Check for updates

Magic NeRF lens: interactive fusion of neural radiance fields for virtual facility inspection

Ke Li[1,2], Susanne Schmidt[2], Tim Rolff[2], Reinhard Bacher[1], Wim Leemans[1] and Frank Steinicke[2]*

[1]Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany, [2]Human-Computer Interaction Group, Universität Hamburg, Hamburg, Germany

Virtual reality (VR) has become an important interactive visualization tool for various industrial processes including facility inspection and maintenance. The capability of a VR application to present users with realistic simulations of complex systems and immerse them in, for example, inaccessible remote environments is often essential for using VR in real-world industrial domains. While many VR solutions have already been developed to support virtual facility inspection, previous systems provide immersive visualizations only with limited realism, because the real-world conditions of facilities are often difficult to reconstruct with accurate meshes and point clouds or typically too time-consuming to be consistently updated in computer-aided design (CAD) software toolkits. In this work, we present *Magic NeRF Lens*, a VR framework that supports immersive photorealistic visualizations of complex industrial facilities leveraging the recent advancement of neural radiance fields (NeRF). We introduce a data fusion technique to merge a NeRF model with the polygonal representation of it's corresponding CAD model, which optimizes VR NeRF rendering through magic-lens-style interactions while introducing a novel industrial visualization design that can support practical tasks such as facility maintenance planning and redesign. We systematically benchmarked the performance of our framework, investigated users' perceptions of the magic-lens-style visualization design through a visual search experiment to derive design insights, and performed an empirical evaluation of our system through expert reviews. To support further research and development of customized VR NeRF applications, the source code of the toolkit was made openly available.

KEYWORDS

virtual reality, neural radiance field, data fusion, human-computer interaction, extended reality toolkit

1 Introduction

The rapid development of virtual reality (VR) technology has enormous potential for supporting industrial processes such as product design, safety training, and facility inspection (Dai et al., 1997). One major advantage of VR is its ability to provide immersive and interactive 3D visualizations of complex systems. The possibility for users to spatially interact with industrial designs and thoroughly inspect these facilities in a telepresence environment can significantly improve productivity, task engagement, and workflows (Büttner et al., 2017). Moreover, at critical infrastructures such as particle accelerators and nuclear power plants, where human onsite visits to the facilities are limited

due to safety hazards and operation constraints (di Castro et al., 2018; Dehne et al., 2017), VR visualization systems that accurately represent the complex facility conditions and fully immersive users in the inaccessible remote environments is crucial for online facility inspection and maintenance planning. However, existing industrial VR systems typically visualize the virtual facilities using a polygonal representation of their computer-aided design (CAD) models (Raposo et al., 2006) or meshes reconstructed from photogrammetry (Schönberger and Frahm, 2016), RGBD cameras (Zollhöfer et al., 2018), or Light Detection and Ranging (LiDAR) measurements Gong et al. (2019). These 3D representations often provide only limited realism in modeling complex geometries of real-world conditions, introducing incompleteness or inaccuracy in their visual appearances and restricting their potential to support real-world tasks such as detailed facility inspection, quality control, and maintenance planning (Dai et al., 1997).

In recent years, neural radiance fields (NeRFs) (Mildenhall et al., 2020) have emerged as a novel volumetric 3D representation that can replicate and store the intricate details of our complex realities through training a Fourier feature neural network (Tancik et al., 2020) with a set of sparse 2D images and their camera poses as input. As NeRF can generate highly realistic 3D reconstructions with a relatively small amount of input data, it offers a new approach to establishing visualization frameworks for virtual facility inspection. However, developing interactive VR experiences for industrial NeRF models faces several challenges. Unlike traditional render pipelines based on the rasterization of geometric primitives, a NeRF model needs to be rendered through an expensive volumetric ray marching process which includes a large number of feed-forward queries to the neural network. As a result, the render performance scales largely with resolution and the size of the rendered volume. Due to the substantial volume of real-world industrial facilities, the temporal complexity of NeRF rendering hinders its implementation in real-time, stereoscopic, high-resolution, high-frame-rate industrial VR applications. Furthermore, although there are vast amount of recent work on NeRF (Barron et al., 2021; Müller et al., 2022), prior efforts have mainly concentrated on demonstrating proof-of-concept experiment results, rather than delivering a user-friendly toolkit for visualizing and interacting with real-world NeRF models in immersive VR environments.

As a first step towards adapting VR NeRF to real-world industrial processes and applications, we present *Magic NeRF Lens*, a visualization toolkit for virtual facility inspection leveraging NeRF as photorealistic representations of complex real-world environments. Our framework provides a native render plugin that enables interoperability between a low-level network inference engine (Müller et al., 2022) and a high-level game engine (Li et al., 2022a), making customized VR NeRF experience creation possible through typical VR application development workflows using the Unity game engine. To reduce the temporal complexity of VR NeRF rendering, we introduce a data fusion technique that combines the complementary strengths of volumetric rendering and geometric rasterization. As Figure 1A illustrates, the photorealistic rendering of a NeRF model is merged with the polygonal representation of its corresponding CAD models, creating a 3D-magic-lens-style visualization (Viega et al., 1996) and achieving render volume reduction without

sacrificing users' immersion and presence in the VR environments. As Figure 1B illustrates, such data fusion also can be realized through dynamic editing of the NeRF model. Users can flexibly select target render volume through an intuitive 3D drawing interaction by dynamically revealing or concealing a portion of the NeRF model using the polygonal representation of the industrial CAD model as context.

We systematically benchmarked our framework using real-world industrial data, highlighting the advantages of the proposed magic lens interactions in optimizing NeRF rendering within an integrated VR application. Furthermore, we explored users' perceptions of the magic-lens-style visualizations for visual search inspection tasks, thoroughly analyzing their impact on system usability, task performance, and spatial presence to derive design insights. Additionally, confirmatory expert reviews were conducted with five facility control and management specialists at an industrial facility, demonstrating the framework's benefits in supporting real-world virtual facility inspection processes, including facility redesign and maintenance planning. For a better understanding of our methods, interaction designs, and user evaluation, we strongly encourage readers to refer to our Supplementary Video:https://youtu.be/2U4X-EaSds0.

In summary, the contribution of this work includes:

1. The development and open-source implementation of an immersive visualization toolkit to support general virtual facility inspection tasks leveraging photorealistic NeRF rendering Githubrepository:https://github.com/uhhhci/immersive-ngp.
2. The design and implementation of two magic-lens-style interaction and visualization methods for optimizing VR NeRF rendering through data fusion.
3. Systematic evaluation of the performance of our framework and visualization design through technical benchmarking, user study, and expert reviews.

2 Related work

2.1 Immersive visualization for virtual facility inspection

Facility inspection and maintenance are important routines in the operations of complex industrial facilities such as particle accelerators (Dehne et al., 2017) and nuclear power plants (Sarmita et al., 2022). In such cases, VR visualization systems are crucial in enabling users to perceive and interact with remote environments that are often inaccessible to onsite human visits due to safety hazards and regulations (di Castro et al., 2018). While numerous VR solutions have been proposed to facilitate the virtual inspection process, prior systems predominantly visualize industrial facilities using traditional 3D representations based on explicit geometric primitives (Raposo et al., 2006) or point clouds (Gong et al., 2019). A common approach in previous VR solutions involves converting CAD models into polygonal mesh representations that can be rendered and interacted using conventional game engines (Dai et al., 1997; Raposo et al., 2006; Burghardt et al., 2020). However,
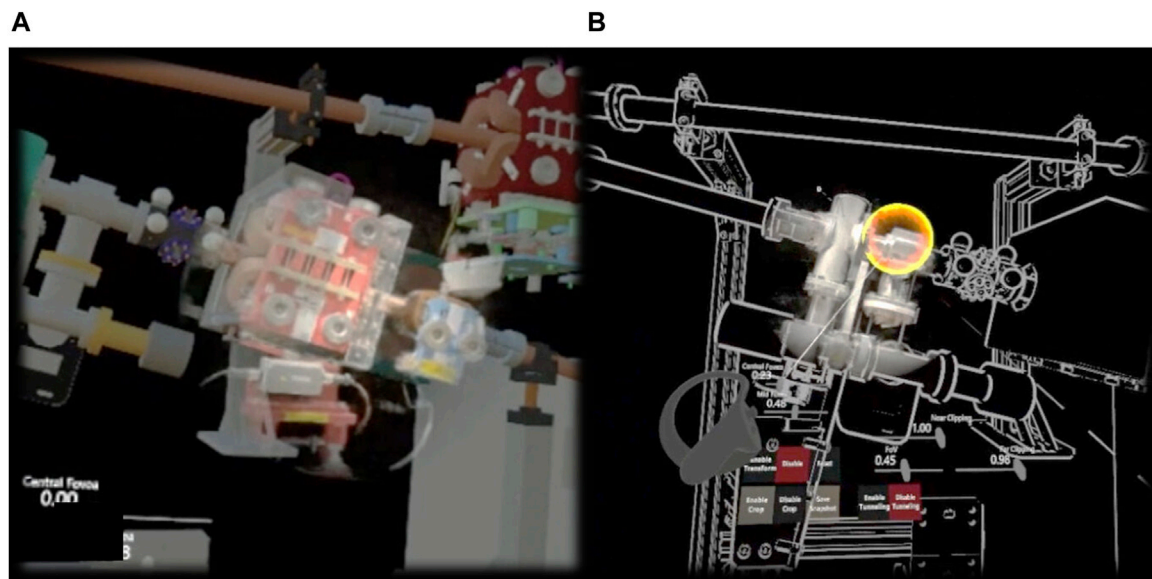
FIGURE 1
Screenshots from the Magic NeRF Lens framework illustrating **(A)** the data fusion of a high-resolution NeRF rendering and the polygonal representation of its CAD model, and **(B)** the 3D NeRF drawing interaction using the polygonal CAD representation as context.

as noted by Dai et al. (1997), complex industrial facilities are often inadequately described by their CAD designs due to the time-consuming process of regularly and precisely updating these models using CAD toolkits. Other lines of work use 3D reconstruction results obtained from RGBD sensors (Zollhöfer et al., 2018), photogrammetry (Remondino, 2011), or LiDAR scanners (Bae et al., 2013) to visualize the remote environment. However, the 3D visualizations generated by these methods exhibit limited accuracy and completeness in modeling the intricate geometries and appearances of real-world industrial facilities. In the immersive VR environment, users' heightened visual sensitivity to inaccuracies and unrealistic 3D visualizations exacerbates these limitations, further reducing their effectiveness in real-world inspection tasks. Our work improves previous systems by enabling high-quality photorealistic visualization of complex industrial facilities leveraging NeRF rendering (Müller et al., 2022) in immersive VR, further opening up opportunities for enhanced spatial understanding, analysis, and decision-making in the industrial usage of immersive virtual inspection systems.

2.2 NeRF and VR NeRF

By training a neural network with only 2D images and their camera poses, a NeRF model can compress scene details into a small network model instead of storing all the color values of the entire voxel grid on disk (Mildenhall et al., 2020; Barron et al., 2021; Müller et al., 2022). Compared to photogrammetry or conventional RGBD sensors, creating a NeRF often requires fewer image inputs but could produce higher visual quality than photogrammetric point clouds, which tend to be erroneous with limited image feature overlap or uniform

textures (Zollhöfer et al., 2018). Compared to active 3D scanners with sub-millimeter accuracy such as those with structured illumination (Zhang, 2018), NeRF can work with "optically uncooperative" surfaces such as metallic or absorbent materials that are common in industrial facilities. While it is also common to combine both image-based approach and active 3D scanning to create a high-quality 3D digital twin (Remondino, 2011), it is noteworthy that the post-processing associated with these methods can take weeks to months, during which time the conditions of the facility may have already changed due to maintenance activities.

As NeRF holds significant potential for various applications, there has been a substantial amount of recent research aimed at enhancing NeRF training (Mildenhall et al., 2020; Müller et al., 2022; Tancik et al., 2022), rendering (Barron et al., 2021; Deng et al., 2021; Müller et al., 2022), and editing capabilities (Lazova et al., 2022; Haque et al., 2023; Jambon et al., 2023) to make it viable for real-time interactive applications. Recent advancements in NeRF representation and compression, utilizing more efficient data structures such as 4D tensors (Chen et al., 2022) and multi-resolution hash tables (Müller et al., 2022), paved the way for real-time NeRF training and rendering (Müller et al., 2022), making the question of how to complement NeRF rendering with user interface systems to support different application domains becomes more and more relevant. However, much of the prior research in machine learning and computer vision for NeRF primarily focused on demonstrating proof-of-concept experiment results, rather than delivering user-friendly toolkits for visualization and interaction with real-world NeRF data in immersive VR applications. For example, popular NeRF visualization toolkits such as NeRF Studio (Tancik et al., 2023) and instant-ngp (Müller et al., 2022) primarily support scene visualization on 2D desktops. Although these toolkits can

potentially enable VR visualization through stereoscopic rendering, they lack further integration into game engines such as Unity, which are the major platforms for VR application development that can enable more interactive and versatile VR content creation. Our framework integrates a NeRF render plugin that enables interoperability between low-level NeRF inference implementation Müller et al. (2022) and high-level game engine, making customized VR NeRF application development more scalable and flexible to a wider range of audiences. Additionally, in line with the open-source approach seen in instant-ngp and NeRF Studio, our framework's development is also open-source to support further research and development of VR NeRF.

Another challenge for VR NeRF development is the enormous amount of network queries required for stereoscopic, high resolution, high frame rate VR rendering (Li S. et al., 2022). One promising direction is to use foveated rendering to reduce render resolution in the peripheral region of the human visual field (Deng et al., 2021). However, the existing foveated NeRF method requires training and recombining the rendering results of separate networks. Moreover, it did not use the efficient multi-resolution hash coding data structure, which can achieve most of the rendering and performance speedup (Deng et al., 2021). Another line of work investigated adapting NeRF into the conventional geometric rasterization render pipeline by converting NeRF models into textured polygons to support photorealistic rendering on mobile devices (Chen et al., 2023). However, the surface estimation process can lead to inaccurate results with specular materials and sparse viewpoints, which degrades the visual quality and robustness of NeRF compared to volumetric rendering. Our data fusion method which combines the complementary strength of volumetric rendering and rasterization preserves NeRF render quality at the fovea region of the human's visual field while achieving render volume reduction through multi-modal data fusion without sacrificing users' immersion in the VR environments.

2.3 Magic lens techniques

Magic lens techniques were first developed by Bier et al. (1993), allowing users to change the visual appearance of a user-defined area of the user interface (UI) by overlaying a transparent lens over the render target. Interactive magic lenses are widely used in modern visualization systems (Tominski et al., 2014), where context-aware rendering of large information spaces is needed to save computational costs (Viega et al., 1996; Wang et al., 2005). In 3D computer graphics, several 3D magic lens effects have been developed to allocate computational resources to more resolution-important features for the visualization of volumetric medical scans (Viega et al., 1996; Wang et al., 2005) or context-aware AR applications (Brown and Hua, 2006; Barčević et al., 2012). Magic lens style visualization is also used in immersive VR HMD rendering to perform sensor fusion with different resolutions, frame rates, and latency (Li et al., 2022b). Our framework adapts 3D magic lens-style interactions for photorealistic NeRF rendering in immersive VR while introducing a novel magic-lens-style visualization design for industrial facilities.

3 System design

3.1 Design goals

Our visualization system is designed as a step towards tackling the facility inspection and maintenance planning challenges at industrial facilities. For example, particle accelerator facilities such as those at the Deutsche Elektronen-Synchrotron (DESY) [1] often includes a complex system with more than 10 million control system parameters and tens of thousands of components that require frequent inspection and maintenance. However, the accelerator must operate continuously for more than 5,000 h per year, during which time on-site human access is not possible, and any unexpected interruptions to operation result in high energy and setup costs. While VR can immerse users in inaccessible remote environments to support various inspection tasks, the design of such visualization systems needs to thoroughly consider various human factors to avoid negative effects of VR usage such as motion sickness while delivering a system with high usability. Moreover, the design of our framework targets creating a more interactive visualization experience that allows users to modify and customize their virtual experiences to enhance user task engagements and provide a wider range of functionalities (Aukstakalnis and Blatner, 1992). With considerations of interactivity, task requirements, and human factors, we formalize the following design goals (**DG**):

DG1: The visualization system can render high-quality photorealistic representations of complex facility environments without introducing undesirable effects such as motion sickness or reducing users' immersion in the virtual environment.

DG2: If requested, the immersive visualization of the virtual facility can be the one-to-one real-world size of the real facility to provide a realistic scale of the industrial environments.

DG3: The visualization toolkit allows users to naturally and intuitively perform facility inspection tasks through both exploratory and interactive VR experiences.

3.2 Interaction techniques

3.2.1 Basic interactive virtual inspection

According to Aukstakalnis and Blatner (1992), the design of a VR application can be categorized into three different levels: passive, exploratory, and interactive. Passive VR experiences such as those with 360° videos enable observations of the remote environment but without giving users control of what they perceive. Exploratory VR experiences allow users to freely change their positions in the virtual environment, however, there are limited actions available to control and modify the virtual content. While passive and exploratory VR is useful in certain facility inspection tasks, the design of our

---

1 https://www.desy.de/index_eng.html
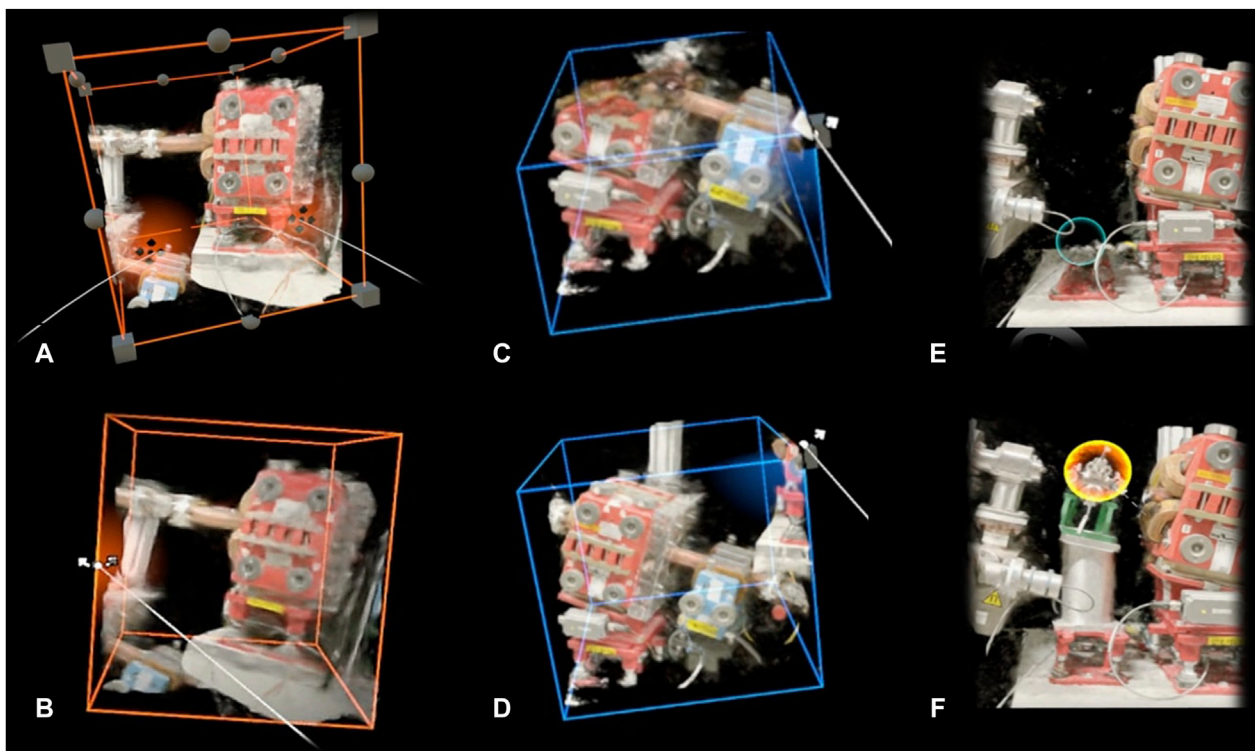
FIGURE 2
Illustration of our system extension to instant-ngp. **(A,B)**: NeRF model manipulation, **(C,D)**: NeRF model crop box manipulation, **(E,F)**: Volume editing via 3D drawing.

framework employs both exploratory and interactive VR designs to improve user task engagements and provide a broader range of virtual inspection functionalities.

As Figures 2A,B illustrates, with consideration of **DG3**, our framework integrates natural and intuitive 3D interactions with the NeRF model, allowing users to perform spatial transformation by translating, rotating, and scaling the NeRF model using VR controllers, making it possible to dynamically explore and manipulate the virtual facilities. As Figures 2C,D illustrates, we also provide the possibility to select target render volume using a volumetric crop box. Users can manipulate the crop box by rotation, translation, and scaling using VR controllers, making it possible to precisely define the region of interest for rendering and ensuring optimal focus in the virtual environment. Figures 2E,F demonstrates an interactive volumetric editing design, where users can remove and erase a portion of the NeRF model through a responsive 3D drawing effect. Such interaction can facilitate detailed customization and fine-tuning of the virtual environments to meet specific design and simulation requirements.

3.2.2 Magic NeRF lens with FoV restrictor

While designs described in Section 3.2.1 enable dynamic manipulation and editing of NeRF models, several considerations from **DG1** and **DG2** are not met. For example, when inspecting facility equipment that covers a substantial area of volume, the increased latency in VR NeRF rendering on devices with limited computational resources will lead to the negative effects of motion sickness. The magic NeRF lens with FoV restrictor is designed to visualize facilities with larger volumes with considerations of various human factors (**DG1**) and inspection task requirements (**DG2**). Applying a FoV restrictor to the NeRF model creates a "VR tunneling" effect (Lee et al., 2017), which is a typical motion sickness reduction technique in VR gaming to restrict the optical flow and sensory conflicts of the peripheral region of the human eye (Seay et al., 2001). However, a FoV restrictor could significantly reduce the user's sense of presence and immersion. To maintain rendering performance without sacrificing users' immersion in the VR environment, we propose using the "Mixed Reality (MR) tunneling" effect (Li et al., 2022b) by merging the NeRF model with the polygonal representation of the CAD model in the peripheral regions of the user's vision field. Such visual data fusion techniques can improve users' overall perception (**DG1**), while the polygonal CAD model can be used as the reference for the NeRF model to achieve a one-to-one real-world size of the physical facility (**DG2**).

Figure 3A shows the schematic relationship between the main VR camera, the NeRF camera, and the active NeRF rendering volume for the magic NeRF lens effect with a FoV restrictor. Figure 3B shows the volumetric crop box of the NeRF model, which is dynamically aligned with the user's viewing direction and the near clipping plane of the VR camera, so that the volume crop box acts as an interactive lens that automatically selects the target rendering volume according to the size of the image plane W and the distance of the far clipping plane L. The parameters W and L are both user-defined values to adjust the total number of sampling points for ray marching and could be adjusted
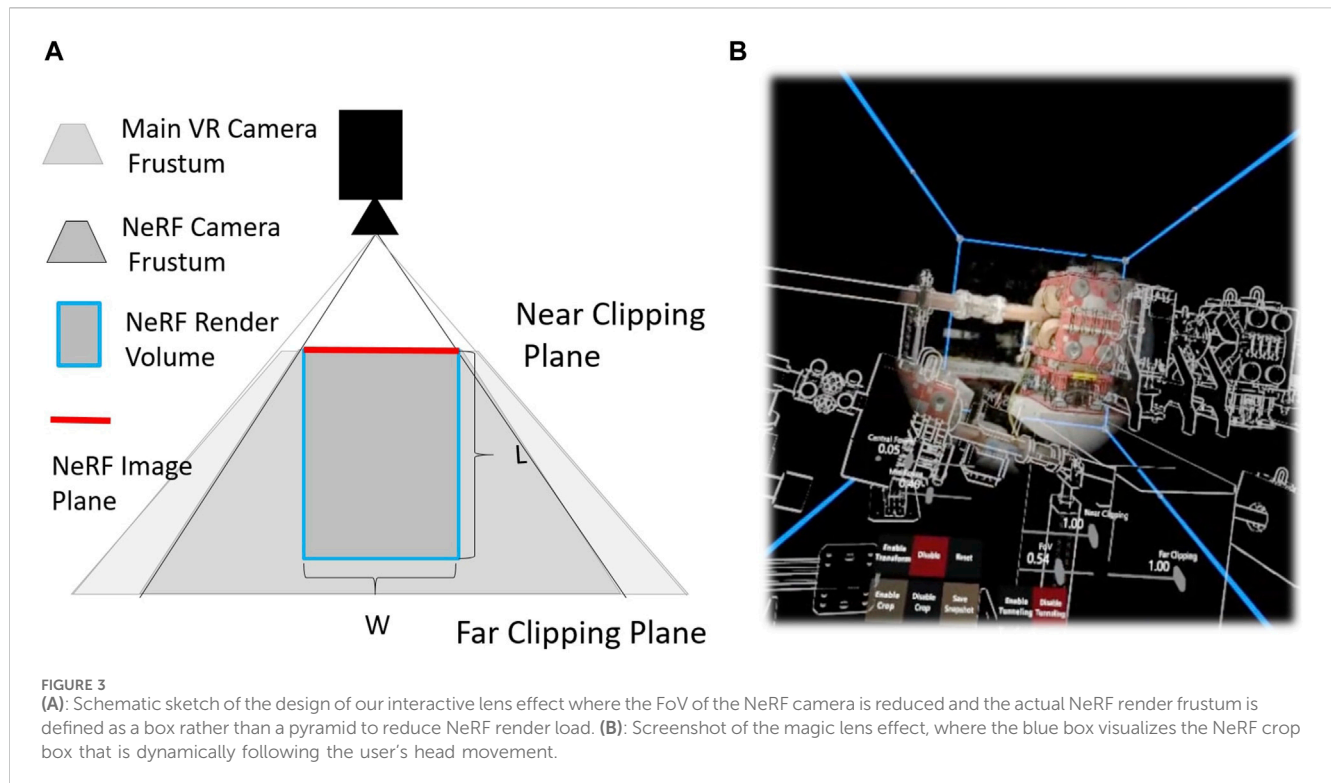
during application runtime to avoid frame rate jitter. For the static MR tunneling effect, the high-resolution NeRF rendering is displayed in the central region of the HMD. However, an eye tracker could be integrated to achieve foveated MR tunneling and reduce the need for frequent head movements while inspecting the facilities (Li et al., 2022b).

3.2.3 Magic NeRF lens with context-aware 3D drawing

The second magic NeRF lens effect is designed to enable more interactive visualization of the facility (**DG3**) while further optimizing VR NeRF rendering performance (**DG1**). For many facility inspection tasks, users typically do not need to see the entire NeRF model. For example, components such as walls, simple electrical boxes, floors, and ceilings are static elements that normally do not need to be inspected and modified and could be visualized simply by the polygonal representations of their CAD models. The magic NeRF lens effect with 3D drawing interaction enables users to dynamically select the target render region using the polygonal presentation of the CAD model as context. As shown in Figure 1B, an edging rendering of the polygonal CAD model visualizes the overall facility environment. Users can adjust the radius of a 3D sphere attached to the VR controller and point the 3D sphere at a spatial location where the NeRF render volume should be revealed on demand by using the edge polygonal CAD rendering as a guideline. With the empty space skipping render techniques that are implemented in most NeRF render pipelines, the network query at the location where the volume density value is zero, indicating that this space is essentially empty, could be automatically ignored for that sample location. As a result, the magic NeRF lens effect with the 3D drawing technique could

potentially improve the overall rendering speed (**DG1**) while providing users with more dynamic and responsive visualization experiences (**DG3**).

4 System implementation

4.1 VR NeRF rendering and interaction implementation

4.1.1 Render plugin

A key technical challenge in implementing interactive VR experiences using neural rendering lies in bridging the interoperability gap between the low-level inference implementations and high-level game engines that are typically used for the VR application development process. Our system implementation tackles this challenge by introducing a native render plugin that enables data sharing between instant-ngp NeRF inference implementation (Müller et al., 2022) and the Unity VR application runtime. As illustrated in Figure 4, a pre-trained NeRF model and its' associated 2D image sequences are loaded into the instant-ngp application and the render device's GPU memory. OpenGL render buffers are created and upscaled by Nvidia's Deep Learning Super Sampling (DLSS) technique to improve image quality and increase the perceived rendered resolution in real-time [2]. A customized CUDA and C++ native plug running in

---

2   https://www.nvidia.com/de-de/geforce/technologies/dlss/

FIGURE 4
The system architecture based on the framework proposed in our arXiv preprint (Li et al., 2022a), depicting the individual processes from left to right: Starting with the input, a view model projection (VMP) matrix is computed from VR input devices and the HMD. The VMP is applied in the native Unity plugin, which provides a communication layer to the instant-ngp backend as well as the final rendering. The instant-ngp (Müller et al., 2022) backend performs the volume rendering through NeRF by updating the provided texture. The pre-processing refers to NeRF model training. (This figure is adapted from Li et al., 2022a).

Unity can access the OpenGL render buffer and render textures of instant-ngp through pre-compiled dynamic link libraries (DLL). We enable efficient data exchange between Unity and instant-ngp by sharing only the texture handle pointers for render event updates. In immersive VR rendering on a head-mounted display (HMD), two render textures are created and placed as screen space overlays in front of the user's eyes. Synchronization problems are avoided by updating both render textures simultaneously in a single render frame. In terms of software, our framework implementation uses Unity Editor version 2019.4.29f1 with the OpenVR desktop and SteamVR runtime.

4.1.2 Manipulating a NeRF model as an object

To support spatial manipulation of a NeRF model in VR, we create a model space for the NeRF model whose spatial properties such as position, rotation, and scale are defined by a volume bounding box. As Figures 2A,B shows, the bounding box is represented as a transparent cube whose translation, rotation, and scale matrices are combined into one transformation matrix that is applied to the view matrix of the instant-ngp camera to render the correct view of the NeRF model. The model-view-projection (VPM) matrix is applied to instant-ngp renderer through the native CUDA and C++ plugin. The object manipulation interaction from the Mixed Reality Toolkit (MRTK) [3] is attached to the cube, allowing users to intuitively rotate, scale, and translate the box and its associated NeRF model using one- or two-handed control.

---

[3] https://github.com/microsoft/MixedRealityToolkit-Unity

4.1.3 Crop box editing

To enable spatial transformation of a volumetric crop box, represented as the axis-aligned bounding box (AABB) in instant-ngp, we attach a second 3D cube game object with an object manipulator where the spatial transformations of the Unity AABB bounding box, such as translation and rotation, are applied to the AABB crop box defined in instant-ngp's coordinate system. As shown in Figures 2C,D, users can extend, reduce, or rotate the AABB render volume using the object manipulation interaction provided by MRTK.

4.1.4 Volumetric editing

We adapt instant-ngp's implementation of a volumetric editing feature in our framework. Similar to instant-ngp's preliminary implementation, a 3D sphere object is attached to the controller at a certain distance to indicate the intended drawing region. As Figure 2E shows, users can interactively select the region where voxels within the sphere should be made transparent by moving the sphere to the target region in 3D space while pressing a button on the VR controller to confirm the drawing action. The same interaction and action could be performed to reveal the volume at the region set to transparent, as shown in Figure 2F. In an instant-ngp NeRF model, the volume density grid stores the transparency of the scene learned by the network. A binary bitfield can be generated based on a transparency threshold and can be used to quickly mask the transparent region from being sampled during raytracing (Müller et al., 2022).

In summary, our system implementation provides many core components for building interactive VR NeRF applications using the Unity game engine to streamline the VR NeRF development process. A full demo video of the current system can be found in the Supplementary Video.
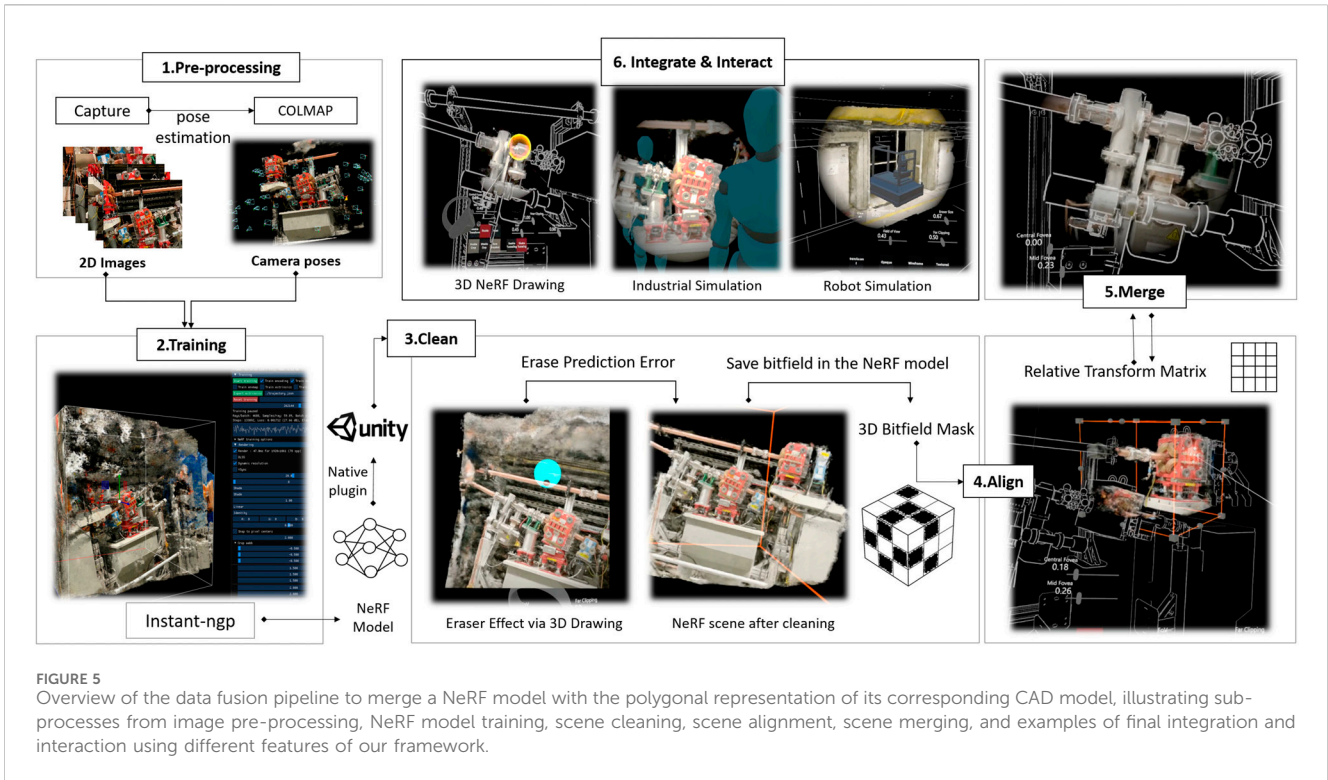
FIGURE 5
Overview of the data fusion pipeline to merge a NeRF model with the polygonal representation of its corresponding CAD model, illustrating sub-processes from image pre-processing, NeRF model training, scene cleaning, scene alignment, scene merging, and examples of final integration and interaction using different features of our framework.

4.2 Data fusion pipeline

In this section, we focus on developing a data fusion pipeline to merge a NeRF model with the polygonal representation of its corresponding CAD model to visualize the facility at a one-to-one real-world size and provide a realistic scale of the industrial environments. We choose CAD models as CAD models of industrial facilities are often created in the early stages of facility design and have complementary features to their corresponding NeRF models. Moreover, CAD models of industrial facilities typically lack realistic textures because the detailed environments are usually too complex to be modeled accurately (Gong et al., 2019). Semantic information typically embedded in CAD models is also not represented in NeRF models without additional expensive network training and voxel-wise segmentation. Other benefits of fusing a polygonal representation of a CAD model with a NeRF model include the possibility of creating more realistic mixed reality (MR) experiences that include depth occlusion and physical interactions using the mesh representation of the CAD model without the need to perform mesh reconstruction of the NeRF model to achieve comparable effects.

Figure 5 provides an overview of the data fusion pipeline, which is divided into the following six steps.

4.2.1 Preprocessing

Since most 2D images do not contain their camera poses, the 2D images must be preprocessed using the conventional Structure from Motion (SfM) algorithm (Schönberger and Frahm, 2016) to estimate the camera poses using software such as COLMAP (Schönberger and Frahm, 2016). However, this step can be skipped for cameras that can track their own poses.

4.2.2 Training

The processed data is trained in the instant-ngp framework to generate an initial estimate of the scene function and an initial occupancy grid whose size is defined by a preset AABB bounding box.

4.2.3 Scene cleaning

The quality of NeRF rendering can be affected by artifacts of real-world 2D images, such as motion blur, lens distortion, or insufficient images around the viewing angle. This often results in false clouds in the 3D NeRF scene, which can degrade the user's viewing experience and needs to be cleaned up before the NeRF model can be applied to practical applications. Therefore, our VR NeRF framework provides the ability to modify the pre-trained instant-ngp NeRF model. The initial density grid estimation is inspected and cleaned using the interactive eraser function described in Section 4.1. Users can manually remove regions with cloudy prediction errors or low render quality. The edited density grid and the binary bitmask of the density grid are stored in the NeRF model for reuse. When reloading the cleaned NeRF model, our framework automatically checks the saved bitmask, so that the erroneous network prediction that has already been removed will no longer be sampled and rendered.

4.2.4 Scene alignment

The user could manually align the cleaned NeRF model with the CAD model using the NeRF object manipulation and crop box editing functions described in Section 4.1. If the NeRF model is adjusted to the one-to-one scale of the real facility, alignment using the entire NeRF model will result in a rendering volume too large to

be efficiently manipulated in VR. Therefore, it is recommended to perform the alignment process by focusing on a small part of the scene for accurate object manipulation.

4.2.5 Scene merging

In the merge step, the user could perceptually validate the alignment by reducing the FoV of the NeRF camera and adjusting the transparency of the shader that renders the NeRF images so that both the CAD drawing and the NeRF rendering are simultaneously visible. The user can iteratively move back and forth between the alignment and fusion steps until the two 3D representations are spatially aligned. The relative transformation matrix between the NeRF model and the CAD model can be saved for reuse.

4.2.6 Integrate and interact

The merged visualization system can be used for various customized virtual inspection applications, such as facility upgrades and redesign, maintenance planning, and robot simulation.

5 Performance benchmarking

While instant-ngp represents a significant advancement in achieving real-time interactive NeRF rendering, previous benchmarking efforts have primarily focused on demonstrating proof-of-concept render results for monoscopic image generation at a fixed FoV Müller et al. (2022). Notably, the system performance of instant-ngp implementations for stereoscopic VR rendering within an integrated VR application for real-world usecases is unknown. In this section, we present a performance benchmark experiment and its results using our magic NeRF lens visualization design. In particular, our experiment aims to provide insights into how parameters such as FoV and VR HMD's pixel density affect VR NeRF rendering performance. This enables us to evaluate the overall performance and efficiency of our system implementation while deriving configuration recommendations for using the magic NeRF lens in rendering volume reduction.

5.1 Experiment design

We examine the performance trend when varying the NeRF rendering FoV from 10°, which covers the foveal region of human vision, to 60°, which covers the average central visual field for most people (Osterberg, 1937). We also vary the PPD value to match the resolution requirements of VR displays of different quality, including medium display resolution (PPD = 15, e.g., Oculus Quest 2), medium to high-end displays (PPD = 20, e.g., Meta Quest Pro), or high-end displays (PPD = 25, e.g., Varjo XR-3). The final NeRF render resolution per eye ($R \times R$) for each configuration is calculated using the following equation:

$$R = FoV \times PPD \times 2 \tag{1}$$

where we multiply the pixel density per degree by 2 to account for the upsampling required for aliasing reduction and edge smoothing via supersampling (Akeley, 2000).

5.2 Materials

We evaluate the performance of our system on a self-generated real-world dataset consisting of 60 2D images at 3689 × 2983 resolution reconstructing a section of a complex particle accelerator facility. The images were captured by particle accelerator operation specialists during the facility's maintenance shutdown period. Its VR NeRF reconstruction is shown in Figure 7. Since the CAD model is designed to be at an exact scale of the real facility, we aligned the NeRF model with its' polygonal representation of CAD model using the data fusion pipeline described in Section 4.2 and were able to measure that the NeRF rendered volume of this NeRF model corresponds to approximately 2.2 m × 1.47 m × 2.08 m in real-world scale, which would require a full FoV VR rendering to examine closely. All of our benchmarks are run on an Nvidia Geforce RTX 3090 GPU and an Intel (R) Core(TM) i7-11700K CPU with 32 GB of RAM. As demonstrated in previous NeRF rendering implementations (Li et al., 2022a; Müller et al., 2022), DLSS has a significant influence on VR NeRF rendering performance. As illustrated in Figure 4, DLSS is a crucial part of the NeRF render pipeline in upscaling the render buffer such that images rendered at low resolutions can be perceived at a higher resolution, thereby improving performance without sacrificing visual quality [4]. As a result, DLSS is enabled for all benchmarking efforts.
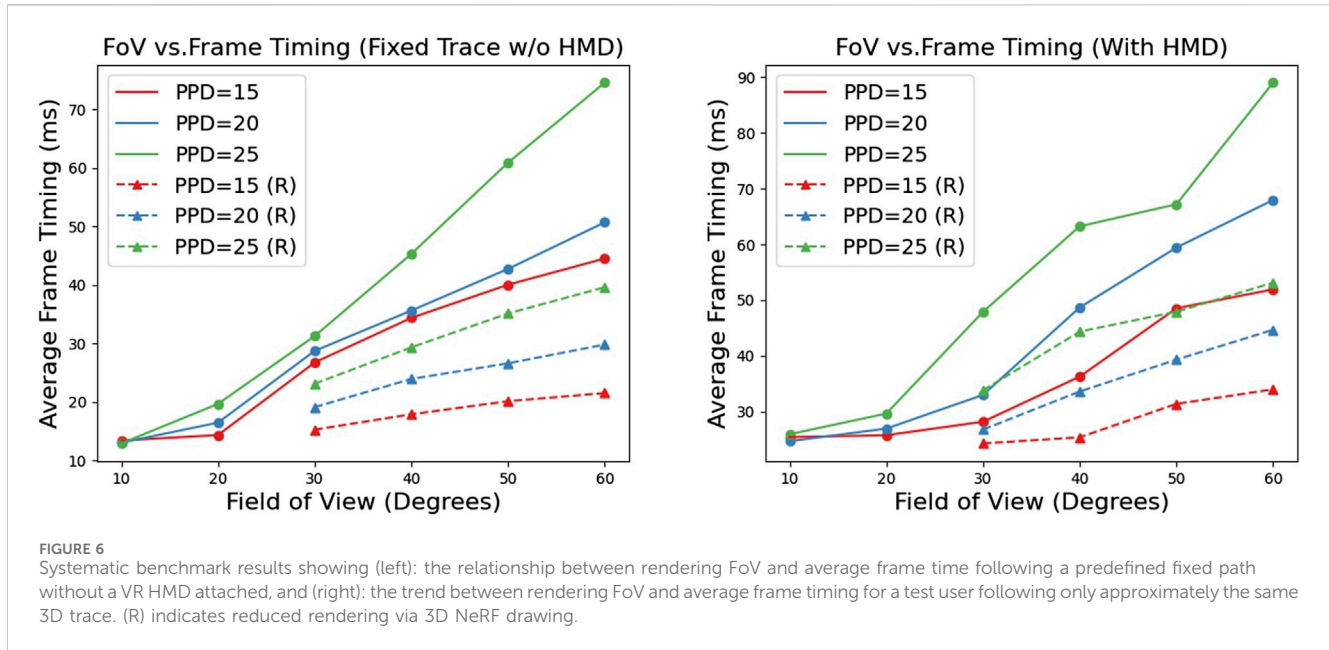
5.3 Systematic benchmark (w/o HMD)

To simulate the system performance of how a user would use our framework to inspect different parts of the NeRF model, we first asked a test user to closely inspect three fixed locations within the NeRF model. A custom script recorded the user's exact 3D trajectory with the camera transformation information at each frame. The relative position of the three target components concerning the polygonal representation of the CAD model was visualized as a schematic sketch, as shown in Figures 7-(Instruction Card). A video recording of the 3D trace is also available in the Supplementary Material.

First, we performed a systematic benchmark where we collected the average frame timing ($\overline{f_t}$) for stereoscopic rendering of different display FoV and PPD configurations by replaying the same pre-recorded 3D trace. Figure 6 (left) plots the average frame timing result for each rendering configuration. As expected, the rendering latency scales linearly with increasing FoV for the basic NeRF lens effect. Additionally, it is shown that our system could theoretically achieve real-time VR rendering of 30 frames per second ($\overline{f_t} < 33.33\ ms$) at less than 50° FoV.

To evaluate whether additional contextual rendering could lead to real performance gains, we create an edited NeRF model by 3D drawing interaction, where we define the 3D density grid so that only the target components are visible. Since FoV settings below 20° could already achieve good real-time performance without reducing the rendering volume, we only examine FoV and IPD settings above 30° in this section. Figure 6 (left) also plots the average frame timing

---

4 For evaluating the impact of DLSS on VR NeRF render performance, we refer readers to our arXiv preprint on immersive-ngp (Li et al., 2022a).

FIGURE 6
Systematic benchmark results showing (left): the relationship between rendering FoV and average frame time following a predefined fixed path without a VR HMD attached, and (right): the trend between rendering FoV and average frame timing for a test user following only approximately the same 3D trace. (R) indicates reduced rendering via 3D NeRF drawing.

for each rendering configuration for rendering the selected regions. As expected, exposing only the target render volume reduces the overall render load due to empty space skipping and early ray termination implementation. This confirms the performance optimization insight that removing unimportant scene details using the 3D NeRF drawing effect could be an option to gain additional performance.

5.4 Empirical user benchmark results (with HMD)

In the second benchmarking experiment, we evaluate the same set of rendering configurations. However, instead of replaying a pre-recorded 3D trace, a test user was asked to repeat each configuration, following roughly the same path to inspect the three target components to evaluate the overall performance of the actual VR system. For high-resolution configurations where real-time performance could not be achieved, the test user was able to use keyboard control instead, while keeping the VR HMD connected to the entire rendering pipeline. Figure 6 (right) shows the general trend between FoV and $\overline{f_t}$. Compared to the systematic performance results, the actual total frame timing is about 5–10 ms higher because of the additional computational resources needed to drive the HMD. However, the general performance trend in an actual system with VR HMD is similar to the systematic benchmark result. Additionally, we could derive that for an actual VR NeRF system, the optimal render resolution is 1200×1200 at $30°$ FoV (PPD = 20) or at $40°$ FoV (PPD = 15) using an RTX 3090 GPU.

6 User study

Understanding human factors is an important step in the development of new VR systems and interaction techniques. In this section, we present a user study experiment that aims to:

1. Systematically evaluate the overall performance and capabilities of the proposed visualization system in supporting actual virtual facility inspection processes through a visual search task,
2. Quantify the user experiences through perceptual metrics such as system usability, perceived motion sickness, and perceived task load to assess the effectiveness of the system implementation,
3. Understand the impact of different magic NeRF lens visualization and interaction styles on the overall system usability, task performance, and spatial presence.

In addition, we conducted confirmatory expert reviews at an industrial facility with five control and management specialists to validate the benefits of our framework in supporting real-world virtual facility inspection tasks.

6.1 Study design

6.1.1 Conditions
We investigate the two magic lens designs with the two most common types of CAD model visualizations in large-scale industrial facilities. As shown in Figure 7, (C1) is the baseline condition with only FoV restriction but no data fusion for comparison, (C2, C4) implement the magic NeRF lens effect with FoV restriction, and (C3, C5) implement the magic NeRF lens effect with 3D drawing. (C2, C3) use a polygonal representation of a CAD model with colored abstract texture, while (C4, C5) use a polygonal representation of the CAD model with only edging rendering.

6.1.2 Materials
The user study was conducted on the same graphics workstation and software configuration used for the system benchmark described in Section 5. For all conditions, the FoV of the NeRF rendering camera was set to $30°$ with a per-eye render resolution of 1200×1200 pixel, which is one of the optimal rendering
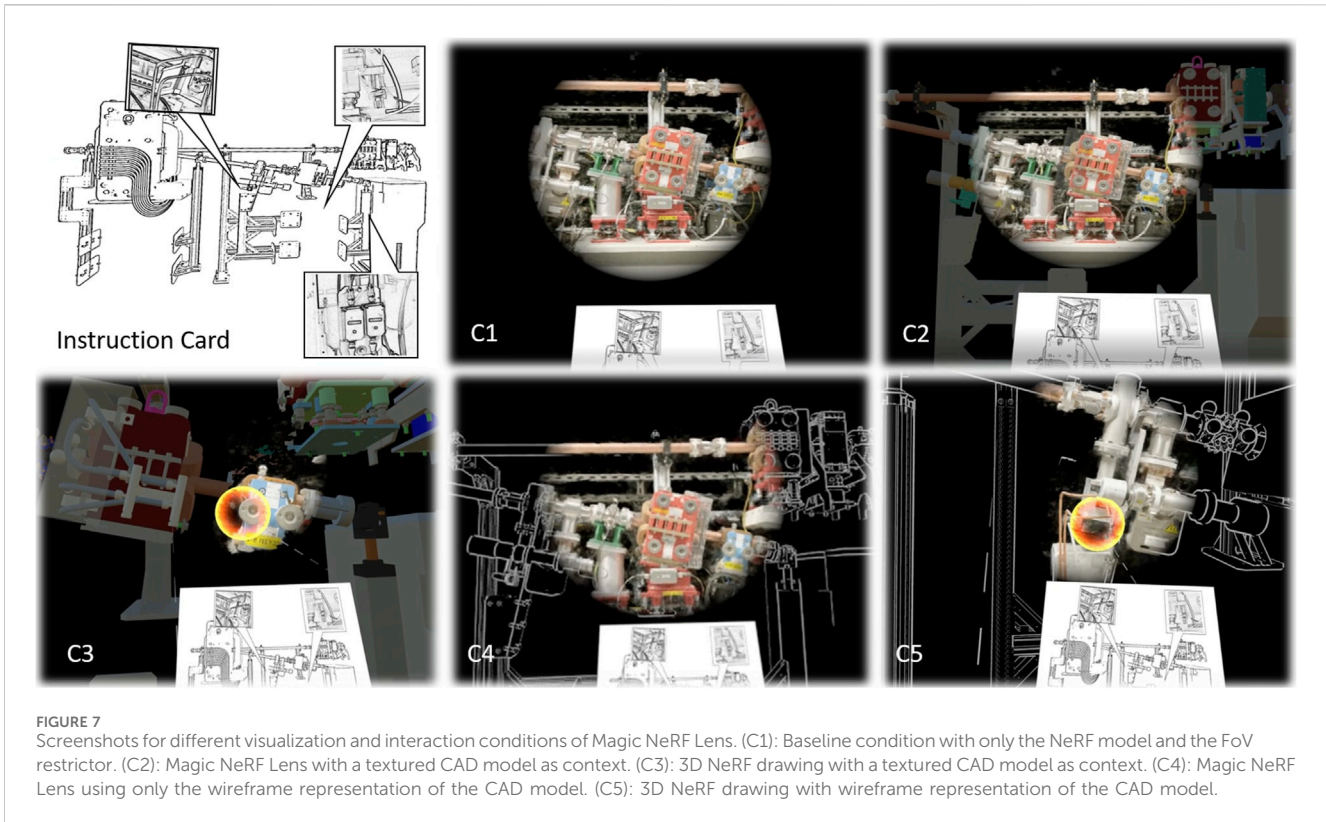
FIGURE 7
Screenshots for different visualization and interaction conditions of Magic NeRF Lens. (C1): Baseline condition with only the NeRF model and the FoV restrictor. (C2): Magic NeRF Lens with a textured CAD model as context. (C3): 3D NeRF drawing with a textured CAD model as context. (C4): Magic NeRF Lens using only the wireframe representation of the CAD model. (C5): 3D NeRF drawing with wireframe representation of the CAD model.

configurations determined in our benchmark experiments. We used an Oculus Quest Pro VR HMD, which has a PDD value of 22. However, our VR NeRF framework is compatible with other VR headsets that support the SteamVR and OpenVR desktop runtime.

6.1.3 Tasks

To simulate how a user would perform virtual inspection tasks, we designed a visual search activity where the user was asked to locate three detailed components in the NeRF model based on a schematic sketch. As shown in Figures 7-(Instruction Card), the schematic sketch consisted of an overview of the CAD model, with each search target highlighted in a dialog box. Arrows are provided to indicate an approximate location where the detailed component might be found. The instructions for each search target were generated from the real-world images and therefore represent updated real-world conditions compared to the abstract CAD model. All search targets are detailed elements that could not be effectively updated in the CAD model due to maintenance activities or the difference between the actual installation and the original design. A pilot test with 3 users confirmed that all target components in each scene could be found in 3–5 min for all proposed conditions. To avoid learning effects in the course of the study, 5 different scenes were prepared, each showing a different part of an actual large-scale industrial facility.

6.1.4 Participants

We invited 15 participants, 4 female, 10 male, and 1 who preferred not to disclose their gender. 5 participants were between 18 and 24 years old, and 10 participants were between 25 and 34 years old. All were students or researchers in HCI, Computer Science, or Physics, with HCI students receiving compensation in the form of course credits. 8 participants use VR systems regularly (at least once a month), and only 2 use them less than once a year, 5 never used VR before.

6.1.5 Procedures

Each participant first signed a consent form and completed a demographic questionnaire. The experimenter then presented a test virtual environment in which participants could practice locomotion and 3D VR drawing controls for a maximum of 3 min. Each participant was then presented with the 5 conditions in order. To counterbalance the order and assignment of scenes to conditions, we used a replicated Latin square design with one treatment factor (condition) and three blocking factors (participant, trial number, and scene). As a result, each condition/scene combination was experienced by 3 participants. During each condition, participants could press a "start" button to begin the study and begin time recording to assess task performance. The experimenter monitored the participants' VR interactions via a secondary screen that mirrored the VR displays. After the participant informed the experimenter when they had found a component, the experimenter confirmed its correctness, after which the participant could begin searching for the next search target. Once all three components were found, the participant could press an "end" button, which marked the end of time recording for the search tasks. The total study time was approximately 60 min.

6.1.6 Hypotheses

We hypothesized that displaying the polygonal representation of the facility's CAD model in the user's

periphery would provide context to the user and thus improve spatial orientation within the scene. Since the use of NeRF drawing tools introduces additional complexity, we further hypothesized that using 3D NeRF drawing will have less usability and higher cognitive load on users than Magic NeRF lens effect through only MR tunneling.

This led to the following hypotheses:

- **H1**: Magic NeRF lens effects (C2-C5) reduce perceived mental demand and effort, and lead to higher performance, both subjectively and objectively, as measured by time spent on the visual search task.
- **H2**: Magic NeRF lens effects with MR tunneling display (C2 & C4) yield higher usability scores than those based on the 3D drawing techniques (C3 & C5).

Based on the results of Li et al. (2022b), we did not expect significant differences between the conditions in terms of presence and cybersickness. Nevertheless, we collected subjective ratings on both metrics to quantify how well users perceive these effects in general.

6.2 Results

We collected multiple objective and subjective measures to assess the user experience as study participants interacted with our magic NeRF lens framework. Tables listing the overall ratings for each subscale of the rating scales are included in the Supplementary Material.

6.2.1 Usability

Usability was measured using the System Usability Scale (SUS) with 10 questions providing ratings between 1 (strongly disagree) and 5 (strongly agree), where higher values correspond to better usability (Brooke, 1996). The overall rating of usability converts to a value range [0.100]. A repeated measures ANOVA with Greenhouse-Geisser correction revealed a significant effect of the visualization method on usability, $F(2.514, 35.202) = 4.269, p = .015, \eta_p^2 = .234$. Post-hoc tests with Bonferroni correction showed a significant difference between C3 and C4 ($p = 0.048$).

6.2.2 Perceived workload

We measured six aspects of perceived workload using the NASA-TLX questionnaire (Hart and Staveland, 1988). Because each aspect was represented by a single Likert scale ranging from 0 to 100 (with 21 levels), we used nonparametric Friedman tests to analyze the responses. There was a statistically significant difference in mental demand depending on the condition, $\chi^2(4) = 17.857, p = .001$. Post-hoc analysis with Wilcoxon signed-rank tests and Bonferroni-Holm p-value adjustment revealed a significant difference between C2 and C5 ($Z = -3.257, p = .011$). Friedman tests also showed significant effects of the visualization method on effort ($\chi^2(4) = 15.366, p = .004$) and frustration ($\chi^2(4) = 10.183, p = .037$), although none of the *post hoc* tests were significant after p-value adjustment. No significant differences were found for physical demand, temporal demand, and performance.

6.2.3 Task performance

Task performance was measured as the time it took participants to find the three locations indicated on the instruction card for each scene. A Shapiro-Wilk test indicated that the residuals of search time were not normally distributed, which was confirmed by visual inspection of the QQ plots. Therefore, a Friedman test was performed, which revealed a significant effect of visualization method on search time ($\chi^2(4) = 23.307, p < .001$). Post-hoc Wilcoxon signed-rank tests with Bonferroni-Holm adjustment of p-values showed significant differences between C2 and C3 ($Z = -3.010, p = .018$), C2 and C5 ($Z = -3.124, p = .014$), C4 and C3 ($Z = -3.237, p = .011$), and C4 and C5 ($Z = -3.351, p = .008$).

6.2.4 Presence

We measured the sense of presence using the Igroup Presence Questionnaire (IPQ) (Schubert, 2003) with all three subscales *Spatial Presence*, *Involvement*, and *Experienced Realism*, as well as a single item assessing the overall "sense of being there." Three ANOVAs for the subscales and a Friedman test for the single item revealed no significant differences.
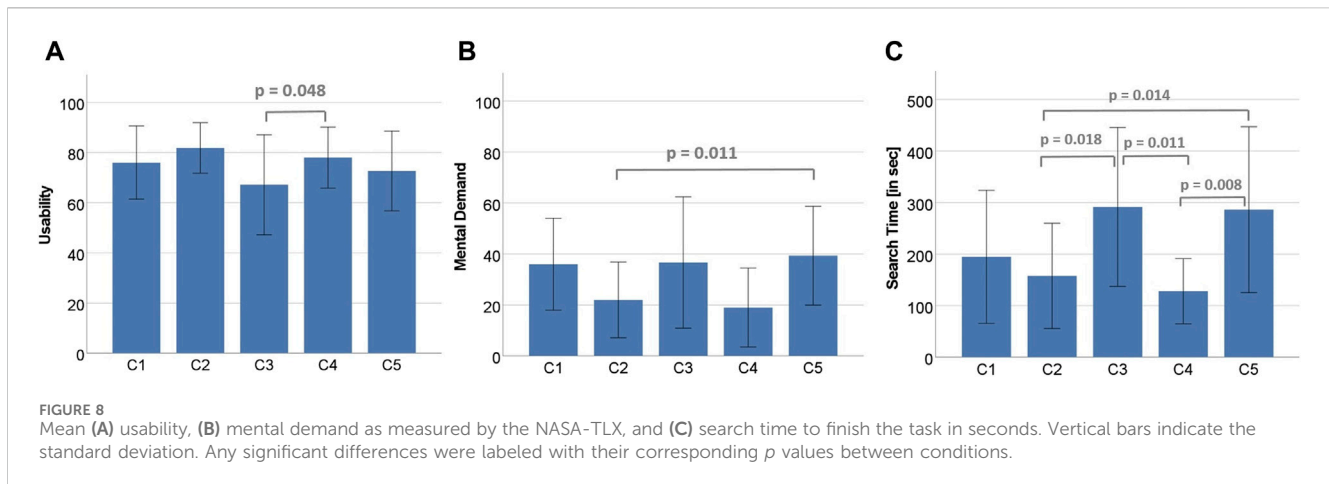
6.2.5 Cybersickness

As a subjective measure of cybersickness, participants rated the perceived severity of 16 symptoms covered by the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993). From the ratings, we calculated subscores for nausea, oculomotor, and disorientation, as well as a total cybersickness score, as suggested by Kennedy et al. (1993). No increase in cybersickness across trials was observed, so we analyzed absolute values rather than relative differences between trials. Since the residuals were not normally distributed, we performed a Friedman test for each of the four (sub) scores. For nausea, a significant difference between conditions was found, $\chi^2(4) = 10.330, p = .035$, but this could not be confirmed by *post hoc* pairwise Wilcoxon signed-rank tests with Bonferroni-Holm correction. For oculomotor, disorientation, and total cybersickness scores, no significant effects were found using Friedman tests.

6.3 Discussion

6.3.1 Spatial orientation

The magic lens effects with FoV restriction (C2, C4) achieved better ratings in terms of mental effort and higher task performance than C1, though not significant, therefore H1 was rejected. Nonetheless, qualitative user feedback indicated that C2 and C4 "make(s) it more confident/easier to navigate around the machines" (N = 2). Without the contextual guidance of CAD models, "it was difficult to get a feeling of scale or to identify different places correctly." Detailed comments mentioning spatial orientation confirmed that these effects further "helped with general orientation" and "provided a sense of [the user's] position." In contrast, the 3D NeRF drawing interaction (C3, C5) performed significantly worse than the 3D NeRF drawing effect (C2, C4), confirming H2. The main problem reported with these conditions was "difficulty to find the correct depth" (N = 4). This could be because most participants were unfamiliar with the complex environments, and using only a 2D schematic sketch as

FIGURE 8
Mean **(A)** usability, **(B)** mental demand as measured by the NASA-TLX, and **(C)** search time to finish the task in seconds. Vertical bars indicate the standard deviation. Any significant differences were labeled with their corresponding *p* values between conditions.

visual instruction does not provide users with sufficient spatial orientation to navigate efficiently to the search target. As potential future use of our framework and techniques needs to accommodate users' diverse skill sets and backgrounds, future application of the magic NeRF lens effect with 3D drawing could provide specific 3D points or markers to help orient users and provide reference points to support users who need additional visual aids and guidance for them to effectively navigate within the intricate industrial systems.

6.3.2 Task load, performance, usability, and cybersickness

Across all conditions, there were no significant differences in perceived cybersickness, with participants experiencing little to no motion sickness. In general, users found the magic lens effects with FoV restrictor (C2, C4) highly usable (SUS_{C2} = 81.83, and SUS_{C4} = 78), with better objective task performance and lower perceived mental effort. However, as assumed by H2, the magic NeRF lens effects with 3D drawing received lower usability scores, with a significant difference between (C3) and (C4). As Figure 8 shows, usability scores were moderately correlated with cognitive load (r = − 458) and task performance (r = − 543). Qualitative feedback indicates that participants often repeatedly unfolded and erased the NeRF models because they were unsure where to find the components, leading to frustration and even framerate jitters when more areas than necessary were revealed. In addition, participants mentioned that the CAD is often outdated and incomplete compared to the NeRF model. This discrepancy makes it even more difficult for users without a facility management background to complete the tasks. Nevertheless, two participants mentioned the 3D NeRF drawing effect as their preferred condition because "it was more fun" and has "a nice property of reducing complexity and putting the focus on the spots you want to investigate". Therefore, we recommend future use of 3D NeRF drawing effect mainly for facility management experts but still make it an option for non-expert users to accommodate different preferences.

6.3.3 Context rendering style

Concerning the rendering style of the polygonal representation of the CAD model (colored solid vs. edge rendering), the conditions yielded similar results in both subjective ratings and task performance. Since users reported different preferences in the open-ended feedback, a customization option could be offered in a practical application.

6.4 Expert feedback

To explore how our framework could be used by practitioners in a real industrial setting, we validated our system through expert reviews at the Deutsches Elektronen-Synchrotron (DESY), where the development of a VR NeRF system for particle accelerator maintenance was first proposed. Five facility managers and control system specialists participated in the reviews. All participants have a leading position in the design, coordination, or control of particle accelerators at DESY, and two of them are also experienced VR experts who have already developed VR systems for facility inspection.

The expert reviews were conducted using an exploratory application that illustrates a section of the NeRF model of a large-scale industrial facility. It provides many flexibility and customization options, allowing the user to freely adjust the NeRF camera's field of view, change the size of the NeRF editing sphere, vary the translucency of the merging effects, as well as the manipulation interactions mentioned in Section 4. The application ran on an Alienware m17 R2 laptop with 16 GB of RAM and an RTX 2080 GPU and was displayed through an Oculus Quest Pro HMD. We reduced the resolution of the application to 800 × 800 pixels to achieve real-time performance.

Overall, participants felt "very confident to use the system" and "it is something (they) could work with," even though the application was running on a laptop with moderate performance and moderate resolution. All expert participants confirmed that using NeRF for virtual facility inspection could benefit their workflow. One facility management expert commented that NeRF is a compelling, low-cost alternative for 3D facility documentation: "I think the system has a good advantage. It is quite nice to project the NeRF model on the CAD model, as it is a lot more effort to take laser scans of the facility." In addition, most expert participants preferred to inspect the facilities in VR because the system "helps them to see if (they) could reach anything" or "if (they) could fit any equipment through the existing environment." In addition, two experts mentioned that having a one-to-one real-

world scale NeRF model aligned with the polygonal CAD model in immersive VR also gives them a better spatial awareness of complex machines than working with a 2D desktop application. For example, they mentioned that the VR NeRF environment could help them assess in advance if "an operator's hand would fit through a narrow gap to handle components" or if special equipment would need to be prepared in advance.

The expert who leads the design and upgrade of their facilities mentioned the benefits of data fusion visualization: "With this system, I see the possibility to test something in theory before you build it in practice. For example, when you have a machine, and you want to test if you have enough space for installing it, it is quite nice you could test everything in the virtual area before you do it in reality." He also mentioned that their CAD models usually only show the initial design of the facility. Once the facility is operational, these CAD models can become incomplete and outdated. He found the contextual 3D NeRF drawing effect particularly helpful in comparing the difference between the original design and the actual implementation, which could even help operators update the original CAD designs accordingly.

For future development of the facility inspection system, one expert suggested the interesting idea of integrating contextual QR code scanning to further support information retrieval from their large inventory database. Sometimes, facility inspection tasks require scanning labels containing QR codes with manufacturing and maintenance information that are attached to all cables and equipment. The ability to retrieve such labels directly from the NeRF model would further streamline facility inspection processes. Concerning safety-critical processes such as immersive robot teleoperation, participants mentioned that although NeRF could provide photometrically accurate results, its geometric accuracy also needs to be verified and compared with conventional 3D sensors. Nevertheless, as our proposed NeRF magic lens effects and data fusion pipeline could be applied to other types of conventional 3D models, we encourage further investigation of experimenting with data fusion with other types of 3D data of large-scale facilities.

7 Limitations and future work

Our framework still has several limitations that could be addressed in future work. First, future research could investigate an automatic CAD-NeRF alignment approach to skip the manual hand-eye calibration process. Detailed investigation of an interactive point-matching algorithm could be a promising approach, taking into account the real-world mismatch between the NeRF model and the CAD model (Besl and McKay, 1992). In addition, optimization techniques such as empty space skipping and early ray termination could also lead to frame rate jitter as the number of network queries becomes view-dependent. Although the use of a FoV restrictor could reduce such effects, such framerate jitter will be more noticeable on medium and low-end graphics hardware. Therefore, we encourage future work to investigate the integration of further optimization techniques, such as foveated rendering, to enable a comfortable VR NeRF system even on low-end graphics devices.

It is also important to note that the field of photorealistic view synthesis is evolving rapidly. Relevant work on representing real-world scenes with 3D Gaussian functions as an alternative to multi-layer

perceptron has emerged in the last 2 months, paving the way for more efficient rendering of radiance fields via rasterization rather than ray marching (Kerbl et al., 2023). We encourage future work to further extend our system to support 3D Gaussian splatting in immersive VR (Kerbl et al., 2023), and to compare the different trade-offs between the two types of representation and rendering for user interaction and system performance.

8 Conclusion

We presented *Magic NeRF Lens*, an interactive immersive visualization toolkit to support virtual facility inspection using photorealistic NeRF rendering. To support the rendering of industrial facilities with substantial volume, we proposed a multimodal data fusion pipeline to visualize the facility through magic-lens-style interactions by merging a NeRF model with the polygonal representation of its' CAD models. We designed two magic NeRF lens visualization and interaction techniques and evaluated these techniques through systematic performance benchmark experiments, user studies, and expert reviews. We derived system configuration recommendations for using the magic NeRF lens effects, showing that the optimal configuration for visualizing industrial facilities at one-to-one real-world size is 20 PPD at 30° FoV, or 15 PPD at 40° FoV within an integrated VR application. Through a visual search user study, we demonstrate that our MR tunneling magic NeRF lens design achieves high usability and task performance, while the 3D NeRF drawing effect is more interactive but requires future integration of more visual guidance to support users who are not familiar with the complex facility environments. Follow-up system reviews with 5 experts confirmed the usability and applicability of our framework in support of real-world industrial virtual facility inspection tasks such as facility maintenance planning and redesign. Finally, we believe that the interdisciplinary and open-source nature of this work could benefit both industrial practitioners and the VR community at large.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The requirement of ethical approval was waived by the According to the rules of the Ethics Commission of the Department of Informatics (Faculty of Mathematics, Informatics and Natural Sciences of Universität Hamburg), a basic questionnaire was completed, and based on the results, consultation with the Ethics Commission for approval was not required. For the studies involving humans because According to the rules of the Ethics Commission of the Department of Informatics (Faculty of Mathematics, Informatics and Natural Sciences of Universität Hamburg), a basic questionnaire was completed, and based on the results, consultation with the Ethics Commission for approval was not required. The studies were conducted in accordance with

the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

KL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. SS: Data curation, Formal Analysis, Writing–original draft, Writing–review and editing. TR: Validation, Writing–original draft, Writing–review and editing. RB: Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–original draft, Writing–review and editing. WL: Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–original draft, Writing–review and editing. FS: Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–original draft, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2024.1377245/full#supplementary-material

References

Akeley, K. (2000). *Realityengine graphics*.

Aukstakalnis, S., and Blatner, D. (1992). Silicon mirage: the art and science of virtual reality.

Bae, H., Golparvar-Fard, M., and White, J. (2013). High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (aec/fm) applications. *Vis. Eng.* 1, 3–13. doi:10.1186/2213-7459-1-3

Barčević, D., Lee, C., Turk, M. A., Höllerer, T., and Bowman, D. A. (2012). "A hand-held ar magic lens with user-perspective rendering," in *2012 IEEE international symposium on mixed and augmented reality (ISMAR)*, 197–206.

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. (2021). "Mip-nerf: a multiscale representation for anti-aliasing neural radiance fields," in *2021 IEEE/CVF international conference on computer vision (ICCV)*, 5835–5844.

Besl, P. J., and McKay, N. D. (1992). Method for registration of 3-d shapes. *Sens. fusion IV control paradigms data Struct. (Spie)* 1611, 586–606. doi:10.1117/12.57955

Bier, E. A., Stone, M. C., Pier, K. A., Buxton, W., and DeRose, T. (1993). "Toolglass and magic lenses: the see-through interface," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*.

Brooke, J. (1996). "Sus: a "quick and dirty" usability scale," in Usability evaluation in industry 189.

Brown, L. D., and Hua, H. (2006). Magic lenses for augmented virtual environments. *IEEE Comput. Graph. Appl.* 26, 64–73. doi:10.1109/mcg.2006.84

Burghardt, A., Szybicki, D., Gierlak, P., Kurc, K., Pietruś, P., and Cygan, R. (2020). Programming of industrial robots using virtual reality and digital twins. *Appl. Sci.* 10, 486. doi:10.3390/app10020486

Büttner, S., Mucha, H., Funk, M., Kosch, T., Aehnelt, M., Robert, S., et al. (2017). "The design space of augmented and virtual reality applications for assistive environments in manufacturing: a visual approach," in *Proceedings of the 10th international conference on PErvasive technologies related to assistive environments*.

Chen, A., Xu, Z., Geiger, A., Yu, J., and Su, H. (2022). *Tensorf: tensorial radiance fields. ArXiv abs/2203.09517*.

Chen, Z., Funkhouser, T., Hedman, P., and Tagliasacchi, A. (2023). "Mobilenerf: exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures," in *The conference on computer vision and pattern recognition (CVPR)*.

Dai, F., Hopgood, F. R. A., Lucas, M., Requicha, A. A. G., Hosaka, M., Guedj, R. A., et al. (1997). "Virtual reality for industrial applications," in *Computer graphics: systems and applications*.

Dehne, A., Hermes, T., Moeller, N., and Bacher, R. (2017). "Marwin: a mobile autonomous robot for maintenance and inspection," in *Proc. 16th int. Conf. On accelerator and large experimental Physics control systems (ICALEPCS'17)*, 76–80.

Deng, N., He, Z., Ye, J., Duinkharjav, B., Chakravarthula, P., Yang, X., et al. (2021). Fov-nerf: foveated neural radiance fields for virtual reality. *IEEE Trans. Vis. Comput. Graph.* 28, 3854–3864. doi:10.1109/tvcg.2022.3203102

di Castro, M., Ferre, M., and Masi, A. (2018). Cerntauro: a modular architecture for robotic inspection and telemanipulation in harsh and semi-structured environments. *IEEE Access* 6, 37506–37522. doi:10.1109/access.2018.2849572

Gong, L., Berglund, J., Fast-Berglund, Å., Johansson, B. J. E., Wang, Z., and Börjesson, T. (2019). Development of virtual reality support to factory layout planning. *Int. J. Interact. Des. Manuf. (IJIDeM)* 13, 935–945. doi:10.1007/s12008-019-00538-x

Haque, A., Tancik, M., Efros, A., Holynski, A., and Kanazawa, A. (2023). *Instruct-nerf2nerf: editing 3d scenes with instructions*.

Hart, S. G., and Staveland, L. E. (1988). Development of nasa-tlx (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi:10.1016/S0166-4115(08)62386-9

Jambon, C., Kerbl, B., Kopanas, G., Diolatzis, S., Leimkühler, T., and Drettakis, G. (2023). "Nerfshop: interactive editing of neural radiance fields," in *Proceedings of the ACM on computer Graphics and interactive techniques 6*.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi:10.1207/s15327108ijap0303_3

Kerbl, B., Kopanas, G., Leimkuehler, T., and Drettakis, G. (2023). 3d Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph. (TOG)* 42, 1–14. doi:10.1145/3592433

Lazova, V., Guzov, V., Olszewski, K., Tulyakov, S., and Pons-Moll, G. (2022). *Control-nerf: editable feature volumes for scene rendering and manipulation*. arXiv preprint arXiv:2204.10850.

Lee, J.-Y., Han, P.-H., Tsai, L., Peng, R.-D., Chen, Y.-S., Chen, K.-W., et al. (2017). Estimating the simulator sickness in immersive virtual reality with optical flow analysis. *SIGGRAPH Asia 2017 Posters*. doi:10.1145/3145690.3145697

Li, K., Rolff, T., Schmidt, S., Bacher, R., Frintrop, S., Leemans, W. P., et al. (2022a). "Bringing instant neural graphics primitives to immersive virtual reality," in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 739–740.

Li, K., Schmidt, S., Bacher, R., Leemans, W. P., and Steinicke, F. (2022b). "Mixed reality tunneling effects for stereoscopic untethered video-see-through head-mounted displays," in *2022 IEEE international symposium on mixed and augmented reality (ISMAR)*, 44–53.

Li, S., Li, H., Wang, Y., Liao, Y., and Yu, L. (2022c). *Steernerf: accelerating nerf rendering via smooth viewpoint trajectory. ArXiv abs/2212.08476*.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). "Nerf: representing scenes as neural radiance fields for view synthesis," in *European conference on computer vision*.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph. (TOG)* 41, 1–15. doi:10.1145/3528223.3530127

Osterberg, G. (1937). Topography of the layer of rods and cones in the human retina. *J. Am. Med. Assoc.* 108, 232. doi:10.1001/jama.1937.02780030070033

Raposo, A. B., Corseuil, E. T. L., Wagner, G. N., dos Santos, I. H. F., and Gattass, M. (2006). "Towards the use of cad models in vr applications," in *Vrcia '06*.

Remondino, F. (2011). Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote. Sens.* 3, 1104–1138. doi:10.3390/rs3061104

Sarmita, W., Rismawan, D., Palapa, R., and Mardha, A. (2022). "The implementation for virtual regulatory inspection in nuclear facility during the covid-19 pandemic in Indonesia," in *The international conference on advanced material and technology (icamt) 2021*.

Schönberger, J. L., and Frahm, J.-M. (2016). "Structure-from-motion revisited," in *Conference on computer vision and pattern recognition (CVPR)*.

Schubert, T. W. (2003). The sense of presence in virtual environments: a three-component scale measuring spatial presence, involvement, and realness. *Z. für Medien.* 15, 69–71. doi:10.1026//1617-6383.15.2.69

Seay, A. F., Krum, D. M., Hodges, L. F., and Ribarsky, W. (2001). "Simulator sickness and presence in a high fov virtual environment," in *Proceedings IEEE virtual reality 2001*, 299–300.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P. P., et al. (2022). "Block-nerf: scalable large scene neural view synthesis," in *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 8238–8248.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., et al. (2020). *Fourier features let networks learn high frequency functions in low dimensional domains. ArXiv abs/2006.10739*.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., et al. (2023). *Nerfstudio: a modular framework for neural radiance field development. arXiv preprint arXiv: 2302.04264*.

Tominski, C., Gladisch, S., Kister, U., Dachselt, R., and Schumann, H. (2014). "A survey on interactive lenses in visualization," in *Eurographics conference on visualization*.

Viega, J., Conway, M., Williams, G. H., and Pausch, R. F. (1996). "3d magic lenses," in *ACM symposium on user interface software and technology*.

Wang, L., Zhao, Y., Mueller, K., and Kaufman, A. E. (2005). The magic volume lens: an interactive focus+context technique for volume rendering. *Vis. 05. IEEE Vis.* 2005, 367–374. doi:10.1109/VISUAL.2005.1532818

Zhang, S. (2018). High-speed 3d shape measurement with structured light methods: a review. *Opt. Lasers Eng.* 106, 119–131. doi:10.1016/j.optlaseng.2018.02.017

Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R., et al. (2018). State of the art on 3d reconstruction with rgb-d cameras. *Comput. Graph. Forum* 37, 625–652. doi:10.1111/cgf.13386