Cartography M.Sc.

Master thesis

Using Information Chunking for Spatial Learning based on Augmented Reality

Indoor spatial information organization and visualization based on chunking method

Jiongyan Zhang

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITY OF TWENTE.

2023

Using Information Chunking for Spatial Learning based on Augmented Reality

Indoor spatial information organization and visualization based on chunking method

submitted for the academic degree of Master of Science (M.Sc.)
conducted at School of Design and Engineering
Technical University of Munich

Author:	Jiongyan, Zhang
Study course:	Cartography M.Sc.
Supervisor:	M.sc Shengkai Wang (TUM)
Reviewer:	Dr.rer.nat. Nikolas Prechtel (TUD)

Chair of the Thesis
Assessment Board: Prof. Dr. Liqiu Meng

Date of submission: 07.09.2023

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Using Information Chunking for Spatial Learning based on Augmented Reality"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 07.09.2023 Jiongyan, Zhang

Acknowledgement

The relentless passage of time, fleeting and transient, has marked the culmination of this brief yet seemingly endless odyssey of academic life. I am immensely gratified to have successfully concluded this chapter with the accomplishment of my graduation project, an endeavor that has enriched my understanding of AR and AR design, bestowing upon me a wealth of concepts and experiential insights.

Here, I extend my foremost gratitude to my mentor, Mr. Wang Shengkai, who, throughout the course of my graduation project, dutifully fulfilled his role as a mentor, providing selfless guidance and unwavering assistance. It is fair to assert that without his unwavering support, completing this graduation project would have been a formidable challenge.

Simultaneously, I wish to express my appreciation to my circle of friends, all of whom are undeniably exceptional talents. Although their research pursuits bore no relevance to my thesis topic, they generously offered their assistance, participating enthusiastically in my user surveys and furnishing me with invaluable data.

Lastly, I would like to extend my gratitude to my colleagues at Cruise, with whom I have spent over a year. They provided me with valuable insights for the completion of my thesis and imparted a deeper understanding of professional programming knowledge. This has greatly contributed to my successful graduation project and the commencement of my career journey towards future.

Abstract

Due to the complexity of indoor spatial structures, the acquisition of spatial information indoors has long been a challenging area of study. Virtual Augmented Reality (AR) technology has introduced novel learning approaches to this field; however, the visualization guidance of numerous virtual landmarks can lead to information overload and organizational confusion. To address this issue, this study employs a method known as 'Spatial Chunking,' aimed at restructuring the informational architecture of indoor spaces to provide improved learning cues and experiences. The Spatial Chunking method leverages the inherent orderliness of indoor spatial structures, organizing information through color-based categorical and hierarchical visualization. Following the study's completion, a controlled experiment is conducted, with results indicating that, compared to a singular emphasis on virtual landmark reinforcement, the adoption of Spatial Chunking significantly enhances the quality of spatial learning. This research offers valuable insights into refining methods for indoor spatial learning and underscores the potential applicability of Spatial Chunking in AR-based learning.

Keywords: Augmented Reality, information organization, spatial learning, sptial chunking

Contents

List of Figures

List of Tables

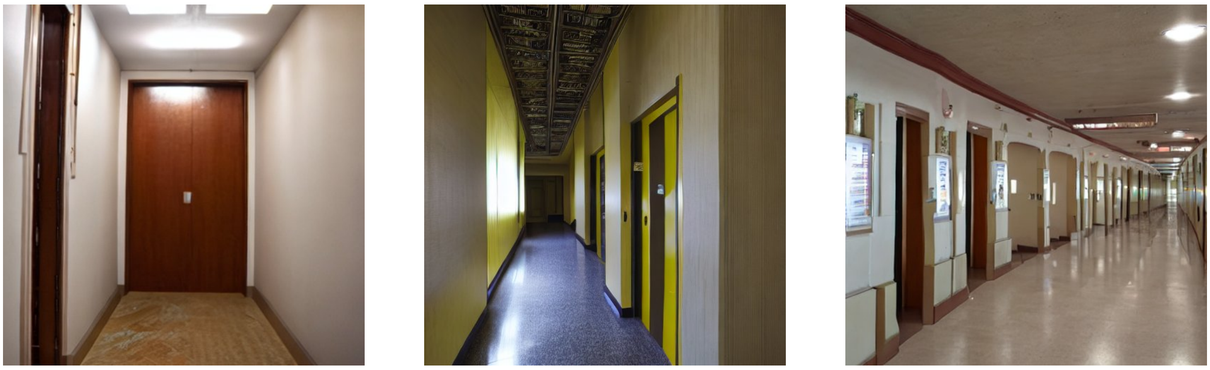
1 Introduction

1.1 Background and motivations

Essential for indoor positioning and navigation, Augmented Reality (AR) and Mixed Reality (MR) technologies have demonstrated their effectiveness in guiding users to their intended destinations (Joshi, Hiwale, Birajdar, & Gound, 2020; J. Kim & Jun, 2008; Mulloni, Seichter, & Schmalstieg, 2012). These innovative technologies seamlessly enhance wayfinding experiences by superimposing virtual landmarks and indicators onto the physical environment. However, recent investigations have prompted concerns regarding the potential drawbacks of excessive reliance on AR and virtual reality (VR) for indoor navigation. Scholars such as Gramann, Hoepner, and Karrer-Gauss (2017) and Ruginski, Creem-Regehr, Stefanucci, and Cashdan (2019) assert that an overdependence on these technologies in navigation may diminish users' awareness of their surroundings, subsequently hindering their capacity to grasp and interpret the spatial arrangement. Additionally, inadequately designed landmarks have been linked to compromised user safety (Fang, Li, & Shaw, 2015; May & Ross, 2006), and poorly crafted MR navigation systems can lead to user confusion, significantly impacting their spatial learning (B. Liu, Ding, Wang, & Meng, 2022b). While AR and VR have undeniably transformed indoor applications and human spatial navigation, addressing their potential adverse effects on spatial cognition is imperative. This is essential to prevent users from excessively relying on virtual landmarks and to cultivate their familiarity with indoor environments. In critical scenarios, users' perception and mastery of indoor space play a pivotal role, as they depend on their spatial knowledge rather than blindly following assisted navigation systems (B. Liu & Zhan, 2021). Consequently, it is vital to strike a balance between the usability of navigation systems for wayfinding and their influence on spatial learning, as underscored in recent research (Brügger, Richter, & Fabrikant, 2019; Wen, Deneka, Helton, & Billinghurst, 2014)

To tackle the constraints of human short-term memory, it is vital to consider that the capacity to remember multiple objects is inherently limited (Brener, 1940; Cowan, 2001). To enhance users' spatial learning ability, spatial chunking emerges as a viable solution (B. Liu & Zhan, 2021). Spatial chunking involves breaking down specific types of memory into smaller, related units based on certain rules, ultimately enhancing memory efficiency. Researchers posit that spatial chunking achieves this by compressing memory encoding through retrieval from long-term memory, thereby lightening the load on working memory and bolstering spatial memory performance (Thalmann, Souza, & Oberauer, 2019). The primary objective of this research is to explore whether MR-based spatial chunking assistance can effectively facilitate users' spatial learning during indoor navigation. Drawing on the principles of spatial chunking and human memory, the study aims to foster users' active spatial learning by strategically organizing visual cues within the MR system, countering their potential to become oblivious to the surrounding environment. This approach seeks to strike an optimal balance between providing navigation assistance and encouraging users' active engagement with their spatial surroundings during indoor navigation tasks. Ultimately, this endeavor aims to improve users' overall spatial learning performance and diminish their reliance on virtual landmark indicators.

In conclusion, AR and MR technologies have proven invaluable as supplementary tools for indoor positioning and navigation, significantly benefiting users in reaching their destinations. However, the potential negative effects on spatial cognition and learning require thorough consideration. Achieving an optimal balance between navigation assistance and active spatial engagement is essential to ensure that users retain their ability to perceive and learn indoor spaces effectively. Spatial chunking emerges as a promising approach to enhance users' spatial learning during indoor navigation tasks. By thoughtfully integrating spatial cues within the MR system, users can gain a deeper understanding of their environment and reduce their reliance on virtual landmarks. Continued research and implementation of AR and MR technologies will maximize their potential to facilitate indoor navigation while preserving users' spatial cognition and learning abilities.



(a) Low information (b) Middle information (c) High Information

Figure 1: Spatial information cases (Sample images come from the Stable Diffusion Model [https://stablediffusionweb.com/])

From Figure 1, we can know that the information load of indoor spaces is often imbalanced. In some cases, the targets in the current spatial region are scarce, making it difficult for users to acquire meaningful spatial knowledge. In other cases, the targets in the space are dense and complex, leading to cognitive overload and hindering effective learning. Our research aims to address this issue by exploring appropriate methods for spatial chunking, in alignment with human cognitive patterns, to enhance users' learning efficiency.

The arrangement of visual cues is a complex and broad concept. One promising approach is to employ spatial information chunking from multiple perspectives, such as color, shape, texture, and position. These chunks rely on spatial and semantic proximity, organizing information and visualizing various aspects, with unpredictable effects. This presents challenges for chunk-based AR spatial learning. Addressing how to organize, utilize, and present the correlation of indoor spatial objects is a crucial point urgently requiring attention in this paper. Spatial information chunking offers an efficient way to manage data in indoor environments. By grouping related visual cues based on their relationships, users can navigate more effectively. For instance, similar colors or shapes may represent different zones, while cues close together might indicate spatial connections. Implementing this approach requires considering factors like the perspective for chunking data. Moreover, optimizing the organization and presentation of visual cues is vital. Grouping related cues and ensuring each chunk conveys meaningful information allows for effective processing and memory. Selecting suitable cues is essential, making sure they are relevant to users' navigation goals and representative of the indoor spatial objects. Considering the impact of visual complexity is also crucial. While chunking improves memory efficiency, overly complex cues may overload users' cognitive capacity. Striking a balance between information richness and manageability is necessary. In conclusion, spatial information chunking from multiple perspectives shows promise for enhancing AR spatial learning in indoor environments. However, challenges exist in determining perspectives, organizing cues, and selecting relevant information. Addressing these aspects contributes to the advancement of chunk-based AR spatial learning for indoor navigation scenarios.

1.2 Research questions

1.2.1 Research Objectives

The primary aim of this research is to develop an MR visualization system rooted in spatial chunking principles, aimed at guiding users in comprehending the current indoor environment and enhancing their ability to retain detailed information. Spatial chunking, a well-established psychological technique, capitalizes on the limitations of human memory to expand its capacity. As per the findings of Kahn, Tan, and Beaton (1990), humans have been shown to excel at recalling information when it is organized into chunks of 7±2 elements. Furthermore, chunking serves as a valuable tool in the realm of spatial cognition and navigation, aiding individuals in effectively surmounting memory constraints and optimizing memory functions (Krukar, Anacta, & Schwering, 2020; Shank, 2018). Consequently, the primary research objective can be further subdivided into the following three specific sub-objectives:

1. The search for an effective method for spatial chunking has led researchers to explore various approaches. Many existing spatial chunking methods rely on semantics, such as direction, turning points, or functional areas (J. Sargent, Dopkins, Philbeck, & Chichka, 2010; J. Q. Sargent, Zacks, Hambrick, & Lin, 2019). While these methods possess significance and operational ease, they often struggle to balance information between chunks, limiting users' ability to fully utilize their memory capacity. Other studies have attempted to combine quantitative and qualitative methods to achieve a more systematic spatial division, aiming to help users perceive and remember more without cognitive overload. One promising approach involves using information metrics as spatial chunking standards. Researchers can directly chunk space based on semantic meaning, ensuring an even distribution of information load within each chunk. Spatial entropy (Batty, 1974), a measure of the disorder of current spatial information, and Fisher information metric based on SLAM robots (Verbelen, De Tinguy, Mazzaglia, Çatal, & Safron, 2022) are commonly used spatial metrics calculated based on information theory. However, these chunking methods may be relatively rigid, partitioning space into distinct blocks without adequately organizing the underlying relationships among objects within the space. In modern architectural design, indoor space layouts are often deliberate, exhibiting inherent correlations and hierarchies (Reddy, Chakrabarti, & Karmakar, 2012; Suh & Cho, 2021; Zhou & Liu, 2021). Therefore, our design emphasizes organically chunking objects within the current space based on different features, such as shape, color, or texture. This flexible approach aims to avoid rigid segmentation that might lead to semantic disconnections and hinder users' perception of the overall spatial structure. By considering the inherent relationships and hierarchies within the indoor environment, our proposed method seeks to enhance users' spatial cognition and memory retention during navigation tasks.

2. To effectively design visual guidance for information chunks, it's essential to consider their inherent challenges. Each information chunk holds a distinct set of data, making it challenging for individuals to intuitively grasp the spatial distribution and commit it to memory, as pointed out by Stites, Matzen, and Gastelum (2020). Consequently, the research necessitates the development of tailored visual aids. These visual aids should not only convey meaningful semantics but also align with cognitive principles regarding memory load management. Simultaneously, they should actively engage users' cognitive faculties, as user-driven cognition tends to yield more favorable memory outcomes, as highlighted by Moore and Zirnsak (2017). In cases where individuals possess superior spatial memory abilities, they may naturally focus on and remember the various virtual

landmarks within the current information chunk. However, this effect may not be as pronounced for average individuals or those with lower spatial memory capabilities. Hence, the implementation of suitable information filtering mechanisms becomes imperative to prevent information overload, as suggested by Zhao, Simpson, Wallgrün, Sajjadi, and Klippel (2020). In the context of the experiment, the design of virtual landmarks should be strategically balanced. These landmarks should captivate the user's attention while also accommodating individuals with limited spatial learning aptitude.

3. The research will include a control experiment to validate the hypothesis. Two groups will be established: one without spatial chunking and the other with spatial chunking, enabling a comparison of users' spatial memory effects. Additionally, the experiment aims to gather quantitative data through questionnaires and data analysis, supporting the final results theoretically. The experiment will begin by assessing users' spatial learning abilities, sense of direction, and familiarity with the experimental site. Subsequently, users will navigate the experimental site while using AR devices equipped with spatial learning software to enhance their learning process. Afterward, users will provide feedback on their perception of the AR system, along with the quantity and accuracy of their spatial memory. The control experiment will provide valuable insights into the impact of spatial chunking on users' spatial memory and overall navigation experience, contributing to the advancement of AR-based spatial learning systems.

1.2.2 Research Questions

Based on the research objectives mentioned above, several research questions are introduced here for further consideration and exploration:

1. Delve into the quest for an efficient method of spatial chunking to divide the area users are required to acquaint themselves with. How should this chunking process be executed? Is a strict adherence to predefined criteria necessary for the balanced segmentation of the present indoor space, or should we embrace a more intricate approach to aptly structure and classify information within the given space, ultimately leading to chunking techniques grounded in spatial information filtration? A rational strategy should be formulated for guiding the design of the AR interface system, with the aim of diminishing the information burden on users while maintaining their holistic perception of the entire space.

2. The research aims to identify an appropriate visualization approach for chunk-based learning that can effectively steer users in spatial comprehension. What is the optimal way to create visual landmarks that enhance users' cognitive processes? In mixed reality (MR) devices, virtual landmarks play a significant role in capturing users' attention, underscoring the importance of landmark design. How can we strike a balance in managing the information load experienced by users? Individuals with varying spatial memory capacities may respond differently to specific information quantities. This investigation will delve into techniques for information filtering and the development of distinct information load levels tailored to individual users.

3. What approaches should be employed to formulate experiments, assessments, and the corresponding data analysis protocols for volunteers? Following the creation of the mixed reality (MR) system geared towards spatial learning, the investigation aims to assess users' performance, seeking to ascertain whether the implementation of spatial chunking yields a beneficial impact on spatial memory. Subsequent to the development of the MR spatial

learning system, the study will assess user memory outcomes to determine whether the utilization of spatial chunking methods aids in enhancing spatial memory.

1.3 Thesis structure

The paper is divided into five chapters, aiming to elucidate the design details of our current AR indoor spatial learning system by introducing the design concepts and showcasing our design outcomes to the readers. Subsequently, we present the experimental details and post-analysis of data to demonstrate the superior indoor spatial learning effectiveness of our current design. Specifically, the chunking method assists in organizing indoor spatial information and aiding users in better memorizing various details within the indoor environment.

Chapter 1 provides an introduction to the research background and motivation, while highlighting the challenges that our study needs to address, thereby establishing the core of the entire discussion.

Chapter 2 primarily summarizes existing relevant research findings, discussing their relevance to our current study and providing references and insights that can contribute to our research. It also identifies limitations or areas requiring improvement in these approaches, offering an overall overview of the relevant field.

Chapter 3 focuses on the devices utilized in our experiments and the design principles of the current AR-assisted system. Additionally, this chapter presents the visual effects of the system to provide a more intuitive visual experience.

Chapter 4 details the experimental procedures, including the organization of the experiments, recruitment of volunteers, and data collection. Following the experimentation phase, this chapter quantitatively and qualitatively analyzes the collected data, comparing the conditions of the experimental group and the control group to determine the corresponding advantages of spatial learning.

Lastly, Chapter 5 concludes with a summary and discussion of the subsequent steps.

2 Related Work

This chapter provides a comprehensive review of relevant research findings and state-of-the-art (SOTA) methods. It begins by reviewing the support and enhancement of learning abilities through AR, shedding light on the underlying principles. The chapter then introduces chunking-related learning methods, discussing their applications in psychology and how they support the learning process. Furthermore, we discuss the current research on AR-based spatial learning and the achievements within this field, highlighting areas that can be leveraged and applied. Moreover, this chapter delves into the quantitative analysis methods of sketch maps, as sketch maps serve as important indicators for evaluating users' spatial learning abilities. However, they pose challenges in terms of quantitative analysis. Therefore, such discussions are invaluable for our experiments and research. Finally, this chapter concludes by weighing the pros and cons of various aspects of the current research, in order to provide support for our ongoing study.

2.1 AR Cartography

Integrating Augmented Reality (AR) into cartography has become a prominent research topic. By replacing conventional paper or electronic maps, AR immerses users in the real environment, combining visual cues to provide innovative navigation and spatial orientation experiences. Schmalstieg and Reitmayr (2007) explore the potential and challenges of Augmented Reality (AR) as a novel cartographic medium, along with the reciprocal impact and contributions of AR and cartography. They argue that AR has the capacity to offer users more immersive and engaging map experiences, while also requiring resolution of certain technical and social issues. They also highlight that AR can learn from cartography in terms of methods and techniques, while simultaneously presenting new opportunities for cartography as a discipline. Bobrich and Otto (2002) propose a method to enhance cartographic visualization through AR technology, which involves integrating virtual objects with real maps to offer additional dimensions and interactivity. They believe that AR can provide a novel approach for cartographic applications, enabling the presentation and manipulation of digital geospatial data on top of simulated maps. Carbonell Carrera and Bermejo Asensio (2017b) discuss the advantages of AR technology in terrain interpretation and spatial orientation, and whether AR can improve learners' spatial orientation skills. The article suggests that, with appropriate design, AR can offer a novel way of interacting with three-dimensional terrain representations, thereby aiding learners in determining their own and target locations within the environment. Dickmann, Keil, Dickmann, and Edler (2021) emphasize that the core of AR technology lies in the integration of virtual objects with real-world scenes, thereby providing users with additional spatial information. They also highlight that different AR techniques can influence the visualization and perception of AR elements in three-dimensional space. The article asserts that understanding these technological foundations is an essential prerequisite for harnessing AR in cartography. Carbonell Carrera and Bermejo Asensio (2017a) hold the view that AR technology can serve as an innovative teaching tool to enhance students' understanding and manipulation of three-dimensional terrain information. They also highlight that AR can be integrated with devices such as tablet computers, allowing users to interact with digital terrain models (DTMs) through gestures, thereby providing a novel way of engaging with terrain representations.

Research on AR in cartography is still in its infancy. Anastopoulou et al. (2023) conduct an analysis of fifteen scientific papers involving different Levels of Detail (LoD) management techniques in AR environments, considering dimensions such as data type, technology type, and user behavior. The findings indicate that the applications in this domain are still at an early stage and relatively limited in number. However, they provide a comprehensive synthesis of existing knowledge and highlight the exciting and dynamic challenges that lie ahead in future research. Indeed, research on AR presents its own challenges. Dong et al. (2021)'s findings reveal differences in participants' attention towards environmental objects when using AR compared to 2D maps. AR users exhibit less visual attention towards buildings but more towards people. Additionally, the results of route drawing show that AR users had more difficulty forming clear route memories. Similarly, Carbonell-Carrera, Jaeger, and Shipley (2018)'s study finds that two-dimensional contour maps have lower motivational impact on students compared to AR, which increases students' interest and engagement. However, students perceive that two-dimensional contour maps improve their abilities despite requiring more effort. Interestingly, in terms of spatial reasoning ability, there is no significant difference observed between two-dimensional contour maps and AR. Keil, Edler, and Dickmann (2019)'s perspective on holograms suggests that they have the potential to influence people's perception and interaction with geographic space, thereby enhancing direction, navigation, and spatial knowledge. However, the article also highlights a crucial challenge related to the early development stage of AR hardware, particularly concerning the position accuracy and stability of holograms. Addressing this issue is essential for conducting standardized spatial cognition experiments and applications in the future. As AR technology continues to advance, overcoming such technical limitations will be instrumental in fully harnessing the benefits of holographic representations for spatial cognition and navigation. These results highlight the complexities and nuances of incorporating AR into mapping and learning experiences, suggesting the need for further exploration and refinement to optimize its effectiveness in different contexts.

2.2 AR assisted learning

Augmented reality (AR) is a technology that overlays virtual information on the real environment, providing users with rich visual and interactive experiences. This section will review some related literature, analyzing the advantages and challenges of AR technology in the education field.

In recent years, AR technology has been widely applied and researched in the education field, aiming to improve learners' learning outcomes and motivation. Especially in the domain of scientific learning, AR can assist learners in comprehending abstract and complex concepts or phenomena that are difficult to observe directly. It also facilitates a better explanation of scientific content knowledge (Cheng & Tsai, 2013; Sahin & Yilmaz, 2020; Sylaiou et al., 2015). Numerous empirical studies have demonstrated that integrating AR technology into science curricula (such as physics, chemistry, earth science, biology, mathematics, etc.) can enhance students' scientific learning outcomes. For instance, it can improve content comprehension and foster interest in science (Radu, 2014). Furthermore, it can increase motivation and engagement in scientific learning (Cai, Chiang, & Wang, 2013; Diegmann, Schmidt-Kraepelin, Eynden, & Basten, 2015; Goff, Mulvey, Irvin, & Hartstone-Rose, 2018), as well as improve academic performance (Akçayır & Akçayır, 2017).

Akçayır and Akçayır (2017) lead a systematic review of AR technology in the education field, summarizing the impact of AR technology on learning, the advantages and limitations of AR technology, and the future research directions. They found that AR technology can improve learners' cognitive, affective, and behavioral performance, while also facing some technical, educational, and ethical issues that need further research and discussion. Küçük, Yỳlmaz, and Gökta (2014) explore the effect of AR technology on students' academic achievement and motivation in English course studying, using an experimental group and a control group design, comparing the learning effects of using AR technology and traditional materials. They found that the experimental group students were superior to the control group students in cognitive level, affective attitude, and memory retention, indicating that AR technology can enhance the teaching effect.

In addition to formal educational settings, informal science institutions (ISIs) such as science centers, science museums, zoos, botanical gardens, and aquariums play a crucial role as venues for utilizing AR devices to facilitate public understanding and engagement in science (Bao & Engel, 2019; Capozzi, Lorizzo, Modoni, & Sacco, 2014; Chen, Zhou, & Zhai, 2023; Yoon & Wang, 2014). A systematic review conducted by (Chen et al., 2023) examine 22 relevant studies and concluded that AR or VR, by supplementing real objects with virtual materials, can directly present abstract concepts through simulation and emulation methods. This approach can better assist users in their learning process.

In general, the enhancement of human learning abilities through AR is supported by psychological theories. AR can improve learners' engagement and interest. Multiple studies have shown that AR can stimulate learners' curiosity and desire for exploration in scientific content, thereby increasing their participation and interest in classroom or informal learning settings (Chen et al., 2023; Huang, Chen, & Chou, 2016; Radu, 2014). AR can meet learners' basic psychological needs, such as autonomy, competence, and relatedness, by providing choices, feedback, interactivity, and social support. This, in turn, enhances their intrinsic motivation and autonomous drive (Lai, Chen, & Lee, 2019). Simultaneously, AR can optimize learners' cognitive load. Cognitive Load Theory (CLT) posits that humans have limited working memory capacity, and AR technology can optimize working memory load by reducing irrelevant information or increasing relevant information, thereby enhancing cognitive efficiency and learning outcomes (Kalyuga, 2011; Q. Liu, Yu, Chen, Wang, & Xu, 2021). AR can also enhance learners' spatial abilities. Spatial abilities refer to the perception, analysis, and manipulation of spatial relationships, shapes, directions, and movements. It plays a crucial role in science education (Uttal et al., 2013). AR can assist learners in understanding spatially complex or abstract concepts or phenomena by providing three-dimensional, dynamic, and interactive visual representations (Ibáñez & Delgado-Kloos, 2018).

In summary, a specially designed AR system can effectively assist humans in various aspects of learning. This enhancement in learning abilities is attributed to the AR system's ability to effectively organize cognitive load and facilitate coordinated brain function at the psychological level. As a result, it promotes the optimal utilization of human memory and learning capabilities.

2.3 Chunking assisted learning

Chunking is a process of segmenting certain information into meaningful chunks based on specific semantic rules and psychological principles. It serves to better organize information and enhance learning efficiency. Chunking has been widely studied and applied in various learning domains and scenarios, as it can help learners overcome the limitations of short-term and working memory, enhance the organization and retrieval of information, and facilitate the acquisition and recall of complex knowledge (Gobet, 2017; Thalmann et al., 2019). In the early stage of the chunking theory, Miller (1956) suggest that human memory capacity iss limited to about seven chunks of information, and that chunking could increase the amount of information per chunk by using familiar or meaningful patterns. Later research expanded the notion of chunking and explored its mechanisms and effects in different contexts. For example, Chase and Simon (1973) find that expert chess players could recall chess positions better than novices because they could chunk the pieces into meaningful configurations based on their knowledge and experience.

Chunking primarily utilizes inherent cognitive patterns in humans and assists in memory processes from a psychological perspective. Gobet et al. (2001); Laird, Rosenbloom, and Newell (1984) propose two forms of chunking: purposeful, strategic, and controlled chunking, and automatic, continuous, and perception-related chunking. They also present evidence for chunking and introduced the EPAM/CHREST mechanism as a psychological foundation for chunking-assisted learning. Through this mechanism, chunking at different levels of memory and learning enables the effective acquisition of knowledge in various domains, such as chess, music, mathematics, and physics. Fonollosa, Neftci, and Rabinovich (2015), on the other hand, investigate the dynamical principles of chunking by employing a competitive mode dynamical model to explain the role of chunking in cognitive sequences. Their work provides more solid evidence and a theoretical foundation for chunking-based learning theories. Additionally, from a dynamical perspective, they also offer an explanation for certain mental disorders and cognitive impairments in humans. It is precisely due to these psychological theoretical foundations that chunking-based learning has been applied in various domains and has achieved certain effectiveness. For example, Buschke (1976) argue that in language learning, learners spontaneously organize their learning in a chunking manner, organizing their current knowledge based on certain chunking rules. Zhang, Ding, Stegall, and Mo (2012) also find that the chunking method can compensate for the innate limitations of visual working memory through experiments. Through appropriate chunking design and guidance, students with these limitations can overcome difficulties in learning geometric mathematical knowledge and achieve better learning outcomes.

It is evident that chunking is an effective method for assisting learning in many domains, and this method can occur spontaneously or through guidance. However, there is ongoing research and debate on how chunking is guided and enhances human learning capabilities. Thalmann et al. (2019) review the impact of chunking on working memory and its interaction with long-term memory and attention. This research indicates that chunking can enhance the capacity, duration, and encoding efficiency of working memory while reducing the risk of interference and forgetting. Additionally, this study discusses the neural basis and developmental changes associated with chunking, as well as its applications in different domains. Nassar, Helmers, and Frank (2018) propose a chunking model of visual working memory, suggesting that similar features are co-encoded through center-surround dynamics to enhance memory capacity and performance. They demonstrate the advantages and limitations of chunking through experiments and simulations, as well as individual differences. They argue that chunking reveals the flexibility and complexity of capacity limitations in visual working memory. Jones (2012) think that chunking is a better explanation for cognitive developmental changes in children than the development of short-term memory capacity or processing speed (though the help from the latter two is not denied). They contend that chunking can account for individual, age-related, and cross-cultural differences observed in children's performance across various tasks. Moreover, they highlight that chunking can aid children in constructing more complex and abstract knowledge structures while enhancing their meta-cognitive abilities. Thus, this research elucidates chunking as an inherent attribute of human development. Therefore, there are multiple interpretations regarding how chunking specifically influences humans. However, it is certain that humans may spontaneously employ chunking techniques during the learning process. Furthermore, in educational and instructional settings, systems can utilize chunking or induce users to engage in chunking to significantly enhance the efficiency of knowledge acquisition. This provides psychological guidance for the development of our system in this project.

2.4 AR chunking spatial learning

Spatial learning is a daily behavior and methodology formed by humans in their exploration and cognition of the world, serving as one of the fundamental pillars of knowledge acquisition. Spatial learning is considered an indispensable ability in addition to spatial navigation, which helps humans to understand space and make autonomous responses in emergency situations (B. Liu & Zhan, 2021). Spatial thinking, as a broader topic than spatial ability, encompasses the ability to select or create spatial representations suitable for tasks. Ishikawa and Newcombe (2021) also argue that navigation, as a specific type of spatial thinking, requires understanding of our position and orientation relative to the surrounding environment. They assert that navigation is a fundamental and essential survival skill that can be enhanced through education and training. In fact, landmark knowledge, route knowledge, and survey knowledge are the basis for mastering spatial knowledge and occur during the process of spatial learning (K. Kim & Bock, 2021). Therefore, learning about space is not only helpful but also necessary. Ishikawa (2021) discuss the significance of spatial learning and navigation in everyday human spatial behavior. They emphasize the substantial individual differences observed in this domain. They also highlight that providing maps or verbal instructions may not effectively assist individuals with navigation difficulties, and navigation software may potentially weaken users' spatial awareness. As geographic spaces continue to evolve, researchers need to carefully consider the cognitive processes and experiences of individuals using technology and living within society.

MR/AR devices play a crucial role in spatial learning. Although navigation functions are generally believed to impair human spatial learning initiative (Brügger et al., 2019), visually guided MR/AR systems designed to enhance users' spatial cognition and memory abilities have shown promising results. For example, Hammady, Ma, Strathern, and Mohamad (2020) design an MR navigation application for *MuseumEye* at the Egyptian Museum using HoloLens, which validated the feasibility of MR system-guided learning. Furthermore, Buchner, Buntins, and Kerres (2022) believe that AR systems prevent users from switching back and forth between traditional paper & electronic maps and real visual field, enhancing users' attention to the surrounding environment. Users also have more positive feedback on the role of AR devices (Dong et al., 2022; B. Liu, Ding, & Meng, 2021). In addition, some development guidelines based on MR/AR systems have increased the applicability of these devices (B. Liu, Ding, Wang, & Meng, 2022a; B. Liu & Meng, 2020; Rokhsaritalemi, Sadeghi-Niaraki, & Choi, 2020). Further experiments also prove that the immersive or virtual environment based on the subtle design can lead people to better recognize the surrounding environment while benefiting spatial learning and navigation tasks. Hedge, Weaver, and Schnall (2017) creat a digital representation of the Fulda Gap region and design a series of experiments to test the effects of different navigation strategies and types of spatial knowledge on spatial learning and wayfinding, finding that using this system improves the performance of male users in survey knowledge tests. Furthermore, AR serves as a suitable assessment tool for measuring spatial short-term memory abilities. Juan, Mendez-Lopez, Perez-Hernandez, and Albiol-Perez (2014) design the ARSM task, which utilizes AR technology to assess short-term spatial memory difficulties in children. This task demonstrates superior effectiveness compared to traditional testing methods. Simultaneously, in the realm of complex spatial environments, AR can enhance human cognition of the current space. Goldiez, Ahmad, and Hancock (2007) investigate the effects of different Augmented Reality (AR) display strategies on human performance in simulated "search and rescue" navigation tasks. They discovere that specifically designed AR systems exhibit potential advantages in such scenarios. Additionally, they propose design guidelines to facilitate the future performance improvement of AR systems.

The strategy of spatial chunking provides new inspiration for spatial learning. Research suggests that when humans or advanced animals engage in spatial learning, their brains spontaneously perform spatial chunking to assist memory (Burte & Montello, 2017; Lee, 2023; Meck & Williams, 1997; J. Sargent et al., 2010). However, this memory mode proves inadequate when dealing with complex spatial structures, while it is also a crucial difference between individuals with high and low sense of direction (SOD) (Stites et al., 2020). Therefore, assisted spatial chunking is necessary (Stieff, Werner, DeSutter, Franconeri, & Hegarty, 2020). Current research mainly focuses on qualitatively chunking space based on spatial structures or other semantics (Klippel, Hansen, Richter, & Winter, 2009; Krukar et al., 2020; Nuhn & Timpf, 2022; J. Q. Sargent et al., 2019). Meanwhile, some researches also propose the concept of spatial entropy or other information measures, which can serve as a reference for chunking based on quantitative chunking methods (Altieri, Cocchi, & Roli, 2018, 2019; Batty, 1974; Boeing, 2019;

Calmet & Calmet, 2005; Verbelen et al., 2022).

2.5 Quantitative sketch map analysis

In the experiment, the primary means of assessing users' spatial memory is to have them draw sketches at the conclusion of the trial. However, sketches themselves are non-quantitative in nature. Different users employ diverse sketching methods, making the final results challenging to align and standardize. Blaser (1998) posit that sketches of geographic spatial scenes are much simpler than initially expected. However, these sketches still possess a certain level of expressiveness and can convey specific geographic spatial information. Moreover, sketches are considered a valuable method for interacting with computers. Thus, in the experiment, we aim to gain insights into what information the current sketches have learned. Billinghurst and Weghorst (1995) explore the applicability of sketch maps as external representations of individuals' cognitive maps of virtual environments. The study compare map optimality and the number of object categories within the same world, while also employing relative object position ratios to compare maps from different worlds. The research demonstrate that sketch maps accurately reflected volunteers' topological relationships within their current cognitive maps. Importantly, this study is primarily based on dense maps. Chipofya, Wang, and Schwering (2011) also acknowledge the significance of sketches, deeming them as easily manipulable objects that benefit ordinary users in organizing their geographic spatial knowledge. However, it is important to note that sketches are results derived from observation rather than precise measurements, which introduces certain distortions, generalizations, and instabilities. Therefore, a subsequent series of sketch-based registration operations becomes necessary to standardize sketches from certain perspectives, enabling quantitative representation and comparison.

To address this issue, some research has considered employing procedural and modular methods to quantitatively describe the quality of sketch maps. Skubic, Blisard, Bailey, Adams, and Matsakis (2004) approach the analysis of route map sketches from a qualitative perspective. The researchers aim to generate multi-level interviews based on spatial relationships to acquire essential knowledge about the paths. This method facilitates the formatted organization of the knowledge structure within the route map sketches. Schwering, Krukar, Manivannan, Chipofya, and Jan (2022) summarize the challenges related to the completeness, generality, and spatial accuracy of sketch maps, stating that these aspects pose difficulties in the analysis of sketch maps. Therefore, subsequent efforts should be directed towards designing appropriate workflows and software to support the analysis of sketch maps based on these three key points. Gardony, Taylor, and Brunyé (2016) attempt to standardize sketch map analysis techniques using an open-source software package called GMDA (Gardony Map Drawing Analyzer). They employ this tool to analyze sketch map instructions from both a global perspective, evaluating overall landmark configuration and providing a holistic analysis, and a single landmark perspective, which influences the overall score. The software's workflow is systematically validated through appropriate operations and its reliability is demonstrated through simulation and experimental data. Krukar, van Eek, and Schwering (2023) explore how to comprehensively analyze sketch map types, measurement accuracy, and spatial knowledge differences based on different drawing task instructions. They also propose future research directions. The study emphasizes the significance of instructions, highlighting the importance of providing accurate and clear drawing guidance during sketch map drawing tasks. Meanwhile, Tu Huynh and Doherty (2007) design a method that utilizes tablet computers and Geographic Information Systems (GIS) to collect, draw, and explore the process of sketch map creation. They also analyze the relationship between the order and types of sketch map elements drawn by participants and the cognitive map theory. The study suggests that sketch maps can be utilized to gather data on volunteers' spatial perception abilities. This lays the groundwork for our future work. Sketch maps are highly abstracted and distorted representations of the physical environment. Based on empirical research, Schwering et al. (2014) identify certain unaffected and reliable alignment techniques. These aspects of sketch maps involve the topological, directional, and sequential information of street segments, intersections, landmarks, and blocks. They employ existing qualitative representations to formalize these aspects into a qualitative constraint network.

Challenges still persist in this area, as some diversity studies have revealed cognitive differences between males and females when drawing sketch maps. Men tend to employ orientation or measurement strategies, while women lean towards route strategies (Huynh, Doherty, & Sharpe, 2010). These differences also introduce certain research complexities. Moreover, several preliminary studies have also demonstrated the various uncertainties inherent in sketch maps. This undoubtedly presents challenges for our research. Therefore, selecting an appropriate method for quantifying sketch maps is of paramount importance.

2.6 Summary

Our research builds upon some other preliminary studies and focuses on using landmarks to enhance users' spatial memory within the current environment. This topic presents challenges due to the complexity of indoor spaces, where similar spatial structures can lead to disorientation, and the lack of global landmarks hinders individuals' overall spatial awareness, often resulting in getting lost indoors (Hirtle & Bahm, 2015; Ho, Tsai, Hsu, Chang, & Lai, 2017; Huang, Shu, Yeh, & Zeng, 2016; Tan, Lee, & Lam, 2020). To address this issue, we aim to employ augmented reality (AR) virtual landmarks as an aid to facilitate spatial learning. Previous research indicates that virtual landmarks can attract individuals' attention to themselves and nearby targets, exhibiting the potential to enhance the learning process (B. Liu et al., 2021). However, an excessive number of landmarks may lead to visual overload, as individuals might struggle to effectively process spatial information due to the presence of numerous unorganized landmarks. Therefore, we endeavor to apply the concept of spatial chunking to the design of the AR system, aiming to effectively organize information and enable users to remember more spatial knowledge by following certain patterns and information organization principles.

To achieve spatial chunking, conventional research often adopts qualitative approaches, segmenting the current space based on its spatial structure and semantics. Alternatively, quantitative methods are used to calculate the amount of spatial information, enabling a more precise partitioning of space. However, these methods often result in rigid spatial divisions, neglecting the integrity and coherence of the space. Therefore, our approach involves organizing information within the entire space to achieve spatial chunking, ensuring a more holistic and cohesive representation. Meanwhile, our research focuses on information decomposition and filtering, presenting users with an appropriate information volume for different levels of SOD (sense of direction), enabling users to remember as much as possible and maximizing spatial learning efficiency. Our design fully utilizes the characteristics of MR devices, both leveraging the attractive power of virtual landmarks in MR devices and presenting the real environment to users. It enables users to mobilize their senses when performing their personal tasks, thereby learning about the surrounding environment from a top-down perspective at the psychological level (Desimone & Duncan, 1995) and enhancing the learning effect.

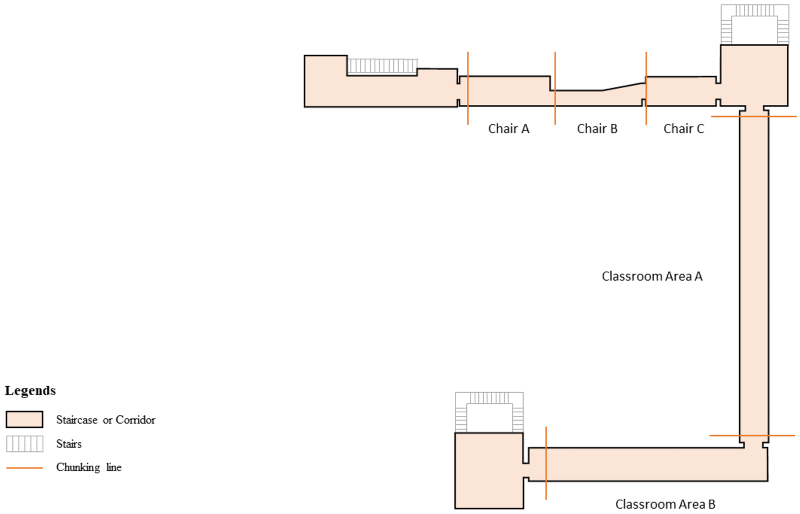
3 Methodology

In this section, we will present the workflow currently utilized to provide a clearer overview of our research approach. Additionally, we will showcase our hardware and software platforms, address pertinent technical issues, and elucidate our design principles. Our overarching design principle is to employ spatial partitioning methods to visually organize the objects within the current space, aiming to enhance user experience. There are multiple approaches to spatial partitioning, as mentioned in Section 2.4, where we discussed qualitative or quantitative partitioning methods. In the design of qualitative partitioning, we can divide the space based on its spatial semantics, such as corners, for instance. This type of partitioning seeks to deconstruct the global space in a manner congruent with human spatial cognition, reducing the memory load required for memorizing the entire space at once. Alternatively, we can divide the space into several regions based on landmark semantics, each region having a fixed semantic meaning, such as a storage area or a professor's office. While this method is natural and simple, it may not achieve the desired effect for certain long corridors with numerous objects in our chosen experimental setting. These corridors lack clear spatial semantics for segmentation, and the multitude of objects can lead to severe information overload. Furthermore, using landmark-based segmentation might not be effective, as different department offices in long corridors might be interspersed rather than aggregated within a single area. Another innovative approach involves using spatial information metrics, such as spatial entropy, to measure the complexity of the current space. Subsequently, the space is evenly divided based on the computed information complexity, ensuring that each partition has approximately similar information content, thus avoiding information imbalance resulting from manual segmentation. However, this approach could overlook natural spatial division points (e.g., corners, entrances), disrupting human spatial memory patterns. For instance, people often mentally separate spaces that are visible from those that are not, a phenomenon that this information metric approach might disregard. Given these considerations, neither of these two spatial partitioning methods is suitable for our current experimental setting or other organized indoor spaces. Hence, we adopt a more flexible spatial partitioning strategy, involving the segmentation of space based on the type and hierarchy of spatial objects. These partitions are then visually distinguished through various visualization techniques to convey the partitioning effect effectively. Finally, we will display sample images of the final design to illustrate the outcomes of our efforts.

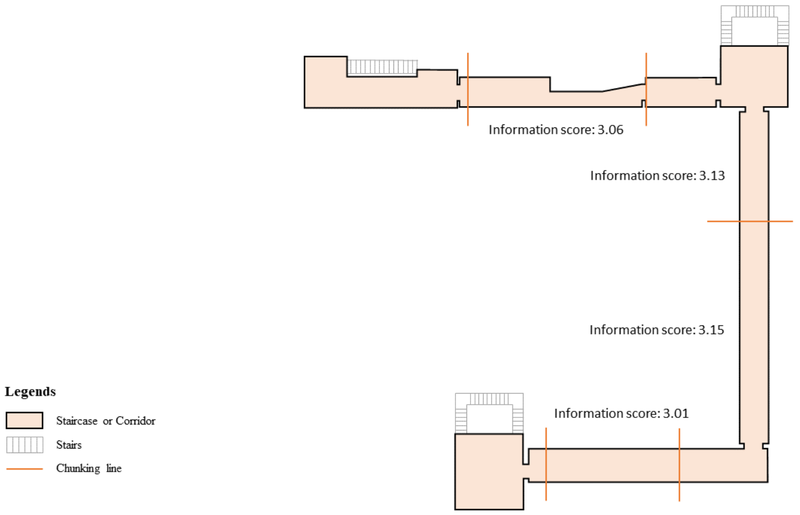
Figure 2a illustrates semantics-based chunking, where spatial divisions are made based on spatial or landmark semantics. However, this approach may inadvertently fragment discontinuous functional zones or lead to localized information overload. Conversely, Figure 2b depicts information-based chunking. This method achieves equitable partitioning of areas by evaluating their information content, thereby preventing information overload. For instance, it might involve encompassing indoor corners within a single chunk, contravening the spatial cognitive habit of the average person, who typically regards corners as the conclusion of a region.

3.1 Workflow

Based on the research question presented above, our first step involves surveying the experimental site. Subsequently, we design an AR system using a series of hardware and software devices, adhering to specific visualization rules. The design process entails modeling and interactive visualization operations. Once the design is complete, user research and experiments are conducted to validate the system's superior capabilities in spatial learning and elucidate the underlying reasons for this phenomenon. Our workflow shows as Figure 3.



(a) Semantic chunking



(b) Quantitative chunking

Figure 2: Cases of semantic and quantitative chunking

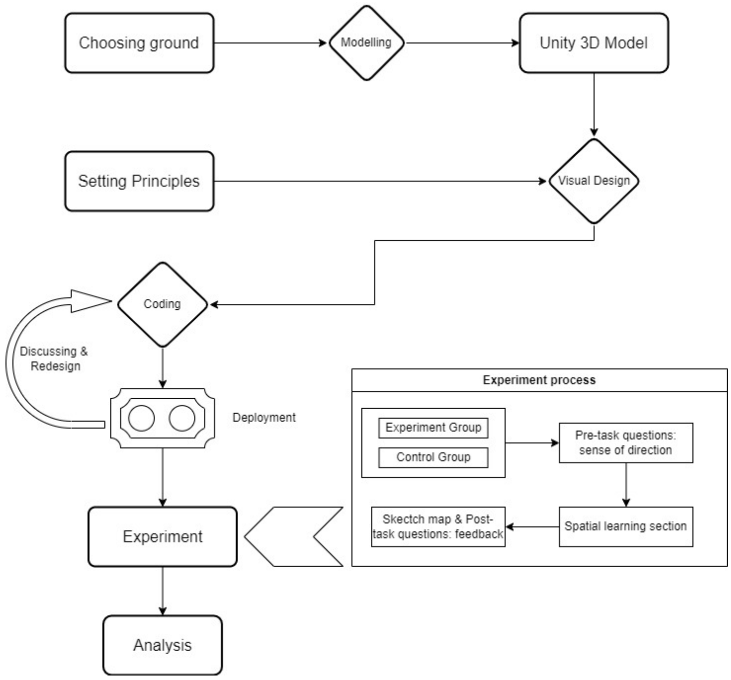


Figure 3: Workflow

We have selected the experimental site, which is located on the first floor of Building 0509 in the main campus of the Technical University of Munich. The area forms a U-shaped layout and comprises several office spaces, including Cartography, Remote Sensing, Traffic Construction, and OpenLab. This area presents a variety of scenarios, ranging from simple ones near OpenLab to more complex regions, such as the Cartography office area. The latter contains numerous similar office doors and various intricate facilities, which can potentially induce visual load and information confusion for an average person. Indeed, this area presents a diverse range of complexities in its scenes and facilities, with certain spatial patterns like offices with in the same chair often clustering together. These characteristics make it an excellent choice as an experimental site. The varying complexities and spatial regularities offer an ideal environment for conducting our research and evaluating the effectiveness of the AR system for spatial learning. After selecting the site, we will proceed with spatial modeling in Unity. The modeling process involves creating 3D representations of the walls, floor, doors of individual offices, and various accessories such as fire extinguishers and trash bins. Each object will be assigned distinct shapes and materials to cater to visualization requirements. It is essential to maintain a 1:1 scale during the modeling process to accurately represent the real-world environment in the AR system. This level of detail and accuracy in the modeling will contribute to creating an immersive and realistic spatial learning experience for users. Exactly, after completing the spatial modeling, we will apply specific spatial chunking rules and incorporate additional interactive visual elements in Unity. By utilizing C# scripting in Unity, we can access and modify the properties of different objects within the model. This allows us to implement the predefined visualization rules and design the code accordingly to achieve the desired effects. Through the implementation of spatial chunking and visual interactions, we aim to enhance the organization and comprehension of spatial information within the AR system. Users will benefit from a structured and coherent representation of the complex environment, enabling them to better learn and navigate through the space. This process necessitates iterative design and extensive discussions to achieve satisfactory outcomes.

Once the design is completed, the system will be deployed onto the corresponding AR terminals. Such a system can be utilized for conducting subsequent experiments. We will establish an experimental group (equipped with a comprehensive spatial chunking-enabled AR spatial learning system) and a control group (using an AR spatial learning system without spatial chunking functionality) to compare the superiority of our spatial chunking approach. Additionally, volunteers will be invited to fill out questionnaires to assess their spatial perception abilities and provide feedback on the system. The users' spatial learning outcomes will primarily be reflected through the sketches they create. The sketches will be standardized through spatial alignment to better evaluate the quantity, accuracy, and precision of the landmarks recalled by users. The final results will allow for a comparison between the experimental and control groups to identify differences.

3.2 Device and platform

In this project, we primarily utilize the Hololens 2 as the main hardware device. The Hololens 2 builds upon the advancements of its first-generation counterpart, offering numerous performance improvements. It can be used for various innovative immersive experiences, such as AR and VR, while providing excellent hand gesture and eye movement tracking capabilities, which further enhances its applications and research significance (Guo & Prabhakaran, 2022). The versatile features of Hololens 2 allow its application across various domains, including but not limited to healthcare and education (Miller Koop et al., 2022; Palumbo, 2022; Park, Hunt, Nadolski, & Gade, 2020; Pose-Díez-de-la Lastra et al., 2022). In our project, Hololens 2 enables spatial self-localization and calibration, allowing users to immerse themselves in an augmented reality-assisted environment. Once calibrated, the models built in Unity can be accurately projected into the real world, providing users with a seamless and authentic experience. This capability of Hololens 2 ensures precise alignment between the virtual content and the physical surroundings, enhancing the realism and effectiveness of the AR system. With the ability to seamlessly blend digital content into the real world, Hololens 2 empowers our project to create a compelling and interactive spatial learning platform for users.

Our software platform utilizes Unity, a professional cross-platform game engine capable of developing 2D and 3D games or applications (Jitendra, Srinivas, Surendra, Rao, & Chowdary, 2021). With Unity, we can design indoor environments and corresponding objects, apply unique materials, and implement interactive operations using code. To deploy the application to Hololens 2, we use Windows Universal Platform (UWP) for development, targeting the ARMx64 architecture for deployment. To bridge the gap between Unity and Hololens 2, we leverage the Mixed Reality Toolkit (MRTK) package, which provides essential Mixed Reality settings and components, allowing seamless integration of Unity applications with the Hololens 2 hardware (Ong, Siddaraju, Ong, & Siddaraju, 2021). MRTK ensures that our Unity-based application is optimized and compatible with the specific capabilities of the Hololens 2 device, enabling us to deliver a compelling mixed reality experience to our users. In MRTK, we align Unity's main camera to the logical position of the Hololens 2 headset and enable spatial anchor functionality, which accurately provides position and orientation information for subsequent experimental operations. Additionally, we activate a set of Data Providers and Pointers functionality, allowing users to interact with different objects based on their head movements and orientations. These features enhance the user experience by enabling intuitive and natural interactions within the mixed reality environment.

Unity serves as a cross-platform game engine and development environment, offering an array of tools and features designed to empower creators in crafting diverse interactive experiences. (Version 2020.3.15f2, education licence)

MRTK, developed by Microsoft, is a Mixed Reality Toolkit that constitutes an open source initiative. It operates within the Unity framework. MRTK facilitates seamless interoperability across a diverse spectrum of software and hardware platforms. (MRTK 3, MIT license)

HoloLens 2, a product of Microsoft, is a mixed reality head-mounted display that projects virtual three-dimensional imagery into the glasses through the utilization of holograms, achieving a fusion with the real world. (Hololens 2, Microsoft license)

Figure 4: Device and platform

3.3 Classification chunking

Classifying the learning objects is an effective method of spatial chunking. For instance, in the experimental area, there are offices of various disciplines, often in spatial proximity but intermingled. Without any visual cues, individuals may perceive the office doors as indistinguishable, making it challenging to remember the distribution of different disciplines' offices. Therefore, providing necessary visual cues is crucial. We assign distinct colors to the doors of offices based on their respective disciplines or chair types. As a result, even if the users cannot recall the specific content represented by each door, a quick glance allows them to roughly remember the blocks corresponding to different disciplines' offices, forming a preliminary impression of that area. In this way, individuals with lower spatial awareness and orientation abilities can grasp the overall spatial knowledge. Conversely, individuals with strong spatial learning abilities can use this spatial chunking to efficiently organize information, invoking their working memory to learn detailed knowledge in a more structured manner.

In indoor spaces, rooms (i.e., doorways) are the most crucial elements because for most people, self-positioning and navigation indoors are primarily based on certain rooms as pivots. However, other supplementary elements are also significant, as they can help individuals deepen their cognitive awareness of the current space. In certain situations, these elements may also serve as targets that individuals seek indoors, such as fire hydrants, trash bins, etc. In the current experimental area, there are stereoscopes, fire hydrants, display boards, models, trash bins, and other elements. These objects are also included in the spatial modeling process to enrich the spatial content and increase the spatial information payload. Since these objects have relatively simple classification attributes (they are similar objects generally and lack semantic differences), a unified color scheme is applied to these virtual landmarks globally. Through this approach, users can also develop a categorical understanding of the different supplementary objects in the current space based on the colors displayed.

We will categorize the experimental area to a certain extent. Firstly, we will classify each room. The experimental area includes offices from different departments, such as Hydrology Department, Cartography Department, Photogrammetry and Remote Sensing Department, Transportation Engineering Department, Traffic Construction Department, OpenLab, Restrooms, etc. Each of these rooms will be identified with its unique color. All other subsidiary small objects will also be assigned distinctive colors, such as fire hydrants, trash bins, stereoscopes, show boards, and some rare independent objects (they can be represented with the same color due to their rarity, while also reducing the color space density to avoid overcrowding and decrease color discrimination). Additionally, we will use red as the highlight color, so when a user gazes at a virtual landmark, it will be highlighted, displaying relevant information. The table 1 illustrates our color design.

Table 1: Landmark classification table

Category	**RGB**	**Color**
Cartography Dept	(0.13f , 0.54f , 0.13f , 1.0 f)	
Photogrammetry Dept	(0.8 f , 0.55f , 0.22f , 1.0 f)	
Traffic Eng.	(0.0 f , 0.54f , 0.54f , 1.0 f)	
Transportation Cons.	(0.62f , 0.16f , 0.94f , 1.0 f)	
OpenLab	(0.7 f , 0.13f , 0.13f , 1.0 f)	
Restrooms	(1.0 f , 1.0 f , 0.0 f , 1.0 f)	
Fire Hydrants	(1.0 f , 0.5 f , 0.5 f , 0.5 f)	
Trash Bins	(1.0 f , 0.5 f , 0.0 f , 0.5 f)	
Stereoscopes	(0.0 f , 1.0 f , 0.0 f , 0.5 f)	
Show boards	(0.0 f , 0.5 f , 1.0 f , 0.5 f)	
Independent Objects	(0.8 f , 0.7 f , 0.0 f , 0.54f)	
Highlight Color	(1.0 f , 0.0 f , 0.0 f , 1.0 f)	
Transparent Color	(0.0 f , 0.0 f , 0.0 f , 0.0 f)	

3.4 Hierarchy chunking

In AR, virtual landmarks can be enhanced by emphasizing their contrast with the surrounding environment and complemented by shapes that carry relevant semantics. By carefully designing the appearance and characteristics of these virtual landmarks, they become more distinct and recognizable, aiding users in better perceiving and remembering their spatial context. However, a large number of landmarks can lead to memory overload, resulting in a decline in memory efficiency. Therefore, we have implemented a hierarchical structure to address this issue. In specific spatial contexts, we designate certain objects as primary landmarks, considering them the most crucial elements within the current space. To make them more conspicuous, we assign them deeper colors. Related landmarks, which have been categorized alongside these primary landmarks during the classification step, are deemed to possess lower semantic importance than the primary landmarks. Consequently, we assign them lighter colors, visually making them less attention-grabbing compared to the primary landmarks. With this design, we anticipate that individuals will prioritize the observation of primary landmarks, enabling them to grasp the most important elements within the spatial environment and establish a preliminary understanding of the spatial framework (akin to the logic of categorization-based spatial chunking). Subsequently, individuals with additional cognitive capacity may choose to further observe and learn about secondary landmarks. These secondary landmarks share the same color tone as the primary landmarks but have a lower brightness level, facilitating the establishment of visual associations. In summary, our hierarchical approach is designed to guide individuals' attention, enhance spatial learning, and mitigate memory overload, thereby optimizing the efficiency of the AR system. The assignment of primary and secondary virtual landmarks follows a logical rationale.

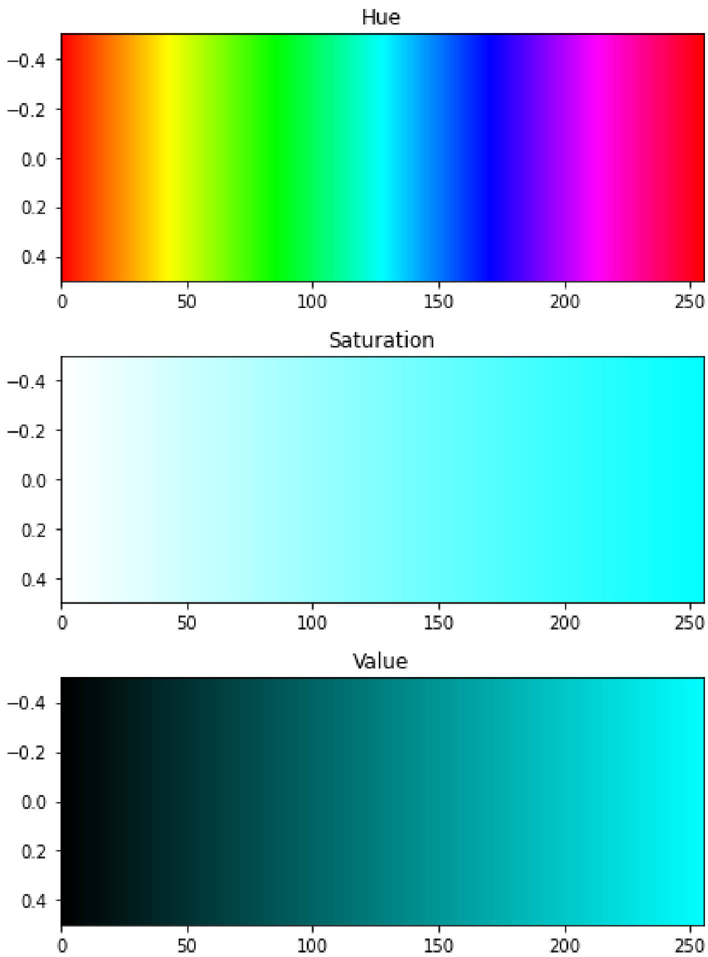


Figure 5: HSV example: we mainly use the S channel to distinguish the main object and the secondary object, as the S channel can clearly make the color faded.

Certain ancillary small objects, such as fire hydrants, have a consistent semantic nature, and thus do not require hierarchical classification. On the other hand, each office within the experimental site serves as a prominent navigation and orientation target for users, exhibiting a certain level of semantic complexity. Therefore, they are suitable for hierarchical categorization. In the context of office areas, we designate the offices of the Chairs of respective Programs as primary landmarks, as they can represent the semantic area of the respective program. Moreover, remembering the location and spatial configuration of the Chair's office is beneficial for navigation when seeking offices related to specific Programs. The spatial hierarchy delineation throughout the experimental site can be conducted following this principle. In the specific implementation of visualizing the hierarchical classification of spatial objects, we chose to manipulate the HSV color space to reflect the saliency of the objects. For primary landmarks, we maintain their original RGB values. For secondary landmarks, their colors are transformed into the HSV color space for processing, with a certain proportion of saturation reduction. This adjustment makes the colors noticeably lighter while preserving their original hue, establishing a visual distinction between primary and secondary landmarks while still allowing them to be associated with each other.

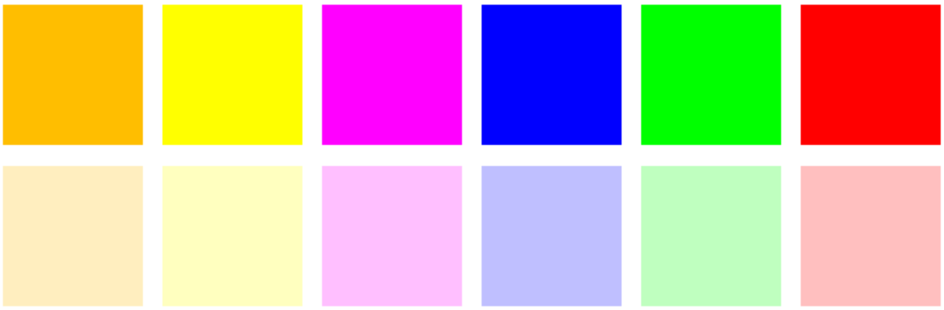


Figure 6: By manipulating in the HSV color space and decrease the S (Saturation value), we can make the color lighter.

Algorithm 1 Fading Color Function

function FADINGCOLOR($color, fadeAmount = 0.25$)
    RGBToHSV($color, initialHue, initialSaturation, initialValue$)
    $fadedValue \leftarrow$ Clamp01($initialSaturation \times fadeAmount$)
    $fadedColor \leftarrow$ HSVToRGB($initialHue, fadedValue, initialValue$)
    $fadedColor.a \leftarrow 0.8$
    return $fadedColor$
end function

Because this master-slave relationship defines a series of coordinated highlight behaviors when objects are fixated upon, and the coordinated highlights disappear together when the fixation ends. Due to the computational limitations of the Hololens 2 client device, to prevent visual sluggishness, this master-slave relationship is defined only for objects within the current visible space, rather than for all objects in the global space, to avoid computational overload.

In addition to color associations, we have also incorporated a highlighting feature. When users wear the Hololens headset, it can detect their current gaze direction. If the gaze direction corresponds to a specific virtual landmark, that landmark will be highlighted in red and display its corresponding name. This approach allows users to obtain more information about the objects they are interested in and assists in spatial object categorization. When users focus on primary landmarks, associated secondary landmarks will also be highlighted together. This is done to enhance users' familiarity with similar objects, strengthen their understanding of categorical and hierarchical spatial divisions. On the other hand, when users focus on secondary landmarks, only those landmarks will be highlighted individually. This effort is made to reduce cognitive load and emphasize the significance of primary landmarks, guiding users' attention towards them effectively.

3.5 Effect

The design outcomes generate in the Unity Editor provide a simulated representation of the spatial environment. However, when the design is implemented in the real world, various

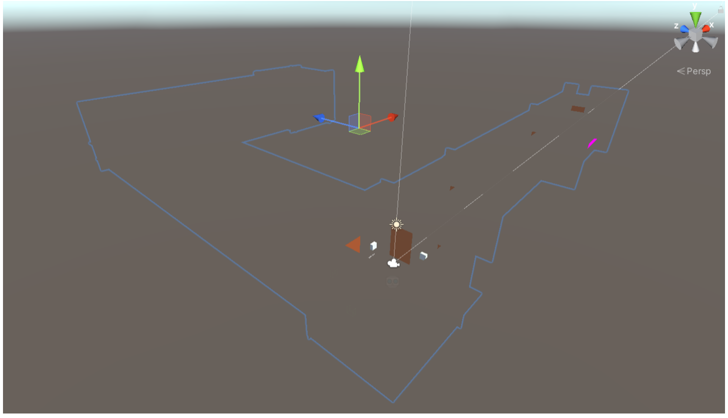
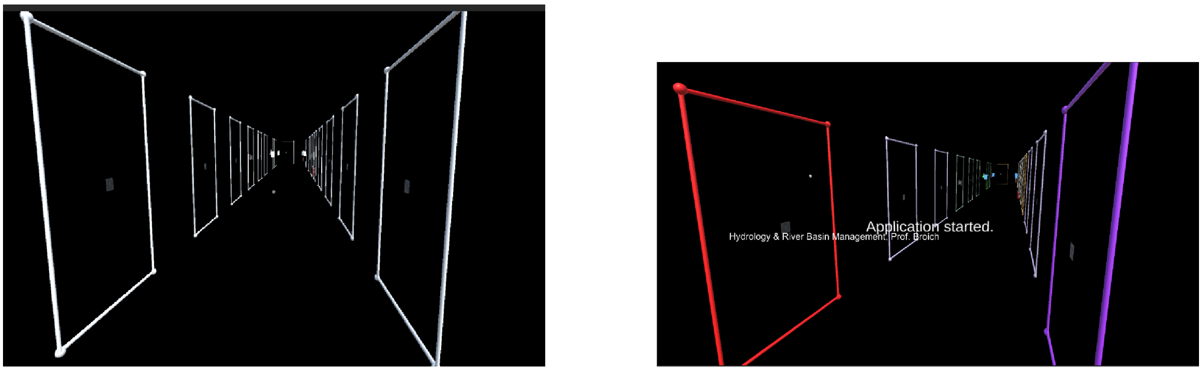


Figure 7: Experiment area model in Unity



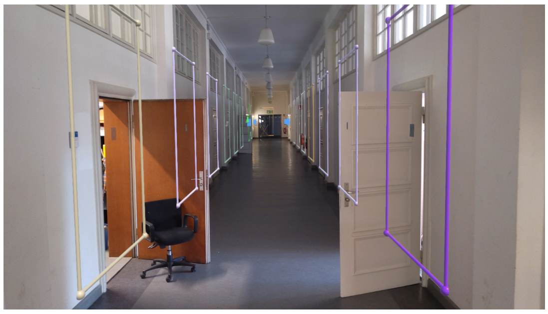
(a) Unity visual effect of the control group

(b) Unity visual effect of the experimental group

Figure 8: Visual effect of the Unity designs

factors, such as the device's screen visualization, the influence of natural light, and other external elements, can impact the visual perception of the virtual elements. Therefore, certain aspects of the visual design, such as color schemes and shapes, need to be further developed and adjusted to ensure a seamless integration with the physical environment. To address the challenges of practical visualization, a key modification is made to the spatial modeling of doors. Instead of using simple door color blocks, the design now employs door frames, which are rectangles encompassing the original door area. This adjustment is made to prevent the occurrence of large color blocks in room-dense corridors, as this could obscure significant portions of the actual scene, making it difficult for users to distinguish between virtual and physical elements. The goal of this alteration is to strike a balance between the virtual and real-world elements, ensuring that the visual cues provided by the augmented reality system complement the physical surroundings rather than overwhelming them. By using door frames, users can better perceive the location and boundaries of virtual elements in relation to the actual environment, enhancing their ability to navigate and understand the spatial layout effectively. Ultimately, these design refinements aim to create an immersive and informative experience for users, optimizing their spatial learning capabilities and reducing visual confusion. The seamless integration of augmented reality into the real-world environment fosters a more intuitive and effective spatial learning process, empowering users to engage with the virtual content while maintaining a clear understanding of their physical surroundings. The visualization effects are shown from Figure 7 to Figure 9.

(a) Real visual effect of the control group



(b) Real visual effect of the experimental group

Figure 9: Real visual effect in Hololens 2

4 Experiment and Analysis

We need to conduct experiments to obtain the corresponding data and demonstrate the effectiveness of our approach. The experiment consists of two groups: the experimental group and the control group, each consisting of 19 participants, mainly with the age from 24 to 27. The experimental group will use the software interface designed according to the methodology presented in Section 3, while the control group will use a visualization interface without spatial partitioning effects, i.e., without classification and hierarchical visualization, with all virtual landmarks presented in a uniform color. The experimental procedures for both groups are essentially the same. Initially, the researchers will introduce the purpose of the experiment to the volunteers and provide information about the basic procedures and equipment. The experiment will then commence, and the volunteers will be asked to sign privacy-related agreements and complete pre-task questions. These questions mainly assess users' background information, sense of direction, self-positioning ability, navigation skills, and familiarity with AR and VR technologies. Subsequently, the researchers will guide the volunteers through the process of spatial alignment before proceeding to explore the test area. After the exploration, the volunteers will sketch their maps, provide feedback on their experiences, and participate in subsequent mini-tests to allow the researchers to collect relevant information. Once all the questionnaires are collected, the researchers will perform data analysis. The data analysis will examine users' background, familiarity with the test area, sense of direction, evaluation of the equipment and software, and other factors. The sketch maps will be standardized to observe users' memory capacity and accuracy. These analyses will comprehensively examine the differences between the experimental and control groups, providing corresponding analysis and interpretations. Our experiment area is Figure 10, along with the corresponding landmark information.

4.1 Experiment Process

4.1.1 Pre-task questionnaires

In the pre-task questionnaire, we will request volunteers to sign relevant privacy agreements before proceeding with the experiment. The experiment will be conducted only after ensuring that the participation of volunteers is voluntary and informed. This step will also gather information about the volunteers' gender, nationality, age, profession, or occupation. In the questionnaire, we will use the SBSOD (Q1-Q15, Santa Barbara Sense of Direction scale, Hegarty, Richardson, Montello, Lovelace, and Subbiah (2002)), which is widely used to assess volunteers' sense of direction, self-navigation, and self-positioning abilities. Additionally, Q16 and Q17 will

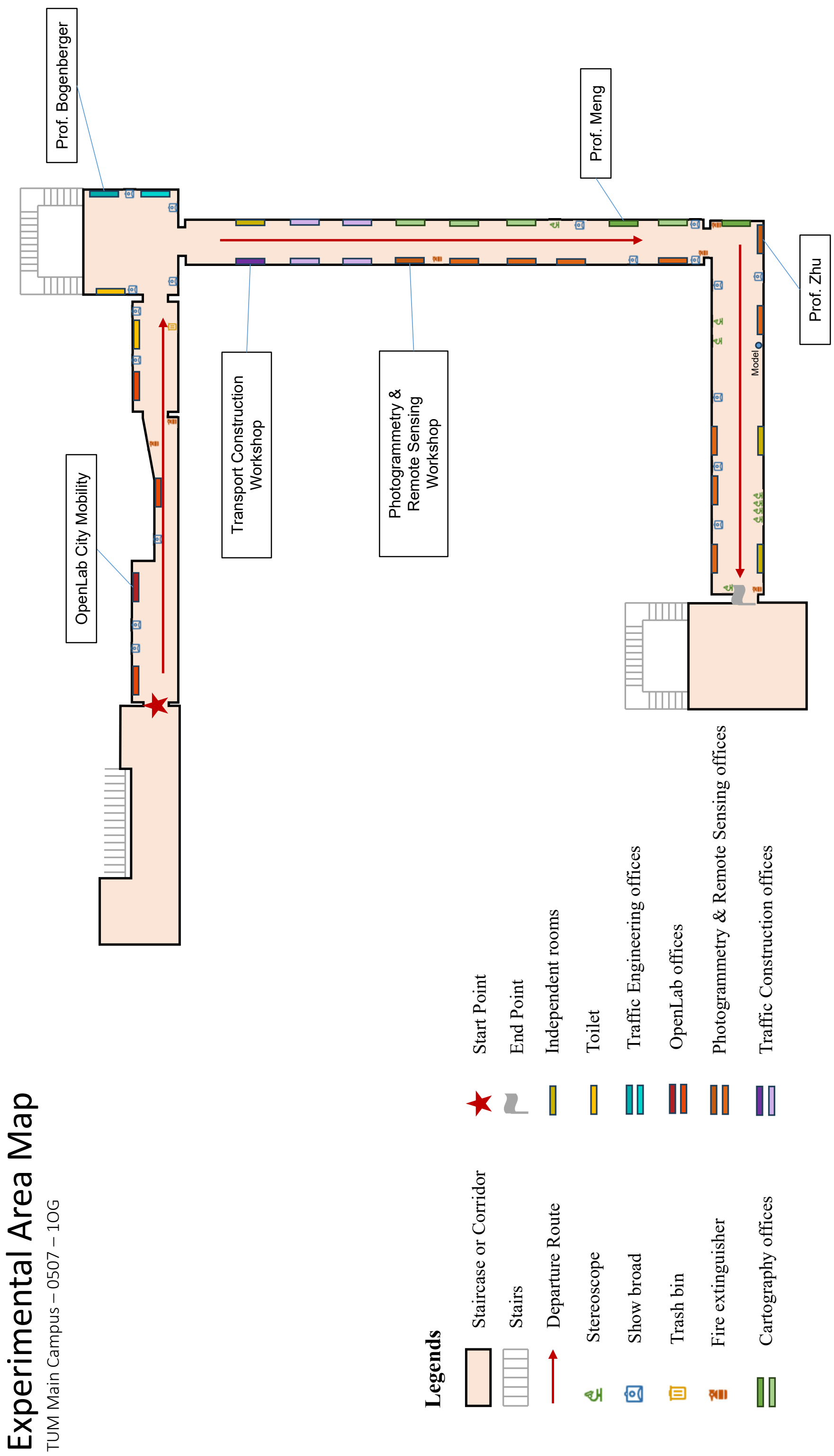


Figure 10: Experiment area map

investigate whether users tend to get lost indoors and outdoors, and inquire about their familiarity with AR and VR technologies. The researchers will be present throughout the questionnaire administration to provide assistance in clarifying any confusing points (e.g., distinguishing between AR and VR, as many people may not have a clear understanding of the difference). Each question will be graded on a seven-point scale ranging from AGREE to DISAGREE, allowing for a more nuanced differentiation of users' self-assessment at various levels.

4.1.2 Spatial learning section

In the spatial learning phase, the participants will be guided to the calibration area and instructed to open the visualization software. Our deployed software incorporates the MRTK's Spatial Anchor feature, allowing the Hololens 2 to recognize the current scene through computer vision feature matching and perform self-positioning and orientation correction. After the calibration is completed, the participants will be led to the starting point. They will be informed about the functionalities of the visualization software and the significance of virtual landmarks before commencing their journey. Participants will proceed at a normal walking pace while observing the surrounding objects and virtual landmarks. Subsequently, participants will continue walking until they reach the endpoint. They will then remove the devices and return to the office to complete the follow-up questionnaires. From Figure 11, we can see that during the

Figure 11: Experiment scene case

experimental procedure, ensuring accurate global positioning and orientation of the Hololens 2 headset is crucial, as our Mixed Reality (MR) system necessitates alignment with the real-world environment. Leveraging computer vision techniques, we enable the Hololens 2 to acquire an understanding of the current scene, thereby achieving indoor passive localization.

4.1.3 Post-task questionnaires

After completing the spatial learning phase, the volunteers will return to the office to complete the follow-up questionnaires. Firstly, the volunteers will be asked to draw a sketch of the experimental area they just walked through based on their memory. The researchers will provide necessary guidance, such as prompting the volunteers to recall the contours of the walked area and suggesting them to start drawing the remembered objects sequentially from the starting point. Adding labels is encouraged, and more detailed labels are preferred (although it is challenging for anyone to remember specific office details after walking through once). Volunteers are also encouraged to draw some ancillary objects, such as fire hydrants, display boards, trash bins, and so on. This process is generally limited to five minutes, and the researchers will time this process. When the user cannot recall more objects, this segment will end. The next segment mainly involves the volunteers' evaluation of the system, including their familiarity with the experimental area, their preference for the equipment and software, and their assessment of how colors, labels, and virtual landmarks assisted their spatial learning and memory. Finally, the volunteers will evaluate their preferences for AR and maps in indoor spatial learning scenarios.

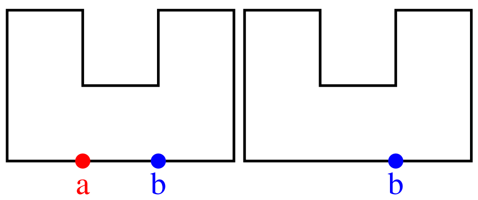
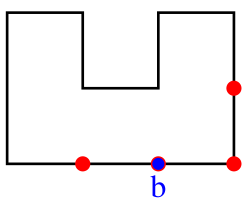
Afterwards, we have designed a scenario-based experiment: assuming there is a controllable fire in the experimental area, and you may wish to respond to the situation by trying to find a fire hydrant to extinguish the fire. Based on your memory, where do you remember there are fire hydrants? This question mainly assesses users' attention and memory towards ancillary objects during spatial learning. The floor plan of the experimental area will be provided. Users are not required to perfectly mark the locations of fire hydrants on the floor plan; instead, they only need to draw a circle to indicate the area where they remember the fire hydrant is located. Finally, volunteers will complete the NASA Task Load Index (Hart and Staveland (1988)) to assess the experimental pressure and assist the researchers in making improvements for subsequent experimental steps. Afterwards, we will ask the volunteers to reconfirm their consent for data collection and sign the agreement once again.

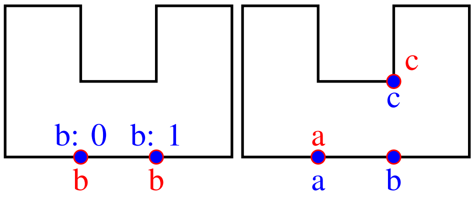
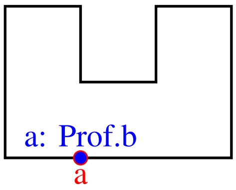
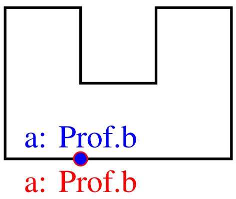
4.2 Data Preparation

The collected questionnaires contain both structured and unstructured data. All data need to be formatted and digitized for record keeping. Three sections of the questionnaire can be directly converted into numerical and formatted data: the questions about spatial orientation in the pre-task questionnaire, the questions evaluating the system in the post-questionnaire, and the data from the NASA Test. As these data are represented numerically to indicate preference levels, the final results are composed of corresponding numerical vectors. Each position in the vector corresponds to a specific question, ensuring semantic interpretation. For unstructured data, especially the sketches drawn by users based on their memory, quantification is particularly challenging. Firstly, different users have varied expression preferences when drawing sketches. Secondly, users possess different levels of information recall for different spatial objects; some might remember the specific professor associated with an office, while others might only recall the presence of an office. Lastly, the neatness of sketches varies; some users produce clear and concise sketches, while others create less organized ones. To address this issue, we have prompted users to initially outline their sketches (U-shaped corridor maps). This allows us to easily determine the spatial orientation in their mental concept after identifying the departure point. Our sketch standardization primarily focuses on the relative positions of nearby objects, information labeling, and the approximate accuracy of several correctly labeled objects' global positions. However, stringent requirements are not imposed on distance relationships or directional relationships between each object. For instance, if the object "trash bin" appears at the end of the first segment of the route, as long as the user places the marker for the trash bin at the front end of that segment, it is not necessary to provide precise relative or absolute coordinates. However, obvious errors, such as placing the trash bin at the starting point of the first segment or close to the departure side of the OpenLab (while it should be further towards the front of the OpenLab), will be considered spatial topological errors and marked incorrect. In addition to the placement of landmarks themselves, the depth of information is also examined. We have established the following rules:

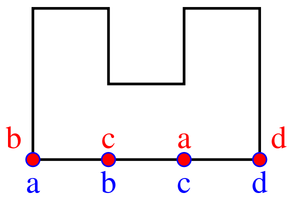
1. Complete placement errors or no labeling score 0 points;
2. Correct placement that lets the reader understand the user does remember the presence of an object but not its specific information, or recalls the information for several objects in the area but cannot precisely label each object (e.g., remembering that the offices in this area belong to the Cartography department but not knowing the specific details), scores 1 point;
3. Accurate labeling of the object and providing basic information about it for the reader to comprehend its type and basic details scores 2 points;
4. Accurate labeling of the object and providing detailed information for the reader to understand the type and specific details of the object (e.g., knowing which professor's office it is) scores 3 points.

Table 2: The red point is the landmark that the user marks and notes; the blue point is the standard map objects along with its information.

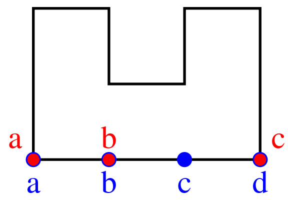
Situation	Score
	0: Completely erroneous labeling or omission of labeling will both be recorded as a score of zero.
	0: Users may randomly label numerous unnamed landmarks in a certain area based on their impression that it contains numerous elements. However, if they cannot accurately indicate the corresponding landmark locations and content, nor provide meaningful spatial reference information, their score is also considered as zero in such cases.

	1: Users fall into two categories: In the first scenario, a user can label a landmark but cannot recall its specific content; however, they are aware of the category to which the landmark belongs, which warrants a score of 1. Alternatively, users who cannot recall the specific content or category of a landmark, yet provide sufficiently accurate surrounding annotations, demonstrating their comprehension of the spatial distribution within that area, also receive a score of 1.
	2: Users are capable of accurately placing marks and providing corresponding labels, although lacking in providing more detailed information. Such annotations can demonstrate the users' remembrance of the marked object and its relative spatial location, serving as global anchors for evaluating the spatial and semantic accuracy of other landmarks with less informational content.
	3: Users not only demonstrate accurate recall of the respective landmark's location but also provide detailed information, thereby ensuring semantic clarity as well.

Furthermore, some ambiguous labeling behaviors cannot be considered correct. In corridors with numerous rooms, some users with unclear memory might only recall that the corridor contains numerous rooms, without remembering any specific ones. Consequently, they might haphazardly draw room symbols to indicate the presence of many rooms in that corridor. Such behavior will not be awarded scores according to the second criterion. Firstly, this type of feedback does not reflect that users have acquired any useful spatial information. Additionally, this information does not allow the reader to identify which specific rooms correspond to the randomly labeled ones. Based on this, we assess each sketch, list various objects, and assign scores to each one. Since objects are categorized as primary (rooms) and secondary (such as fire extinguishers, trash bins), we have performed classification. However, the aforementioned quantification of information solely evaluates semantic accuracy. For spatial accuracy, such as sequential relationships, we have employed the concept of edit distance. We label the objects on both sides of each corridor with letters, then create strings representing the order and information of objects drawn by users, calculating the edit distance between the standard string and the composed string to measure users' spatial accuracy. This completes the quantification of sketch-related information. In the questionnaire, to assess users' memory of non-essential targets, an indoor floor plan was provided, requiring users to encircle the approximate location of the fire extinguisher based on their memory. As users were not required to precisely pinpoint the fire extinguisher's location, but rather to encircle it with a circle indicating its general vicinity, during the final analysis, if the circled area drawn by the user encompassed or was near the fire extinguisher, it was considered as a correct encirclement.



(a) Users exhibit the ability to recall landmarks and their corresponding content within a given area, but demonstrate misplacement of these elements.



(b) Omission Error: Users omit spatial information regarding certain landmarks, which becomes evident when sufficient spatial references are present in the vicinity.

Figure 12: The right sequence is: ***abcd***, and the consequence is ***bcad*** and ***abzc*** (*z* generally refers to default filling.)

4.3 Data Analysis

In the data analysis phase, we initiate by conducting an analysis of background data. This analysis assesses both the experimental and control groups from the perspectives of scene familiarity, familiarity with Mixed Reality (MR) technology, and self-evaluation of spatial orientation. This ensures a balanced foundation in crucial background aspects between the two experimental groups. Subsequently, we delve into sketch analysis, scrutinizing it from two angles: semantic scoring and spatial scoring. Semantic scoring is performed through assigning scores based on semantics, while spatial scoring employs the concept of edit distance to measure sketch quality. We analyze the disparities between the two groups in terms of these metrics. Concurrently, the experimental area is partitioned into three segments to assess user performance across distinct zones. Moving on to the user evaluation of the system, we examine participants' overall perception of the system and their specific evaluations of its design. This examination elucidates user preferences for the two divergent systems—the experimental and control systems—and helps elucidate reasons for disparities in sketch quality. Group analysis of stress, as determined by the NASA questionnaire, reflects the adequacy of the experimental design and gauges the multifaceted stress impact of the two systems on users. Analysis of the annotation of fire extinguishers reveals users' attitudes toward learning non-primary targets within indoor spaces. Ultimately, we perform a correlation analysis to elucidate the degree to which certain user backgrounds and preferences correlate with final outcomes. This comprehensive assessment aids in evaluating the design strengths and weaknesses of the systems.

4.3.1 Background analysis

The background analysis encompasses users' familiarity with the environment, their familiarity with MR devices, as well as their orientation skills and navigational assessments. This is because these three background aspects could directly influence the final outcomes: individuals with a better sense of direction might find it easier to remember spatial information; those familiar with the experimental area naturally have an advantage over those who are not; and individuals more acquainted with MR technology tend to use our equipment more effortlessly. Thus, we aim to ensure similarity in these three aspects of background between the experimental and control groups. And these three kinds of inquiries correspond to various questions within the original questionnaire: For responses involving multiple questions, we normalize the numerical values of each question's answers, rendering all differences comparable across two one-dimensional vectors. To assess the disparities between these vector datasets, we employ the Levene variance

Table 3: Background inquiry questions

Aspect	Questionnaire	Question
Environment familiarity	POST TASK	Q1
MR familiarity	PRE TASK	Q18-Q19
SOD score	PRE TASK	Q1-Q4, Q6, Q8, Q11-Q12, Q14-Q17

Table 4: Difference of background between Experimental Group and Control group.

Group	Levene P-Value	Diff P-Value
SOD	0.796	0.632
Scene Familiarity	0.173	0.480
MR Familiarity	0.187	0.112

comparison test to determine whether the assumption of approximate variances between the two groups of data holds true. If the hypothesis of similar variances is upheld, the Student's t-test is employed. Conversely, if this hypothesis is rejected, the Welch's t-test is utilized. The Levene test can be employed to examine the equality of variances across multiple samples. It initially computes the absolute deviations within the samples and subsequently performs variance analysis to determine if substantial between-group differences exist. This method is applicable to both normally and non-normally distributed data, thus boasting a broad scope of applicability. Generally, a p-value below 0.05 is commonly regarded as significant evidence for rejecting the null hypothesis (Levene, 1960):

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i \left(Z_{i.} - Z_{..}\right)^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(Z_{ij} - Z_{i.}\right)^2} \tag{1}$$

In this context, k represents the number of distinct groups to which the sample cases belong, N signifies the total number of examples. Meanwhile, Z refers to the absolute value of the difference between the examples and the mean or median (Levene et al., 1960). Based on the assumption of variance equality, different choices for hypothesis testing are made: When variance is considered to be dissimilar or unequal, the Mann-Whitney U test is employed. This method is widely used when the assumption of equal variances is not met and there is no need to ascertain the normal distribution of data (Mann & Whitney, 1947); When variance is considered to be similar or equal, the Student's t-test is utilized (Mishra, Singh, Pandey, Mishra, & Pandey, 2019). The resulting p-values from both methods are typically assessed against a threshold of 0.05. The null hypothesis (H_0) posits that no significant difference exists between the two sets of data. If the obtained p-value is less than 0.05, the null hypothesis is rejected. This series of methods serve as our means to assess the distinctiveness of the obtained data, finding widespread utility in both inter-group and intra-group comparisons. Through this approach, we endeavor to identify background differences among participants in the experimental and control groups. We aggregate quantified responses to questions pertaining to sense of direction, scene familiarity, and MR familiarity. This yields three corresponding vectors for the experimental and control groups. Each vector pair is subjected to Levene's test, followed by the selection of an appropriate testing method. Subsequently, p-values are computed, and a threshold of 0.05 is employed to determine whether the null hypothesis should be rejected.

The P-value threshold for the hypothesis testing of both variance differences and mean differences

Table 5: Mean semantic scores of overall and segments

Area	Exp. important	Con. important	Exp. un-important	Con. un-important
Overall	0.93	0.20	0.64	0.21
Block 1	1.01	0.15	0.35	0.04
Block 2	0.62	0.23	0.85	0.34
Block 3	1.08	0.38	0.73	0.24

is set at the empirical value of 0.05. Based on the data from Table 4, following the Levene's test, Student's t-test is applied to all three groups of data. And the resulting P-values also indicate the inability to reject the hypothesis of equal means, indicating minimal differences among these three datasets. This suggests that the participants involved in the experiment, whether from the control or experimental group, exhibit comparable overall backgrounds, which can imply that potential biases in background conditions do not significantly impact the experimental outcomes.

4.3.2 Sketch map analysis

The analysis of hand-drawn sketches constitutes a focal point of this study. Hand-drawn sketches encapsulate users' recollection of various environmental features within the experimental setting following their interaction with the system, thereby reflecting their spatial memory quality. The evaluation of sketches is approached from two aspects. The first approach employs a quantitative scoring system, whereby distinct scores are assigned to individual landmarks based on the accuracy of their depiction and the level of detail provided by users. This scoring scheme serves to depict the quality of users' hand-drawn maps. Important landmarks and un-important landmarks can be analyzed separately. The second approach involves an evaluation based on spatial edit distance analysis. Each landmark along a given route is attributed a unique letter, and users' annotations are transformed into corresponding strings. Subsequently, the edit distance between the user-generated string and the standard string is computed. A higher edit distance signifies a greater number of modifications required to transform the user-generated string into the standard string, thereby indicating poorer sketch quality derived from users' hand-drawn maps. Simultaneously, the U-shaped experimental area is partitioned into three distinct segments, and each of these segments is to analyze whether there is difference for segment situation between the control and experimental group.

In Table 5, drawing from the holistic and segmented quantitative analysis of sketches, the scores presented in the table signify the mean assigned scores based on the entirety of each group. Notably, across all aspects, the experimental group consistently outperforms the control group in the table. Based on the results presented in the table, both in terms of the overall magnitude and when considering segmented analysis, the experimental group outperforms the control group. By depicting the score distributions of each group using box plots, as illustrated in Figure 13, it becomes evident that the experimental group achieves significantly higher scores for both important and un-important landmarks compared to the control group. Within each group, the experimental group's scores for important landmarks notably surpass those for un-important landmarks. Conversely, in the control group, unimportant landmarks acquire a slightly higher distribution of scores than important landmarks. Furthermore, the data distribution of scores for un-important landmarks in both groups displays higher variability, indicating substantial discrepancies in users' memory of un-important landmarks.

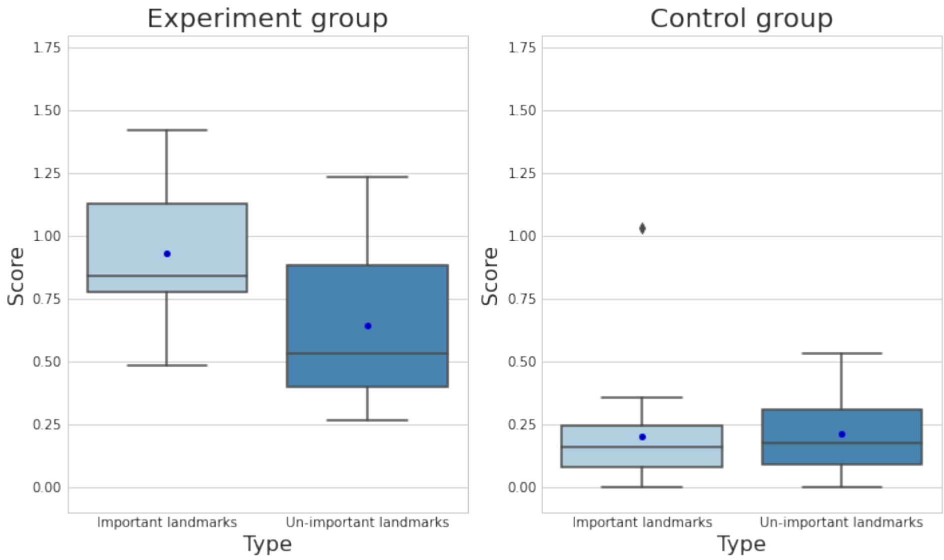


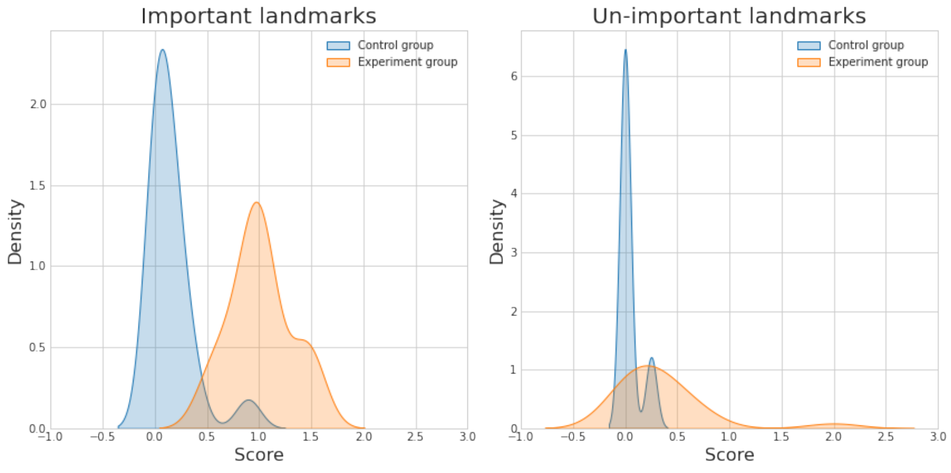
Figure 13: Distribution plots of user sketch quality assessment scores based on important and un-important landmarks for both the experimental and control groups.

Table 6: Comparison of Inter-group and Intra-group Mean Differences between Experimental and Control Groups

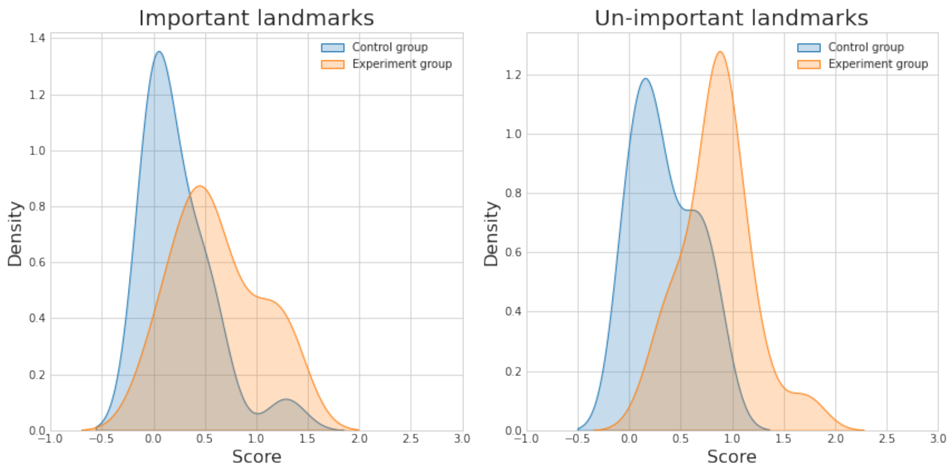
Group	Levene P-Value	Diff P-Value
Exp. Important & Exp. Un-important	0.490	0.003
Con. Important & Con. Un-important	0.632	0.892
Exp. Important & Con. Important	0.222	< 0.001
Exp. Un-important & Con. Un-important	0.008	< 0.001

In order to gain a more specific understanding of inter-group and intra-group score differences, we also conducted variance and mean difference tests, resulting in the outcomes presented in Table 6. Regarding the intra-group analysis, a significant disparity is observed between the scores of important and un-important landmarks within the experimental group. This observation, in conjunction with the data presented in Table 5, suggests that the scores for important landmarks are notably higher than those for un-important landmarks within the experimental group. In contrast, within the control group, such a distinction is not prominently evident, indicating that scores for both important and un-important landmarks are relatively same. In the context of inter-group analysis, the experimental group exhibits substantial differences compared to the control group across all aspects, with users' recall performance notably surpassing that of the control group.

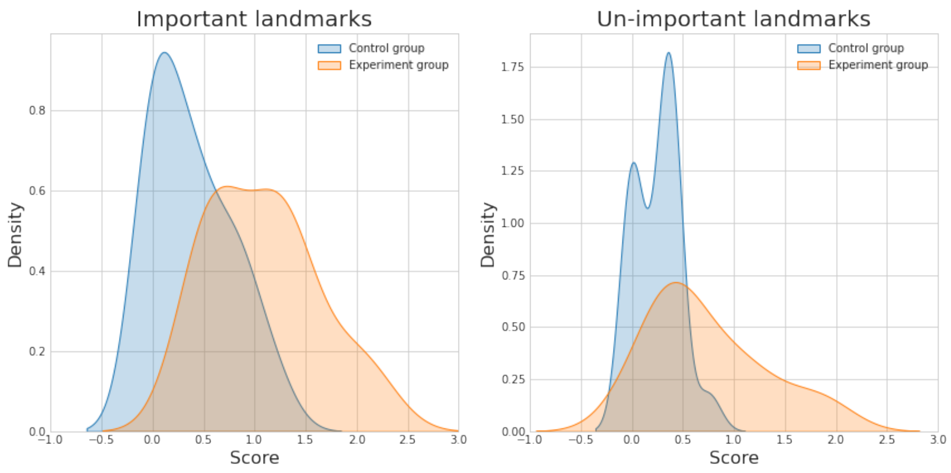
Blockwise analysis also reveals that both for scores related to important landmarks and un-important landmarks, the experimental group outperforms the control group. From Figure 14, it can be observed that in the density visualizations of user ratings, the density of ratings in the experimental group typically exhibits a unimodal distribution, whereas in the control group, the density of ratings usually appears as either unimodal or bimodal distributions. Overall, the distribution of user ratings in the experimental group appears more dispersed, while in the control group, it tends to be more concentrated. In terms of peak values, the scores from the peak region of the experimental group is higher than that of the control group, explaining the visual



(a) Block 1



(b) Block 2



(c) Block 3

Figure 14: Blockwise Analysis: scores in each segment

Table 7: Difference analysis between each segment (p-values of Student's t-test or Mann-Whitney U's t-test)

Segment	Exp. Important & Con. Important	Exp. Un-important & Con. Un-important
Block 1	< 0.001	< 0.001
Block 2	< 0.001	< 0.001
Block 3	< 0.001	0.001

superiority of the experimental group's performance over the control group in each block. For important landmarks, in Block 1 and Block 3, which correspond to the initial and final regions of the experimental area, the experimental group performs better. However, in Block 2, the middle section of the experimental area, their performance is relatively poorer. In contrast, the control group performs poorly in Block 1, and there are a few occurrences of higher scores in both Block 2 and Block 3. As for un-important landmarks, the experimental group's performance is better in Block 2 compared to Block 1 and Block 3. On the other hand, the control group's performance in Block 1 remains the weakest, with a concentration of low scores at the highest peak of the bimodal distribution. The p-values resulting from the segment-wise analysis of differences for each block, as listed in Table 7, reveal a substantial advantage in favor of the experimental group over the control group. Hence, users in the experimental group demonstrate outstanding performance in each segment.

We also conduct a differential analysis of the performance of the experimental group and the control group in various blocks. Upon examination, significant differences are observed in the recall of both important and un-important landmarks by the experimental group in various blocks ($p-\text{value} < 0.05$). However, for the control group, the scores for the recall of important landmarks remain relatively consistent across different blocks ($p-\text{value} = 0.09$), while the scores for non-important landmarks exhibit significant variations ($p-\text{value} < 0.05$). Combining this with Table 14, it can be understood that the experimental group performs poorly in recalling important landmarks in Block 2, and both the experimental and control groups exhibit lower performance in recalling un-important landmarks in Block 1.

For the analysis of edit distances, as it involves the topological relationships and sequential positions of landmarks, no distinction is made between important and un-important landmarks in the grouping. The average distance values are calculated for the overall area and for each block, listed as Table 8, showing that for all aspects the experimental group indicates better sptial performance than the control group.Following the calculation of average edit distances for each segment, a box plot representation is obtained, as illustrated in Figure 15. From the graph, it can be observed that the edit distance scores of the control group are higher than those of the experimental group. The Levene's test for both groups yields a p-value of 0.046, indicating the unequal variances assumption cannot be upheld. Therefore, the Mann-Whitney U's test is employed, resulting in a p-value less than 0.001, indicating a significant difference between the two groups. This suggests that not only in terms of quantitative measures but also in terms of spatial quality, the sketches of the experimental group outperform those of the control group. In the case of the control group, the edit distance scores are concentrated in the high score range, suggesting generally lower spatial quality of sketches in this group. On the other hand, the distribution of edit distance scores for the experimental group is primarily concentrated in the low score range with a few outliers, indicating some variability in spatial quality within the experimental group, but with the majority still concentrated in the low-score region.

Table 8: Average values of the edit distance score for the overall area and each block of the experimental group and the control group

Area	Control group	Experimental group	Diff p-value
Overall	0.93	0.69	< 0.001
Block 1	0.91	0.68	< 0.001
Block 2	0.95	0.69	< 0.001
Block 3	0.92	0.70	< 0.001

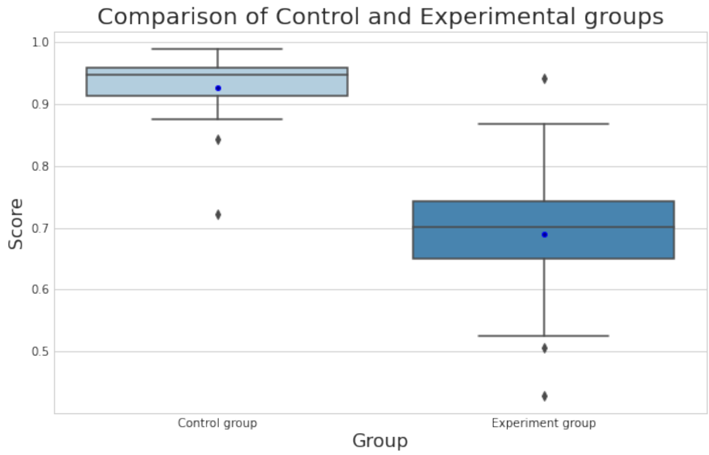
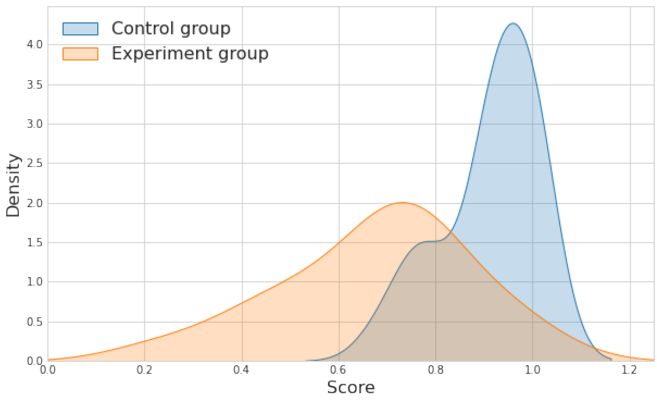


Figure 15: Box plot representation of edit distance scores

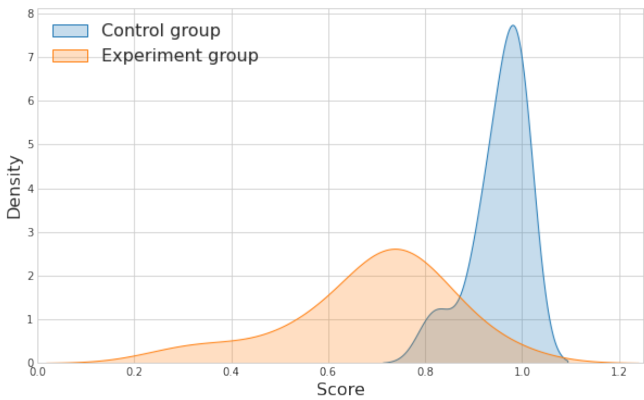
Similarly, conducting a segmented analysis on the edit distance, we obtain the distribution of edit distance scores for each block, as illustrated in Figure 16. For each block, the distribution of edit distance scores for the two groups is visualized by plotting score densities and interpolating using KDE. The results indicate that in each block, the control group exhibits higher edit distance scores, suggesting a lower level of spatial accuracy reflected in their sketches. From the graph, it can be observed that the distribution of edit distance scores generally follows a unimodal pattern, with the peaks of the control group corresponding to higher scores than those of the experimental group. This trend reflects the higher spatial distribution quality of sketches in the experimental group compared to the control group. For the experimental group, the poorest performance is observed in Block 3, which lacks low edit distance scores. Although the peaks of the experimental group are consistent across the three blocks on the graph, and the means of the three areas are not significantly different, Block 3 exhibits a deficiency in low edit distance scores. As for the control group, their performance is consistent across the three blocks, yet there is a segment of low scores in Block 1 and 2, indicating there are small numbers of relatively slightly better performance in those blocks. However, the overall examination reveals that there is no significant difference among each block in the edit distance score ($p-$ value > 0.05).

4.3.3 Small objects recall quality

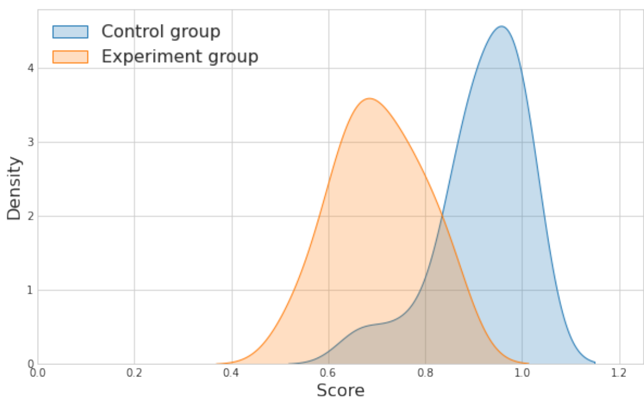
The small experiment in this section aims to assess users' memory of inconspicuous targets. The fire extinguisher is an example of an inconspicuous yet crucial object in emergency situations, making it a typical subject for examination. Both the experimental and control groups are instructed to mark the locations they remember for fire extinguishers on a map, and the results are



(a) Block 1



(b) Block 2



(c) Block 3

Figure 16: Blockwise Analysis: edit distance scores in each segment

represented in a box plot as shown in Figure 17. The figure demonstrates that the experimental group exhibits a higher degree of memory for fire extinguisher locations compared to the control group. To assess the significance of this difference, a Student's t-test is conducted on the two sets of data, under the condition of homogeneity of variances, yielding a result p-value of 0.019. This indicates a significant difference between the two groups, providing evidence that the experimental group's memory performance regarding fire extinguishers is significantly superior to that of the control group.

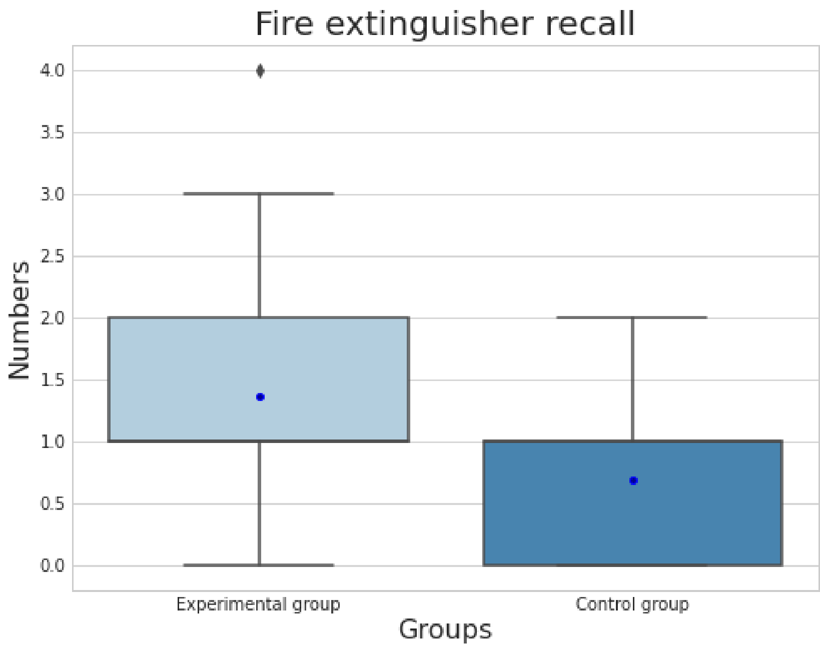


Figure 17: Box plot of fire extinguisher recall quality

4.3.4 User evaluation analysis

Following the sketch analysis, we collect user evaluations of the system. Since both the experimental and control groups do not have the opportunity to experience products from the opposite group and the evaluation involves subjective judgments, the inter-group comparison lacks significant differences and specific meaning. The questionnaire primarily request users to assess the preferences, memory assistance, positional assistance, and orientation cognition effects of color, labels, and landmark design. Specifically, "COLORS" represents landmark colors, including transparency; "LABELS" represents label text and icons; and "LANDMARKS" represents aspects such as shape, location, and orientation. For the three elements of color, labels, and landmark design, we aim to investigate whether users perceive these designed elements as important and, in terms of preference, memory assistance, positional assistance, and orientation cognition, which aspect holds greater significance for the users. To ascertain whether these three elements play a predominant role and in which aspect they have the most significant impact, we primarily employ ANOVA analysis and Tukey HSD analysis. ANOVA serves as a relatively preliminary analysis, comparing the variability among means of different groups to determine if within-group variables exhibit notable differences and providing corresponding confidence

Table 9: Evaluation scores for the experimental group

Element	Preference	Memory	Positioning	Orientation
COLOR	4.95	3.95	3.37	3.47
LABEL	5.37	5.88	5.42	5.05
LANDMARK	5.53	6.11	6.00	5.26

Table 10: Evaluation scores for the control group

Element	Preference	Memory	Positioning	Orientation
COLOR	4.37	3.84	3.37	3.58
LABEL	5.32	5.37	4.58	4.79
LANDMARK	5.21	5.53	5.16	4.68

levels:

$$F = \frac{BMSS}{WMSS}, \quad p - \text{value} = P(F > F_{observed}|H_0) \tag{2}$$

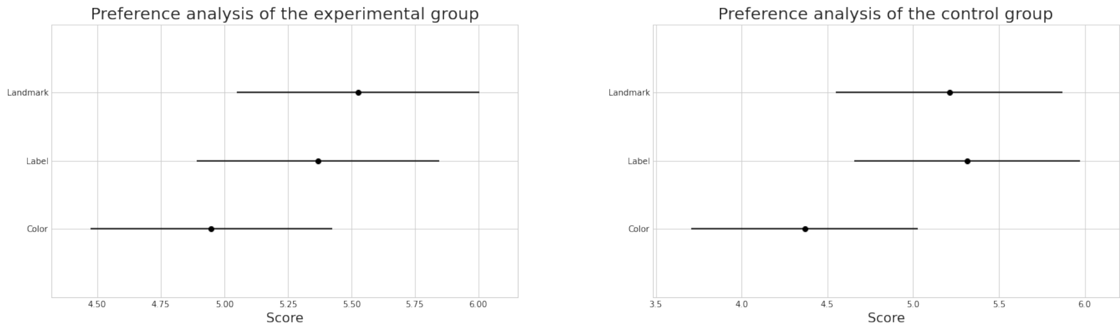
Among them, $BMSS$ is the between means sum of squares, while the $WMSS$ is the within means sum of squares, and F is the test statistical value (Fisher, 1970). Upon determining the significant differences among the current group variables, it is essential to identify which variables exhibit substantial disparities and pinpoint the most prominent one among these variables. This necessitates the utilization of Tukey HSD analysis, which facilitates the detection of pairwise differences between variables and highlights the most significant one through visual representation. This method is essentially analogous to the t-test, yet it corrects for familywise error rate, mitigating the risk of Type I errors, and not sensitive to the data normality (Tukey, 1949). However, the reliability of these two tests is contingent upon the requirement that the data exhibit a normal distribution and possess homogeneity of variances. Homogeneity of variances can be assessed using the Levene's test, while the normality of distribution is examined through the Shapiro-Wilk method. The null hypothesis of the Shapiro-Wilk test assumes that the data is drawn from a population that follows a normal distribution.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

W is the test statistic, x represents the sample, and a is derived from the covariance matrix (Shapiro & Wilk, 1965). Through these steps, we sequentially examine each aspect. We firstly calculate the mean value of each aspect for three elements as Table 9 and Table 10. However, these means are quite similar with little variation, making it difficult to identify prominent variables. Hence, normality and homogeneity of variances tests are conducted, followed by the output of ANOVA and Tukey HSD results.

1. Preference

   Commencing with the experimental group, the Shapiro-Wilk test is applied to the three elements. The results reveal that the corresponding p-values for all three elements are not uniformly less than 0.05, thereby not satisfying the assumption of normal distribution. Furthermore, the p-value for the test of homogeneity of variances is 0.433, indicating a semblance of homogeneity in variance. Subsequent multifactor ANOVA testing yields a p-value of 0.325, suggesting an absence of significant differences within the groups. Similarly, the variables within the control group also deviate from a normal distribution, and the ANOVA analysis similarly indicates an absence of significant differences within the group. The corresponding Tukey HSD plot is depicted in Figure 18. While there are no significant differences between inter-group variables, comparatively, both the control group and the experimental group find the design concerning colors dissatisfactory.



(a) Experiment group

(b) Control group

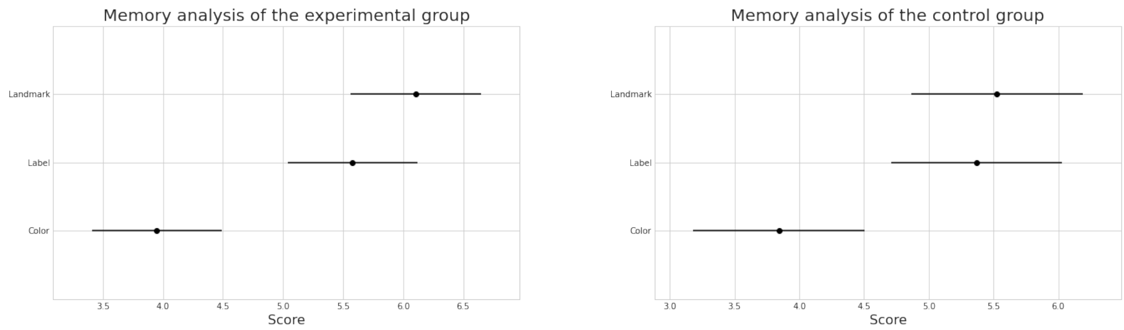
Figure 18: Tukey HSD plot on *how do you like the element you saw*

2. Spatial memory

   The investigation into whether users perceive these three elements as aiding their spatial memory is ongoing, with the examination of corresponding differences. The variable distribution within the experimental group does not entirely conform to a normal distribution, while that of the control group does; the Levene's test yields p-values of 0.089 for the experimental group and 0.016 for the control group, suggesting homogeneity of variances for the experimental group but not for the control group. The results of a multifactor ANOVA analysis for the experimental and control groups are both less than the threshold 0.05, respectively, indicating significant differences in intra-group variables. Despite the non-normality in the experimental group's data, we still generate the corresponding Tukey HSD plot, as illustrated in Figure 19. The figure reveals that the issue with color design stands out more compared to landmark and label design. Given the lower mean value associated with color design, it can be inferred that concerning spatial memory, there is a notably significant concern with color design relative to the other two aspects, which might necessitate further analysis and improvements in subsequent stages.

3. Spatial locating

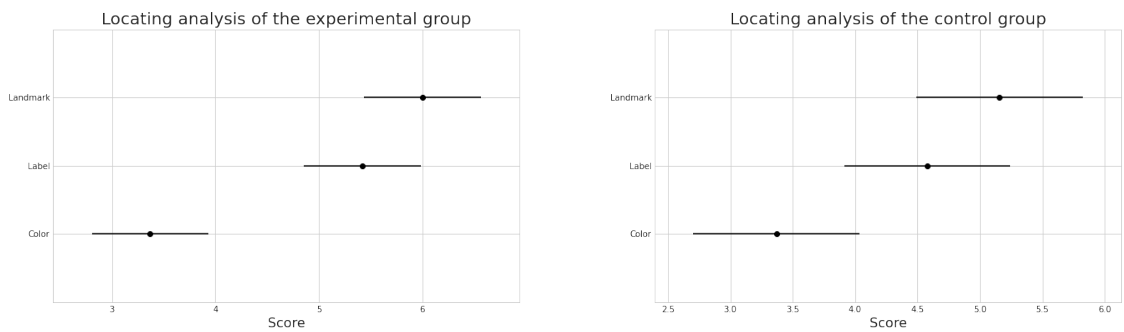
   We also evaluate the impact of these designs on users' spatial orientation. The variable distributions for both the experimental and control groups do not adhere to a normal distribution. Upon conducting the Levene's test, the experimental group yields a p-value of 0.224, while the control group's p-value is 0.715, indicating homogeneity of variances. The multifactor ANOVA reveals p-values less than 0.001 for the experimental group and 0.714 for the control group, suggesting significant internal differences for the experimental group but not for the control group. Despite the non-conformity of normality in both sets of data, a Tukey HSD chart is presented here, as depicted in Figure 20. From the graph, it can be inferred that color design remains a salient concern in this evaluation. Users perceive that landmark design and label design contribute to a better sense of self-orientation.



(a) Experiment group

(b) Control group

Figure 19: Tukey HSD plot on *how do you think each element helps you remember the area*



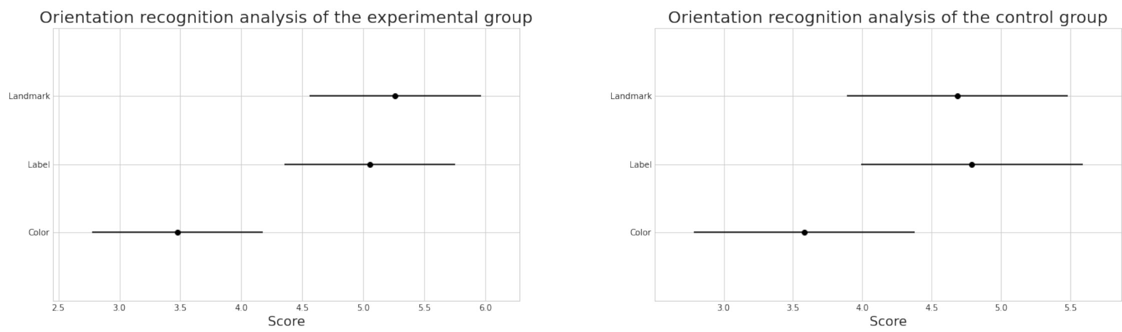
(a) Experiment group

(b) Control group

Figure 20: Tukey HSD plot on *how do you think each element helps you locate your position*

4. Orientation recognition

   We investigate the self-assessment of user orientation cognition concerning these three elements. The variable distributions of both experimental and control groups do not adhere to normal distribution. Conducting the Levene test yields a p-value of 0.849 for the experimental group and 0.915 for the control group, indicating homogeneity of variance. The multifactor ANOVA reveals a p-value of 0.006 for the experimental group, suggesting significant internal differences. In contrast, the control group has a p-value of 0.137, implying minor internal differences. Despite both groups' data exhibiting non-normality, we present the Tukey HSD plot in Figure 21. In the control group, Tukey HSD also shows small differences in variables between groups. In the experimental group, color is still a prominent variable and is in a relatively disadvantaged position.



(a) Experiment group

(b) Control group

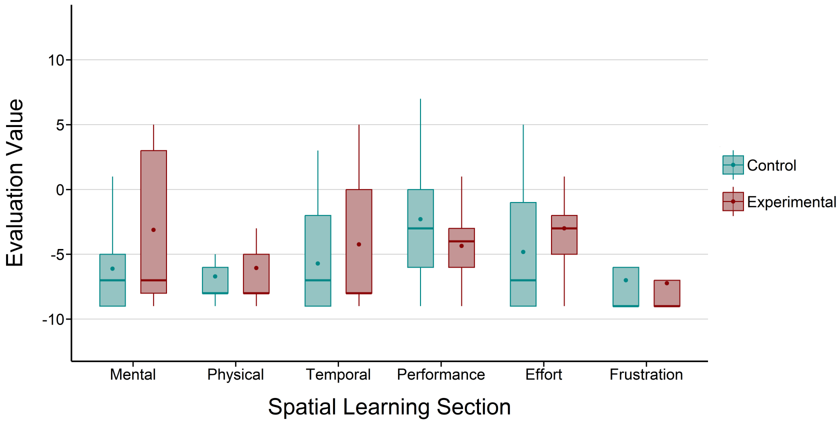
Figure 21: Tukey HSD plot on *how do you think each element helps you identify your orientation*

In summary, in the four aspects mentioned above, users generally think that the color design is significantly weaker than the landmark design and the label design in most cases. This result will affect our investigation of the color design and the subsequent research analysis work. After follow-up interviews, most people report that the color design has a significant role in landmark semantic classification; however, in semantic layering, due to the influence of natural light and the deficiency of device visualization, the color is distorted, making users unable to feel the depth of color, and would tend to think that the dark and light colors are different colors. For example, the dark green in the device may be biased towards blue, and be mistaken by some users as dark blue.

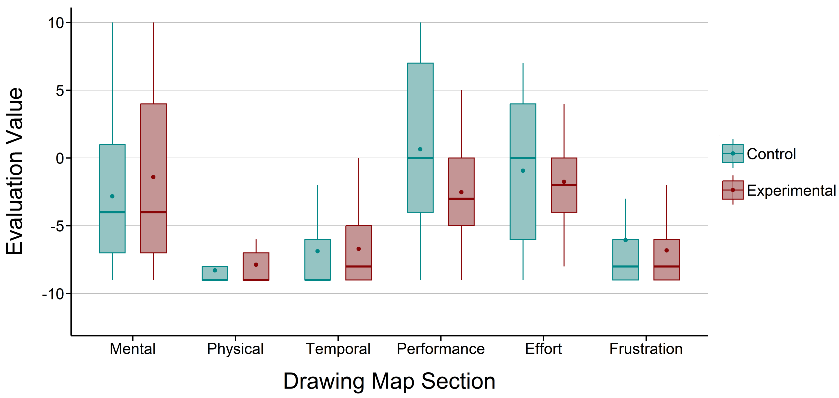
In most cases, individuals comprehend indoor spatial structures through indoor maps, including physical maps commonly posted in various locations on campus or electronic maps. Therefore, we also explore whether users who have experienced AR devices hold higher expectations and interest in understanding indoor spaces through AR. We conduct a brief study within both the experimental and control groups, gathering user perceptions of AR and maps. In the experimental group, the average interest rating for AR is 5.63, while for maps, it is 4.53. Levene's test indicates homogeneity of variance between the two groups, and a Student's t-test yields a p-value of 0.019, indicating significant differences. Further analysis within the control group reveals an average interest rating of 5.58 for AR and 4.11 for maps. After conducting tests for homogeneity of variance and Student's t-test, the p-value is 0.009, indicating substantial differences. Hence, it can be inferred that both the control and experimental groups hold high expectations for utilizing AR in understanding indoor spaces.

4.3.5 Workload evaluation

The NASA Load Test primarily assesses the stress experienced by users during the experiment, encompassing psychological stress, physical stress, time constraints, self-assessment of performance, self-exertion assessment, and evaluation of negative emotions. Users are required to rate their stress levels during both the Learning and Recalling phases. This feedback not only provides insights into the challenges users face during the experiment, aiding in the interpretation of their performance, but also contributes to the evaluation of the experiment's overall workload, facilitating improvements in subsequent experimental designs. Among all the metrics, lower scores indicate more positive ratings, while higher scores represent a more negative perception.



(a) Spatial learning section



(b) Recalling section

Figure 22: Boxplot of NASA Test Load evaluation on two aspects

Comparing these data sets, it can be observed through difference tests that there are no significant differences ($p - \text{value} > 0.05$) in the distribution of ratings for the same attributes between the experimental group and the control group. However, visualizing the data distribution through box plots still reveals some clues. A fascinating observation emerges from the data, as depicted in Figure 22: during both the learning and recall phases, the experimental group reports a greater degree of stress in terms of mental, temporal, and physical demands when compared to the control group. However, when assessing their own performance, whether in spatial learning or recall stages, the experimental group demonstrates increased confidence, believing their performance to surpass that of the control group; this aligns with the factual outcomes. Compared to the control group, the experimental group believe that they notably perceive a more organized stream of information while wearing the AR system, allowing for the creation of finer sketches. However, during the recall phase, the control group tends to invest more effort in recollecting the landmarks, even if the final results are inferior to those of the experimental group. Both groups exhibit a relatively low level of negative emotions. Post-experiment interviews reveal that some members of the experimental group tend to experience positive guidance from the visualization system, enhancing their awareness of visual pressure and information load in the spatial environment. Conversely, due to the nearly uniform design of landmarks, certain members of the control group struggle to perceive the information load in the spatial environment. They find it challenging to distinguish which landmarks require their attention, as these landmarks appear monotonous, leading to their reluctance to engage in detailed observation.

4.3.6 Correlation analysis

After conducting comparative analyses on the existing data, we proceed to quantify users' background information and perform regression analysis by correlating it with their quantified sketching scores. The objective is to explore whether users' performance in the current experiment correlates with their backgrounds. The regression analysis employs three key background factors as independent variables: users' familiarity with the test environment, their sense of direction (SOD), and their familiarity with Mixed Reality (MR) technology. The dependent variables consist of users' quantified scores for both important and non-important landmarks, as well as their editing distances. For this regression analysis, several classic linear and non-linear models are utilized as tools for assessing correlations:

1. Multivariate Linear Regression: A simple and classic regression model used to investigate whether a linear relationship exists between the independent and dependent variables. This method employs the least squares approach to estimate the coefficient matrix, aiming to optimize the coefficients of the model. The estimation pattern for the coefficient matrix is as follows:
$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y \tag{4}$$
Where X represents the observed independent variables and y represents the dependent variable. In cases of multicollinearity among the observed variables, this method enables the estimation of the coefficient matrix to obtain an optimal solution (Trenkler et al., 1996).

2. Polynomial Regression: This method is analogous to multiple linear regression, with the distinction that it involves calculating the values of each independent variable from zeroth to the highest order based on the given degree. The coefficients matrix is then computed using the least squares method. Here we use 2 as our degree.

3. Ridge Regression: Also known as Tikhonov regularization. When solving overdetermined problems, the covariance matrix of the coefficient matrix may become close to or equal to a singular matrix. In such cases, using the least squares method can lead to instability. Therefore, in Ridge Regression, Tikhonov introduces a regularization term to mitigate bias:
$$J(\theta) = L(\theta, d) + \lambda R(\theta) \tag{5}$$
$\lambda R(\theta)$ represents the regularization term and the corresponding parameter. This approach helps mitigate issues related to singular values (Tikhonov, 1963).

4. Lasso Regression: It offers flexibility in covariate selection, inducing smoother changes in coefficient values compared to ridge regression. Lasso actively engages in feature selection and regularization, enhancing the interpretability of the model (Tibshirani, 1996).

5. Decision Tree: A non-linear method that partitions the current data into subsets using designated hyperplanes, selecting appropriate features for regression decisions.

To assess the regression outcomes, we employ the coefficient of determination, also known as the R-squared, as a tool to examine the model's performance. Its semantics involve examining the proportion of variance in the dependent variable that can be explained by the independent variables. This allows us to evaluate whether the current model possesses a certain level of explanatory power (Steel & Torrie, 1960). Our dependent variables consist of the important landmark scores, un-important landmark scores, and edit distance scores of both the experimental and control groups. Hence, six rounds of regression analyses are conducted. Initially, we construct a scatterplot matrix to observe the relationships between variables, as depicted in Figure 23. From the graph, it can be observed that there is no evident direct relationship between the variables. To better analyze the collinearity between independent and dependent variables and enhance the effectiveness of regression analysis, a correlation coefficient heatmap is constructed for numerical analysis.

From Figure 24, it can be observed that there is no significant collinearity relationship between any pair of independent variables. Therefore, various regression models can be attempted based on this observation, and the results are presented from Table 11 to 16. The final regression results indicate that the commonly used regression models perform poorly on the dataset. There is no clear correlation between the experimental group and the control group in relation to different dependent variables. Since the R Squared values are negative, the models' fit is even worse than that of random data. To further investigate whether participants' backgrounds are related to sketch quality, the Spearman correlation analysis is conducted to explore whether there is a significant relationship between independent and dependent variables.

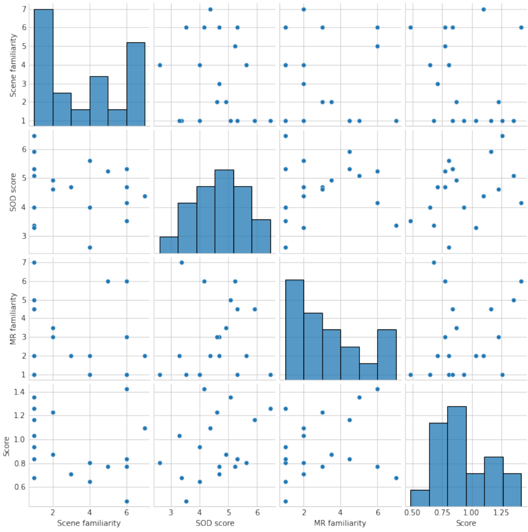
Table 11: Experimental Group: important landmarks

Model	Mean Squared Error	R-squared
Linear Regression	0.199	-3.508
Polynomial Regression	1.039	-22.509
Ridge Regression	0.188	-3.264
Lasso Regression	0.112	-1.542
Decision Tree Regression	0.134	-2.026

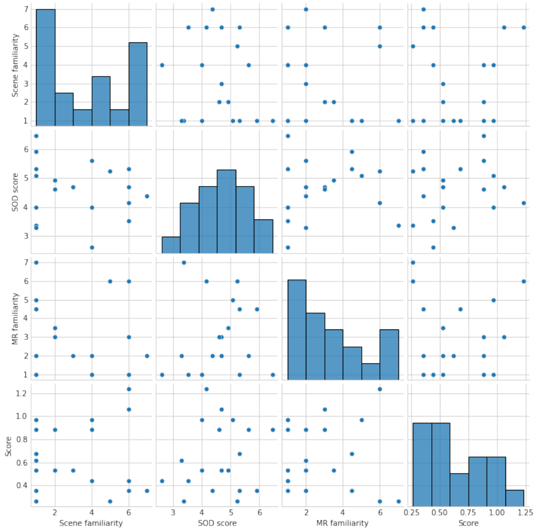
Table 12: Experimental Group: un-important landmarks

Model	Mean Squared Error	R-squared
Linear Regression	0.205	-2.582
Polynomial Regression	0.456	-6.963
Ridge Regression	0.184	-2.212
Lasso Regression	0.069	-0.207
Decision Tree Regression	0.235	-3.110

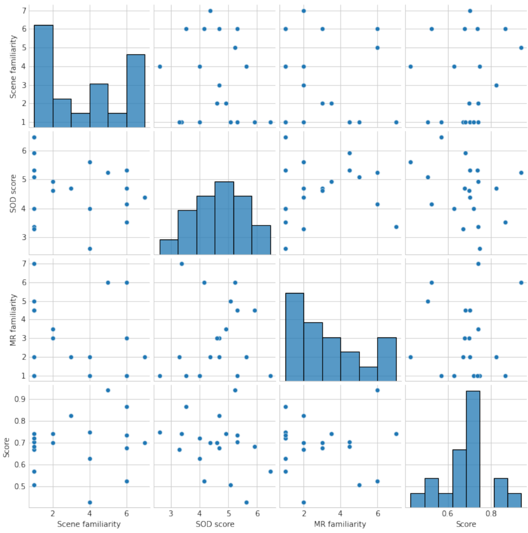
The Spearman correlation coefficient is a non-parametric measure of correlation used to assess whether there is a monotonic relationship between two variables. The Spearman correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative monotonic relationship, 1



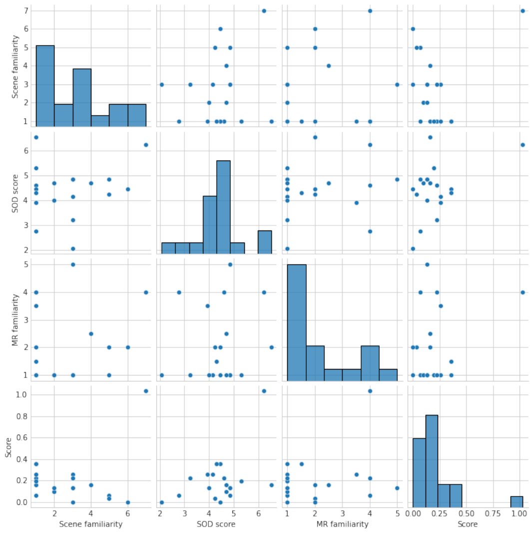
(a) Experiment group: Important landmark



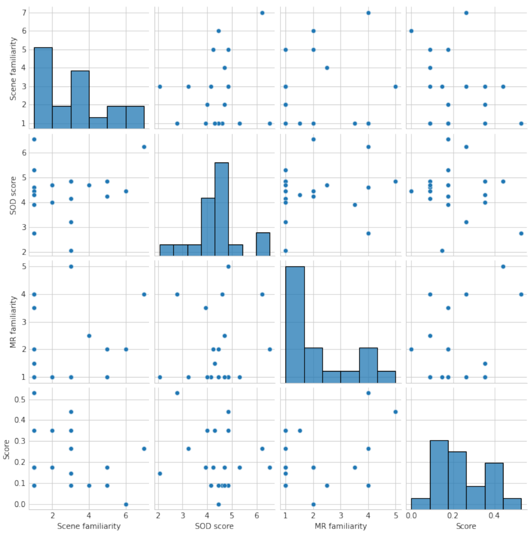
(b) Experiment group: Un-important landmark



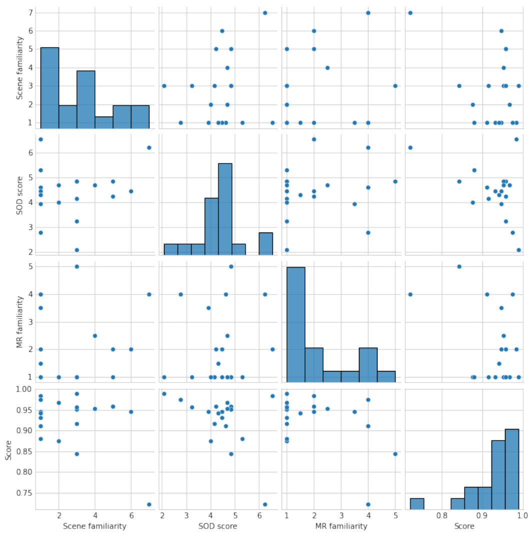
(c) Experiment group: Edit distance



(d) Control group: Important landmark

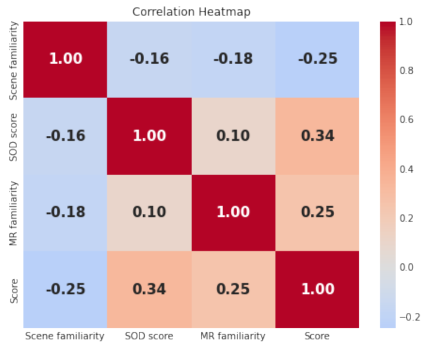


(e) Control group: Un-important landmark

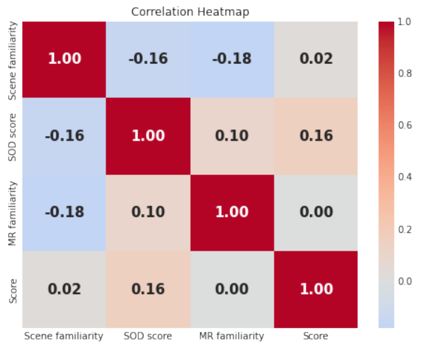


(f) Control group: Edit distance

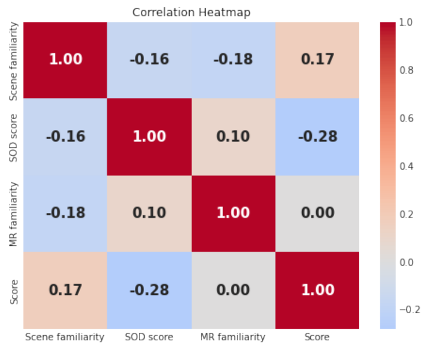
Figure 23: Scatterplot matrix of independent and dependent variables



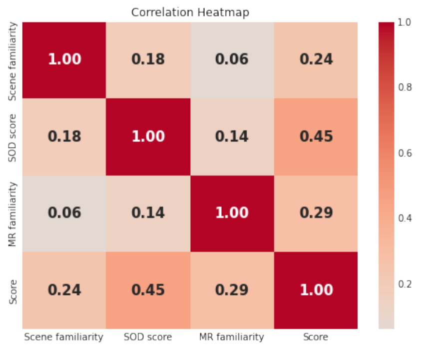
(a) Experiment group: Important landmark



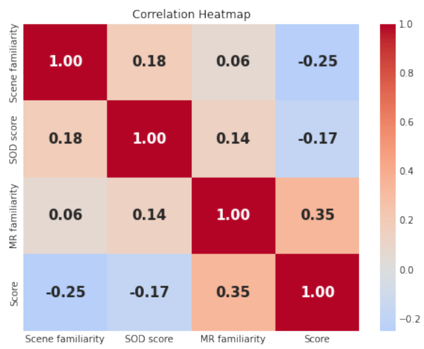
(b) Experiment group: Un-important landmark



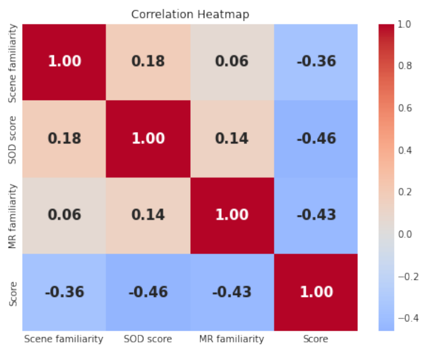
(c) Experiment group: Edit distance



(d) Control group: Important landmark



(e) Control group: Un-important landmark



(f) Control group: Edit distance

Figure 24: Correlation heatmap of independent and dependent variables

Table 13: Experimental Group: Edit distance

Model	Mean Squared Error	R-squared
Linear Regression	0.030	-0.779
Polynomial Regression	0.170	-9.233
Ridge Regression	0.028	-0.699
Lasso Regression	0.018	-0.106
Decision Tree Regression	0.029	-0.429

Table 14: Control Group: important landmarks

Model	Mean Squared Error	R-squared
Linear Regression	0.019	-3.170
Polynomial Regression	0.017	-2.582
Ridge Regression	0.019	-3.003
Lasso Regression	0.011	-1.262
Decision Tree Regression	0.031	-5.671

Table 15: Control Group: un-important landmarks

Model	Mean Squared Error	R-squared
Linear Regression	0.040	-7.722
Polynomial Regression	0.075	-15.491
Ridge Regression	0.038	-7.296
Lasso Regression	0.018	-2.999
Decision Tree Regression	0.079	-16.429

Table 16: Control Group: edit distance

Model	Mean Squared Error	R-squared
Linear Regression	0.001	-2.274
Polynomial Regression	0.002	-5.187
Ridge Regression	0.001	-2.207
Lasso Regression	0.001	-1.050
Decision Tree Regression	0.001	-2.738

indicates a perfect positive monotonic relationship, and 0 indicates no monotonic relationship. This correlation coefficient can also measure curved relationships (Daniel, 1990).

Table 17: Spearman correlation p-values

	Scene familiarity	SOD	MR familiarity
Experimental&Important	0.240	0.213	0.384
Experimental&Un-important	0.895	0.767	0.936
Experimental&Edit distance	0.486	0.248	0.830
Control&Important	0.146	0.491	0.715
Control&Un-important	0.362	0.810	0.422
Control&Edit distance	0.935	0.262	0.411

Consequently, the Spearman correlation coefficients p-value between each set of data's independent variables and dependent variables are computed, yielding Table 17. Under the consideration to set the threshold as 0.05, in the table, all of data exhibit weak monotonic correlations. Therefore, it can be asserted that within the AR-based spatial learning paradigm we have designed, improvements can be achieved in users with different scene familiarity, MR familiarity and SOD, while these enhancements have little correlation with users' backgrounds.

5 Discussion and Conclusion

5.1 Summary

In this paper, we first introduce the potential benefits of MR-based virtual landmarks for indoor navigation, positioning, and spatial cognition, emphasizing their role in facilitating spatial learning. However, the challenge of organizing intricate yet visually monotonous indoor spatial information arises, compounded by an overload of landmarks. Therefore, we introduce the Chunking method to organize and segment spatial information. After exploring spatial semantic chunking and spatial information chunking, we employ spatial classification and hierarchical visualization techniques to design our AR system. Following the system's design, a control group without spatial classification and hierarchical visualization is set up to assess the effectiveness of our system. Concurrently, we conduct experiments involving both experimental and control groups, each consisting of 19 participants, to examine the efficacy of spatial learning. Upon completing the experiments, data analysis is carried out. We comprehensively analyze participants' backgrounds, evaluate their spatial learning outcomes at both magnitude and spatial levels, and compare and reflect on their subsequent feedback. The experiment and corresponding data analysis validate that our current chunking method, which involves classification and layering, yields superior outcomes compared to scenarios without chunking. This approach effectively addresses the challenge of managing vast amounts of information and significantly enhances spatial learning. The study also involves correlational analysis between participants' backgrounds and their final outcomes, demonstrating the general applicability of our system's efficacy in enhancing spatial cognition across diverse user profiles.

5.2 Discussion

5.2.1 Explanation to the final result

Our study, including the design, experiments, and subsequent analysis, has convincingly showcased the effectiveness of AR systems employing spatial chunking in enhancing spatial learning when compared to AR visualization systems lacking information organization. However, delving into the underlying reasons for this phenomenon is imperative. It is widely accepted that the reorganization of intricate and disordered information can establish a structured foundation for spatial memory, thereby bolstering learning outcomes. Nevertheless, as we delve into Section 4.3.5, we stumble upon some intriguing observations.

Firstly, it is important to note that the spatial chunking methodology is intended to restructure existing spatial information, reducing information overload and enhancing learning capacity. Curiously, our experimental group reports elevated levels of stress compared to the control group. Despite this, the experimental group displays greater self-confidence in their performance, a trend consistent with our data analysis findings.

Building on this, Cheng (2017) assert that a low perceived information load does not necessarily translate into a positive learning intent. In our study, we distinguish between information load and the perception of information load as two distinct concepts. Information load signifies the current quantity of organized information, whereas the perception of information load pertains to the user's subjective experience of this information volume. This perspective aligns with İbili (2019), who argue that the primary role of AR is to reduce irrelevant information load while augmenting relevant information load, underscoring the critical importance of AR system design.

In the control group, nearly all landmarks adhered to a uniform design pattern. Pascoal and Guerreiro (2017) elucidate that inundating AR environments with excessive contextual information can render it meaningless, as AR's core objective is to emphasize relevant information. Consequently, in such scenarios, the control group may struggle to discern complex and disordered information within the current scene, possibly leading to a lack of focused thinking—a phenomenon akin to the Dunning-Kruger effect (Mazor & Fleming, 2021).

Contrastingly, even though the experimental group perceives a greater level of pressure, it is important to note that what they perceive is useful and well-organized information, affording them an advantage in their final performance. For the control group, users experience a sense of disorientation amidst the cluttered information. Consequently, they not only fail to perceive potential associations within the space but also exhibit a lack of inclination toward proactive learning. In contrast, the experimental group is capable of perceiving rational information cues generated by virtual landmarks. This engenders a proactive learning attitude, enabling them to swiftly comprehend the fundamental spatial structure and experience the acquisition of spatial knowledge. However, a judicious reduction in the perception of information load during the design phase is essential. An excessive perception of information load can have adverse effects on learning, ultimately diminishing learning efficiency (Chu, 2014; Wang, Fang, & Gu, 2020).

Simultaneously, block-wise exploration is also a noteworthy point for discussion. Firstly, concerning the edit distance scores, there are no significant differences observed among the various blocks. This is because, overall, even in the experimental group, the number of annotations for landmarks is quite low. In situations where the number of landmarks is low, spatial topological errors are less likely to occur, resulting in relatively stable edit distance scores across different blocks.

Regarding the semantic scores for important landmarks, the presence of a large number of landmarks in Block 2 leads to a performance decrease for both groups. Despite our efforts in information organization, users are still affected by a certain degree of information overload in this block, resulting in relatively poor performance. In contrast, the control group's scores are consistently low across all blocks, exhibiting poor performance in each one.

As for the semantic scores for non-important landmarks, we observe that both the control and experimental groups perform worse in Block 1. Block 1 lack higher scoring segments compared to the other blocks. Users tend to initially focus their attention on visually prominent primary landmarks when wearing AR devices, leading to a delay in attention to un-important landmarks.

Correlation analysis also provides a new perspective. We explore the influence of background factors on the final result assessment by fitting various commonly used linear and nonlinear models to examine the relationship between influencing factors and the final results. However, these models yield poor results. Subsequently, in Spearman's rank correlation test, it is observed that there is a lack of monotonic consistency in the relationship between the independent variables and the dependent variables. This outcome suggests that our current system design is not significantly influenced by user backgrounds, demonstrating the system's general applicability.

5.2.2 Achievements

Overall, we have successfully designed an indoor spatial learning assistance system based on the concept of spatial chunking. Subsequently, through comprehensive data analysis following the experiments, we have provided evidence supporting the effectiveness of our design. Our study has designed an AR-based indoor spatial learning assistance system with the aim of aiding individuals in organizing potentially related information within indoor spaces, thereby enhancing their ability for spatial learning. We employ the Chunking technique to flexibly segment the information present within the indoor environment. Different categories of objects are distinguished using various colors, with primary landmarks being highlighted in bright colors and non-primary landmarks depicted in lighter shades. Following the design and deployment of the system onto mobile AR devices, we conducted comparative experiments involving experimental and control groups. Subsequent data analysis after the experiments has demonstrated that key metrics that could potentially influence the experimental outcomes, such as spatial orientation, familiarity with the environment, and proficiency with Mixed Reality (MR), remain balanced between the experimental and control groups. Given this, we assign our designed AR system to users, allowing them to explore our experimental environment. We then evaluate users' spatial learning quality based on semantic landmark scoring of the obtained recall sketches and spatial edit distances. The study reveals that, regardless of whether it is for important landmarks or less significant ones, and whether it is in terms of semantic or spatial scoring, the experimental group's average scores surpass those of the control group, and statistical tests confirm that the differences are significant. This indicates that the incorporation of spatial chunking, as opposed to not using chunking, effectively enhances individuals' spatial learning efficiency. Simultaneously, we also conduct tests on users' memory of some inconspicuous landmarks, such as fire extinguishers. The results indicate that, in comparison to the control group, the experimental group can significantly remember more fire extinguishers. This suggests that the design of spatial chunking also demonstrates advantages in memory retention for small objects. Furthermore, feedback obtained through post-experiment surveys reveals that participants generally perceive AR as superior to traditional paper or electronic maps in the context of indoor spatial cognitive learning. We conclude our study with a correlation analysis. The results reveal that, for our AR system, several prominent background variables such as familiarity with MR, familiarity with the environment, and directional aptitude do not significantly impact the final quality of spatial memory. This suggests that our designed system effectively caters to users with varying backgrounds, substantiating its considerable versatility.

5.2.3 Limitations

However, our study still has certain limitations that necessitate further refinement and reflection in order to enhance its generalizability and deepen our insights:

1. The experiment requires a larger pool of participants. Although the current participants exhibit a degree of balance in several key background indicators, a broader array of volunteers in future research will introduce greater diversity in background characteristics. This rich source of data will assist us in conducting more diversified differential studies, exploring the enhancements offered by the Chunking-based AR spatial learning system. It will also enable us to assess the system's generalizability and universality.

2. Furthermore, our experimental environment is relatively limited. Our testing site represents a typical university campus professorial office area, which contains somewhat monotonous content. In future research, expansion to other types of environments should be considered, such as classroom areas within schools, exhibition areas, and the exploration of experiments in different types of large buildings such as malls and office complexes. Deploying AR systems designed based on Chunking principles in various architectural settings is beneficial for assessing the universality of our current approach and for addressing a range of potential new challenges.

3. The design of AR systems requires more careful consideration. For instance, in the aspect of color design, color is a vital tool for our chunking process. However, many users express significantly higher levels of dissatisfaction with color design compared to landmark and label design. Through subsequent feedback interviews, users suggest that using color depth to represent the effectiveness of primary landmarks in reflecting the hierarchical structure of landmarks may not be very effective. This is associated with real-world external factors such as natural light interference and the actual display issues of AR device screens. Therefore, system design should take into account practical deployment issues and attempt to use more robust elements to execute the Chunking concept.

4. Furthermore, a more nuanced approach to questionnaire design is essential for our research. Currently, our experiments predominantly rely on sketch-based assessments to evaluate the quality of users' spatial learning. However, it is evident that this approach has its limitations, particularly in providing a holistic understanding of spatial memory. To address this limitation, we need to incorporate a wider array of test questions that encompass diverse aspects of spatial cognition. This could include quantitative tests related to users' orientation and distance estimation abilities, as well as questions that delve into the cognitive processes underlying their spatial learning experiences. Such an approach would not only enrich our dataset but also enable us to gain deeper insights into the multifaceted phenomena associated with spatial learning in the context of augmented reality.

5. While our experimental group outperforms the control group in the memory of small objects, overall, users in both groups exhibit subpar memory retention for certain inconspicuous

landmarks. Subsequent research should endeavor to balance memory for significant and inconspicuous landmarks through more diverse design approaches.

It's worth considering the practical implications of our findings. As we refine our AR-guided spatial learning system, we should also contemplate real-world deployment challenges. For instance, the feedback received from users regarding the use of color for chunking raises important questions about the usability and visibility of AR systems in various environmental conditions. Exploring alternative, more robust design elements for implementing chunking may be a valuable avenue for future research. In conclusion, while our current research has provided valuable insights, there is still ample room for improvement and exploration. By refining our questionnaire design, expanding the scope of our assessments, and considering real-world deployment challenges, we can advance our understanding of AR-guided spatial learning and its practical implications.

5.2.4 Scene transferability

Our current research has achieved significant results, but it is limited to office areas within university campuses. Generalizing and extending our methodology to other large buildings would be beneficial, as it would allow our research to be applicable in different scenarios and promote the industrial implementation of AR-based spatial learning systems. Although we have not experimented with our methodology and system in other buildings, we still provide some guiding recommendations. We believe that modern architectural designs adhere to certain rules that aim to facilitate the functionality of buildings and rationalize the layout of indoor spaces, avoiding confusion for occupants. However, indoor spaces often lack global landmarks to guide individuals, and different areas tend to appear uniform, lacking distinctive and prominent features. As a result, some logically and contextually relevant spatial objects may not be immediately noticeable to people. Therefore, our task is to employ visualization techniques to present these relationships and reorganize spatial information based on the orientation abilities of different individuals, accommodating their cognitive load capacity. Here, we present several iconic public buildings and offer design suggestions based on our methodology:

1. Museum of Art and History: In art and history museums, guiding visitors through various exhibition halls is crucial. To achieve this, we can implement unique virtual landmarks and color coding to help visitors understand different art movements and historical periods. The application of spatial chunking can be employed to facilitate the recognition of distinct exhibition areas and enhance the visitors' experience. Notable virtual landmarks and color-coded virtual objects can be used to categorize and hierarchically organize information for each exhibition hall. For instance, in a history museum, a visual timeline can be used to guide users through the history exhibition hall, aiding in quickly grasping the organization of the exhibits. In an art museum, paintings from different art movements can be categorized at the first level, followed by chunking the works of corresponding artists, providing visitors with informative navigation.

2. Conference Center: Conference centers typically feature vast indoor spaces interconnected by various-sized meeting rooms. In such venues, participants need to swiftly locate their desired meetings and event spaces. To assist them, virtual signposts, gradient-color-coded visualizations of schedules, and annotations on meeting rooms can be created. These visual aids can provide users with an intuitive understanding of the proximity and timing of each event. A color-coding system can help users comprehend the spatial organization of the building.

3. Library: Libraries consist of diverse sections, reading rooms, and numerous bookshelves. Users often seek specific books or study areas. In this context, virtual landmarks can be placed in different reading zones, and interactive virtual maps can be integrated. As libraries typically have well-defined areas, some users may rely solely on the building's design to navigate. Thus, the system can emphasize visual assistance by providing rough global category landmarks to help users determine their general search direction. For users who prefer or require more intricate navigation, finer-grained landmarks can be offered for complex or design-dependent areas. Interactive maps, as well as navigation assistance at both coarse and fine levels, can cater to various user needs.

4. Shopping Mall: Shopping centers encompass malls of various sizes and multifaceted functional zones, with a constant flow of visitors, making location and navigation challenging. Shoppers often need to find restrooms, elevators, emergency exits, as well as specific stores and entertainment facilities promptly. Virtual icons and interactive navigation tools, accompanied by bold color zoning, can be employed to provide clear guidance. For instance, using translucent color blocks to cover different commercial or functional zones can offer users direct visual cues.

5. University Campus: On university campuses, students and visitors need to locate academic buildings, student centers, and dining areas. To facilitate this, virtual landmarks can be established, and color-coded virtual objects can be used to distinguish different academic departments and administrative offices, aiding efficient navigation around campus. Additionally, the complexity arising from multiple distinct buildings on a campus can be addressed by incorporating global chunking landmarks: for significant areas that serve as global orientation points, users can have global visibility.

By applying our methodology to these diverse public buildings, we aim to enhance spatial perception and cognitive mapping for users, making their experiences more intuitive and efficient within these spaces.

5.3 Conclusion

Our study introduces a spatial chunking approach to address the issue of information overload in indoor spatial learning through Augmented Reality (AR). We employ a color-based categorical and hierarchical visualization method, aiming at guiding individuals towards more efficient indoor spatial learning, while preserving finer details of the indoor environment.

Subsequent experimental results indicate a significant improvement in users' spatial learning efficiency when employing the spatial chunking method, as opposed to a naive and uniform virtual landmark design. This finding underscores the critical role of information organization in AR-based learning and highlights the impact of information load management on learning outcomes. Moreover, our correlation analysis suggests that the enhancement in spatial learning performance is minimally correlated with user background factors, thereby supporting the universality of our research findings.

Although the study has identified some shortcomings in aspects like the design of virtual landmark colors and lack of efficiency to assist un-important landmark learning, our study offers valuable insights for enhancing practical indoor spatial learning and provides a foundational framework for future investigations into the intricate relationships among information load, information organization, and learning outcomes. It also offers a forward-looking methodological approach for conducting similar experiments in different contexts. We encourage researchers and designers in the spatial learning field of AR technology to consider the adoption of the spatial chunking approach to optimize AR learning experiences and consequently improve users' learning outcomes.

6 Appendix

6.1 Questionnaire

Informed Consent Form

Dear Participant,

thank you for your participation in the study.

In this study, you’ll be asked to perform certain tasks wearing Microsoft HoloLens and to answer some questions. The whole process will take around 30-60 minutes. You are free to stop, quit the study and retract your data at any time during the study with no further consequences. **Privacy:** Original data obtained from this study will be anonymized and only processed in aggregate. In such form, it might be published in academic journals, presentations or other media, but never in a way that would allow individual identification. One week after the study it might no longer be possible to retract your data from such aggregated analyses.

If you have any questions, please contact Jiongyan Zhang via jiongyan.zhang@tum.de.

▢ I confirm I volunteered to participate in this study.
▢ I confirm I was allowed to ask questions and that I was provided with responses.
▢ I confirm I was presented with this document prior to the beginning of the study.
▢ I confirm and I understood my right to quit the study at any time.
▢ I confirm I was informed that a conversation with researcher will be recorded after navigation during the study.

Date

Signature of researcher

Signature of participant

If you would like to be informed about future studies, please let us know your email address:

Email address (optional)

Pre-task questionnaire

Gender: ☐Male ☐Female ☐Other Nationality: Age:

Field you are working/studying in:

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. Circle "1" if you strongly agree that the statement applies to you, "7" if you strongly disagree, or some number in between if your agreement is intermediate. Circle "4" if you neither agree nor disagree.

1= strongly agree / none, 7 = strongly disagree / a lot of

1	I am very good at giving directions.	strongly agree	1	2	3	4	5	6	7	strongly disagree
2	I have a poor memory for where I left things.	strongly agree	1	2	3	4	5	6	7	strongly disagree
3	I am very good at judging distances.	strongly agree	1	2	3	4	5	6	7	strongly disagree
4	My “sense of direction” is very good.	strongly agree	1	2	3	4	5	6	7	strongly disagree
5	I tend to think of my environment in terms of cardinal directions(N, S, E, W).	strongly agree	1	2	3	4	5	6	7	strongly disagree
6	I very easily get lost in a new city.	strongly agree	1	2	3	4	5	6	7	strongly disagree
7	I enjoy reading maps.	strongly agree	1	2	3	4	5	6	7	strongly disagree
8	I have trouble understanding directions.	strongly agree	1	2	3	4	5	6	7	strongly disagree
9	I am very good at reading maps.	strongly agree	1	2	3	4	5	6	7	strongly disagree
10	I don’t remember routes very well while riding as a passenger in a car.	strongly agree	1	2	3	4	5	6	7	strongly disagree
11	I don’t enjoy giving directions.	strongly agree	1	2	3	4	5	6	7	strongly disagree
12	It’s not important to me to know where I am.	strongly agree	1	2	3	4	5	6	7	strongly disagree
13	I usually let someone else do the navigational planning for long trips.	strongly agree	1	2	3	4	5	6	7	strongly disagree
14	I can usually remember a new route after I have traveled it only once.	strongly agree	1	2	3	4	5	6	7	strongly disagree
15	I don’t have a very good “mental map” of my environment.	strongly agree	1	2	3	4	5	6	7	strongly disagree
16	I usually get lost indoors.	strongly agree	1	2	3	4	5	6	7	strongly disagree
17	I usually get lost outdoors.	strongly agree	1	2	3	4	5	6	7	strongly disagree
18	I have __ experience with Augmented Reality.	none	1	2	3	4	5	6	7	a lot of
19	I have __ experience with Virtual Reality.	none	1	2	3	4	5	6	7	a lot of

TASK1: Please draw and label all the spatial information you remember, including rooms, room names, stereoscopes, fire distinguisher, garbage can and anything else you can recall.

TASK 2: Please complete the following questionnaire ("1" = not at all, "7" = very much).

COLORS stands for the color of landmarks including transparency, LABELS stands for the label text and icon, LANDMARKS stands for the shape, position and orientation etc.

1. Before this walking, how familiar were you with the study area?

   not at all 1 2 3 4 5 6 7 very much

2. How do you like the device/hardware?

   not at all 1 2 3 4 5 6 7 very much

3. How do you like the interface?

   not at all 1 2 3 4 5 6 7 very much

4. How do you like each element you saw?

Colors	not at all	1	2	3	4	5	6	7	very much
Labels	not at all	1	2	3	4	5	6	7	very much
Landmarks	not at all	1	2	3	4	5	6	7	very much

5. How do you think each element helps you remember the area?

Colors	not at all	1	2	3	4	5	6	7	very much
Labels	not at all	1	2	3	4	5	6	7	very much
Landmarks	not at all	1	2	3	4	5	6	7	very much

6. How do you think each element helps you locate your position?

Colors	not at all	1	2	3	4	5	6	7	very much
Labels	not at all	1	2	3	4	5	6	7	very much
Landmarks	not at all	1	2	3	4	5	6	7	very much

7. How do you think each element helps you identify your orientation?

Colors	not at all	1	2	3	4	5	6	7	very much
Labels	not at all	1	2	3	4	5	6	7	very much
Landmarks	not at all	1	2	3	4	5	6	7	very much

8. How do you think of the spatial information overload in this system?

   not at all 1 2 3 4 5 6 7 very severe

9. Do you think the landmark design has a certain sense of hierarchy and classification?

   not at all 1 2 3 4 5 6 7 very much

10. Do you think this kind of sense of hierarchy and classification assists you to remember the space?

    not at all 1 2 3 4 5 6 7 very much

11. To what extent does the landmark design help you to judge the distance between landmarks?

not at all	1	2	3	4	5	6	7	very much

12. To what extent does the landmark interfere with your eyesight?

not at all	1	2	3	4	5	6	7	very much

13. To what extent does the landmark interfere with your observation to the real world?

not at all	1	2	3	4	5	6	7	very much

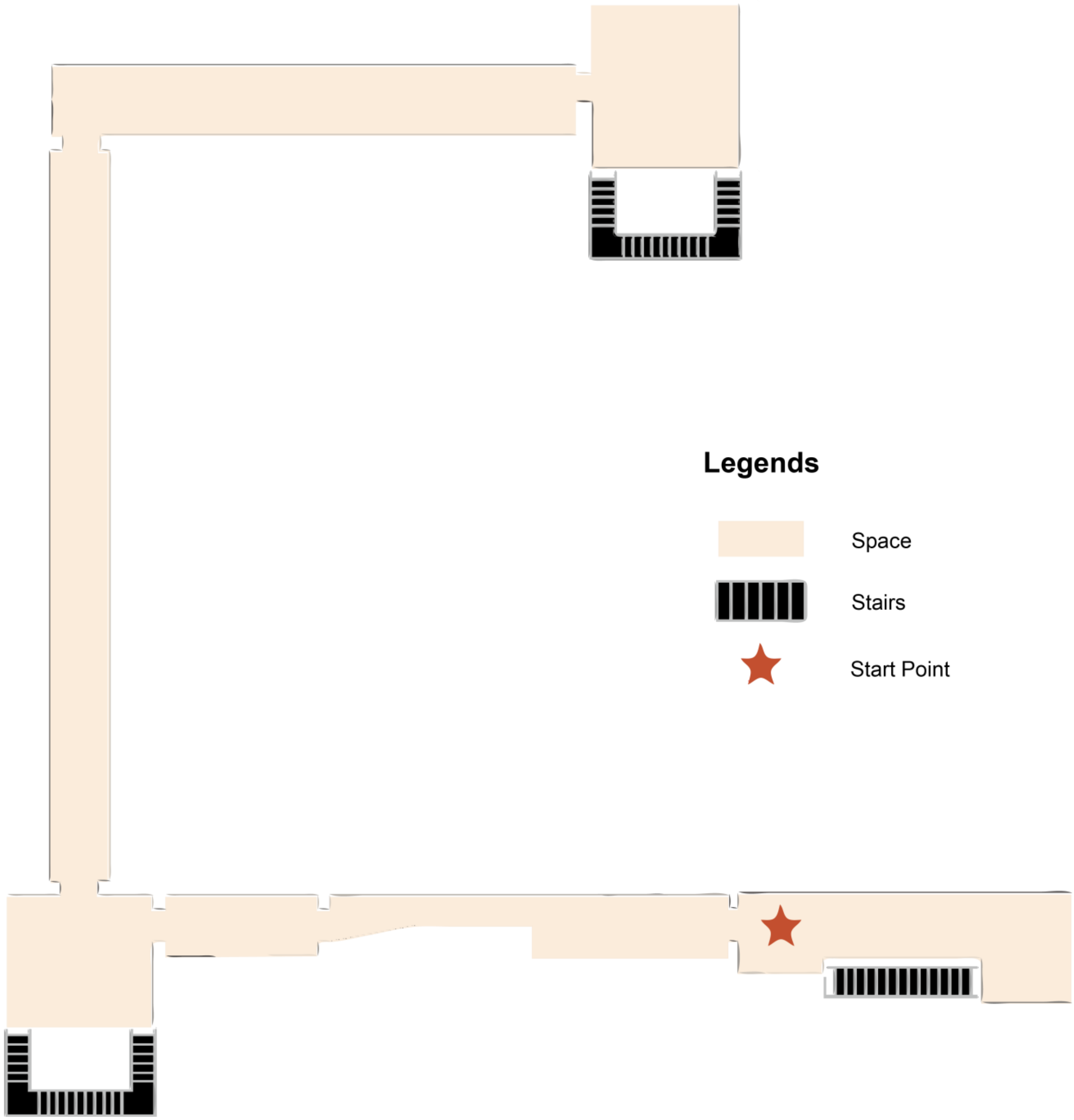
14. To what extent does the design help you with understanding the local spatial layout?

not at all	1	2	3	4	5	6	7	very much

15. Please compare your preference between AR and map: which help you understand the indoor space more?

Map	not at all	1	2	3	4	5	6	7	very much
AR	not at all	1	2	3	4	5	6	7	very much

TASK 3: Just imagine there is a fire alarm, and you need to extinguish a controllable fire: where do you think you could find the fire extinguisher?



Please use different symbols:
Spatial Learning Section: "X"
Task Quiz Section: "O"

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses workload on five 7-points scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

1. Mental Demand: How mentally demanding was the task?

Very Low | Very High

2. Physical Demand: How physically demanding was the task?

Very Low | Very High

3. Temporal Demand: how hurried and rushed was the pace of the task?

Very Low | Very High

4. Performance: How successful were you in accomplishing what you were asked to do?

Perfect | Failure

5. Effort: How hard did you have to work to accomplish your level of performance?

Very Low | Very High

6. Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low | Very High

Please confirm the authenticity of data in this document (including pre-task questionnaire, post task questionnaire and task quiz).

▢ I confirm the authenticity of all data in the document (including pre-task questionnaire, post task questionnaire and task quiz).

Date	Signature of researcher
	Signature of participant

References

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational research review*, *20*, 1–11.

Altieri, L., Cocchi, D., & Roli, G. (2018). A new approach to spatial entropy measures. *Environmental and ecological statistics*, *25*, 95–110.

Altieri, L., Cocchi, D., & Roli, G. (2019). Advances in spatial entropy measures. *Stochastic Environmental Research and Risk Assessment*, *33*, 1223–1240.

Anastopoulou, N., Kokla, M., Tomai, E., Cheliotis, K., Liarokapis, F., Pastra, K., & Kavouras, M. (2023). Cartographic perspectives on spatial and thematic levels of detail in augmented reality: a review of existing approaches. *International Journal of Cartography*, 1–19.

Bao, M., & Engel, S. A. (2019). Augmented reality as a tool for studying visual plasticity: 2009 to 2018. *Current Directions in Psychological Science*, *28*(6), 574–580.

Batty, M. (1974). Spatial entropy. *Geographical analysis*, *6*(1), 1–31.

Billinghurst, M., & Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings virtual reality annual international symposium'95* (pp. 40–47).

Blaser, A. (1998). Geo-spatial sketches. *University of Maine, Department of Spatial Information Science and Engineering and National Center for Geographic Information and Analysis*.

Bobrich, J., & Otto, S. (2002). Augmented maps. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, *34*(4), 502–505.

Boeing, G. (2019). Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, *4*(1), 1–19.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*(5), 467.

Brügger, A., Richter, K.-F., & Fabrikant, S. I. (2019). How does navigation system behavior influence human behavior? *Cognitive research: principles and implications*, *4*, 1–22.

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303.

Burte, H., & Montello, D. R. (2017). How sense-of-direction and learning intentionality relate to spatial knowledge acquisition in the environment. *Cognitive Research: Principles and Implications*, *2*, 1–17.

Buschke, H. (1976). Learning is organized by chunking. *Journal of Verbal Learning and Verbal Behavior*, *15*(3), 313–324.

Cai, S., Chiang, F.-K., & Wang, X. (2013). Using the augmented reality 3d technique for a convex imaging experiment in a physics course. *International Journal of Engineering Education*, *29*(4), 856–865.

Calmet, X., & Calmet, J. (2005). Dynamics of the fisher information metric. *Physical Review E*, *71*(5), 056109.

Capozzi, F., Lorizzo, V., Modoni, G., & Sacco, M. (2014). Lightweight augmented reality tools for lean procedures in future factories. In *Augmented and virtual reality: First international conference, avr 2014, lecce, italy, september 17-20, 2014, revised selected papers 1* (pp. 232–246).

Carbonell Carrera, C., & Bermejo Asensio, L. A. (2017a). Augmented reality as a digital teaching environment to develop spatial thinking. *Cartography and geographic information science*, *44*(3), 259–270.

Carbonell Carrera, C., & Bermejo Asensio, L. A. (2017b). Landscape interpretation with augmented reality and maps to improve spatial orientation skill. *Journal of Geography in Higher Education*, *41*(1), 119–133.

Carbonell-Carrera, C., Jaeger, A. J., & Shipley, T. F. (2018). 2d cartography training: has the time come for a paradigm shift? *ISPRS International Journal of Geo-Information*, *7*(5), 197.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55–81.

Chen, J., Zhou, Y., & Zhai, J. (2023). Incorporating ar/vr-assisted learning into informal science institutions: A systematic review. *Virtual Reality*, 1–17.

Cheng, K.-H. (2017). Reading an augmented reality book: An exploration of learners' cognitive load, motivation, and attitudes. *Australasian Journal of Educational Technology*, *33*(4).

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of science education and technology*, *22*, 449–462.

Chipofya, M., Wang, J., & Schwering, A. (2011). Towards cognitively plausible spatial representations for sketch map alignment. In *Spatial information theory: 10th international conference, cosit 2011, belfast, me, usa, september 12-16, 2011. proceedings 10* (pp. 20–39).

Chu, H.-C. (2014). Potential negative effects of mobile learning on students' learning achievement and cognitive load—a format assessment perspective. *Journal of Educational Technology & Society*, *17*(1), 332–344.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

Daniel, W. (1990). Spearman rank correlation coefficient. applied nonparametric statistics (2nd edn). pws-kent, boston. 358–365.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193–222.

Dickmann, F., Keil, J., Dickmann, P. L., & Edler, D. (2021). The impact of augmented reality techniques on cartographic visualization. *KN J. Cartogr. Geogr. Inf*, *71*, 285–295.

Diegmann, P., Schmidt-Kraepelin, M., Eynden, S., & Basten, D. (2015). Benefits of augmented reality in educational environments-a systematic literature review.

Dong, W., Qin, T., Yang, T., Liao, H., Liu, B., Meng, L., & Liu, Y. (2022). Wayfinding behavior and spatial knowledge acquisition: Are they the same in virtual reality and in real-world environments? *Annals of the American Association of Geographers*, *112*(1), 226–246.

Dong, W., Wu, Y., Qin, T., Bian, X., Zhao, Y., He, Y., . . . Yu, C. (2021). What is the difference between augmented reality and 2d navigation electronic maps in pedestrian wayfinding? *Cartography and Geographic Information Science*, *48*(3), 225–240.

Fang, Z., Li, Q., & Shaw, S.-L. (2015). What about people in pedestrian navigation? *Geo-Spatial Information Science*, *18*(4), 135–150.

Fisher, R. A. (1970). Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution* (pp. 66–70). Springer.

Fonollosa, J., Neftci, E., & Rabinovich, M. (2015). Learning of chunking sequences in cognition and behavior. *PLoS computational biology*, *11*(11), e1004592.

Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2016). Gardony map drawing analyzer: Software for quantitative analysis of sketch maps. *Behavior research methods*, *48*, 151–177.

Gobet, F. (2017). Entrenchment, gestalt formation, and chunking.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in cognitive sciences*, *5*(6), 236–243.

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2018). Applications of augmented reality in informal science learning sites: A review. *Journal of Science Education and Technology*, *27*, 433–447.

Goldiez, B. F., Ahmad, A. M., & Hancock, P. A. (2007). Effects of augmented reality display settings on human wayfinding performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *37*(5), 839–845.

Gramann, K., Hoepner, P., & Karrer-Gauss, K. (2017). Modified navigation instructions for spatial navigation assistance systems lead to incidental spatial learning. *Frontiers in psychology*, *8*, 193.

Guo, H.-J., & Prabhakaran, B. (2022). Hololens 2 technical evaluation as mixed reality guide. *arXiv preprint arXiv:2207.09554*.

Hammady, R., Ma, M., Strathern, C., & Mohamad, M. (2020). Design and development of a spatial mixed reality touring guide to the egyptian museum. *Multimedia Tools and Applications*, *79*, 3465–3494.

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Hedge, C., Weaver, R., & Schnall, S. (2017). Spatial learning and wayfinding in an immersive environment: the digital fulldome. *Cyberpsychology, Behavior, and Social Networking*, *20*(5), 327–333.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425–447.

Hirtle, S. C., & Bahm, C. R. (2015). Cognition for the navigation of complex indoor environments. *Indoor wayfinding and navigation*, 1–12.

Ho, T.-W., Tsai, C.-J., Hsu, C.-C., Chang, Y.-T., & Lai, F. (2017). Indoor navigation and physician-patient communication in emergency department. In *Proceedings of the 3rd international conference on communication and information processing* (pp. 92–98).

Huang, T.-C., Chen, C.-C., & Chou, Y.-W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, *96*, 72–82.

Huang, T.-C., Shu, Y., Yeh, T.-C., & Zeng, P.-Y. (2016). Get lost in the library? an innovative application of augmented reality and indoor positioning technologies. *The Electronic Library*, *34*(1), 99–115.

Huynh, N. T., Doherty, S., & Sharpe, B. (2010). Gender differences in the sketch map creation process. *Journal of Maps*, *6*(1), 270–288.

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for stem learning: A systematic review. *Computers & Education*, *123*, 109–123.

İbili, E. (2019). Effect of augmented reality environments on cognitive load: pedagogical effect, instructional design, motivation and interaction interfaces. *International Journal of Progressive Education*, *15*(5), 42–57.

Ishikawa, T. (2021). Spatial thinking, cognitive mapping, and spatial awareness. *Cognitive Processing*, *22*(Suppl 1), 89–96.

Ishikawa, T., & Newcombe, N. S. (2021). *Why spatial is special in education, learning, and everyday activities* (Vol. 6). Springer.

Jitendra, M. S., Srinivas, A. S., Surendra, T., Rao, R. V., & Chowdary, P. R. (2021). A study on game development using unity engine. In *Aip conference proceedings* (Vol. 2375).

Jones, G. (2012). Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in psychology*, *3*, 167.

Joshi, R., Hiwale, A., Birajdar, S., & Gound, R. (2020). Indoor navigation with augmented reality. In *Iccce 2019: Proceedings of the 2nd international conference on communications and cyber physical engineering* (pp. 159–165).

Juan, M.-C., Mendez-Lopez, M., Perez-Hernandez, E., & Albiol-Perez, S. (2014). Augmented reality for the assessment of children's spatial memory in real settings. *PloS one*, *9*(12), e113751.

Kahn, M. J., Tan, K. C., & Beaton, R. J. (1990). Reduction of cognitive workload through information chunking. In *Proceedings of the human factors society annual meeting* (Vol. 34, pp. 1509–1513).

Kalyuga, S. (2011). Informing: A cognitive load perspective. *Informing Science*, *14*, 33.

Keil, J., Edler, D., & Dickmann, F. (2019). Preparing the hololens for user studies: An augmented reality interface for the spatial adjustment of holographic objects in 3d indoor environments. *KN-Journal of Cartography and Geographic Information*, *69*(3), 205–215.

Kim, J., & Jun, H. (2008). Vision-based location positioning using augmented reality for indoor navigation. *IEEE Transactions on Consumer Electronics*, *54*(3), 954–962.

Kim, K., & Bock, O. (2021). Acquisition of landmark, route, and survey knowledge in a wayfinding task: in stages or in parallel? *Psychological Research*, *85*(5), 2098–2106.

Klippel, A., Hansen, S., Richter, K.-F., & Winter, S. (2009). Urban granularities—a data structure for cognitively ergonomic route directions. *GeoInformatica*, *13*, 223–247.

Krukar, J., Anacta, V. J., & Schwering, A. (2020). The effect of orientation instructions on the recall and reuse of route and survey elements in wayfinding descriptions. *Journal of environmental psychology*, *68*, 101407.

Krukar, J., van Eek, A., & Schwering, A. (2023). Task-dependent sketch maps. *Spatial Cognition & Computation*, 1–30.

Küçük, S., Yỳlmaz, R. M., & Gökta, Y. (2014). Augmented reality for learning english: Achievement, attitude and cognitive load levels of students. *Education & Science/Egitim ve Bilim*, *39*(176).

Lai, A.-F., Chen, C.-H., & Lee, G.-Y. (2019). An augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology*, *50*(1), 232–247.

Laird, J. E., Rosenbloom, P. S., & Newell, A. (1984). Towards chunking as a general learning mechanism. In *Aaai* (pp. 188–192).

Lee, S. A. (2023). Navigational roots of spatial and temporal memory structure. *Animal Cognition*, *26*(1), 87–95.

Levene, H. (1960). Robust tests for equality of variances. *Contributions to probability and statistics*, 278–292.

Levene, H., et al. (1960). Contributions to probability and statistics. *Essays in honor of Harold Hotelling*, *278*, 292.

Liu, B., Ding, L., & Meng, L. (2021). Spatial knowledge acquisition with virtual semantic landmarks in mixed reality-based indoor navigation. *Cartography and Geographic Information Science*, *48*(4), 305–319.

Liu, B., Ding, L., Wang, S., & Meng, L. (2022a). Designing mixed reality-based indoor navigation for user studies. *KN-Journal of Cartography and Geographic Information*, *72*(2), 129–138.

Liu, B., Ding, L., Wang, S., & Meng, L. (2022b). Misleading effect and spatial learning in head-mounted mixed reality-based navigation. *Geo-spatial Information Science*, 1–15.

Liu, B., & Meng, L. (2020). Doctoral colloquium—towards a better user interface of augmented reality based indoor navigation application. In *2020 6th international conference of the* *immersive learning research network (ilrn)* (pp. 392–394).

Liu, B., & Zhan, Z. (2021). Work-in-progress——improve spatial learning by chunking navigation instructions in mixed reality. In *2021 7th international conference of the immersive learning research network (ilrn)* (pp. 1–3).

Liu, Q., Yu, S., Chen, W., Wang, Q., & Xu, S. (2021). The effects of an augmented reality based magnetic experimental tool on students' knowledge improvement and cognitive load. *Journal of computer assisted learning*, *37*(3), 645–656.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.

May, A. J., & Ross, T. (2006). Presence and quality of navigational landmarks: effect on driver performance and implications for design. *Human factors*, *48*(2), 346–361.

Mazor, M., & Fleming, S. M. (2021). The dunning-kruger effect revisited. *Nature Human Behaviour*, *5*(6), 677–678.

Meck, W. H., & Williams, C. L. (1997). Perinatal choline supplementation increases the threshold for chunking in spatial memory. *Neuroreport*, *8*(14), 3053–3059.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Miller Koop, M., Rosenfeldt, A. B., Owen, K., Penko, A. L., Streicher, M. C., Albright, A., & Alberts, J. L. (2022). The microsoft hololens 2 provides accurate measures of gait, turning, and functional mobility in healthy adults. *Sensors*, *22*(5), 2009.

Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. *Annals of cardiac anaesthesia*, *22*(4), 407.

Moore, T., & Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual review of psychology*, *68*, 47–72.

Mulloni, A., Seichter, H., & Schmalstieg, D. (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* (pp. 212–215).

Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological review*, *125*(4), 486.

Nuhn, E., & Timpf, S. (2022). Landmark weights-an alternative to spatial distances in shortest route algorithms. *Spatial Cognition & Computation*, 1–27.

Ong, S., Siddaraju, V. K., Ong, S., & Siddaraju, V. K. (2021). Introduction to the mixed reality toolkit. *Beginning Windows Mixed Reality Programming: For HoloLens and Mixed Reality Headsets*, 85–110.

Palumbo, A. (2022). Microsoft hololens 2 in medical and healthcare context: State of the art and future prospects. *Sensors*, *22*(20), 7709.

Park, B. J., Hunt, S. J., Nadolski, G. J., & Gade, T. P. (2020). Augmented reality improves procedural efficiency and reduces radiation dose for ct-guided lesion targeting: a phantom study using hololens 2. *Scientific Reports*, *10*(1), 18620.

Pascoal, R. M., & Guerreiro, S. L. (2017). Information overload in augmented reality: The outdoor sports environments. In *Information and communication overload in the digital age* (pp. 271–301). IGI Global.

Pose-Díez-de-la Lastra, A., Moreta-Martinez, R., García-Sevilla, M., García-Mato, D., Calvo-Haro, J. A., Mediavilla-Santos, L., ... Pascau, J. (2022). Hololens 1 vs. hololens 2: improvements in the new model for orthopedic oncological interventions. *Sensors*, *22*(13), 4915.

Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and ubiquitous computing*, *18*, 1533–1543.

Reddy, S. M., Chakrabarti, D., & Karmakar, S. (2012). Emotion and interior space design: an ergonomic perspective. *Work*, *41*(Supplement 1), 1072–1078.

Rokhsaritalemi, S., Sadeghi-Niaraki, A., & Choi, S.-M. (2020). A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, *10*(2), 636.

Ruginski, I. T., Creem-Regehr, S. H., Stefanucci, J. K., & Cashdan, E. (2019). Gps use negatively affects environmental learning through spatial transformation abilities. *Journal of Environmental Psychology*, *64*, 12–20.

Sahin, D., & Yilmaz, R. M. (2020). The effect of augmented reality technology on middle school students' achievements and attitudes towards science education. *Computers & Education*, *144*, 103710.

Sargent, J., Dopkins, S., Philbeck, J., & Chichka, D. (2010). Chunking in spatial memory. *Journal of Experimental Psychology: Learning, memory, and cognition*, *36*(3), 576.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., & Lin, N. (2019). Event memory uniquely predicts memory for large-scale space. *Memory & cognition*, *47*, 212–228.

Schmalstieg, D., & Reitmayr, G. (2007). Augmented reality as a medium for cartography. In *Multimedia cartography* (pp. 267–281). Springer.

Schwering, A., Krukar, J., Manivannan, C., Chipofya, M., & Jan, S. (2022). Generalized, inaccurate, incomplete: how to comprehensively analyze sketch maps beyond their metric correctness. In *15th international conference on spatial information theory (cosit 2022).*

Schwering, A., Wang, J., Chipofya, M., Jan, S., Li, R., & Broelemann, K. (2014). Sketchmapia: Qualitative representations for the alignment of sketch and metric maps. *Spatial cognition & computation*, *14*(3), 220–254.

Shank, P. (2018). What research tells us about chunking content. *ELearning Industry*.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Skubic, M., Blisard, S., Bailey, C., Adams, J. A., & Matsakis, P. (2004). Qualitative analysis of sketched route maps: translating a sketch into linguistic descriptions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *34*(2), 1275–1282.

Steel, R. G. D., & Torrie, J. H. (1960). Principles and procedures of statistics: with special reference to the biological sciences. *(No Title)*.

Stieff, M., Werner, S., DeSutter, D., Franconeri, S., & Hegarty, M. (2020). Visual chunking as a strategy for spatial thinking in stem. *Cognitive research: principles and implications*, *5*, 1–15.

Stites, M. C., Matzen, L. E., & Gastelum, Z. N. (2020). Where are we going and where have we been? examining the effects of maps on spatial learning in an indoor guided navigation task. *Cognitive Research: Principles and Implications*, *5*, 1–26.

Suh, J., & Cho, J. Y. (2021). A triangular relationship of visual attention, spatial ability, and creative performance in spatial design: An exploratory case study. *Journal of Interior Design*, *46*(3), 11–27.

Sylaiou, S., Mania, K., Liarokapis, F., White, M., Walczak, K., Wojciechowski, R., . . . Patias, P. (2015). Evaluation of a cultural heritage augmented reality game. *Cartographies of Mind, Soul and Knowledge*, 153–174.

Tan, S. Y., Lee, K. J., & Lam, M. C. (2020). A shopping mall indoor navigation application using wi-fi positioning system. *Int. J*, *9*, 4483–4489.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Sov Dok*, *4*, 1035–1038.

Trenkler, G., et al. (1996). Methods of multivariate analysis. wiley series in probability and mathematical statistics: Probability and mathematical statistics section: Alvin c. rencher (1995): Isbn 0471-571-520,[pound sign] 58.00, xvi+ 627 pp. *Computational Statistics & Data Analysis*, *22*(3), 334–335.

Tu Huynh, N., & Doherty, S. T. (2007). Digital sketch-map drawing as an instrument to collect data about spatial cognition. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *42*(4), 285–296.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99–114.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*, *139*(2), 352.

Verbelen, T., De Tinguy, D., Mazzaglia, P., Çatal, O., & Safron, A. (2022). Chunking space and time with information geometry..

Wang, C., Fang, T., & Gu, Y. (2020). Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia. *Computers & Education*, *143*, 103683.

Wen, J., Deneka, A., Helton, W., & Billinghurst, M. (2014). Really, it's for your own good... making augmented reality navigation tools harder to use. In *Chi'14 extended abstracts on human factors in computing systems* (pp. 1297–1302).

Yoon, S. A., & Wang, J. (2014). Making the invisible visible in science museums through augmented reality devices. *TechTrends*, *58*, 49–55.

Zhang, D., Ding, Y., Stegall, J., & Mo, L. (2012). The effect of visual-chunking-representation accommodation on geometry testing for students with math disabilities. *Learning Disabilities Research & Practice*, *27*(4), 167–177.

Zhao, J., Simpson, M., Wallgrün, J. O., Sajjadi, P., & Klippel, A. (2020). Exploring the effects of geographic scale on spatial learning. *Cognitive research: principles and implications*, *5*, 1–18.

Zhou, Q., & Liu, Z. (2021). Research on the optimization strategy of shopping mall spatial layout in hefei based on space syntax theory. *Complexity*, *2021*, 1–9.