotating volumetric captures, thereby enhancing interactive and engaging viewing experiences of 3D volumetric videos.

CCS CONCEPTS

- Human-centered computing → Mixed / augmented reality.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DIS '24, July 1–5, 2024, IT University of Copenhagen, Denmark

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0583-0/24/07...\$15.00

<https://doi.org/10.1145/3643834.3661631>

KEYWORDS

Volumetric Video; Authoring Interface; Mixed Reality; Augmented Visual Effects; Object-Centric Annotation

ACM Reference Format:

Jian Liao, Kevin Van, Zhijie Xia, and Ryo Suzuki. 2024. RealityEffects: Augmenting 3D Volumetric Videos with Object-Centric Annotation and Dynamic Visual Effects. In *Designing Interactive Systems Conference (DIS '24), July 1–5, 2024, IT University of Copenhagen, Denmark*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3643834.3661631>

1 INTRODUCTION

In recent years, *augmented videos* [9, 49, 79]—live or recorded 2D videos enhanced with embedded visual effects—have gained an increasing popularity in human-computer interaction (HCI). By seamlessly integrating visual effects with physical motion, augmented videos provide more interactive and engaging viewing experiences, similar to augmented and mixed reality, but on a screen. Traditionally, creating such augmented videos requires significant time and expertise using professional video-editing software like Adobe Premiere Pro. However, recent HCI research has enabled interactive and improvisational authoring experiences, simplifying the creation of these augmented live or recorded videos in various applications, such as sports analysis (e.g., *VisCommentator* [9]), classroom education (e.g., *RealitySketch* [87]), storytelling (e.g., *Interactive Body-Driven Graphics* [79]), interactive data visualization (e.g., *Augmented Chironomia* [33]), live presentation (e.g., *RealityTalk* [58]), and entertainment (e.g., *PoseTween* [60]).

However, these works primarily focus on augmented 2D videos, and to the best of our knowledge, no prior work has explored augmented 3D volumetric videos. Especially with the recent releases of sophisticated mixed reality headsets like the Apple Vision Pro,

spatial and 3D volumetric videos become an emerging entertainment medium in the mainstream consumer market. Despite the recent proliferation of 3D volumetric capture technologies, such as point-cloud rendering or reconstructed 3D capture with depth cameras or LiDAR sensors, editing and augmenting these volumetric videos remains challenging. Existing tools like *4Dfx* [1], *DepthKit Studio* [20], and *HoloEdit* [84] offer only basic video touch-ups and timeline manipulation. Consequently, users must either edit these effects frame-by-frame, similar to the traditional 2D video-editing techniques, or program the behavior of visual effects within a 3D game environment, such as Unity or Unreal Engine.

In this paper, we present RealityEffects, a desktop authoring interface that supports the real-time and interactive creation of augmented 3D volumetric videos. To augment the volumetric 3D scene, users can simply select and bind captured physical objects with annotated visual effects. The system then automatically tracks physical objects such that the embedded visual effects can move and respond dynamically with the corresponding physical motion and body movement. We call this approach *object-centric augmentation*, which can significantly reduce the time and cost of creating augmented volumetric videos. Unlike 2D videos, the augmented 3D scene allows free-viewpoint movement, enabling immersive viewing experiences.

To design our system, we collected 120 video examples utilizing the video-edited object-centric augmentation. Based on the observed common augmentation techniques, we contribute a taxonomy of object-centric augmentations for 3D volumetric videos, which includes annotated labels, highlighted objects, ghost effects, and trajectory visualization. Along with the novel direct manipulation authoring, RealityEffects extends the idea of previously explored volumetric augmentation (e.g., *Remixed Reality* [59]) to support more comprehensive visual effects that can be used in a wide range of applications, such as sports analysis, physics education, classroom tutorials, and live presentations. We evaluated our system with a lab-based usability study ($N=19$). Our study results suggest that object-centric augmentation is a promising way to lower the barrier to editing and annotating volumetric captures while allowing flexible and expressive video augmentation.

Finally, our paper contributes to:

- (1) A taxonomy and design space of object-centric augmentation for 3D volumetric captures, based on the analysis of existing object-centric 2D video augmentation techniques.
- (2) RealityEffects, a tool for creating augmented 3D volumetric videos that leverage a novel direct manipulation technique to bind dynamic visual effects with corresponding physical motion.
- (3) Application demonstration and user evaluation of RealityEffects, which suggests the untapped potential of augmented volumetric captures for more interactive and engaging viewing experiences.

2 RELATED WORK

2.1 Volumetric Capturing and Editing

2.1.1 Volumetric Capture and Its Applications. Volumetric captures or videos refer to the technique of capturing 3D space and subsequently viewing it on a screen with free-viewpoint movement.

These techniques have been explored since the 1990s (e.g., *Virtualized Reality* [44]), but recent research has greatly advanced this domain in both high-quality 3D reconstruction (e.g., *Fusion4D* [21], *Montage4D* [22], *Relightables* [32], *VolumeDeform* [39]) and more accessible volumetric capturing with mobile phones (e.g., *Kinect Fusion* [40], *DepthLab* [23], *Polycam* [72]). With recent advances in commercially-available depth cameras like Kinect, volumetric captures have been used in various applications such as telepresence (e.g., *Holoportation* [68], *JackInSpace* [50], *Project Starline* [52], *PhotoPortals* [51]), remote collaboration (e.g., *RemoteFusion* [2], *Mini-Me* [70], *On the Shoulder of Giant* [71], *Virtual Makerspaces* [73]), remote hands-on instruction (e.g., *Loki* [91], *BeThere* [81], *3D Helping Hands* [89]), and immersive tutorials for physical tasks (e.g., *MobileTutAR* [7], *ProcessAR* [15], *My Tai Chi Coaches* [34]). Past research has utilized static or live 3D reconstructed scenes for remote MR collaboration, facilitating more immersive interactions with remote users [27, 88, 90]. Alternatively, live 3D reconstruction has been used to facilitate co-located communications for VR users (e.g., *Slice of Light* [94], *Asynchronous Reality* [26]). These captured 3D geometries are also used for anchoring virtual elements (e.g., *SnapToReality* [67], *SemanticAdapt* [12]), creating virtual contents (e.g., *SweepCanvas* [57], *Window-Shaping* [38]), or generating virtual environments (e.g., *VRoamer* [10], *Oasis* [82, 83]) by leveraging object detection and semantic segmentation of volumetric scenes (e.g., *SemanticPaint* [92], *ScanNet* [17]).

2.1.2 Augmenting and Editing Volumetric Capture. More closely related to our work, past work has also explored further blending the virtual and physical worlds by augmenting captured volumetric scenes or the real world. By using VR/MR devices, systems can alternate the captured scene by erasing physical objects (e.g., *SceneCtrl* [101], *Diminished Reality* [13, 63]) or replacing them with virtual ones (e.g., *RealityCheck* [35], *TransformMR* [45]). Alternatively, previous work has used the depth information to blend virtual augmentation into the real-world with projection mapping (e.g., *IllumiRoom* [42], *RoomAlive* [41], *Room2Room* [69], *Dyadic Projected SAR* [4], *OptiSpace* [24]). Systems like *Mixed Voxel Reality* [74], *Remixed Reality* [59], and *Virtual Reality Annotator* [76] further advance this approach by augmenting the volumetric scene by leveraging both spatial manipulation (copy, erase, move), temporal modification (record, playback, loop), and volumetric annotation (sketches) with a VR headset and live 3D reconstruction.

While these works partially demonstrated the visual augmenting of captured scenes, supported augmentation techniques remain simple (e.g., appearance change for color or texture). Moreover, since their focus is on the immersive experience of these modified scenes, the authoring aspect of these volumetric scenes and videos is not well explored in the literature. Our focus is rather on the authoring interface, which can support more comprehensive visual augmentation for the volumetric scenes. This is because the current work on authoring tools or video-editing tools for volumetric capture is either focused on static scenes (e.g., *DistanciAR* [95]), timeline manipulation (e.g., *4Dfx* [1]), or simple video touch-ups (e.g., *DepthKit Studio* [20], *HoloEdit* [84]). In contrast, RealityEffects enables more expressive visual augmentation for dynamic volumetric scenes by leveraging *object-centric augmentation*, which we take inspiration from 2D video authoring, as described next.

2.2 Authoring Augmented 2D Videos

In the context of 2D videos or mobile AR interfaces, *augmented videos* refer to a live or recorded video in which embedded visuals are seamlessly coupled with captured physical objects [9, 49, 79]. Systems like *PoseTween* [60], and *Interactive Body-Driven Graphics* [79] demonstrate the interactive authoring tools for generating responsive graphics that can move with the corresponding body movement in the live or recorded video. Such visual augmentation can provide more engaging experiences for live presentations (e.g., *RealityTalk* [58], *Augmented Chironomia* [33]) sports training (e.g., *VisCommentator* [9], *EventAnchor* [19], *YouMove* [3]), storytelling (e.g., *RealityCanvas* [97]), and education (e.g., *HoloBoard* [31], *Sketched Reality* [43]). Moreover, augmented videos are also useful media for prototyping AR experiences (e.g., *Pronto* [56], *Rapido* [55], *Teachable Reality* [62]) or remote collaboration (e.g., *In-Touch with the Remote World* [28, 29]).

Traditionally, these videos require professional video-editing skills, but HCI researchers have investigated end-user authoring tools to lower the barrier of expertise. In particular, taking inspiration from object-based video navigation techniques [46, 64–66, 77, 93], Goldman et al. [30] and Silvia et al. [80] explored object-centric video annotation, which allows users to add dynamic annotation based on the tracked object in the 2D video. More recently, systems like *RealitySketch* [87], *RealityCanvas* [97], *VideoDoodles* [99], and *Graphiti* [78] have further expanded the object-centric augmentation for dynamic AR sketching interfaces. However, to the best of our knowledge, no prior work has explored these techniques for 3D volumetric videos, which introduce the additional interaction challenge of selecting or aligning objects in 3D scenes [37, 61]. This paper contributes to the first object-centric augmentation for *3D volumetric video*, along with a taxonomy of possible augmentation design.

2.3 Object-Centric Immersive Visualization

Our design for spatial annotations and visual effects is also inspired by various object-centric immersive visualization and visual analytics techniques [18]. Previous works have explored various spatio-temporal visualization techniques, such as spatial and semantic object annotation (e.g., *ReLive* [36], *Skeletonotator* [54]), trajectories of objects (e.g., *MIRIA* [6]), trajectories of human motion (e.g., *AvatAR* [75], *Reactive Video* [16], *DemoDraw* [14]), ghost effects (e.g., *GhostAR* [8]), object and location highlights (e.g., Kepplinger et al. [47]), and heatmap visualizations (e.g., *HeatSpace* [25], *Eagle-View* [5]). These free-viewpoint movements and multi-viewpoint analyses can greatly improve the way we watch and analyze object- and body-related movements with deeper insights [5, 48, 85, 100]. While our tool is inspired by these works, our focus lies on the *authoring aspect* of these dynamic effects and visualizations, rather than developing novel visualization systems. For example, we designed our system in a way that end-users can easily select, bind, and visualize motion data without any pre-defined programs or configurations. We believe our tool along with the direct manipulation authoring approach, allows flexible and customizable volumetric video editing that can be used for broader applications beyond these visual analytics tools.

3 A TAXONOMY OF OBJECT-CENTRIC AUGMENTATION

3.1 A Taxonomy Analysis

To better understand common practices and techniques for object-centric augmentations, we first collected and analyzed a set of 120 existing videos available on the Internet, most of which were created using professional video-editing software. These examples showcase a variety of techniques and collectively contribute to a preliminary taxonomy of object-centric visual augmentation, helping the design of end-user systems for authoring these effects.

3.1.1 Definition of Object-Centric Augmentation. To design our system feature, we first need to understand and investigate common practices for *object-centric augmentation*. In this paper, object-centric augmentation refers to “*a class of virtual elements 1) that are embedded and spatially integrated with objects in a scene, and 2) whose properties change, respond, and animate based on the behaviors of physical objects in the scene*”. Here, *virtual elements* can include text, images, visual effects, and visualization; *objects* can be physical objects, parts of the human body, or environments; and *properties* can encompass location, orientation, scale, and other visual properties.

3.1.2 Motivation and Goal. While object-centric augmentations are frequently used in many professional videos and several works explore this domain [30], these works lack a taxonomy analysis [30, 60, 78, 87] or focus on more specific domains such as presentations [58], storytelling [79] or robotics [86], leaving a gap in the holistic understanding of possible designs, even for 2D videos and, certainly, for 3D volumetric videos. The goal of this taxonomy analysis is to provide initial insights into object-centric augmentations. We have adapted methods from similar prior research papers [58, 62, 97] to provide an initial and preliminary taxonomy of a representative subset of common practices, recognizing that conducting a systematic visual search of videos is more challenging than conducting a systematic search of research papers.

3.1.3 Corpus and Dataset. To collect the video examples, the authors (A1, A2, and A4) manually searched popular video and image search platforms (e.g., YouTube, Pinterest, Vimeo, Behance, and Google Images), primarily relying on visual searches, as these videos are not associated with a specific keyword like “*3D visual effects*”. After some initial filtering, we started to identify some patterns in the visuals we collected, and with the help of the similar image suggestion feature on Pinterest, we expanded more visual search criteria like annotations, highlights, augmented effects, labels, floating text, floating screen, analysis, visualization, and motion. We also did a reverse search to find the videos, through this process, we first collected 200 videos. Note that there is a much smaller proportion of examples for 3D volumetric videos, none of them are volumetric videos, while most of them feature 3D visual effects. Then, the authors (A1, A2, and A4) filter out by focusing only on the object-centric augmentation (e.g., removing videos that use entirely virtual effects without physical objects or visual effects that are not associated with the physical objects). After the filtering process, we obtained 120 videos that contain object-centric augmentation based on our definition,

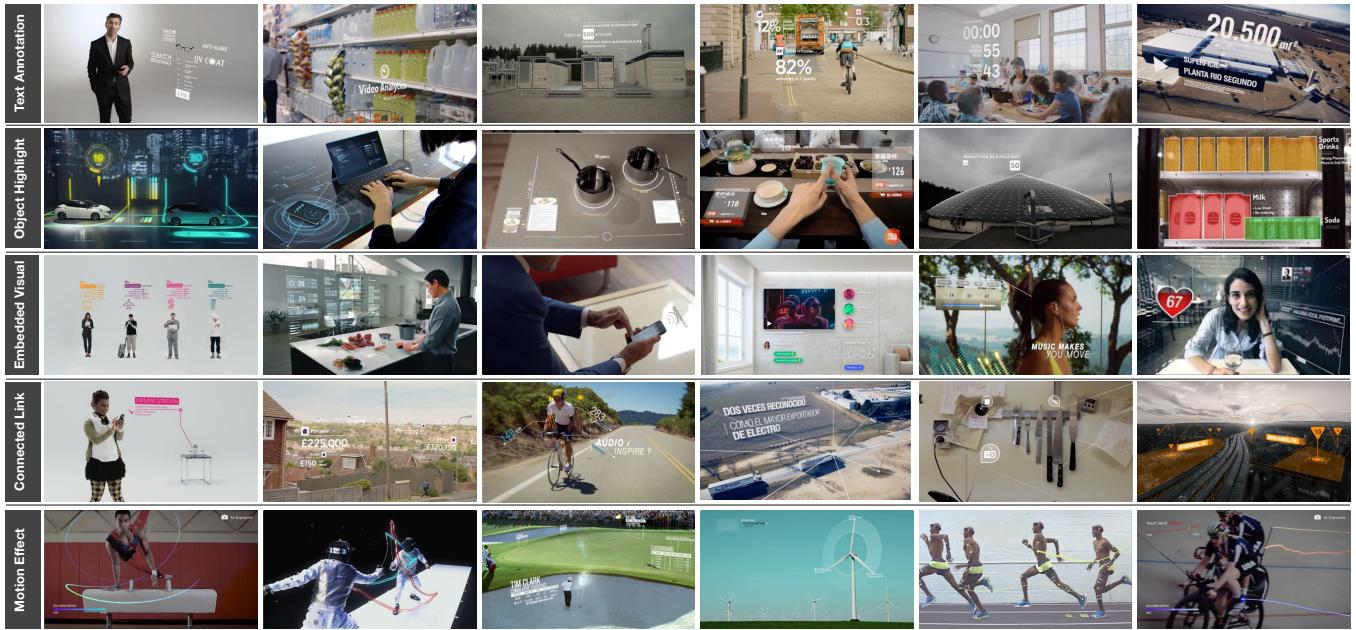


Figure 2: Design Space Analysis: We collected a set of 120 existing edited videos and images and observed the five most common design techniques, namely, Text Annotation, Object Highlight, Embedded Visual, Connected Link, and Motion Effect. The screenshots are copyrighted by each video creator. We listed the link for each video in the Appendix.

3.1.4 Coding Methodology. We analyzed all 120 collected videos to identify snippets displaying annotations and visual effects, capturing screenshots of object-centric visual effects from each. Through this process, one of the authors (A1) led the collection of screenshots, with assistance from another author (A2), resulting in a total of 336 screenshots, averaging 2.8 screenshots per video. We chose screenshots over full videos as a coding corpus because each video may contain different techniques, and these screenshots served as representative keyframes for our taxonomy analysis. Subsequently, we conducted open coding to identify a preliminary taxonomy of object-centric visual effects. With the 336 collected representative images, author A1 led the initial open coding process to identify a first approximation of the dimensions and categories, then iterated with other authors (A2 and A4) on digital whiteboards (Miro and Google Slides). In this process, the three authors independently reviewed the collected screenshots and refined the taxonomy initially identified by A1. Subsequently, all authors reflected on the initial design space to discuss the consistency and comprehensiveness of the categorization. Finally, after systematic coding by authors A1 and A2, which involved individual tagging for the complete dataset, we reviewed the tagging to resolve discrepancies and obtain final coding results. All authors then reflected on the design space and finalized the categorization by merging, expanding, and removing categories.

3.1.5 Limitations. We acknowledge several limitations in our current methodologies, including corpus selection and taxonomy analysis. First, our selected videos may not represent a comprehensive and exhaustive corpus. While we aimed to collect as diverse a

dataset as possible, the nature of our visual search, rather than a systematic keyword search, limits our ability to claim comprehensive representation. Second, the taxonomy analysis might have benefited from the involvement of the video creators to better capture the design space from their perspectives. Despite these limitations, we believe this taxonomy can help identify common practices and techniques for object-centric augmentation, benefiting both our own and other HCI research.

3.2 Design Space of Object-Centric Augmentation

Based on the analysis, we identified the following five most common augmentation techniques: 1) text annotations, 2) object highlights, 3) embedded visuals, 4) connected links, and 5) motion effects (Figure 2).

3.2.1 Text Annotation. Text annotation is one of the most common techniques identified. It involves attaching textual labels or descriptions to physical objects. These can be *static descriptions*, providing information about the object, or *dynamic data and parameters*, such as speed, distance, or price, akin to embedded data visualization [96]. The attached objects can be *graspable physical items*, parts of the *human body*, or *stationary locations* like buildings or furniture.

3.2.2 Object Highlight. Object highlight is a technique used to visually attract an audience's attention to a specific object. For example, object highlight techniques include changing the *color* of the object, highlighting the *contour* of the object, adding highlighting *marks* to the object, or changing the *opacity* of other objects. These object

highlights can be applied to either *2D surfaces*, such as showing a colored circle on the ground around cars, phones, or pans, or *3D objects*, such as displaying a bounding box and sphere or 3D mesh of the target object.

3.2.3 Embedded Visual. Embedded visuals are 2D images or visual information attached to describe objects, similar to text annotation but through static visuals or animations. Embedded visuals include *simple icons* to describe the object, *2D images and photos* to show the associated information, *animation* to visually describe the behavior, *screens* to display the associated website or user interfaces, and *charts or graphs* to visualize the associated data.

3.2.4 Connected Link. Connected links are lines that indicate the relationship between two elements. These connected lines can be *object to virtual elements*, linking text annotations or embedded visuals to a specific object to indicate which object is being described. Alternatively, the connected lines can be *object to object*, explaining the relationship and association between multiple physical objects, such as indicating network communication between multiple IoT devices or visualizing the connection between different body parts like arms or legs. These connected links can dynamically move and animate whenever the physical objects move.

3.2.5 Motion Effect. Motion effects are techniques used to visualize the motion of physical objects. Most commonly, *motion trajectories* are used to show the path a specific object moves, such as illustrating the trajectory of a golf swing, baseball batting swing, and body movement in gymnastics. Alternatively, some videos leverage slow-motion morphing effects or *ghost effects* to depict the trajectory of the entire body or objects, similar to the famous bullet-time effects in the movie *Matrix*.

3.2.6 Others. While much less common, we have also observed several other effects, such as particle effects and virtual 3D animation. However, since object-centric augmentation already leverages the dynamic motion of physical objects, simple visual augmentation can significantly make the video more expressive and enrich the viewing experience.

4 REALITYEFFECTS SYSTEM

4.1 System Overview

This section introduces RealityEffects, a desktop authoring interface designed to support the real-time and interactive creation of augmented 3D volumetric videos, whether live or recorded. The goal of our system is to enable users to create augmented volumetric videos through direct manipulation, without the need for programming, by leveraging an object-centric augmentation approach. Given the design space exploration outlined above, RealityEffects allows users to easily embed text, visuals, highlight effects, and 3D objects, which can be bound to physical objects and bodies captured in the volumetric video. The following workflows are supported by RealityEffects:

- Step 1.** Track a captured object or body part by clicking the tracking points from the desktop 3D scene.
- Step 2.** Add visual effects that are automatically bound to the selected physical object.

Step 3. Obtain the dynamic data and parameters of the real-world motion.

Step 4. Bind and visualize the obtained dynamic parameter to create responsive graph plots or associated animation.

4.2 System Implementation

As shown in Figure 3, RealityEffects is implemented across three main modules: streaming, processing, and augmenting. The entire application is written in JavaScript using React.js, React Three Fiber, and Electron.js. It runs on a desktop Windows machine, and we recommend using a desktop machine equipped with graphics cards to speed up rendering. The source code for our system implementation is available on GitHub¹.

4.2.1 Streaming Module. The streaming module utilizes the off-the-shelf Azure Kinect depth camera SDK to capture volumetric point-cloud data. The data feed includes both RGB and Depth data in separate channels, each with a resolution of 640×576 and a refresh rate of 30 FPS. Both channels share the same (x,y) coordinate data structure, enabling us to retrieve the depth information for any (x,y) tracking point. The obtained RGB-D data is then passed to the processing module.

4.2.2 Processing Module. With the RGB-D data feed, the application performs 3D scene reconstruction by rendering the 3D point cloud data directly using Three.js, where $z = Depth(x,y)$ and $RGB = Color(x,y)$. We utilize MediaPipe Pose Estimation for body tracking and OpenCV for object tracking. The application calculates the centroid by averaging the (x,y) values, retrieves the depth information with the (x,y) coordinates, and registers the centroid as the attachable object in the authoring interface for further augmentation.

4.2.3 Augmenting Module. With the attachable object from the processing module, RealityEffects allows users to select objects from pose estimation and color tracking and augment them with object-centric annotations and dynamic visual effects. The object-centric annotations are essentially Three.js coded objects such as static text labels and bounding boxes, with a bloom pass to create glowing highlight effects. The dynamic visual effects are parameterized object motions that allow us to create visualizations from the motion and parameters, such as trajectory, position, distance, and angle. Users can augment the moving object with motion effects like a trajectory (a series of points) and a trailing effect², and augment the motion with embedded visualizations using an iframe to create charts and interactive widgets. The augmentation can be applied to a real-time camera data feed, allowing users to review their own performance as it's being annotated, which unlocks several application scenarios like sport analysis and e-commerce live streaming. Users can also freely move or zoom the camera in 3D space through mouse movements. When using a recorded volumetric video, the system supports simple pause and play functionalities. Since we only use a single Kinect camera, capturing the entire room is challenging. Therefore, we also scan the room with a static 3D scanner (iPad Pro 12-inch with LiDAR camera and 3D Scanner App) and place it as a 3D volumetric background asset

¹<https://github.com/jlia0/RealityEffects>

²<https://drei.pmnd.rs/?path=/docs/misc-trail--docs>

(glTF file) only for visual aesthetic purposes in most of the figure and video demonstrations.

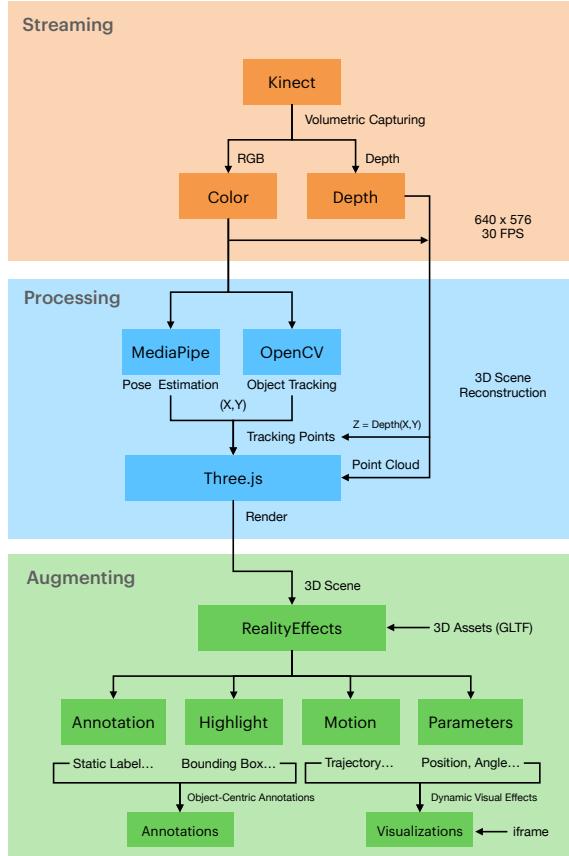


Figure 3: RealityEffects consists of three modules – (1) Streaming: Azure Kinect SDK provides the depth camera data feed, (2) Processing: Mediapipe and OpenCV for body and color tracking (3) Augmenting: Three.js for visual rendering.

Step 1. Object Selection and Tracking

The first step is to select a captured physical object. For object-centric augmentation, all embedded annotations and visual effects should be tightly coupled with physical objects. Therefore, our system first allows the user to specify which objects to track and bind. To specify the object, the user can enter the selection mode and simply click the object in the scene. Then, the system automatically adds a tracking point in the 3D scene and tracks its location. For the tracking point, the system supports three categories: 1) *physical object*, 2) *body*, 3) *stationary physical environment*. The performance of object tracking can be found in Table 1. We evaluate the accuracy by counting the time duration of tracking target losses over a fixed period of a captured video while we freely move the objects around in space and at different angles. While this is a fairly simple evaluation, we found that pose estimation is fairly robust and close to its acclaimed benchmarks, whereas color tracking is unstable, especially for objects with reflective materials. Future improvements

are needed using more robust methods such as *SAM-Track* [11] and *Track Anything* [98].

	Pose Estimation	Color Tracking
Accuracy	91.3%	65.7%

Table 1: Pose Estimation and Color Tracking Accuracy

4.2.4 Physical Object. First, for the colored physical object, the system tracks the object's 3D position based on the combination of color tracking and point-cloud information. When the user clicks an object, the system gets the current RGB value of the clicked points in 2D screen. Then, the system captures a similar colors based on an upper and lower threshold range of RGB (± 10) to obtain the largest contour in the scene, based on Node OpenCV library. Given the detected object in the 2D scene, the system raycasts to the volumetric scene to obtain the associated point-cloud depth information, which allows us to get the coordinated 3D position in the scene.

4.2.5 Body. For post estimation and body tracking, we simply use Mediapipe to get the estimation of 33 body tracking points. We also tried Kinect-built-in body tracking feature, but the performance was not satisfying because of high latency and low accuracy. Similar to color tracking, the system allows the user to directly select one of the tracking points of the body skeleton, and the system automatically calibrate the 2D coordinates with the depth information. When the user enters the body selection mode, then the system shows the twenty body skeleton points which the user can select. When the user selects a certain skeleton parts, then it becomes highlights and starts tracking in the 3D scene.

4.2.6 Stationary Location. For stationary location, the user can simply select a location in the scene and use a ray cast to obtain the stationary 3D position in the physical environment, such as floor or wall. The user can also place it in mid-air by moving the point with a mouse. In this selection, the tracked point is stationary, thus there is no dynamic movement. However, this tracked location can be used as a reference point, such as a distance from a certain location.

Step 2. Virtual Object Binding

Once the system starts tracking the selected object, then the user can add virtual objects that can be bound to the tracked object. Informed by the taxonomy analysis, the system supports the following four virtual 3D objects: 1) text annotation, 2) object highlight, and 3) embedded visual.

4.2.7 Text Annotation. First, the user can bind the text label to the associated physical object in the volumetric 3D scene. To place a text annotation, the user specifies the associated object and then clicks the text label button. Then, the system starts showing the 2D text label floating around the tracked object. Since the attached text label is bound to the object, the text label position moves when the object moves. The user can change the text value by typing the name in the menu window. The user can also add a dynamic value by using a variable, based on the JavaScript variables such

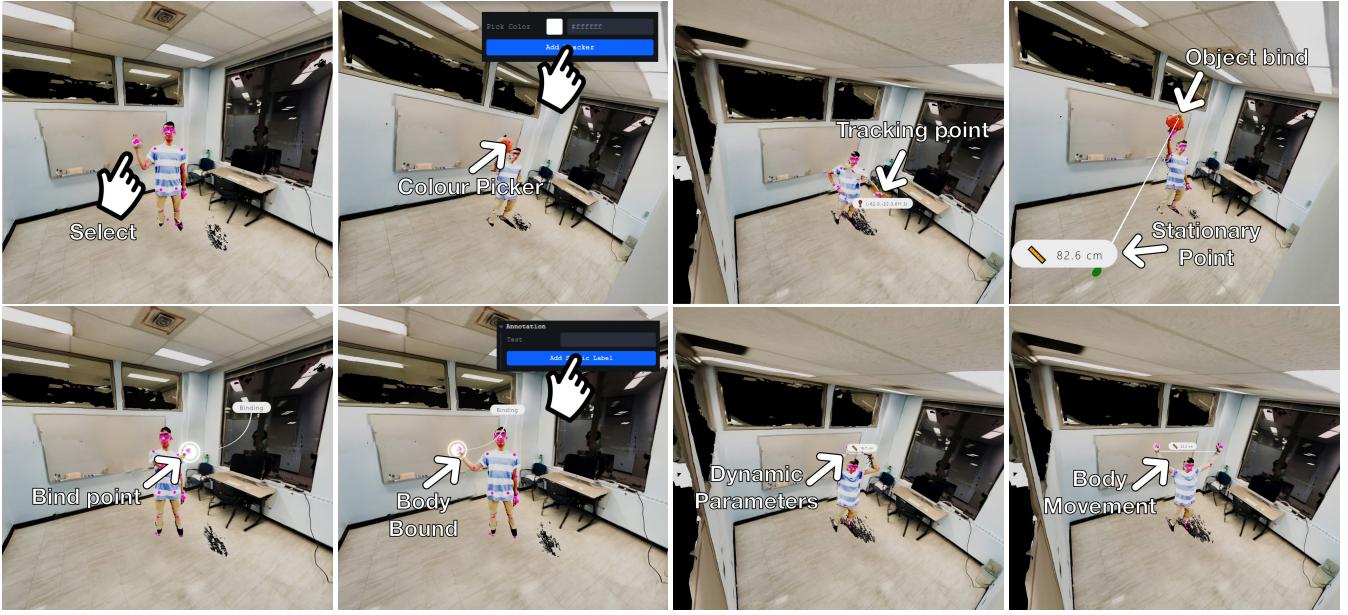


Figure 4: Authoring Workflow: Collection of examples from RealityEffects’s workflow to demonstrate features such as dynamic parameters, highlighting and annotations.

as Date.now(), or a user-defined variable based on the dynamic parameters, such as position, speed, angle, distance, etc., as we describe in Step 3. While the text is a 2D object, it always moves its orientation to face the camera. The user can also disable this to

4.2.8 Object Highlight. The user can also add an object highlight bound to the tracked object. The system supports two basic object highlight options: 1) 3D primitive shapes, such as bounding box, bounding sphere, and bounding cylinder, and 2) 2D highlight shapes such as colored circles or rectangles. To add the object highlights, the user first selects the tracked object and then chooses the object highlight button in the menu. Then, the system lets the user choose the shape of the highlight (default: 3D sphere), then the object highlight is added to the scene. Unlike text annotation, the object highlight is placed in the center of the tracked object. The user can also change the scale, offset, orientation, and color accordingly through a direct manipulation interface.

4.2.9 Embedded Visual. The user can also add embedded 2D visuals. Informed by the taxonomy analysis, the system supports images, icons, videos, and embedded websites as the associated visual aids. From the technical point of view, all the embedded visual is implemented as embedded iframe in Three.js. Therefore, the image, YouTube video, or website can be embedded as an iframe by specifying the URL or local file. To add the embedded visual, the user can also select the object and enter the embedded visual menu. Then the user can enter the URL or file directory. Once loaded, the added visual elements start following based on the object’s movement. Again, the user can also change the size, orientation, and opacity of these elements. Since the embedded visual is an interactive HTML, the user can also interact with the screen such as buttons or links. By leveraging this feature, we can also embed

dynamic graphs and charts by associating the dynamic parameter, as discussed in Step 4. By default, the 2D visual always changes its orientation to face the camera, but the user can also change it by disabling it.

Step 3. Parameterize the Real World

The user can also parameterize the real world to obtain the dynamic data value associated with the captured motion. The system obtains these real-time values based on 3D reconstructed information. The system supports the following parameterized values: 1) X, Y, and Z position of the tracked object, 2) speed of the tracked object, 3) distance between two tracked objects, 4) angle between three tracked objects, and 5) 2D area of three or more tracked objects.

4.2.10 Position. The system can obtain the 3D position of the tracked object by simply getting the current position value. The user can use this dynamic value in text labels or dynamic graphs by using the specific variable. In the system, the user can use this value by using the variable like obj_1.x.

4.2.11 Speed. The system also obtains the speed for all the tracked objects, by calculating

$$\text{Speed} = \frac{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}}{t_1 - t_0} \quad (1)$$

where $t_1 - t_0$ is every 0.5 second (every 15 frames with 30 FPS). The user can access this information by using the variable like obj_1.speed.x

4.2.12 Distance. When the user selects multiple objects, the system also calculates the distance between the two objects. If the user also wants to show the line between the two points, the user can enter the geometry menu, and then add the connected line between two

tracked points. Then, the line geometry automatically changes its length and orientation based on the tracked objects. The endpoint of the line is not the dynamic object but can be also a stationary location such as a specific point in the scene. The user can use the distance information by using the variable like `distance_1`

4.2.13 Angle. When selecting the three tracked objects, then the system also calculates the angle between the two lines. In this case, the angle is calculated given the two 3D vectors a, b through $\theta = \cos^{-1}[(a \cdot b)/\|a\| \cdot \|b\|]$, where \cos^{-1} is arc cosine and $a \cdot b$ is a dot product. The user can use the angle data by using the variable like `angle_1`

4.2.14 Area. In the same way, the user can also obtain the dynamic parameter for the area of three points (the area of a connected triangle) or four points (the area of a connected rectangle). The user can use the area by using the variable like `area_1`

Step 4. Visualize the Dynamic Motion

By default, the user can create object-centric augmentation by simply binding virtual elements to the tracked object in Step 2. However, the user can even create more expressive dynamic effects by using or visualizing the dynamic parameters based on the variables defined in Step 3. To that end, the system supports the following three parameter-based dynamic visual effects: 1) dynamic text annotation, 2) dynamic visual appearance, and 3) dynamic graph. The system also supports two motion-related visual effects: 4) motion trajectory, and 5) ghost effects.

4.2.15 Dynamic Text Annotation. Dynamic text annotation is the text annotation described in Step 1 but uses the parameterized value in the 3D scene. For example, if the user types the text value as `PositionX: ${obj_1.x}$`, then the system shows the parameterized value in the text label, which is shown as like `PositionX: 34.23`.

4.2.16 Dynamic Visual Appearance. Similarly, the user can also bind the dynamic parameter to the visual property of the embedded virtual objects, such as scale, rotation, position, opacity, and color. For example, if the user associates the scale of the virtual object with the position of the tracked object, then the embedded virtual object's size changes in response to the position of the tracked object.

4.2.17 Dynamic Graph. The user can also show the dynamic graph by associating the dynamic value with the charts. As we mentioned, we can embed the interactive 2D data visualization with iframe. We prepare several basic graphs such as line charts, bar charts, or pie charts, based on the Chart.js library. For example, if the user associates the y value of the line graph as an angle of between the arm and body, then the system shows the real-time line chart to show the tracked parameter.

4.2.18 Motion Trajectories. Alternatively, the system can also show the motion effects with several prepared visual effects. For example, the user can show the motion trajectory of the tracked object. To do so, the user selects the motion trajectory option in the menu, then the user selects the object. Then, the system starts the trajectory path of the motion, based on the object location. To implement this,

we simply place a small sphere in the position of the tracked object for each frame, then disappear for a certain duration (5 seconds).

4.2.19 Ghost Effects. Finally, the system also supports the ghost effects by duplicating the tracked object's geometry. To do this, we simply clone the entire tracked object for every second, so that the user can see the ghost effect.

5 APPLICATIONS

5.1 Product Showcase and Advertisement

Social e-commerce, which gained prominence during COVID, has popularized remote selling and virtual sales. Our system can be utilized for e-commerce live streaming or recorded product showcases. For example, Figure 5 illustrates a virtual sale presentation using our system. Initially, the presenter showcases the camera, and then the user can annotate the product with labeled annotations. By using the embedded website feature, the user can also add an Amazon link directly in the 3D scene. These embedded websites are interactive and clickable, enabling the audience to directly access the shopping website.



Figure 5: Product Showcase: Use case scenario demonstrating a sales pitch for a handheld camera, where labels, visualizations, and highlights are used to enhance the product's appeal and provide information.

5.2 Tutorial and Instruction

When conducting experiments in a lab, safety is crucial. RealityEffects can assist in maintaining safety standards, for instance, in a chemical lab where preventing cross-contamination of chemicals is essential. A user can define a specific space or surface for RealityEffects to monitor. Based on the data, such as the duration of interaction or movement, the system can augment visualizations to display a heatmap showing levels of contamination seconds after the area or surface is touched.

5.3 Physical Training and Sport Analysis

Our system is also suitable for sports analysis. By augmenting volumetric videos of sports activities, RealityEffects can enhance the understanding of athletic actions. For example, in a soccer game, the system can annotate or highlight players to increase their visibility, focus on individual players by binding objects to them, or use highlighting features. Features such as object-object binding or trajectory augmentation can display visual lines between players to indicate player positioning or the trajectory of the ball during the game. Additionally, the system can generate user-defined data visualizations to display statistics such as player speed or heatmaps of areas with frequent movement or activity.



Figure 6: Chemistry Lab Training: Use case scenario highlighting safe practices, necessary precautions, and potential lab dangers related to cross-contamination.



Figure 7: Maker Space Introduction: Use case scenario providing an orientation to a creative space. Labels and highlights are utilized to identify equipment and safety measures.



Figure 8: Physical Training and Analysis: Use case scenario demonstrating a physical workout routine by measuring repetitions and bodily motion to display them as visualizations.

These applications demonstrate the versatility of RealityEffects in enhancing interactive and dynamic visual experiences across various domains, from commercial showcases to educational settings and athletic training.

6 USER STUDY

To evaluate the effectiveness and user satisfaction of our system, we conducted a lab-based usability study with 19 participants (13 males, 6 females; aged between 19 and 29) from our local community, consisting of university students and working professionals on campus. Each participant was compensated with a \$10 Amazon gift card for their involvement in the user study. Our study was structured around the “usage evaluation” framework proposed by Ledo et al. [53]. The primary purpose of conducting usability studies with end-users is to assess the creative freedom, ease of use, learnability, and overall usability of the system. We also measured which design features were useful in helping participants achieve their goals. Given that our system introduces a novel authoring tool for 3D volumetric videos, we lack a direct comparison against established baselines. To overcome this challenge, we employed a combination of lab-based usability studies and in-depth interviews with users. This approach allowed us to uncover the strengths and weaknesses of our system and provided valuable insights that will inform future research.

6.1 Method and Study Protocol

6.1.1 Method. The study was structured into two sessions: the first aimed at evaluating the prototype’s usability to ascertain its effectiveness, ease of use, and learnability through structured tasks and a survey. Before the first session, we inquired about participants’ experience with 3D graphics software development tools like Unity 3D, Blender, and Unreal Engine. Identifying experienced participants helped in gaining deeper insights during the follow-up conversational interviews. The second session involved an in-depth interview to discuss the system’s benefits, limitations, and potential improvements.

6.1.2 Study Protocol. The user study was designed to measure specific usability factors such as learnability, satisfaction, and ease of use. The total duration for the study was between 45 to 60 minutes per participant, structured as follows:

- **Introduction (3-5 minutes):** Participants were introduced to the project’s goals and the underlying technology. An online whiteboard presentation outlined the system’s design and features, and participants were briefed on the concept of volumetric video, setting the stage for the tasks they would perform.

- **Demonstration and Application (24-30 minutes):** The demonstration phase was split into two parts to cover different aspects of the system. Initially, participants followed a guided tutorial with

slides on how to attach static annotations for a product advertisement scenario. This task aimed to assess the system's learnability and ease of use. Subsequently, participants engaged in a more complex task involving motion tracking, simulating a physical training scenario to evaluate the system's performance under dynamic conditions. This helped in assessing the robustness and responsiveness of the system.

- Survey (15-20 minutes): Finally, participants completed a Google Form questionnaire to provide feedback on their experience. The survey included questions designed to measure user satisfaction and identify usability issues, thereby providing qualitative and quantitative data to support the usability assessment.

6.2 Results

6.2.1 Demographics. We asked the participants at the beginning of our survey to better learn about their background and demographics towards mixed reality experience, 3D graphics software, video editing experience, and volumetric video experience. The collected demographic information is shown in Figure 9. Non-surprisingly, many of our participants do not have 3D graphics development experience or volumetric video experience. Many of them also mentioned that it was their first time hearing about 3D video or 3D volumetric video.

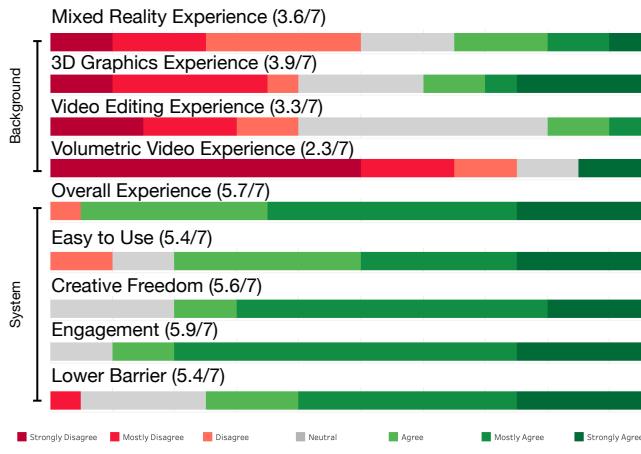


Figure 9: User Study Results – A graph summarizing the 7-point Likert scale responses of demographics and overall experiences from 19 participants.

6.2.2 Overall Experiences. As shown from the Figure 9, which summarized

The Figure 9 outlines the study results. Overall, the vast majority of participants had positive responses. Participant scores of the overall experience averaged 5.7/7. "The user experience was fun and interesting. It was pretty intuitive as well"(P9) and "It is such a fresh and interesting experience for me to play a 3D reality product." (P13). Some participants also had optimistic views towards the system *Seems interesting and could have potential uses for video editing software*(P16).

6.2.3 Ease of Use. The system was determined to be fairly easy to use, averaging 5.4/7. P8 found "*It was intuitive*" and said "*the user panel was accessible*". However, a (P5, P10, P13) found it hard to make selections. However, others(P9, 19) with experience with 3D software found it easy. P13 who was unfamiliar with using software with 3D spaces found it difficult to navigate. P3 put up concerns with certain demographics unfamiliar with 3D software and that *it might seem overwhelming*. However, they did find the streamlined interface made *animation and editing so much faster* P3.

6.2.4 Flexibility and Creative Freedom. The creative freedom of the system was reported to be flexible, with an average score of 5.6/7. Regarding the variety of the features, P1 declared that they could *imagine multiple uses for them*. Another said *The trail feature and ghost effect inspired my creativity* P10. Body motion features resonated more and had positive feedback in terms of creative freedom, P19 said *the ghosting and trailing effects were features they would personally use for creative reasons*.

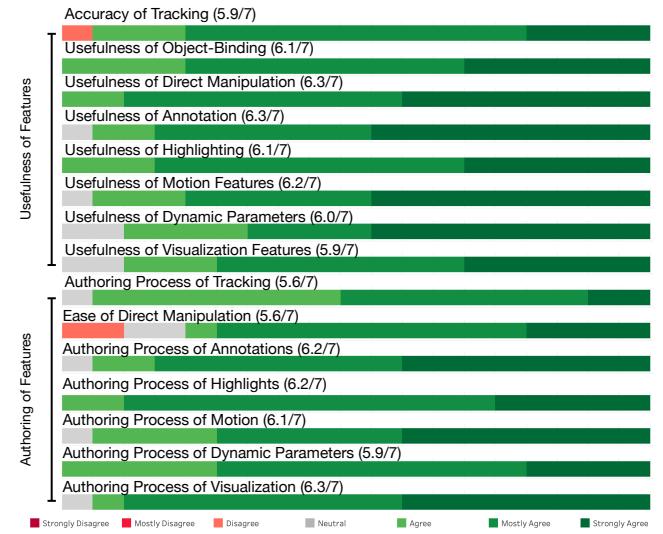


Figure 10: User Study Results – A graph summarizing the 7-point Likert scale responses of the usefulness of features and authoring experiences from 19 participants.

6.2.5 Usefulness of features. One declared the tracking and binding features were necessary, as *labels and highlights would likely not work without them*. (P2) From the questionnaire, highlighting had positive feedback and that the torus *seems very useful for highlighting objects in the scene* (P8, P11)

6.2.6 Potential Applications and Use Scenarios. When asked about the potential applications and use cases, many participants see the potential for sports analysis (P3, P10) and data visualization (P7, P14). For example, P3 says "*On the note of sports, like for example, in ballet, your position is very important. So if you're doing plays like, for example, videos where you highlight certain points like knees they have to be at a certain angle.*" As the participants say, our tool allows the user not only to analyze the sports from different angles, but also to measure the trajectory or posture in improvisational

and interactive ways. Also, the participants see the future potential of video streaming. For example, one of P4 says that *"I would use this in my stream in some way to incorporate effects. And maybe even have audience members triggering different things in my space"* The feedback points to a future when 3D volumetric videos become mainstream. The participants see that tools like RealityEffects allow such a video streaming medium more interactive and engaging.

6.2.7 Limitations and Future Work. While the feedback was generally positive, several limitations were noted. P7 mentioned that the tool's functionality is currently limited, particularly the types of 3D shapes available for highlighting objects, which are restricted to simple geometric forms. Many participants pointed out the need for improvement in tracking accuracy, especially with the color-based tracking system's susceptibility to slight environmental changes, such as lighting conditions or shadow occlusion. Future work should explore alternative tracking methods for 3D objects, given that volumetric object tracking remains an active research area and advances in this field could significantly enhance object-centric 3D video augmentation.

System limitations also include the requirement for physical interaction or assistance from another person, as stated by P4. P3 elaborated on this by mentioning the extensive setup required for video capture, including the need for an open area, camera rig, and trackable objects. Additionally, the tool currently lacks complex time manipulation features found in traditional video-editing tools. Integrating our features into volumetric video-editing platforms could create richer experiences.

Participants also criticized the quality of 3D capturing. Using only a single Azure Kinect depth camera limits the capture area and fails to record occluded regions. Although we integrated a static 3D scene as a background to mitigate this, the limited capture area restricts applications requiring dynamic movement across larger areas. A potential solution could involve using multiple Kinect cameras, similar to approaches like Remixed Reality [59]. While this would allow more immersive visualization, it would also increase computational demands and the complexity of the calibration process. Future work should consider incorporating multiple depth cameras to support activities requiring broader interaction spaces.

Direct manipulation in RealityEffects enables users to feel an immediate connection with the digital content, fostering a sense of control and ownership over the creative process. However, we recognize the importance of considering alternative approaches that could complement or enhance the user experience. Automatic annotation, for example, could offer efficiency benefits by reducing the manual effort required to label and annotate volumetric data. This method could automatically identify and label objects within a scene using advanced machine learning algorithms, which would be particularly useful in complex scenes or for users who require quicker workflows. A suggestion-based interface is another compelling alternative that could blend the strengths of direct manipulation with the efficiencies of automation. By providing users with intelligent suggestions based on context, previous actions, or common patterns, this approach could accelerate the editing process while still allowing users the freedom to make final decisions. Future work could explore these alternatives to support a wider range of user preferences.

Currently, we focus on desktop authoring interfaces due to the complexity of interactions and manipulations involved. However, future investigations could explore opportunities within immersive environments using mixed reality or virtual reality headsets. Such environments would present unique design and technical challenges, such as selecting objects and streaming large amounts of data between the host computer and the headset. Addressing these issues could lead to innovative solutions for immersive augmentation, and we are keen on developing these capabilities for devices like the Hololens.

7 CONCLUSION

This paper presents RealityEffects, a desktop authoring interface designed to edit and augment 3D volumetric videos with object-centric annotations and visual effects. We introduce a novel approach to augment captured physical motion with embedded and responsive visual effects. The primary contribution of this paper is the development of a taxonomy of augmentation techniques. We demonstrate various augmentation techniques, including annotated labels, highlighted objects, ghost effects, and trajectory visualization. The results of our user study indicate that our direct manipulation techniques significantly lower the barrier to annotating volumetric videos. Based on the feedback received, we also discuss potential future work.

ACKNOWLEDGMENTS

This work was partially supported by NSERC Discovery Grant and JST PRESTO Grant Number JPMJPR23I5, Japan.

REFERENCES

- [1] 4DViews. [n.d.]. 4Dfx. <https://www.4dviews.com/volumetric-software>.
- [2] Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*. 235–242.
- [3] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 311–320.
- [4] Hrvoje Benko, Andrew D Wilson, and Federico Zannier. 2014. Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 645–655.
- [5] Frederik Brudy, Suppachai Suwanwatcharakat, Wenyu Zhang, Steven Houben, and Nicolai Marquardt. 2018. Eagleview: A video analysis tool for visualising and querying spatial interactions of people and devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. 61–72.
- [6] Wolfgang Büschel, Anke Lehmann, and Raimund Dachselt. 2021. Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [7] Yuanzhi Cao, Anna Fuste, and Valentin Heun. 2022. MobileTutAR: a Lightweight Augmented Reality Tutorial System using Spatially Situated Human Segmentation Videos. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.
- [8] Yuanzhi Cao, Tianyi Wang, Xun Qian, Pawan S Rao, Manav Wadhawan, Ke Huo, and Karthik Ramani. 2019. GhostAR: A time-space editor for embodied authoring of human-robot collaborative task with augmented reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 521–534.
- [9] Zhutian Chen, Shuainan Ye, Xiangtong Chu, Haijun Xia, Hui Zhang, Huamin Qu, and Yingcai Wu. 2021. Augmenting sports videos with viscommentator. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 824–834.
- [10] Lung-Pan Cheng, Eyal Ofek, Christian Holz, and Andrew D Wilson. 2019. Vroamer: generating on-the-fly VR experiences while walking inside large, unknown real-world building environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 359–366.

- [11] Yangming Cheng, Liulei Li, Yuanyou Xu, Xiaodi Li, Zongxin Yang, Wenguang Wang, and Yi Yang. 2023. Segment and track anything. *arXiv preprint arXiv:2305.06558* (2023).
- [12] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 282–297.
- [13] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–16.
- [14] Pei-Yu Chi, Daniel Vogel, Mira Dontcheva, Wilmett Li, and Björn Hartmann. 2016. Authoring illustrations of human movements by iterative physical demonstration. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 809–820.
- [15] Subramanian Chidambaram, Hank Huang, Fengming He, Xun Qian, Ana M Villanueva, Thomas S Redick, Wolfgang Stuerzlinger, and Karthik Ramani. 2021. Processar: An augmented-reality-based tool to create in-situ procedural 2d/3d ar instructions. In *Designing Interactive Systems Conference 2021*. 234–249.
- [16] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive video: adaptive video playback based on user motion for supporting physical activity. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 196–208.
- [17] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5828–5839.
- [18] Philip DeCamp, George Shaw, Rony Kubat, and Deb Roy. 2010. An immersive system for browsing and visualizing surveillance video. In *Proceedings of the 18th ACM international conference on Multimedia*. 371–380.
- [19] Dazhen Deng, Jiang Wu, Jiachen Wang, Yihong Wu, Xiao Xie, Zheng Zhou, Hui Zhang, Xiaolong Zhang, and Yingcai Wu. 2021. EventAnchor: reducing human interactions in event annotation of racket sports videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [20] DepthKit. [n.d.]. DepthKit Studio. <https://www.depthkit.tv/depthkit-studio>.
- [21] Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escalano, Christoph Rhemann, David Kim, Jonathan Taylor, et al. 2016. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (ToG)* 35, 4 (2016), 1–13.
- [22] Ruofei Du, Ming Chuang, Wayne Chang, Hugues Hoppe, and Amitabh Varshney. 2018. Montage4d: Interactive seamless fusion of multiview video textures. (2018).
- [23] Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, et al. 2020. DepthLab: Real-time 3D interaction with depth maps for mobile augmented reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 829–843.
- [24] Andreas Fender, Philipp Herholz, Marc Alexa, and Jörg Müller. 2018. Optispaces: automated placement of interactive 3D projection mapping content. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [25] Andreas Fender, David Lindlbauer, Philipp Herholz, Marc Alexa, and Jörg Müller. 2017. Heatspace: Automatic placement of displays by empirical analysis of user behavior. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 611–621.
- [26] Andreas Rene Fender and Christian Holz. 2022. Causality-preserving Asynchronous Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–15.
- [27] Lei Gao, Huidong Bai, Gun Lee, and Mark Billinghurst. 2016. An oriented point-cloud view for MR remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*. 1–4.
- [28] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 197–205.
- [29] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 449–459.
- [30] Dan B Goldman, Chris Gonterman, Brian Curless, David Salesin, and Steven M Seitz. 2008. Video object annotation, navigation, and composition. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 3–12.
- [31] Jianguo Gong, Teng Han, Siling Guo, Jiannan Li, Siyu Zha, Liuxin Zhang, Feng Tian, Qianying Wang, and Yong Rui. 2021. Holoboard: A large-format immersive teaching board based on pseudo holographics. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 441–456.
- [32] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escalano, Rohit Pandey, Jason Dourgarian, et al. 2019. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)* 38, 6 (2019), 1–19.
- [33] Brian D Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented Chironomia for Presenting Data to Remote Audiences. *arXiv preprint arXiv:2208.04451* (2022).
- [34] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi coaches: an augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*. 1–4.
- [35] Jeremy Hartmann, Christian Holz, Eyal Ofek, and Andrew D Wilson. 2019. Realitycheck: Blending virtual environments with situated physical reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [36] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmquist, and Harald Reiterer. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *CHI Conference on Human Factors in Computing Systems*. 1–20.
- [37] Nathaniel Hudson, Celena Alcock, and Parmit K Chilana. 2016. Understanding newcomers to 3D printing: Motivations, workflows, and barriers of casual makers. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 384–396.
- [38] Ke Huo and Karthik Ramani. 2017. Window-shaping: 3d design ideation by creating on, borrowing from, and looking at the physical world. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. 37–45.
- [39] Matthias Innmann, Michael Zollhöfer, Matthias Nießner, Christian Theobalt, and Marc Stamminger. 2016. Volumedeform: Real-time volumetric non-rigid reconstruction. In *European conference on computer vision*. Springer, 362–379.
- [40] Shahram Izadi, David Kim, Ottmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 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*. 559–568.
- [41] Brett Jones, Rajinder Sodhi, Michael Murdoch, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 637–644.
- [42] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 869–878.
- [43] Hiroki Kaimoto, Kyzyl Monteiro, Mehrad Faridan, Jiatong Li, Samin Farajian, Yasuaki Kakehi, Ken Nakagaki, and Ryo Suzuki. 2022. Sketched reality: Sketching bi-directional interactions between virtual and physical worlds with ar and actuated tangible ui. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–12.
- [44] Takeo Kanade, Peter Rander, and PJ Narayanan. 1997. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia* 4, 1 (1997), 34–47.
- [45] Mohamed Kari, Tobias Grosse-Puppendahl, Luis Falconeri Coelho, Andreas Rene Fender, David Bethge, Reinhard Schütte, and Christian Holz. 2021. TransformMR: Pose-aware object substitution for composing alternate mixed realities. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 69–79.
- [46] Thorsten Karrer, Moritz Wittenhagen, and Jan Borchers. 2009. Pocketdragon: a direct manipulation video navigation interface for mobile devices. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–3.
- [47] Daniel Keplinger, Günter Wallner, Simone Kruglstein, and Michael Lankes. 2020. See, Feel, Move: player behaviour analysis through combined visualization of gaze, emotions, and movement. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [48] Simon Kloiber, Volker Settgast, Christoph Schinko, Martin Weinzerl, Johannes Fritz, Tobias Schreck, and Reinhold Preiner. 2020. Immersive analysis of user motion in VR applications. *The Visual Computer* 36, 10 (2020), 1937–1949.
- [49] Thomas H Kolbe. 2004. Augmented videos and panoramas for pedestrian navigation. In *Proceedings of the 2nd Symposium on Location Based Services & TeleCartography 2004, 28–29th of January 2004 in Vienna*.
- [50] Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*. 1–9.
- [51] André Kunert, Alexander Kulik, Stephan Beck, and Bernd Froehlich. 2014. Photoports: shared references in space and time. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1388–1399.
- [52] Jason Lawrence, Dan B Goldman, Supreeth Achar, Gregory Major Blasovich, Joseph G Desloge, Tommy Fortes, Eric M Gomez, Sascha Häberling, Hugues Hoppe, Andy Huibers, et al. 2021. Project Starline: A high-fidelity telepresence system. (2021).

- [53] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for HCI toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [54] Bokyoung Lee, Michael Lee, Pan Zhang, Alexander Tessier, and Azam Khan. 2019. Semantic human activity annotation tool using skeletonized surveillance videos. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. 312–315.
- [55] Germán Leiva, Jens Emil Grønbæk, Clemens Nylandsted Klokmosæ, Cuong Nguyen, Rubaiat Habib Kazi, and Paul Asente. 2021. Rapido: Prototyping Interactive AR Experiences through Programming by Demonstration. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 626–637.
- [56] Germán Leiva, Cuong Nguyen, Rubaiat Habib Kazi, and Paul Asente. 2020. Pronto: Rapid augmented reality video prototyping using sketches and enactment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [57] Yuwei Li, Xi Luo, Youyi Zheng, Pengfei Xu, and Hongbo Fu. 2017. SweepCanvas: Sketch-based 3D prototyping on an RGB-D image. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 387–399.
- [58] Jian Liao, Adnan Karim, Shivesh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-Time Speech-Driven Augmented Presentation for AR Live Storytelling. *arXiv preprint arXiv:2208.06350* (2022).
- [59] 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*. 1–13.
- [60] Jingyuan Liu, Hongbo Fu, and Chiew-Lan Tai. 2020. Posesetween: Pose-driven tween animation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 791–804.
- [61] Roberto A Montano-Murillo, Cuong Nguyen, Rubaiat Habib Kazi, Sriram Subramanian, Stephen DiVerdi, and Diego Martinez-Plasencia. 2020. Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 53–62.
- [62] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parhami, and Ryo Suzuki. 2023. Teachable reality: Prototyping tangible augmented reality with everyday objects by leveraging interactive machine teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [63] Shohei Mori, Sei Ikeda, and Hideo Saito. 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications* 9, 1 (2017), 1–14.
- [64] Cuong Nguyen, Yuzhen Niu, and Feng Liu. 2012. Video summaragator: An interface for video summarization and navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 647–650.
- [65] Cuong Nguyen, Yuzhen Niu, and Feng Liu. 2013. Direct manipulation video navigation in 3D. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1169–1172.
- [66] Cuong Nguyen, Yuzhen Niu, and Feng Liu. 2014. Direct manipulation video navigation on touch screens. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 273–282.
- [67] Benjamin Nuernberger, Eyal Ofek, Hrvoje Benko, and Andrew D Wilson. 2016. Snaptoreality: Aligning augmented reality to the real world. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1233–1244.
- [68] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*. 741–754.
- [69] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1716–1725.
- [70] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
- [71] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. 2019. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–17.
- [72] PolyCam. [n.d.]. Polycam. <https://poly.cam/>.
- [73] Iulian Radu, Tugce Joy, and Bertrand Schneider. 2021. Virtual makerspaces: merging AR/VR/MR to enable remote collaborations in physical maker activities. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.
- [74] Holger Regenbrecht, Katrin Meng, Arne Reepen, Stephan Beck, and Tobias Langlotz. 2017. Mixed voxel reality: Presence and embodiment in low fidelity, visually coherent, mixed reality environments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 90–99.
- [75] Patrick Reipschläger, Frederik Brudy, Raimund Dachselt, Justin Matejka, George Fitzmaurice, and Fraser Anderson. 2022. AvatAR: An Immersive Analysis Environment for Human Motion Data Combining Interactive 3D Avatars and Trajectories. In *CHI Conference on Human Factors in Computing Systems*. 1–15.
- [76] Claudia Ribeiro, Rafael Kuffner, and Carla Fernandes. 2018. Virtual reality annotator: A tool to annotate dancers in a virtual environment. In *Digital Cultural Heritage: Final Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage, ITN-DCH 2017, Olimje, Slovenia, May 23–25, 2017, Revised Selected Papers*. Springer, 257–266.
- [77] Stephanie Santos, Fanny Chevalier, Ravin Balakrishnan, and Karan Singh. 2013. Direct space-time trajectory control for visual media editing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1149–1158.
- [78] Nazmus Saquib, Faris Huq, and Syed Arefin Haque. 2022. graphiti: Sketch-based Graph Analytics for Images and Videos. In *CHI Conference on Human Factors in Computing Systems*. 1–15.
- [79] Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, and Wilmot Li. 2019. Interactive body-driven graphics for augmented video performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [80] João Silva, Diogo Cabral, Carla Fernandes, and Nuno Correia. 2012. Real-time annotation of video objects on tablet computers. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. 1–9.
- [81] Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maiocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188.
- [82] Misha Sra, Sergio Garrido-Jurado, and Pattie Maes. 2017. Oasis: Procedurally generated social virtual spaces from 3d scanned real spaces. *IEEE transactions on visualization and computer graphics* 24, 12 (2017), 3174–3187.
- [83] 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*. 191–200.
- [84] Arcturus Studios. [n.d.]. HoloEdit. <https://arcturus.studio/holoedit/>.
- [85] Yuki Sugita, Keita Higuchi, Ryo Yonetani, Rie Kamikubo, and Yoichi Sato. 2018. Browsing group first-person videos with 3d visualization. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. 55–60.
- [86] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In *CHI Conference on Human Factors in Computing Systems*. 1–33.
- [87] Ryo Suzuki, Rubaiat Habib Kazi, Li-Yi Wei, Stephen DiVerdi, Wilmot Li, and Daniel Leithinger. 2020. Realitysketch: Embedding responsive graphics and visualizations in AR through dynamic sketching. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 166–181.
- [88] Matthew Tait and Mark Billinghurst. 2015. The effect of view independence in a collaborative AR system. *Computer Supported Cooperative Work (CSCW)* 24, 6 (2015), 563–589.
- [89] Franco Tecchia, Leila Alem, and Weidong Huang. 2012. 3D helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*. 323–328.
- [90] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.
- [91] Balasaravan Thorai-Kumaravel, Fraser Anderson, George Fitzmaurice, Björn Hartmann, and Tovi Grossman. 2019. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174.
- [92] Julien Valentin, Vibhav Vineet, Ming-Ming Cheng, David Kim, Jamie Shotton, Pushmeet Kohli, Matthias Nießner, Antonio Criminisi, Shahram Izadi, and Philip Torr. 2015. Semanticpaint: Interactive 3d labeling and learning at your fingertips. *ACM Transactions on Graphics (TOG)* 34, 5 (2015), 1–17.
- [93] Benjamin Walther-Franke, Marc Herrlich, Thorsten Karrer, Moritz Wittenhagen, Roland Schröder-Kroll, Rainer Malaka, and Jan Borchers. 2012. Dragimation: direct manipulation keyframe timing for performance-based animation. In *Proceedings of Graphics Interface 2012*. 101–108.
- [94] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. 2020. Slice of light: Transparent and integrative transition among realities in a multi-HMD-user environment. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 805–817.
- [95] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2021. Distanciar: Authoring site-specific augmented reality experiences for remote environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

- [96] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. 2016. Embedded data representations. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 461–470.
- [97] Zhijie Xia, Kyzyl Monteiro, Kevin Van, and Ryo Suzuki. 2023. RealityCanvas: Augmented Reality Sketching for Embedded and Responsive Scribble Animation Effects. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
- [98] Jinyu Yang, Mingqi Gao, Zhe Li, Shang Gao, Fangjing Wang, and Feng Zheng. 2023. Track anything: Segment anything meets videos. *arXiv preprint arXiv:2304.11968* (2023).
- [99] Emilie Yu, Kevin Blackburn-Matzen, Cuong Nguyen, Oliver Wang, Rubaiai Habib Kazi, and Adrien Bousseau. 2023. Videodoodles: Hand-drawn animations on videos with scene-aware canvases. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–12.
- [100] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective matters: Design implications for motion guidance in mixed reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 577–587.
- [101] 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*. 427–436.

8 APPENDIX

The following videos, which are a subset of the 120 videos analyzed for this taxonomy, are used as examples of object-centric augmentation techniques in Figure 2. Each letter indicates the category of the augmentations: T for text annotation, O for object highlight, E for embedded visual, C for connected link, and M for motion effect. Each number represents the order from left to right. The screenshots in Figure 2 are copyrighted by each video creator.

- [T1] Clearly Contacts "Saving Money" © Copyright by Giant Ant <https://vimeo.com/10904876>
- [T2] NORTH: Analytics for the real world – Symphoni © Copyright by PwC Digital Experience Center <https://vimeo.com/121175225>
- [T3] GRTgaz biomethane © Copyright by la famille <https://vimeo.com/40092864>
- [T4] Zoopla TV advert "Smart Knows" © Copyright by Zoopla <https://www.youtube.com/watch?v=jkADFJdYakY>
- [T5] inBloom vision video © Copyright by Intentional Futures <https://vimeo.com/60661666>
- [T6] DREAN // Motion Tracking + layouts © Copyright by Estudio Ánimo <https://vimeo.com/68242831>
- [O1] Live from Tokyo: 2018 Nissan LEAF Launch © Copyright by George P. Johnson <https://www.youtube.com/watch?v=EoMU3SuZ-uw>
- [O2] Device UI in Realtime © Copyright by Dennis Schaefer <https://vimeo.com/165467760>
- [O3] Whirlpool Interactive Cooktop © Copyright by The Hobbs Report <https://www.youtube.com/watch?v=Efj6gKw3wKc>
- [O4] Alibaba brings AR, VR, and virtual influencers to online shopping © Copyright by TechNode <https://www.youtube.com/watch?v=xLQAxYMYxIu>
- [O5] GRTgaz biomethane © Copyright by la famille <https://vimeo.com/40092864>
- [O6] NORTH: Analytics for the real world – Symphoni © Copyright by PwC Digital Experience Center <https://vimeo.com/121175225>
- [E1] Ericsson - Business Users Survey - Commercial © Copyright by Erik Nordlund, FSF <https://vimeo.com/20168424>
- [E2] NTT Data - Future Experiences © Copyright by Designit <https://vimeo.com/142118168>
- [E3] Crafting Brands for Future Life © Copyright by Ben Collier-Marsh <https://vimeo.com/196708386>
- [E4] Mixed Reality - Home Kit © Copyright by Sertan Helvacı <https://dribbble.com/shots/6172560-Mixed-Reality-Home-Kit>
- [E5] Scosche myTrek :: 2011 [Evlab] © Copyright by Greg Del Savio <https://vimeo.com/27620294>
- [E6] Sight © Copyright by Eran May-Raz and Daniel Lazo <https://www.youtube.com/watch?v=OstCyV0nOGs>
- [C1] Ericsson - Business Users Survey - Commercial © Copyright by Erik Nordlund, FSF <https://vimeo.com/20168424>
- [C2] Zoopla TV advert "Smart Knows" © Copyright by Zoopla <https://www.youtube.com/watch?v=jkADFJdYakY>
- [C3] Scosche myTrek :: 2011 [Evlab] © Copyright by Greg Del Savio <https://vimeo.com/27620294>
- [C4] DREAN // Motion Tracking + layouts © Copyright by Estudio Ánimo <https://vimeo.com/68242831>
- [C5] La Boulangerie Delannay © Copyright by Julien Loth <https://vimeo.com/45055294>
- [C6] Thomson // Reuters © Copyright by Rushes Creative, Domhnall Ó Maoleoin, BT CORCORAN, Tania Nunes, and Guy Hancock <https://www.behance.net/gallery/54032303/Thomson-Reuters>
- [M1] Writing Performance in the Language of Light © Copyright by GE Lighting, a Savant company <https://www.youtube.com/watch?v=G9cBpSRT500>
- [M2] Yuki Ota Fencing Visualized Project - MORE ENJOY FENCING (English Ver.) © Copyright by fencing visualized project <https://www.youtube.com/watch?v=h2DXCAWIgU>
- [M3] IBM PGA © Copyright by LOS YORK <https://vimeo.com/40882289>
- [M4] FeelCapital corporate video © Copyright by democràcia <https://vimeo.com/98023574>
- [M5] Nike: Pegasus 31 © Copyright by Tad Greenough <https://vimeo.com/132446809>
- [M6] Writing Performance in the Language of Light © Copyright by GE Lighting, a Savant company <https://www.youtube.com/watch?v=G9cBpSRT500>