Tesseract: Querying Spatial Design Recordings by Manipulating Worlds in Miniature

Karthik Mahadevan
karthikm@dgp.toronto.edu
University of Toronto
Canada

Qian Zhou
qian.zhou@autodesk.com
Autodesk Research
Canada

George Fitzmaurice
george.fitzmaurice@autodesk.com
Autodesk Research
Canada

Tovi Grossman
tovi@dgp.toronto.edu
University of Toronto
Canada

Fraser Anderson
fraser.anderson@autodesk.com
Autodesk Research
Canada

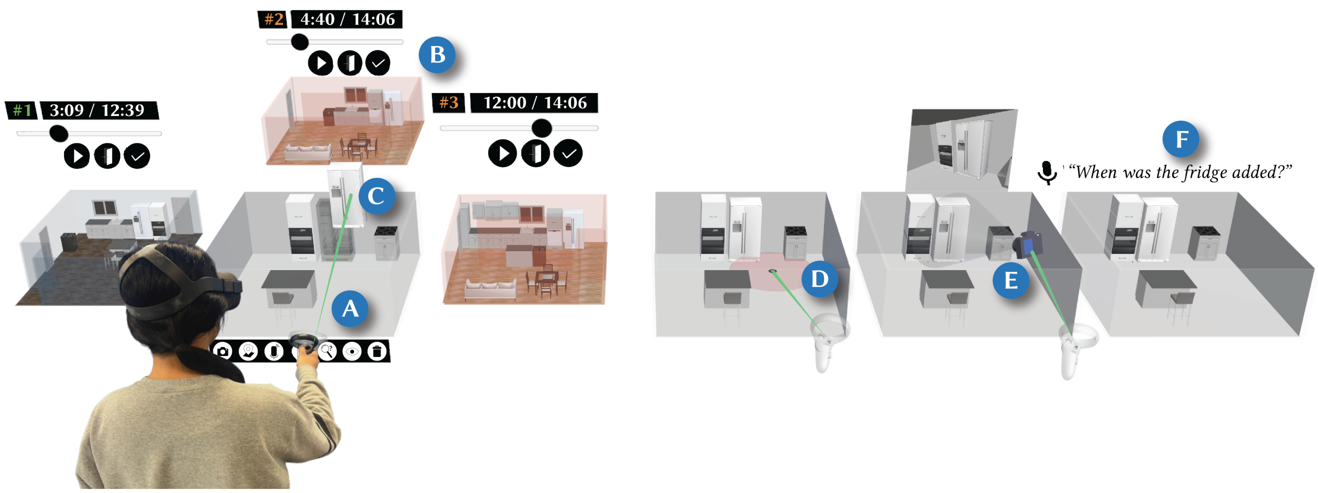

Figure 1: We present Tesseract, a novel Worlds-in-Miniature-based VR system that allows users to search through space and time in spatial design recordings. Users can form queries using the *Search Cube* interface (A). Queries are retrieved as *spatial clips* (B) that users can preview or re-watch in 1:1 scale. Tesseract provides four *querying tools* to enable searching spatial design recordings. Users can manipulate objects into the search cube to perform *object search* (C) or define the behavior of recorded people such as their proximity to objects through *proximity search (D)*, their viewpoints through *viewpoint search* (E), or their speech through *voice search* (F) to retrieve interesting moments related to design activities.

ABSTRACT

New immersive 3D design tools enable the creation of *spatial design recordings*, capturing collaborative design activities. By reviewing captured spatial design sessions, which include user activities, workflows, and tool use, users can reflect on their own design processes, learn new workflows, and understand others' design rationale. However, finding interesting moments in design activities can be challenging: they contain multimodal data (such as user motion and logged events) occurring over time which can be difficult to specify when searching, and are typically distributed over many sessions or recordings. We present Tesseract, a Worlds-in-Miniature-based system to expressively query VR spatial design recordings. Tesseract consists of the Search Cube interface acting as a centralized *stage-to-search* container, and four querying tools for specifying multimodal data to enable users to find interesting moments in past design activities. We studied ten participants who used Tesseract and found support for our miniature-based stage-to-search approach.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580876

CCS CONCEPTS

• **Human-centered computing → Interaction techniques**; **Interactive systems and tools**.

KEYWORDS

Querying spatial design recordings, Worlds-in-Miniature

ACM Reference Format:
Karthik Mahadevan, Qian Zhou, George Fitzmaurice, Tovi Grossman, and Fraser Anderson. 2023. Tesseract: Querying Spatial Design Recordings by Manipulating Worlds in Miniature. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3580876

1 INTRODUCTION

Designers often review their work or the work of others during the design process [41]. For example, architects review building designs before construction to ensure all stakeholders have their requirements satisfied. Designers may review work for different purposes—for example they may reflect on their design process while working on or after completing a design [61]. They may want to understand how another designer achieved a specific result by seeing their workflow. Or, they may encounter a design and wish to understand the design rationale behind certain decisions; in interior design, this ranges from objective decisions about where to place appliances based on guidelines [51] to stylistic choices.

Prior work has explored ways to support designers across stages of the design process in 2D recordings that contain data such as voice, text, screenshots, and command histories. For instance, Think-Aloud Computing enables the reflection of others' work sessions when programming [38], Chronicle supports the playback of document workflow histories [24], and Winder helps reason about design decisions made by team members working asynchronously on visual documents [35].

With the advent of virtual reality (VR) tools aimed at collaborative spatial design (e.g., The Wild [70] and Arkio [2]), *spatial design recordings* or 3D recordings are now possible to create and use in the design process. Spatial design recordings are spatial recordings that contain information about past spatial design activities. For instance, Nvidia recently released VR Capture and Replay (VCR) that accurately captures and replays VR sessions for supporting design reviews and troubleshooting [52]. Spatial design recordings contain design activities that are multimodal, take place over time, and typically occur over multiple sessions. However, tools for searching these recordings are limited. Prior work has mostly proposed solutions for *navigating* spatial recordings individually rather than *searching* through databases of spatial design recordings; for instance, selecting previously logged events on a timeline to re-watch them [10, 68] or directly manipulating scene objects—moving them around or toggling their states—to navigate to moments of change [40].

Despite the potential of spatial design recordings as design aids, finding moments of interest can be challenging for many reasons. Spatial design recordings contain multimodal data about people and their design activities in 3D which occur over time, including but not limited to their motion, gaze, gesture [37], interactions with objects, and conversation. Prior work on multimodal retrieval focuses on the 2D representation of data [34, 65], such as combining text and image to retrieve items from a multimedia database with videos and images [65]. However, it is unclear how users can search spatial design recordings that contain multimodal spatiotemporal data to find interesting moments. Further, design activities often take place over multiple sessions captured across multiple recordings. However, today's tools focus on *navigating* single recordings rather than *searching* across many recordings.

Given the limited support for searching through spatial design recordings, we aim to close this gap through Tesseract, a novel Worlds-in-Miniature (WiM)-based system. Tesseract can support designers to search spatial design recordings to find moments of interest for different purposes such as *reflecting on previous designs*, *learning new workflows*, and *uncovering design rationale*. Tesseract consists of a centralized WiM-based *stage-to-search* container, called the *search cube*. The Search Cube interface (a 3D analog to today's 2D search box) lets users stage queries through four *querying tools* designed to search multimodal data found in spatial design recordings. Users can stage queries by directly manipulating objects into the search cube and defining geometric relationships between them using *object search*. Users can also define the spatial behavior of recorded people by manipulating their position, movement, and what they are looking at through *proximity search* and *viewpoint search*, and can retrieve *conversations* and logged interaction *events* with *voice search*.

We asked 10 participants to use Tesseract to understand whether the Search Cube interface and querying tools can be used to retrieve past design activities found in spatial design recordings, contextualized in the domain of interior design. Participants used Tesseract's Search Cube interface and querying tools to find target clips in past recordings. Participants were able to use Tesseract to express their search intent and retrieve past design activities, which highlights the potential of our WiM-based stage-to-search querying approach. This work advances the state-of-the-art in spatial querying and makes the following contributions:

- Tesseract, a novel system consisting of a Worlds-in-Miniature (WiM)-based Search Cube interface to support searching spatial design recordings,
- A set of querying tools accessible through Tesseract and the Search Cube interface that enable the retrieval of multimodal data embedded in spatial design recordings, and
- Findings from a preliminary user evaluation that highlight usage patterns that participants demonstrated to express their search intent and opportunities for designing future querying tools.

2 RELATED WORK

We begin by breaking down the categories of spatial recordings to situate Tesseract as a system for searching spatial design recordings. Next, we summarize prior art aimed at supporting the design process in 2D and 3D recordings which inspired the design of Tesseract. Finally, we categorize prior techniques for querying 3D data which helped us identify miniature-based stage-to-search as a querying metaphor to search spatial design recordings.

2.1 Spatial Recordings

Spatial recordings can be organized into three categories: *reality capture*, *mixed reality capture*, and *virtual capture*. *Reality capture* records what happens in the real-world using sensors (such as cameras). Examples include motion capture used in video games and film making [67], skeleton data for AR-based physical activity training [1, 11] or to understand social interactions [39], and recently, 3D volumetric capture such as of sports [63]. *Mixed reality capture* records what happens in the real-world along with in-application events, device interactions, or virtual augmentations. Recent tools such as DistanciAR [69] enable remote authoring of real-world experiences where local environments can be captured by a user and augmented with physical objects by remote users (such as for interior design). Similarly, ScalAR supports the adaptation of virtual layout designs to varying physical environments [58]. Lastly, *virtual capture* involves the capture of virtual environments where one or more users interact. This can include recordings of video game sessions [18], VR user studies [42] and design activities [10]. Recently, tools supporting spatial design have gained popularity, such as for architecture, engineering, and construction [2, 70] and for general design activities (such as Nvidia's VCR) [52]. In this work, we focus on virtual capture; specifically, through Tesseract, we target the querying of spatial design recordings where one or many recorded persons interact with and modify virtual environments while engaged in design activities.

2.2 Temporal Navigation of 2D and 3D Recordings

Researchers have proposed tools and techniques to navigate 2D and 3D recordings which may contain design history from many sources such as screen capture, audio, and command histories. For 2D recordings, systems such as DocWizards provide a new way to explore documentation by guiding an unfamiliar user through a previously recorded procedure [3]. ExperiScope lets designers and experimenters revisit user evaluations of interaction techniques with visual traces to assist understanding user patterns [25]. Chronicle captures document workflow histories in a painting application to enable applications such as understanding how a recorded user achieved a particular visual outcome [24]. Winder supports asynchronous collaboration on visual documents through linked tapes to support understanding design rationale [35].

For 3D recordings, Skeletonographer provides an annotation tool to quickly browse human movement data [39]. MIRIA [8] and MRAT [50] provide approaches to browse recorded data both in-situ and externally (e.g., on 2D screens in MRAT). ReLive combines an immersive analytics VR view with a non-immersive desktop view to aid the analysis of mixed reality studies [28]. For virtual capture containing annotated data, individual recordings can be re-watched asynchronously [10] or in a synchronous, shared manner [68]. Other interfaces allow users to understand human motion in VR by automatically extracting temporal points of interest [36], or directly manipulate scene objects to navigate to particular moments within the same recording [40]. AsyncReality explores how to record and playback real-world events that take place when a user is occupied in virtual reality by preserving causality [14].

Our work shares the same motivation of supporting users to replay design recordings to understand design rationale [35] or attempt to learn new workflows [24]. However, we emphasize that *searching* is a different task from *navigating* the timelines of recordings. Temporal navigation techniques such as direct manipulation [40] enable users to browse and quickly replay a single recording to find interesting moments when objects move around or a change is tracked, while searching such as with Tesseract supports users to retrieve a clip that contains the design activity of interest from many spatial recordings. Once a clip is retrieved, users can use existing and familiar timeline navigation techniques to quickly re-watch the clip. While various timeline navigation techniques have been proposed to browse individual 2D and 3D spatial recordings, it is unclear how users can *search* through spatial recordings to retrieve moments of interest. Our work aims to bridge this gap with a focus on querying spatial design recordings that capture collaborative design activities.

2.3 Interaction Techniques to Search 3D Data

Significant prior work has explored querying 3D data with different input modalities, including text-to-search, sketch-to-search, and stage-to-search. Text is well-established for searching static 3D databases. Researchers have proposed matching search phrases to pre-populated properties of models for retrieval [17]. Other approaches enhance the tagging process by re-using tags of geometrically similar models to populate untagged models [21]. Despite the popularity of text as a search modality, it is hard to specify the rich details that 3D models typically contain. This has led to a number of alternative input modalities to search and retrieve 3D data.

Visual search is a well explored mechanism for querying 3D data; for instance through a visual querying language [4] or sketch-based search. Early interfaces utilized 3D sketch tools [17, 29] as well as pixel-based programs to re-project target models into 2D space to make comparisons [57]. More sophisticated data-driven computer vision approaches have recently gained popularity [13, 43, 73], of which some target VR use [19, 44]. EagleView introduced a visual-query interface that lets users directly manipulate a video stream to perform queries on the underlying spatial data [6]. In the physical world, data miming utilizes user gestures to describe and retrieve 3D objects from a database [27]. Visual stimulus such as providing images [22, 23] or 3D models has also been proposed to find similar 3D models [7, 53]. There have also been attempts at combining search modalities for retrieval; for instance, ShapeFindAR lets users combine in-situ AR sketches with textual queries to find 3D models [64].

Though the above techniques enhance user expressivity in querying, they are designed to find individual static 3D assets such as 3D models. On the other hand, *stage-to-search* lets users describe layouts of 3D objects to search scene databases to find similar scenes [16], provide new layout recommendations [45], produce previously unseen but plausible scenes [15], or simply find objects in a large virtual scene [56]. In these systems, stage-to-search increases the user's expressive capabilities since they can visually specify all the objects in the scene and their relationships to each other to form layouts.

For spatiotemporal data, prior work has investigated techniques to enable explorations of large parameterized design spaces; for instance, generating sample animation sequences and allowing users to interactively explore possibilities for visual effects design through selection [5]. More recently, Unified Many-Worlds Browsing enables users to define and drag spatiotemporal queries to find interesting scenarios in ensembles of physics-based phenomena such as for animating [20]. Lastly, recent work has also proposed methods to get user feedback on proposed designs to learn design adjectives or models of user intent [62] which has been shown to be effective across many domains in 2D and 3D design.

Spatial design recordings differ from prior static 3D data due to the addition of a time dimension. Further, they contain combinations of different data types. Beyond 3D objects, they also contain recorded users and their design activities including their conversations. Hence, recent techniques for exploring large parameterized design spaces containing spatiotemporal data such as by selecting relevant results [5] or dragging a query box [20] may not be sufficient to represent what designers may be doing in these recordings. Conceptually, our work is closest to the techniques that support stage-to-search in a static 3D scene. Applying existing 3D static search techniques [16, 56] to search spatial design recordings would only retrieve a single "frame" and would not support searching for a sequence of frames, making it difficult to search for activities such as designers' locomotion or conversation. In this work, we extend stage-to-search to support multimodal querying of spatial design recordings across frames. We introduce querying tools that let users specify their search intent surrounding multimodal data embedded in these recordings.

2.4 Container-based 3D interactions

Prior work has explored Worlds-in-Miniatures (WiMs) [12, 55] to support interactions in virtual environments. They have been used to support visual comparisons of large-scale data [49] and to perform data analysis [26]. WiMs have also been used for wayfinding in static 3D layouts [48], and recently to locate objects within the same scene [56]. Spacetime proposed using WiMs as *containers* to collaboratively edit 3D scenes as well as navigate a container's history by scrubbing its timeline to replay the recording [71] . This is akin to a standard video navigation technique but does not allow for searching within the container or across containers. In contrast to prior work that has applied containers for scene editing and the navigation of recordings, we propose to use WiMs as a centralized input canvas for users to express their search input via stage-to-search to retrieve interesting moments in spatial design recordings. Once a moment is retrieved as a spatial clip, users can play it back similarly to existing WiM-based playback techniques [71].

3 UNIQUE CHALLENGES OF QUERYING SPATIAL DESIGN RECORDINGS

Spatial design recordings contain 3D information about past spatial design activities. Designers' activities are complex in that they: (1) typically contain multimodal data, (2) occur over multiple frames, and (3) are often long and recorded over multiple sessions. An example is creating an interior design layout for a client which takes place over multiple meetings to review and refine the design. Existing search techniques designed for searching 3D models and static 3D scenes [16, 56] do not support searching across multiple frames. On the other hand, existing timeline navigation techniques for spatial recordings can support scrubbing or jumping across multiple frames [40, 71] but do not support searching multimodal data and users are restricted to browsing one recording at a time. Hence, the challenge of querying spatial design recordings lies in the lack of support for users to express their search intent to retrieve moments that consist of multimodal data across multiple frames in multiple sessions of recordings. Our work aims to fill this gap through an understanding of the unique aspects of spatial design recordings.

The multimodal data in spatial design recordings can be categorized into three *components*: *objects*, *people*, and *interactions*. *Objects* can include but are not limited to 3D models, materials, textures, colors, animations, and entire scenes. As designers compose a new design, they manipulate *objects* and their properties, creating versions of the design at each moment which can be recorded and later retrieved through querying. Since spatial design recordings record *people* as they engage in the design process, their activities can also be captured and retrieved [52]. Activities are recorded across frames, including how people move around the virtual environment, what they are looking at, and verbal dialogue with themselves in a think-aloud fashion or in conversation with other people. Lastly, as spatial design recordings are typically created in feature-rich applications that include tools for manipulating objects [2, 70] (e.g., when a 3D model is moved around or when the distance between models is measured), these *interactions* can be logged and retrieved. Each frame of the spatial design recording contains rich, multimodal data; for instance, a user could be speaking while manipulating objects in the space. Hence, this data must be searchable. More importantly, designers' activities occur over time and contain temporal data so an additional challenge is enabling users to express complex queries that span across frames. Through an understanding of the unique challenges of searching spatial design recordings, we developed the following design goals to guide the design of Tesseract.

DG1. Provide a Centralized Space to Form Queries: The conventional 2D text-based search bar provides a unified interface with which users can query data of different types. We want to leverage the 2D search interaction model users are already familiar with by providing a centralized space to form spatial queries to find moments in spatial design recordings. In addition, the centralized space to form queries must be accessible without removing users from their current design activities.

DG2. Provide Expressive Querying Tools to Convey Search Intent: Spatial design recordings contain spatial and temporal data. However, traditional search modalities such as text do not afford the expressivity needed to query such data. We aim to provide users with querying tools that let them convey their search intent across space and time through the centralized search space. Additionally, users may wish to query recordings for various purposes. They may wish to reflect on a training or collaborative session with another designer post-hoc, may discover a brilliant aesthetic in a design and wish to reproduce it, or may want to understand why certain decisions were taken when seeing the final rendition of a design. Hence the querying tools must allow users to express complex queries to reveal these moments.

DG3. Support the Retrieval of Moments with Different Data Types: Since data from spatial design recordings consist of objects, people, and interactions, the proposed querying tools must provide an adequate mapping between the user's search intent and the underlying data to be retrieved; for instance, objects can be part of design activities but how they are arranged to form a layout is also of significance.

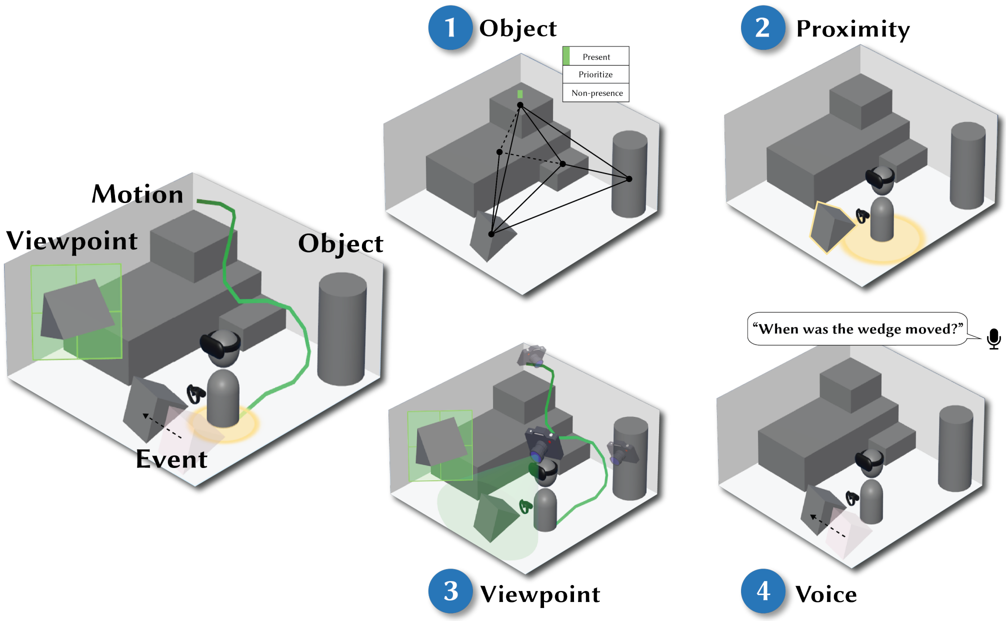

Figure 2: The conceptual idea behind the search cube and proposed querying tools: (1) *Object search* lets users specify layouts of objects and their relationships to one another; (2) *Proximity search* lets users specify the proximity of recorded *people* to *objects*; (3) *Viewpoint search* lets users define what recorded *people* looked at in an instant or over time; (4) *Voice search* lets users describe logged *interaction* events from activities by recorded *people* in the design software or dialogue from conversations that occurred at points during the recording.

4 TESSERACT SYSTEM

To provide a centralized space for querying **(DG1)**, we propose the Search Cube interface, which leverages Worlds-in-Miniature (WiM) to enable users to *stage* spatial queries. We chose WiMs for two reasons. First, they offer a centralized canvas for users to stage a query containing different types of data whilst providing an overview of what has been staged. Second, WiMs have been shown to be more efficient for interacting with objects in mixed reality [33]. For these reasons, WiMs are becoming increasingly popular and have been implemented in recent VR design applications [2, 70]. With a WiM-based *search cube*, users can specify the type of data (objects, people, and interactions) they wish to include in a query. They can add and manipulate objects, define the activities of recorded people such as how they moved around the space, the objects they looked at or that were near them, and the dialogue between them and other recorded people (Figure 2).

To support the increased expressivity of user search intent when forming queries **(DG2)**, we designed four querying tools: *object search, proximity search, viewpoint search,* and *voice search.* The querying tools were designed to support searching the components of data found in spatial design recordings. Each technique is designed to map the user's search intent to the underlying data in spatial design recordings **(DG3)**. Object search supports finding *objects.* Proximity search enables finding *people* and their interactions with objects near them. *Viewpoint search* supports describing the objects a recorded person sees in their field of view. *Voice search* supports retrieving moments with dialogue occurring in think-aloud sessions or conversations between *people*, and logged events based on their *interactions* with tools part of the design software. The querying tools are described below (more implementation details can be found in the supplementary materials):

Object search allows users to add and manipulate objects into the search cube to define relative layouts (Figure 2) that might exist in previous recordings. Each object becomes a part of the query through its *presence* in the search cube. By selecting an object, the user can toggle its *weight* through: *presence*, *prioritize*, and *non-presence* (or exclusion) in the query (Figure 2.1). When multiple objects are inside the search cube, Tesseract detects two *object-object relationships* automatically: *next-to* and *across-from* (dashed and solid lines in Figure 2.1). These are defined based on the forward directions of objects. Tesseract uses relative positions rather than the absolute positions of objects to search for similar layouts agnostic of the floor-plans present in the recordings. Objects next-to each other are within a distance threshold whereas objects across-from face each other. When searching for a relevant moment, some object weights (presence and prioritize) and all object-object relationships receive a score of one if found or satisfied while some object weights (non-presence and prioritize) receive a score of negative one when found or not satisfied respectively. Summing the scores across objects and normalizing them yields a *similarity score* between 0 to 1 which is used to determine the closest matching moment.

Proximity search lets users define the location of recorded people in proximity to specific objects they may be interacting with or visually examining such as when designing for accessibility and support spaces [51]. A resizable *selection disk* represents the area around a recorded person (Figure 2.2) in proximity to nearby objects. Users can manipulate the disk to define the recorded person's location. They can also resize the disk to include more or fewer objects next to the person. To compute a similarity score, Tesseract considers the number of objects in proximity of the disk. The similarity score is the fraction of the number of objects within the disk found at a given moment in a recording divided by the total number of objects within the disk in the search cube. *Proximity search* can also be used to specify an *absolute* location within a room that the recorded person was in (regardless of the objects and the layouts in the room). To access this, users can replace the generic search cube with a specific 3D layout. Then, the selection disk lets the user specify absolute locations. For absolute location, the similarity score is determined by comparing the recorded person's location in the search cube (denoted by the disk) to the location data of people in past recordings. When there are multiple recorded persons, the person with the highest similarity score is used.

Viewpoint search supports querying by letting users define *recorded viewpoints* or areas that a recorded person may have been looking at, either while stationary or moving around. A recorded viewpoint acts as a proxy for a recorded person (Figure 2.3). The viewpoint's position and orientation can be manipulated to filter which objects are visible. To define a sequence of viewpoints that a recorded person demonstrated, *frames* in time can be used. When multiple recorded people are present, users can add more recorded viewpoints representing different people. Each recorded viewpoint (person) has their own trajectory defined using individual frames.

Tesseract considers the objects visible from a recorded viewpoint to compute a similarity score. The simulated viewpoint of a recorded person is generated by placing a virtual camera at the recorded person's position in the recording to determine which objects were visible in its viewing frustum. The similarity score is the ratio of the number of objects in the recorded person's viewpoint divided by the number of objects with the same tags visible from the placed recorded viewpoint in the search cube. When there are multiple frames, Tesseract seeks a sequence of movements of a recorded person with the best normalized sum score. When there are multiple recorded people, Tesseract finds the best normalized sum by enumerating the possible combinations between the recorded viewpoints and persons in the recordings.

Voice search enables users to find *events* logged (Figure 2.4) when recorded people interact with the design software and objects in the virtual environment (such as by adding or deleting 3D models). It also supports finding *conversations* or think-aloud *dialogue* recorded by designers [38]. Tesseract records the user's utterances, with the last uttered phrase becoming the input to the query. For event-based search, keywords from the uttered phrase are matched with logged events (such as add-object). Conversation-based search is performed by converting the search phrase and all conversation dialogue in the recording into sentence embeddings through the a transformer-based model [60]. Cosine similarity is used to find the closest matching recorded dialogue. Tesseract performs event-based and conversation-based search simultaneously, returning the maximum of the two scores.

Users can combine querying tools to perform more complex queries. When combined, the similarity score is an average of the score computed from individual querying tools. For example, users can place and arrange objects to form a particular layout. Then, they can use *voice search* to find dialogue between two designers as the layout was being created. When querying, the combination of object and voice search considers the layout, logged events, and conversation, returning an average of *object search* and *voice search*. *Proximity search* and *viewpoint search* require objects to be included in the search cube. Hence, using them will inherently incorporate object weights and relationships, producing an average of the two similarity scores. Combining querying tools can increase the user's expressive capability which is one of our design goals **(DG2)**.

5 EXAMPLE QUERYING SCENARIOS

We present three example scenarios to demonstrate how Tesseract can support users to find moments of interest in previously captured design activities. The examples are adapted from scenarios identified in Chronicle which investigated the capture, exploration and playback of document workflow histories [24]. The examples are grounded in an interior design context, a popular use case for today's VR collaborative design tools [2, 70].

5.1 Team Support for Understanding Design Rationale

Tesseract can help team members uncover design rationale embedded in asynchronously performed design activities. Philip is a junior designer who just graduated design school and joined an interior design firm. He is assigned to a kitchen renovation project led by Alice, a senior designer. Alice has already partially designed the new kitchen following some design rationale. Her design process was captured using the firm's VR collaborative design software through many recorded sessions. Alice is re-assigned to a different more time-sensitive project so Philip is tasked with completing it. Although Philip learned about kitchen guidelines in school [51], he lacks hands-on experience so any modifications he makes could disrupt the harmony of Alice's design. Alice can explain high-level design goals to Philip, but it is not practical to explain each design decision she has already taken. Philip decides to replay the recorded sessions to understand her design process. However, watching all of Alice's recordings would also be inefficient. Instead, Philip uses Tesseract to stage part of her current design into the search cube with *object search* (Figure 3A, left). Doing so preserves the relative locations of the objects and their relationships. To reveal moments where she was modifying the design, Philip combines this with *voice search* to vocalize, "When is the door moved?". Upon asking, this query reveals a clip with the first moment where Alice is thinking aloud and has just moved the door to the corner to reduce foot traffic around the kitchen's center (Figure 3A, middle). Learning this design rationale, Philip frees up additional space by removing the current island stovetop in the center and adds a new stovetop away from the island to further reduce congestion. Since he is also able to see the different aesthetic choices Alice has experimented with, he understands the style she is attempting. He adds white marble countertops and paints the island a darker color to add contrast, creating a modern aesthetic that matches Alice's existing design (Figure 3A, right). In the next meeting, the client is impressed with the redesign and approves it.

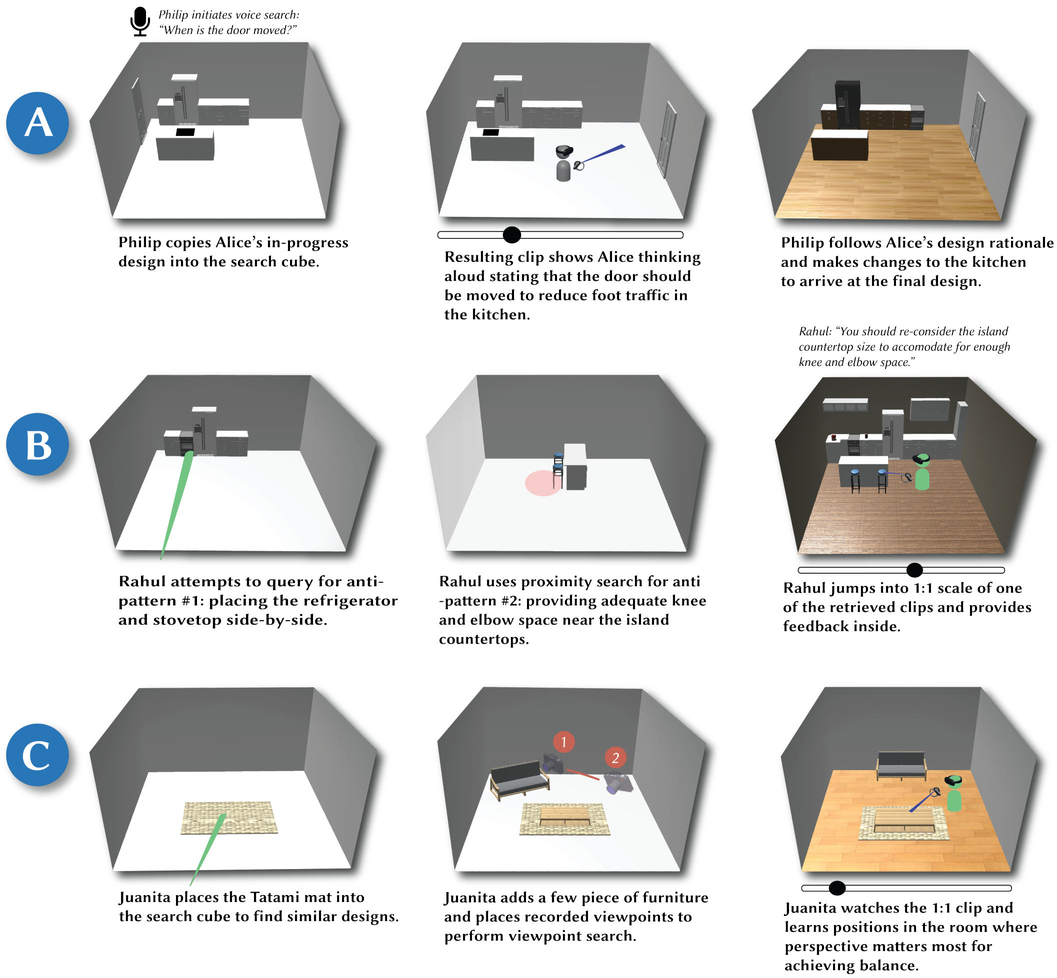

Figure 3: Tesseract can support finding moments of interest in previously captured design activities across a range of scenarios. These can include: (A) Team support for understanding design rationale; (B) Reflecting on Designs; (C) Learning New Workflows.

5.2 Reflecting on Designs

Tesseract can help users reflect on their designs or others' designs. Rahul, an experienced interior designer, is tasked with training a group of newly hired designers. As part of the onboarding process, each designer has been asked to create a kitchen design following interior design guidelines and best practices. In addition to submitting their final designs, they also record their design process in VR as spatial design recordings. Rahul reviews the submitted work and finds problems in some of the designs. He wants to use the recordings to pinpoint what caused the problems, but given his busy schedule, it is impossible for him to review every detail. Instead, Rahul uses his knowledge of best practices to craft queries to identify mistakes in all the submitted designs simultaneously through Tesseract. For example, he places the stovetop and refrigerator side-by-side in the search cube using *object search*, creating an interior design anti-pattern, and finds a few beautiful designs that have prioritized aesthetic without considering function (Figure 3B, left). He then adds a small island counter and seating, and places a recorded person using *proximity search* to identify design activities that assessed whether adequate counter space is provided to support seating as per the NKBA guidelines [51] (Figure 3B, middle). Since best practices suggest that the designer take measurements to ensure adequate support space and validate accessibility requirements, Rahul expects the trainees to have been in the approximate area of the counters at some moment during the recording. Upon querying, the results reveal similar moments from all the submitted spatial design recordings. Rahul is able to individually inspect each design with these issues by jumping into the resulting spatial clip in 1:1 scale and providing comments (Figure 3B, right). Later, the designers see Rahul's personalized feedback on their designs and refine their design processes accordingly.

5.3 Learning New Workflows

Tesseract can be used to learn design workflows. Juanita is an interior designer returning to a renowned firm after a stint in a different industry. Although she brings significant design experience, design

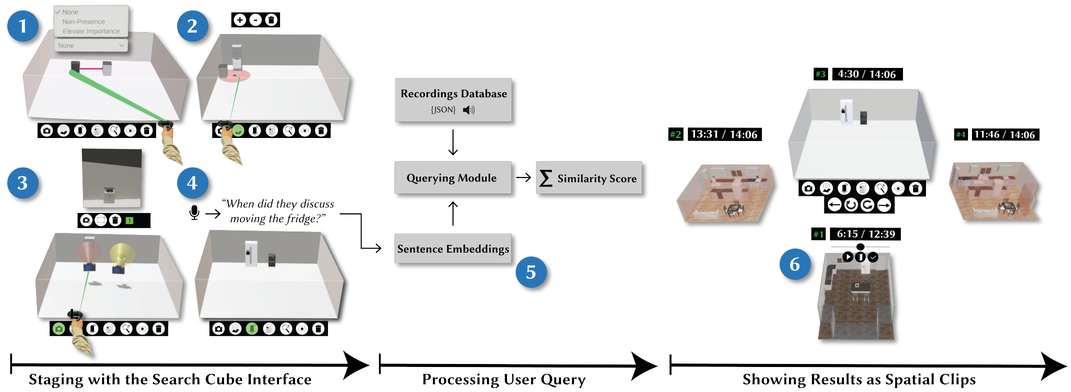

Figure 4: Tesseract system overview - (Left) Search Cube interface: (1) *Object search* where relationships can be specified manually through a menu or detected automatically (pink line denotes objects *across from* each other), (2) *Proximity search* where a selection disk is a proxy for recorded people and their proximity to objects, (3) *Viewpoint search* where each *recorded viewpoint* act as a proxy for a recorded person at some time (or over time through *frames* as the top UI shows), (4) *Voice search* where logged *events* or past *conversation* can be retrieved; (Middle) User queries are compared to a recordings database and possibly (5) sentence embeddings (for *conversational voice search*) to generate a *similarity score*; (Right) Top results (sorted by the time at which they appear in the recording) are populated as (6) *spatial clips* that appear around the Search Cube interface and can be watched or *jumped into* to reveal them in 1:1 scale. The toolbar below the search cube lets the user switch between clips on the same page, reveal new pages of clips, hide the clips, and clear the search cube.

trends change rapidly, with styles gaining and losing favor. She first refreshes her memory by creating a few simple living room designs with the new VR spatial design software used at the firm. There are many well known designers at her firm who specialize in different design aesthetics (including modern, coastal, Scandinavian, and Asian Zen). Their workflows for achieving renditions of their specialized styles are captured as spatial design recordings inside the software. After refreshing her skills, Juanita opens Tesseract and is greeted with a large, constantly evolving database of recordings that she can consume, each labelled by designers proficient in different styles. Wanting to learn the Asian Zen style, she places a Tatami mat in the search cube and performs *object search* (Figure 3C, left). Tesseract returns all the moments that contain the Tatami mat across recordings. She watches some of these moments and realizes that most designers start with the floor and work their way upwards, adding low furniture with simple colors. She adds a few pieces of similar furniture and performs *viewpoint search* with virtual *recorded viewpoints* looking at the design (Figure 3C, middle). From the retrieved moments, she learns that designers usually teleport and look at the layout from different angles to ensure the design is harmonious and balanced from different viewpoints to achieve a sense of tranquility. Juanita *jumps* into one recording that clearly shows the positions in the room where perspectives matter the most (Figure 3C, right). She watches it back in 1:1 scale and gains a strong understanding of the workflow that can be used to achieve this style.

6 IMPLEMENTATION

We built an application to illustrate how Tesseract can be used to query spatial design recordings contextualized in interior design. The application includes a database of 3D models ranging from kitchen appliances to living room furniture. The models represent the *object* component of spatial design recordings. The application supports features such as the addition, deletion, and arrangement of objects. All objects have pre-defined searchable tags such as “stove” or “cabinet” to support variations of the same object (by style or color). Tags are also used to support the comparison of objects when querying (such as when using *object search*). The application was used to create spatial design recordings that participants queried when trying out Tesseract.

Spatial design recordings are created by combining data generated during application use in the form of JSON files, raw audio files, and transcripts of these audio files (using Amazon Transcribe [66]). The JSON files contain movement information about recorded people and objects, as well as in-application events (such as adding or deleting an object). The VR-based Search Cube interface is implemented in Unity and runs on the Oculus Quest 2 headset. Users can navigate the virtual environment and interact with objects using VR controllers. Pressing a button toggles the search cube's visibility. A toolbar underneath the search cube allows users to access various querying tools and features that Tesseract supports including *viewpoint search*, *proximity search*, and *voice search*, copying the current virtual environment into the search cube, performing a query, toggling query results, browsing individual query results and pages of results (after query results populate), and clearing the cube (Figure 4, Right).

The four proposed querying tools were implemented within the interior design application. To enable *object search*, users can add objects from an inventory or the current environment by pointing and selecting objects with the trigger button. Positioning the pointer inside the search cube drops the selected objects, which can then be re-positioned with the pointer, rotated using the thumbstick,

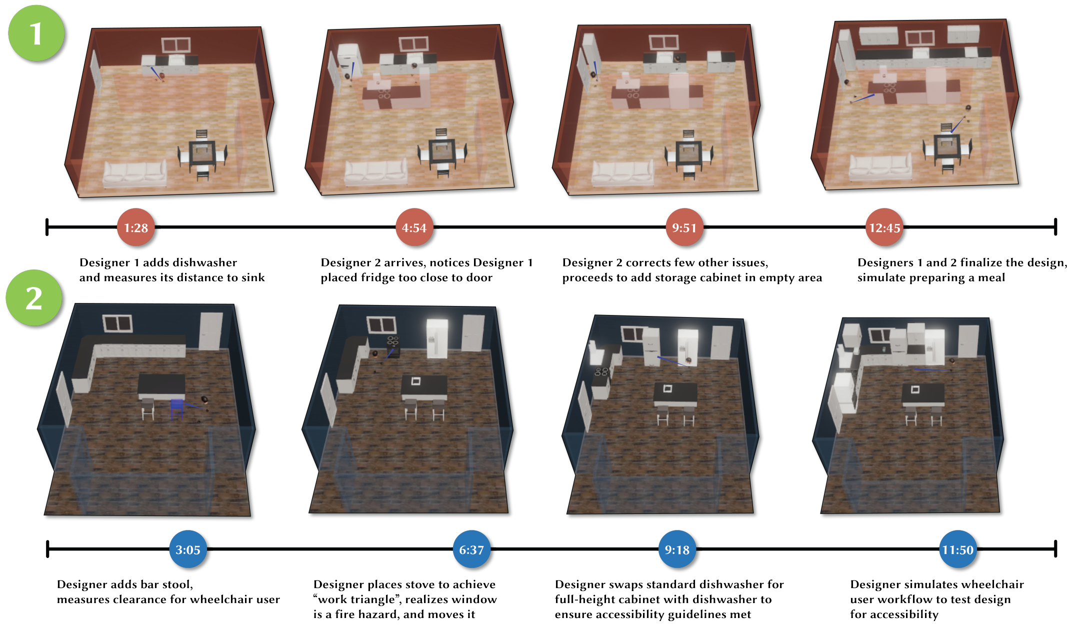

Figure 5: An overview of the spatial design recordings created to help evaluate Tesseract where: (1) is a 14-minutes long recording capturing a designer initially working alone who is later joined by a second senior designer to discuss and review the design; and (2) is a 12-minutes long recording portraying a designer working in a think-aloud fashion to explain their design rationale while designing a kitchen for a client with accessibility requirements.

or deleted using a button. Users can press a button to reveal a drop down menu to control the object's weights. Pressing this button also visually populates object-object relationships that were automatically detected (Figure 4.1). To trigger *proximity search*, users press the proximity search button to reveal the selection disk (plus UI) and place it in the cube, using buttons to adjust its size (Figure 4.2). To trigger *viewpoint search*, users press the viewpoint search button, revealing a UI element. Then they can press buttons to add a *recorded viewpoint* or *frame* to include people and their movement in the environment (Figure 4.3 camera and frame button). They can reposition a viewpoint with the pointer and rotate it using the thumbstick. Lastly, users can press a button on the toolbar to activate *voice search*, revealing a live transcription above the search cube (Figure 4.4). Voice queries are translated to text using the Microsoft Cognitive Services Speech SDK [47].

Results retrieved from queries are called *spatial clips* with a fixed length of 15 seconds. The results are sorted by the order in which they appear in a recording. They are visualized using WiMs arranged a radial layout around the search cube (Figure 4.6). Four spatial clips are displayed at a time (a system parameter). Retrieved clips have a minimum similarity score threshold of 0.5 (a system parameter). Spatial clips are ranked by their similarity score and visualized using a number and color to indicate the score, from green (high = 1) to red (low = 0.5). Spatial clips also reveal a timestamp, a scrubber, and UI buttons. Users can toggle between the populated spatial clips using the rotation buttons or reveal a new page of spatial clips using the arrow buttons. They can *jump* into a spatial clip (with the door button above it) to reveal it in 1:1 scale and explore it further.

7 PRELIMINARY EVALUATION

We conducted a preliminary evaluation with 10 participants to understand how they use Tesseract to stage queries when spatial design recordings are available. The goals of the evaluation were to: (1) understand their usage and perceptions of each querying tool, (2) observe how they combine the querying tools, and (3) explore the benefits and drawbacks of the WiM-based Search Cube interface.

7.1 Reference Spatial Design Recordings

Two spatial design recordings were created using the prototyped interior design application (Figure 5). Each recording portrays scenarios of interior designers working on the design of a virtual kitchen by combining knowledge about kitchen design guidelines [51] and clients' requirements. Each virtual kitchen was furnished by designers using furniture such as chairs, cabinets, and kitchen appliances such as stovetops and dishwashers. The first recording (14 minutes long) captures a designer initially working alone who is later joined by a second senior designer (Figure 5.1). The second recording (12 minutes long) features a designer working in a think-aloud fashion to explain their design rationale in designing a kitchen with accessibility requirements in mind (Figure 5.2). Both recordings contain data about *objects*, *people* and *interactions*, including motion, logged events, conversations, and object manipulations.

7.2 Participants and Procedure

We recruited ten participants (8 male, 2 female) aged 22 to 38 (Mean: 29.3, Median: 27.5) through research networks (XRDRN [72]) and Reddit forum posts [59]. Participants varied in their experience using VR devices (Intermediate - 3/10, Proficient - 6/10, and Expert - 1/10). They received a $50 Amazon gift card after the study. Participants completed the study remotely using their own VR headsets (Oculus Quest 2) while the experimenter observed them through a recorded audio call. Participants completed a consent form prior to the study. Then, Tesseract was introduced through a series of application-embedded video tutorials. Participants were shown short target video clips (up to thirty five seconds long) from the previously created spatial design recordings and tasked to find it by staging queries using Tesseract's querying tools. Participants were provided 5 minutes to find the correct clip for each task. They were encouraged to think-aloud and asked to skip a task if they exceeded the 5 minute allotment.

We provided target clips for participants to find to ensure that they had the same search goal so that we could observe the variations in how they use the querying tools. Seven target clips were chosen based on scenarios where querying design activities can be useful involving: understanding design rationale (4), reflecting on designs (2), and learning new workflows (1). Two target clips assess *object search*—the first to test manipulating objects into the search cube and another to assess weights and relationships, one clip assesses *proximity search*, two clips evaluate *viewpoint search*—one to assess multiple recorded viewpoints and another to assess multiple frames, and one clip assesses *voice search*. The tasks are evaluated in sequence as the querying tools build on top of each other; for instance *viewpoint search* relies on knowing how to manipulate objects in the search cube. After individually trying each tool, participants were tasked with finding a target clip where any querying tool or combination was viable. After trying all the querying tools, participants completed a post-study questionnaire and provided thoughts on the Search Cube interface and querying tools. The evaluation lasted between 75 to 90 minutes.

7.3 Study Tasks

Participants completed 7 querying tasks in which they were asked to find a target clip (Figure 6).

Object search task 1: The designer is trying to find an appropriate location to place a two-door refrigerator.

Object search task 2: The designer is adding island countertops and bar stools and ensuring that there is adequate knee space.

Proximity search task: The designer adds a stove and wants to ensure clearance to the nearby countertops.

Viewpoint search task 1: One designer requests another designer to help them add cabinets to the left corner of the kitchen.

Viewpoint search task 2: The designer inspects the kitchen by simulating a wheelchair user to identify accessibility issues.

Voice search task: One designer identifies that the other designer has placed the refrigerator in the wrong location and violates the kitchen triangle rule in interior design. They ask them to move it to a new location.

Freeform search task: The designer adds two small counters to either side of the stovetop. After adding the counters, the designer realizes that there is not enough landing area for the stove and replaces the counters with larger counters.

8 RESULTS

All participants completed the object search task within 5 minutes. 9/10 participants completed each proximity search task, viewpoint search task, voice search task and freeform search task within 5 minutes. A visual representation of participants' staged queries per task (complete and incomplete) is shown in Figure 6. Overall, participants placed objects at positions consistent with the target clip with minor variations. In the first object search task (Figure 6A), all participants staged a fridge and a kitchen island counter, 6/10 participants staged a bar stool, 4/10 participants included a second kitchen island counter, and 2/10 included a second bar stool. 2/10 participants placed one counter and 1/10 placed more counters. In the second object search task (Figure 6B), all participants staged one kitchen island counter and bar stool each. 9/10 included a second set and 1/10 further added a third island counter. 2/10 participants added three counters and 1/10 participants added more counters.

In the proximity search task, the positions and sizes of the selection disks placed by participants varied (Figure 6C). In the first viewpoint search task, 7/10 participants staged two recorded viewpoints representing two recorded designers and the remaining (3/10) staged one recorded viewpoint representing one designer moving by using two frames (Figure 6D). In the second viewpoint search task, 8/10 participants staged three frames and the rest (2/10) staged two frames with locations and orientations varying each other (Figure 6E). In the voice search task (Figure 6F), 6/10 participants used the word "problem"and 5/10 used the word "triangle", which are the exact words that appeared in the recorded conversation in the target clip. 3/10 participants used "move the fridge" to refer to the recorded event of moving the fridge. In the freeform search task (Figure 6G), 8/10 participants used object search by placing a stove (8/10), one counter (3/10), two counters (2/10), a sink (3/10), and a fridge (4/10). The combinations used were *object-only* (2), *object + proximity* (2), *voice-only* (1), *object + voice* (1), *object + viewpoint* (1), *object + proximity + voice* (1), and *all-combined* (1).

8.1 User Feedback on Querying Tools

Participants were positive about the four querying tools in terms of ease of learning, ease of use, efficiency, results matching expectations, and overall experience (Figure 7).

Object search was well rated among the querying tools. Participants found it intuitive to use *object search*: *"I definitely felt like searching for objects, excluding some, putting some as high importance and finding a relationship felt very intuitive. I mean, we're used to that."* (P2). While the search results mostly matched participants' expectations (Figure 7D), the tool's performance is tied to the choice of objects and their placement in the search cube: *"I don't know if I was lucky, but if you choose the right objects it will be quick but I can see that going either way. Even if you chose objects you thought were relevant maybe the clip you thought was going to pop up did not."* (P7).

Proximity search was well liked due to *"the ease of specifying where the recorded person was located in the scene"* (P8). A few participants (3/10) felt it was less efficient to use (Figure 7C) and

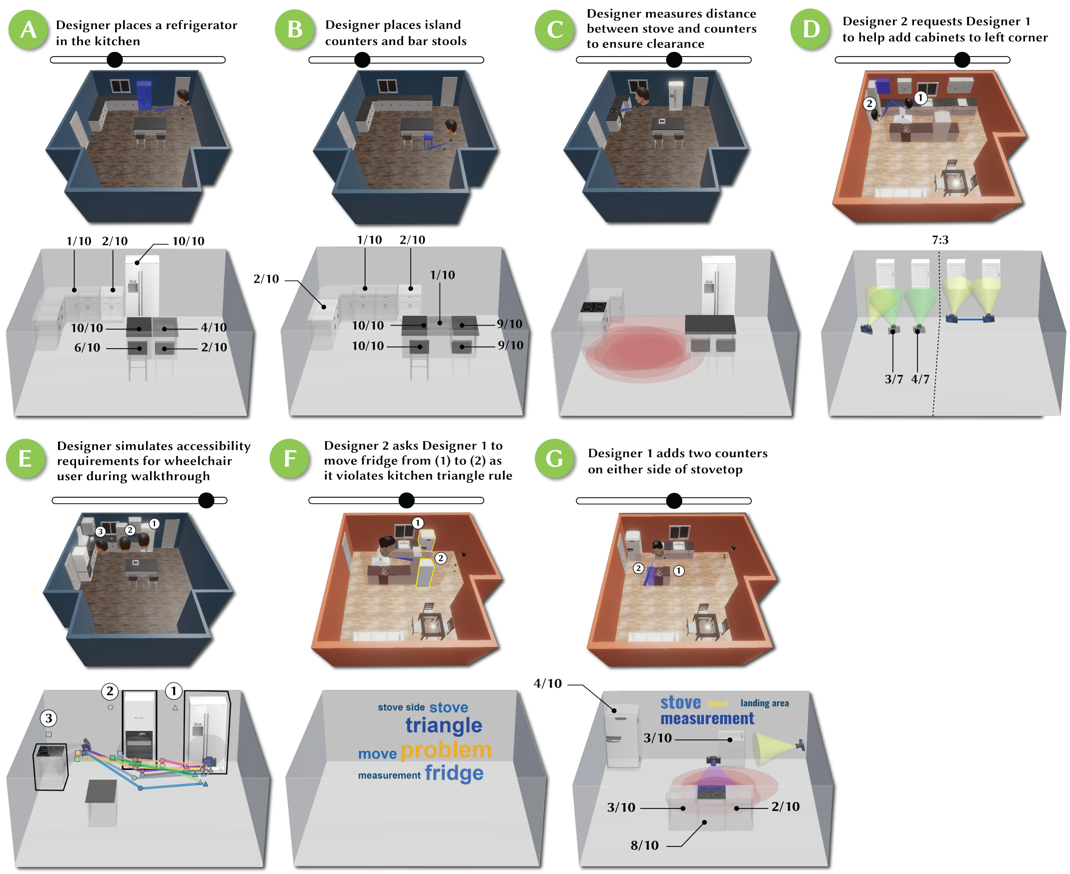

Figure 6: Target clips (Top) and aggregated results of participants' search cubes (Bottom) per task in the evaluation. (Bottom) Results showing an aggregate of the components selected by participants when staging queries. (A) Object search task 1 and (B) Object search task 2: counts of objects added by participants ; (C) Proximity search task: positions of the selection disk placed by participants; (D) Viewpoint search task 1: 7/10 participants used multiple *recorded viewpoints* denoting two recorded people, and 3/10 participants used one recorded person over multiple *frames* (denoted by the blue line); (E) Viewpoint search task 2: placement of 3 *frames* (8/10 participants) or 2 frames (2/10 participants) for a single *recorded viewpoint*; (F) Voice search task: commonly used words (denoted by size) during participants' queries; (G) Freeform search task: aggregate of the counts of objects added (*object search*), selection disk positions (*proximity search*), placements and angles of *recorded viewpoints* (*viewpoint search)* and commonly used words denoted by size (*voice search)* in participants queries.

felt that they had a poor understanding of how it worked: *"In my mind, when I dropped a circle on the ground, I was expecting that the video would probably directly go to that moment. But that did not happen."* (P1).

Viewpoint Search was well liked by most participants: *"With viewpoint search, you can say, in this clip the recorded person is in this orientation."* (P7). While most participants found it efficient to use (Figure 7C), 3/10 participants found it sometimes difficult to precisely manipulate the viewpoint within the search cube. P5 remarked that *viewpoint search* required two steps to perform a query (adding objects to the cube and adding a recorded viewpoint).

Voice Search was received most positively since it was easy to use (Figure 7B). P1 stated: *"It felt natural because I was listening to audio, so it made sense to trigger keywords or say a sentence"*. While most participants found it easy to learn (Figure 7A), some participants (3/10) found it difficult to understand how it worked: *"I was not sure what it was searching for. So, it was a bit opaque to me. I could not tell if it looked for all the words or some and if it found them eventually."* (P2).

8.2 Feedback on the WiM-based Search Cube

Participants found it intuitive to stage a query using the WiM-based search cube: *"I liked how that worked...getting a bird's-eye view of the entire scene and being able to break down what you're looking for."* (P6). While WiMs could make it more intuitive to stage queries, participants stated that manipulating small objects in WiMs can be a challenge: *"It is a great way to perform (queries) in VR. But, it is also very hard to manipulate objects in VR"* (P5). 3/10 participants found it cumbersome to manipulate a recorded viewpoint inside the search cube when using *viewpoint search*.

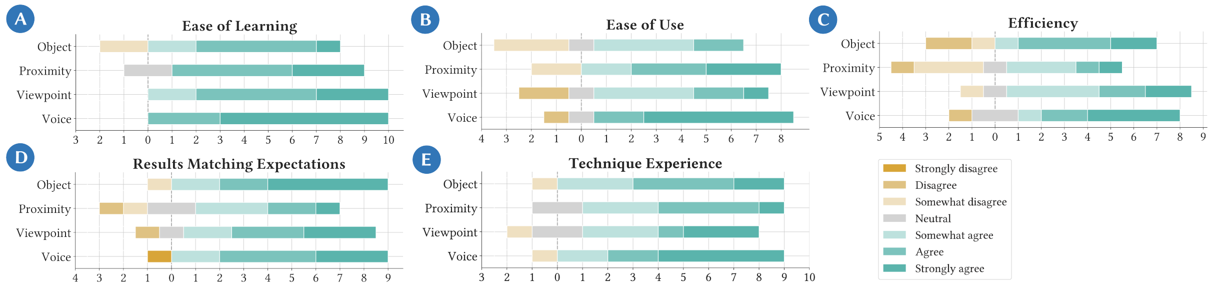

Figure 7: 7-point Likert scale responses to participants' perception of the four proposed querying tools in terms of (A) ease of learning, (B) ease of use, (C) efficiency, (D) results matching expectations, and (E) overall experience.

9 DISCUSSION

From the evaluation results, we discuss the high level usage patterns that participants exhibited while trying Tesseract. We also revisit the extent to which Tesseract meets the design goals that drove its development. We then discuss how future querying tools can build on Tesseract's strengths and weaknesses.

9.1 Tesseract Querying Patterns

We observed and summarize interesting usage patterns that participants exhibited while completing the search tasks. Particularly, we were interested in understanding how participants expressed their search intent using the querying tools, including what objects they used, how they arranged them, and their workflows to find a target clip.

In-Focus vs Peripheral Objects: Most participants included objects that were in focus of the recorded designers (that they were interacting with). For example, all participants added a fridge and bar stool in object search task 1 (Figure 6A) and a bar stool plus island counter in object search task 2 where the recorded designer placed the corresponding objects (Figure 6B). In the freeform search task where the recorded designer measured the distance to the stove (Figure 6G), 8/10 participants included a stove. However, participants also included peripheral objects not mentioned or interacted with. For instance, the island counter in object search task 1 and the fridge in the freeform search task were not interacted with, but 5/10 participants who used object search in the freeform search task and all participants (10/10) in object search task 1 included these objects in their queries. An explanation is these objects may serve as spatial anchors that help describe the moment to be found. If in-focus objects represent the activity, then peripheral objects represent the context. Both object types contribute in finding the desired moment from the recordings. Hence, we suggest that future querying tools account for objects belonging to both the design activity and spatial context more explicitly.

Precise vs Coarse Querying: We found that 1/10 participants replicated the exact layout in the target clip, but most participants (9/10) created partial layouts which typically included three or four objects even when the target clip contained more (Figure 6A, B and G). Participants appeared to add more objects only if their initial query did not yield the correct clip. However, when placing the objects, most participants (9/10) tended to be precise in defining their relationships (e.g., two objects across from each other). A similar trend was observed when they placed recorded viewpoints in the viewpoint search tasks and the selection disks in the proximity search task. Overall, they tended to be *coarse* when selecting the number of objects to include, and *precise* in their placement of these objects in the search cube.

Iterative Query Refinement: We observed that participants often employed the strategy of iterative querying—they refined their queries based on the search results, sometimes by swapping querying tools. For instance, in the freeform search task (Figure 6G), P4 began with a simple *voice search* but was not able to locate the correct clip since this phrase was too common. They refined their query by adding appliances (*object search*) and a recorded viewpoint looking at one of the appliances (*viewpoint search*), which revealed the correct clip. This is similar to 2D text-based search, where users rephrase their query or change keywords when they do not find the desired result. Our proposed querying tools offer different ways for users to express their search intent and can be used in an iterative manner to refine search results through multiple queries.

Browsing Search Results vs Running New Queries: After performing a query, on average, one half of the participants (5/10) re-ran their query if the target clip was not visible on the first page while the other half (5/10) browsed several pages and only updated their query after failing to find the target clip; past work on web search suggests people tend to stop after viewing the first ten search results [31]. Participants commented that they used the search rank and color next to each spatial clip to gauge whether to keep browsing results: *"I think I did prefer to just browse the results because a lot of the time it was like...all of the results were green (good), so I just kind of wanted to keep going."* (P7).

Participants' usage patterns suggest that spatial querying tools should provide flexibility and speed in searching. Tesseract is designed around this philosophy—users can quickly update their queries using the search cube and switch between or combine querying tools. The search results, radially presented as spatial clips with metadata (such as rank and result quality), also assist the quick exploration and comparison of results: *"I noticed that in every search, I looked at all four videos (spatial clips) on the screen and I evaluated them relative to each other."* (P6).

9.2 Revisiting Design Goals

We aimed to fulfill three design goals when designing and implementing Tesseract. Overall, we found that Tesseract meets these goals, and provides an expressive querying interface for users to search spatial design recordings. Here, we revisit the design goals to discuss the opportunities and challenges for designing future querying tools.

DG1. Provide a Centralized Space to Form Queries: The Search Cube interface provides a centralized container through which users can specify what they wish to search for. Since it is represented as a WiM, it takes up minimal visual real estate and does not disrupt the user's current workflow. Designers can run spatial design applications in VR and bring up the Search Cube interface to search for workflows and inspirations from past recordings. Consistent with prior work [33], the WiM-based approach was found to be efficient: *"I think the miniatures are much better than one-to-one scale because in miniatures you can more precisely judge the relations between items. To construct this sort of spatial query in one-to-one scale is quite challenging." (P8).* Most participants (7/10) echoed that having a centralized search cube provided them a good overview of the query being staged, providing support that this design goal is met. However, despite the positive feedback, a challenge with the WiM-based approach is that placing small objects precisely can be hard with VR controllers, which are designed for life-scale interactions. We expect this problem to be alleviated as VR input modalities based on hand and object tracking improve over time.

DG2. Provide Expressive Querying Tools to Convey Search Intent: Tesseract provides four querying tools for users to express their search intent. In the freeform task, participants used different querying tools (or combinations) and were able to find the target clip. Participants expressed that having many querying tools allowed them to pick the one (or combination) that best matched their needs: *"When you're trying to learn something from a video or VR...people could feel that something is more memorable than others at different points. So sometimes we might need to use object proximity. Sometimes we might want to use voice if there was conversation instead. So I think providing different options like this was a very good idea."* (P9). The proposed querying tools can increase user expressivity and support the retrieval of spatiotemporal data across different scenarios of design activities (*reflection*, *learning workflows*, and *understanding design rationale*) as participants showed by retrieving these clips in the evaluation. In some cases, such as with *voice search*, queries can be highly expressive (through natural language) but also vague when used alone. For example, one participant in the freeform search task (Figure 6G) used an inaccurate word "island" when referring to a kitchen counter (Figure 6G). Another participant used the term "stove side" when referring to a direction from the fridge to the stove in the voice search task (Figure 6F). Combining *voice search* with other querying tools (such as *object search*) can reduce ambiguity by letting users place objects and express complex relationships between them. Taken together, the proposed querying tools can increase user expressivity but also lead to challenges for the system as it needs to accurately detect their search intent. Future work can further explore how querying tools can be combined to retrieve moments in different design scenarios.

DG3. Support the Retrieval of Moments with Different Data Types: The proposed querying tools were well received by participants and provide a clear mapping between their input affordances and the underlying data retrieved, achieving a good *expressive match* [54]. For instance, *object search* lets users define complex layouts where 3D models have specific weights and object-object relationships. Similarly, *proximity search* allows users to position recorded users next to objects or layouts of objects which happens frequently during design activities. Despite this, we acknowledge that the mapping between the proposed querying tools and the underlying data can be further strengthened. For instance, *viewpoint search* in its current state only allows for the discrete placement of recorded viewpoints to describe what recorded designers saw, but does not support scenarios where there are continuous changes in their movement over a short period. Future work can integrate *viewpoint search* with existing sketch-based querying interfaces [9] to support querying human movement with complex trajectories. For *object search*, there are also other object-object relationships that can be used to describe complex layouts of objects (e.g., objects on-top of or inside each other) which future work can support.

9.3 Limitations and Future Work

Despite the promise of Tesseract and the associated querying tools, there are some limitations. These challenges are more evident with scale such as when the dataset includes many recordings. The challenges include: (1) the ability of the underlying heuristics to capture the nuances of the user's search intent, (2) how quickly results can be retrieved, and (3) how pertinent search results can be visualized.

Though participants found that the querying tools mostly gave them accurate results after staging a query, the underlying heuristics were simple. Since we used relative positions between objects with a pre-defined forward direction, objects with an arbitrary forward direction such as a symmetric vase are not handled. Applying existing graph kernel-based approaches [15, 16] can improve the heuristics and help to deal with more complex scenes. Since Tesseract's focus is on retrieving designer activities in spatial recordings, leveraging interior design guidelines in addition to the current heuristics could be beneficial; for instance recognizing whether a layout achieves visual balance or supports conversation through the way the seating is arranged [46] could produce fewer but more targeted search results. Further, moving towards recognizing continuous movements and viewpoint changes of designers rather than the current approach where viewpoints are considered discretely could also increase search accuracy. Pre-processing to prepare feature vectors of the recordings a priori [30, 32] rather than querying frame-by-frame could enable faster retrieval of search results. Lastly, instead of hand-crafted heuristics, data-driven methods that leverage deep learning could also be pursued to aid searching. In the current implementation, four spatial clips represent one "page" of search results (akin to a Google search) so if there are many relevant search results, the user needs to browse many pages. To circumvent this, through iterative querying, the user can refine their search intent until the most relevant results appear, reducing the need to browse many pages. However, it would be interesting explore alternative visualizations to visualize a larger number of relevant results instead of four results as in the current prototype.

In the preliminary evaluation participants were given target clips to search for and in most cases (except the freeform search task) the querying tool to employ. However, real world scenarios could involve more freeform usage of the search cube. This would also enhance our understanding of which querying tools work best in combination (though we included a freeform search task). Although the current mapping between the querying tools and data is strong **(DG3)**, there is much more data in spatial design recordings that could be exploited to support querying each component: *objects* can be expanded to include other assets beyond 3D models (such as colors, textures, or animations), data about *people* can include their non-verbal cues such as posture and gestures, and more tool *interactions* can be supported. In addition to incorporating other sources of data, the querying tools can also be enhanced. For instance, proximity search currently does not consider remote interactions such as when users teleport or when they use raycasts to move objects around (though these events can be queried through *voice search*). Hence, future renditions of the *proximity search* could support querying remote interactions with layouts rather than expecting users to be in physical proximity to objects.

We have demonstrated the WiM-based search cube to be an effective approach to find interior design activities in past spatial design recordings. However, spatial design also encompasses other domains such as architecture [2] and automotive design [52]. We think Tesseract could reasonably scale to these domains given their similarities such as the presence of multiple users, their movements, conversations and events logged while using the design software. However, there are some differences such as the scale of the objects being manipulated and the design environment. For instance, designing an entire city block occurs at a larger physical scale than designing a kitchen which the current Search Cube interface does not support. However, it could be easily extended to let users toggle the levels of detail of the search cube to aid staging. Future work is required to explore how Tesseract can support querying the more general class of *virtual capture* spatial recordings.

10 CONCLUSION

New VR tools have unlocked the potential of collaborative spatial design which can be captured as multimodal spatial design recordings and later re-watched. We have presented Tesseract, a novel system that enables users to retrieve moments in past design activities using a Worlds-in-Miniature-based Search Cube interface and four querying tools. The results of a preliminary evaluation show that participants found Tesseract easy to use and were able to stage queries and retrieve moments from past design activities. We hope that Tesseract inspires future work into building querying tools for spatial design recordings as well as spatial recordings.

ACKNOWLEDGMENTS

This research was supported in part by the National Sciences and Engineering Research Council of Canada (NSERC) under Grant IRCPJ 545100 - 18.

REFERENCES

[1] 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 (UIST '13)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/2501988.2502045

[2] Arkio. 2022. Collaborative Spatial Design. Retrieved 2022-09-09 from https://www.arkio.is/

[3] Lawrence Bergman, Vittorio Castelli, Tessa Lau, and Daniel Oblinger. 2005. DocWizards: a system for authoring follow-me documentation wizards. In *Proceedings of the 18th ACM symposium on User interface software and technology*. ACM Press, 191–200. https://doi.org/10.1145/1095034.1095067

[4] Sourav S Bhowmick and Byron Choi. 2022. Data-driven Visual Query Interfaces for Graphs: Past, Present, and (Near) Future. In *Proceedings of the 2022 International Conference on Management of Data*. 2441–2447. https://doi.org/10.1145/3514221.3522562

[5] Stefan Bruckner and Torsten Möller. 2010. Result-driven exploration of simulation parameter spaces for visual effects design. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1468–1476.

[6] Frederik Brudy, Suppachai Suwanwatcharachat, 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*. ACM, Tokyo Japan, 61–72. https://doi.org/10.1145/3279778.3279795

[7] Benjamin Bustos, Daniel A. Keim, Dietmar Saupe, Tobias Schreck, and Dejan V. Vranić. 2005. Feature-based similarity search in 3D object databases. *Comput. Surveys* 37, 4 (Dec. 2005), 345–387. https://doi.org/10.1145/1118890.1118893

[8] 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*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445651

[9] Min-Wen Chao, Chao-Hung Lin, Jackie Assa, and Tong-Yee Lee. 2011. Human motion retrieval from hand-drawn sketch. *IEEE Transactions on Visualization and Computer Graphics* 18, 5 (2011), 729–740. https://doi.org/10.1109/TVCG.2011.53

[10] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. 2019. Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. https://doi.org/10.1145/3359142

[11] 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*. ACM, Virtual Event USA, 196–208. https://doi.org/10.1145/3379337.3415591

[12] Kurtis Danyluk, Barrett Ens, Bernhard Jenny, and Wesley Willett. 2021. A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445098

[13] Fang Wang, Le Kang, and Yi Li. 2015. Sketch-based 3D shape retrieval using Convolutional Neural Networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Boston, MA, USA, 1875–1883. https://doi.org/10.1109/CVPR.2015.7298797

[14] Andreas Rene Fender and Christian Holz. 2022. Causality-preserving Asynchronous Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–15.

[15] Matthew Fisher, Daniel Ritchie, Manolis Savva, Thomas Funkhouser, and Pat Hanrahan. 2012. Example-based synthesis of 3D object arrangements. *ACM Transactions on Graphics* 31, 6 (Nov. 2012), 1–11. https://doi.org/10.1145/2366145.2366154

[16] Matthew Fisher, Manolis Savva, and Pat Hanrahan. 2011. Characterizing structural relationships in scenes using graph kernels. In *In ACM TOG*. https://doi.org/10.1145/1964921.1964929

[17] Thomas Funkhouser, Patrick Min, Michael Kazhdan, Joyce Chen, Alex Halderman, David Dobkin, and David Jacobs. 2003. A search engine for 3D models. *ACM Transactions on Graphics* 22, 1 (Jan. 2003), 83–105. https://doi.org/10.1145/588272.588279

[18] Epic Games. 2022. Fortnite Battle Royale Replay System. Retrieved 2022-09-09 from https://www.epicgames.com/fortnite/en-US/news/fortnite-battle-royale-replay-system

[19] Daniele Giunchi, Stuart James, and Anthony Steed. 2018. 3D sketching for interactive model retrieval in virtual reality. In *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering (Expressive '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3229147.3229166

[20] Purvi Goel and Doug L James. 2022. Unified many-worlds browsing of arbitrary physics-based animations. *ACM Transactions on Graphics (TOG)* 41, 4 (2022),

1–15.

[21] Corey Goldfeder and Peter Allen. 2008. Autotagging to improve text search for 3d models. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries (JCDL '08)*. Association for Computing Machinery, New York, NY, USA, 355–358. https://doi.org/10.1145/1378889.1378950

[22] Alexander Grabner, Peter M. Roth, and Vincent Lepetit. 2018. 3D Pose Estimation and 3D Model Retrieval for Objects in the Wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Salt Lake City, UT, USA, 3022–3031. https://doi.org/10.1109/CVPR.2018.00319

[23] Alexander Grabner, Peter M. Roth, and Vincent Lepetit. 2019. Location Field Descriptors: Single Image 3D Model Retrieval in the Wild. In *2019 International Conference on 3D Vision (3DV)*. 583–593. https://doi.org/10.1109/3DV.2019.00070 ISSN: 2475-7888.

[24] Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2010. Chronicle: capture, exploration, and playback of document workflow histories. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, New York New York USA, 143–152. https://doi.org/10.1145/1866029.1866054

[25] François Guimbretiére, Morgan Dixon, and Ken Hinckley. 2007. ExperiScope: an analysis tool for interaction data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1333–1342. https://doi.org/10.1145/1240624.1240826

[26] Devamardeep Hayatpur, Haijun Xia, and Daniel Wigdor. 2020. Datahop: Spatial data exploration in virtual reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 818–828.

[27] Christian Holz and Andrew Wilson. 2011. Data miming: inferring spatial object descriptions from human gesture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 811–820. https://doi.org/10.1145/1978942.1979060

[28] Sebastian Hubenschmid. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. *Proceedings of the 2022 CHI Conference on Human Factors in Computing System* (2022), 21. https://doi.org/10.1145/3491102.3517550

[29] Takeo Igarashi, Satoshi Matsuoka, and Hidehiko Tanaka. 2006. Teddy: a sketching interface for 3D freeform design. In *ACM SIGGRAPH 2006 Courses (SIGGRAPH '06)*. Association for Computing Machinery, New York, NY, USA, 11–es. https://doi.org/10.1145/1185657.1185772

[30] Ahmet Iscen, Teddy Furon, Vincent Gripon, Michael Rabbat, and Hervé Jégou. 2017. Memory vectors for similarity search in high-dimensional spaces. *IEEE transactions on big data* 4, 1 (2017), 65–77.

[31] Bernard J Jansen and Udo Pooch. 2001. A review of web searching studies and a framework for future research. *journal of the American Society for Information Science and Technology* 52, 3 (2001), 235–246. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1607>3.0.CO;2-F

[32] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data* 7, 3 (2019), 535–547.

[33] Hyo Jeong Kang, Jung-hye Shin, and Kevin Ponto. 2020. A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 275–284. https://doi.org/10.1109/VR46266.2020.00047

[34] Lyndon Kennedy, Shih-Fu Chang, and Apostol Natsev. 2008. Query-adaptive fusion for multimodal search. *Proc. IEEE* 96, 4 (2008), 567–588. https://doi.org/10.1109/JPROC.2008.916345

[35] Tae Soo Kim, Seungsu Kim, Yoonseo Choi, and Juho Kim. 2021. Winder: Linking Speech and Visual Objects to Support Communication in Asynchronous Collaboration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445686

[36] 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, 10-12 (Oct. 2020), 1937–1949. https://doi.org/10.1007/s00371-020-01942-1

[37] Robert M. Krauss, Yihsiu Chen, and Purnima Chawla. 1996. Nonverbal Behavior and Nonverbal Communication: What do Conversational Hand Gestures Tell Us? In *Advances in Experimental Social Psychology*. Vol. 28. Elsevier, 389–450. https://doi.org/10.1016/S0065-2601(08)60241-5

[38] Rebecca Krosnick, Fraser Anderson, Justin Matejka, Steve Oney, Walter S. Lasecki, Tovi Grossman, and George Fitzmaurice. 2021. Think-Aloud Computing: Supporting Rich and Low-Effort Knowledge Capture. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–13. https://doi.org/10.1145/3411764.3445066

[39] Bokyung Lee, Michael Lee, Pan Zhang, Alexander Tessier, and Azam Khan. 2019. An Empirical Study of How Socio-Spatial Formations are Influenced by Interior Elements and Displays in an Office Context. *Proceedings of the ACM on Human-Computer Interaction* 3 (Nov. 2019), 1–26. https://doi.org/10.1145/3359160

[40] Klemen Lilija, Henning Pohl, and Kasper Hornbæk. 2020. Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376604

[41] Yifan Liu, John I Messner, and Robert M Leicht. 2018. A process model for usability and maintainability design reviews. *Architectural Engineering and Design Management* 14, 6 (2018), 457–469. https://doi.org/10.1080/17452007.2018.1512042

[42] Maria-Jesús Lobo, Christophe Hurter, and Pourang Irani. 2020. Flex-ER: A Platform to Evaluate Interaction Techniques for Immersive Visualizations. *Proceedings of the ACM on Human-Computer Interaction* 4 (Nov. 2020), 1–20. https://doi.org/10.1145/3427323

[43] Zhaoliang Lun, Matheus Gadelha, Evangelos Kalogerakis, Subhransu Maji, and Rui Wang. 2017. 3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks. In *2017 International Conference on 3D Vision (3DV)*. 67–77. https://doi.org/10.1109/3DV.2017.00018 ISSN: 2475-7888.

[44] Ling Luo, Yulia Gryaditskaya, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. 2020. Towards 3D VR-Sketch to 3D Shape Retrieval. In *2020 International Conference on 3D Vision (3DV)*. IEEE, Fukuoka, Japan, 81–90. https://doi.org/10.1109/3DV50981.2020.00018

[45] Paul Merrell, Eric Schkufza, Zeyang Li, Maneesh Agrawala, and Vladlen Koltun. 2011. Interactive furniture layout using interior design guidelines. In *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*. ACM Press, Vancouver, British Columbia, Canada, 1. https://doi.org/10.1145/1964921.1964982

[46] Paul Merrell, Eric Schkufza, Zeyang Li, Maneesh Agrawala, and Vladlen Koltun. 2011. Interactive furniture layout using interior design guidelines. *ACM transactions on graphics (TOG)* 30, 4 (2011), 1–10.

[47] Microsoft. 2022. Microsoft Azure. Retrieved 2022-09-09 from https://github.com/Azure-Samples/cognitive-services-speech-sdk

[48] Alessandro Mulloni, Hartmut Seichter, and Dieter Schmalstieg. 2012. Indoor navigation with mixed reality world-in-miniature views and sparse localization on mobile devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*. ACM Press, Capri Island, Italy, 212. https://doi.org/10.1145/2254556.2254595

[49] Jung Who Nam, Krista McCullough, Joshua Tveite, Maria Molina Espinosa, Charles H. Perry, Barry T. Wilson, and Daniel F. Keefe. 2019. Worlds-in-Wedges: Combining Worlds-in-Miniature and Portals to Support Comparative Immersive Visualization of Forestry Data. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 747–755. https://doi.org/10.1109/VR.2019.8797871

[50] Michael Nebeling, Maximilian Speicher, Xizi Wang, Shwetha Rajaram, Brian D. Hall, Zijian Xie, Alexander R. E. Raistrick, Michelle Aebersold, Edward G. Happ, Jiayin Wang, Yanan Sun, Lotus Zhang, Leah E. Ramsier, and Rhea Kulkarni. 2020. MRAT: The Mixed Reality Analytics Toolkit. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376330

[51] NKBA. 2022. NKBA Kitchen & Bath Planning Guidelines. Retrieved 2022-09-09 from https://nkba.org/kitchen-bath-planning-guidelines/

[52] NVIDIA. 2022. NVIDIA Virtual Reality Capture and Replay. Retrieved 2022-09-09 from https://developer.nvidia.com/vcr-early-access

[53] Ryutarou Ohbuchi, Takahiro Minamitani, and Tsuyoshi Takei. 2005. Shape-similarity search of 3D models by using enhanced shape functions. *International Journal of Computer Applications in Technology* 23, 2-4 (Jan. 2005), 70–85. https://doi.org/10.1504/IJCAT.2005.006466 Publisher: Inderscience Publishers.

[54] Dan R Olsen Jr. 2007. Evaluating user interface systems research. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 251–258. https://doi.org/10.1145/1294211.1294256

[55] Randy Pausch, Tommy Burnette, Dan Brockway, and Michael E. Weiblen. 1995. Navigation and locomotion in virtual worlds via flight into hand-held miniatures. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*. ACM Press, Not Known, 399–400. https://doi.org/10.1145/218380.218495

[56] Henning Pohl, Klemen Lilija, Jess McIntosh, and Kasper Hornbæk. 2021. Poros: Configurable Proxies for Distant Interactions in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–12. https://doi.org/10.1145/3411764.3445685

[57] Jiantao Pu, Kuiyang Lou, and Karthik Ramani. 2005. A 2D Sketch-Based User Interface for 3D CAD Model Retrieval. *Computer-Aided Design and Applications* 2, 6 (Jan. 2005), 717–725. https://doi.org/10.1080/16864360.2005.10738335 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/16864360.2005.10738335.

[58] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517665

[59] Reddit. 2022. Reddit - Dive into anything. Retrieved 2022-09-09 from https://www.reddit.com/r/OculusQuest/

[60] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. http://arxiv.org/abs/1908.10084

[61] Donald A Schön. 1987. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions.* Jossey-Bass.

[62] Evan Shimizu, Matthew Fisher, Sylvain Paris, James McCann, and Kayvon Fatahalian. 2020. Design Adjectives: A Framework for Interactive Model-Guided Exploration of Parameterized Design Spaces. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 261–278.

[63] Second Spectrum. 2022. The Next Wave of Seeing Sports. Retrieved 2022-09-09 from https://www.secondspectrum.com/index.html

[64] Evgeny Stemasov, Tobias Wagner, Jan Gugenheimer, and Enrico Rukzio. 2022. ShapeFindAR: Exploring In-Situ Spatial Search for Physical Artifact Retrieval using Mixed Reality. *Proceedings of the 2022 CHI Conference on Human Factors in Computing System* (2022), 12. https://doi.org/10.1145/3491102.3517682

[65] Ivona Tautkute, Tomasz Trzciński, Aleksander P Skorupa, Łukasz Brocki, and Krzysztof Marasek. 2019. Deepstyle: Multimodal search engine for fashion and interior design. *IEEE Access* 7 (2019), 84613–84628. https://doi.org/10.1109/ACCESS.2019.2923552

[66] Amazon Transcribe. 2022. Automatically convert speech to text. Retrieved 2022-09-09 from https://aws.amazon.com/transcribe/

[67] Daniel Vlasic, Rolf Adelsberger, Giovanni Vannucci, John Barnwell, Markus Gross, Wojciech Matusik, and Jovan Popović. 2007. Practical motion capture in everyday surroundings. *ACM transactions on graphics (TOG)* 26, 3 (2007), 35–es. https://doi.org/10.1145/1276377.1276421

[68] Cheng Yao Wang, Mose Sakashita, Upol Ehsan, Jingjin Li, and Andrea Stevenson Won. 2020. Again, Together: Socially Reliving Virtual Reality Experiences When Separated. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376642

[69] 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*. ACM, Yokohama Japan, 1–12. https://doi.org/10.1145/3411764.3445552

[70] The Wild. 2022. VR Collaboration for the Building Industry. Retrieved 2022-09-09 from https://thewild.com/

[71] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling fluid individual and collaborative editing in virtual reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 853–866.

[72] XRDRN. 2022. Help XR research reach the world. Retrieved 2022-09-09 from https://www.xrdrn.org/

[73] Sang Min Yoon, Maximilian Scherer, Tobias Schreck, and Arjan Kuijper. 2010. Sketch-based 3D model retrieval using diffusion tensor fields of suggestive contours. In *Proceedings of the international conference on Multimedia - MM '10*. ACM Press, Firenze, Italy, 193. https://doi.org/10.1145/1873951.1873961