Diminished Reality Techniques for Metaverse Applications: A Perspective From Evaluation

Siru Chen, Lingxin Yu, Yuxuan Liu, Zhifei Ding, Jiacheng Zhang, Xinyue Wang, Jiahao Han, and Richen Liu

***Abstract*—The extended reality (XR) is one of the most widely used approaches for accessing the metaverse world. The metaverse and XR aim to blend the virtual and real parts, offering an immersive and interactive experience. Diminished reality (DR) is a subset of XR that specifically addresses the real-time occlusion, removal, and transparency of objects in the environment. As an immersive technology, DR has been utilized in academia and industry to tackle a wide range of engineering problems. However, there is a little investigative work about DR technique evaluations. In this survey, we categorize the state-of-the-art research into two major categories and six subcategories, providing a novel perspective. We further analyze and evaluate the application effects and performance of these approaches from both quantitative and qualitative perspectives, considering the technical performance and user experience of DR techniques. Finally, we provide an overview of potential future directions for DR applications.**

***Index Terms*—Augmented reality (AR), diminished reality, qualitative evaluation, quantitative evaluation.**

I. Introduction

THE METAVERSE is a virtual iteration of the Internet that encompasses immersive and hyper-realistic 3-D digital worlds. It enables the recreation of the physical world using techniques like digital twins [1], extended reality (XR), and holograms [2]. Alongside the creation of digital models of the physical environment, the metaverse facilitates the existence of virtual information, objects, humans, and environments that transcend physical world. The metaverse enables the blending of real and virtual contents, allowing for more intuitive comparisons and interactive explorations in real-world applications.

DR, a subset of XR, is employed to hide and eliminate objects within the real-time perceived environment [3], [4], [5], [6], [7]. DR was first introduced by Steve Mann as part of the mediated reality framework [8]. The input of DR techniques encompasses not only static image but also video sequences. DR application research faces challenges like maintaining temporal coherence and quality in texture, image, and video restoration [9]. With the rise of the metaverse, new DR methods constantly emerge, integrate AI techniques for improved applications. The DR methods for metaverse applications mainly use three types of techniques. The first type of DR technique is inpainting. It uses texture synthesis to fill in missing parts of a static image. It is often used to fix missing parts of the indoor environment [5], [10] and outdoor scenes [11], [12]. The second type is background recovery. It can help identify and restore background elements in the image that have been occluded or blurred by foreground objects. The technique is applied in DR methods, including image distortion [12] and image-based rendering [13], [14]. The third type is real-time object removal. It can remove or occlude objects through target detection and tracking algorithms [4], [15], [16].

Currently, DR has been utilized in industries. Augmented reality (AR) [17], as a subset of XR, overlays virtual objects into the real environment in real time to seamlessly integrate the virtual and real worlds. In contrast to AR, DR provides a solution for reducing, replacing, and altering perspectives to address the information overload [18] caused by excessive virtual objects in the real world [19], [20], [21]. This overload can be problematic in various applications, where an abundance of irrelevant virtual objects can distract from the intended experience. DR functions, such as reduction, replacement, and altered perspectives, could offer effective solutions to mitigate this issue [22]. DR has also found invaluable applications in surgical assistance, urban planning, gaming and entertainment, interference elimination, and obstacle removal [23], [24], [25], [26], [27], [28].

Despite the increasing utilization of DR in diverse scientific fields, there is a little comprehensive evaluative summaries regarding the practical outcomes of DR. This limitation primarily originates from two evaluation challenges: 1) the generation of precise ground-truth data sets [31] and 2) the need to address the diverse range of DR methods. Subjective analysis of the resulting images remains the prevailing method for evaluating most DR techniques [32]. The work on the subjective perception of object removal and inpainting is limited. There is a limited number of comprehensive summaries of existing DR evaluation methods. To address this gap, we conducted an analysis and provided a summary of DR evaluation approaches.

Manuscript received 13 October 2023; revised 25 January 2024 and 17 May 2024; accepted 14 June 2024. Date of publication 1 July 2024; date of current version 24 October 2024. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62372241; in part by the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant VRLAB2023B05; and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJCX24_0636. *(Corresponding author: Richen Liu.)*

The authors are with the School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China (e-mail: liurichen@gmail.com).

Digital Object Identifier 10.1109/JIOT.2024.3418034

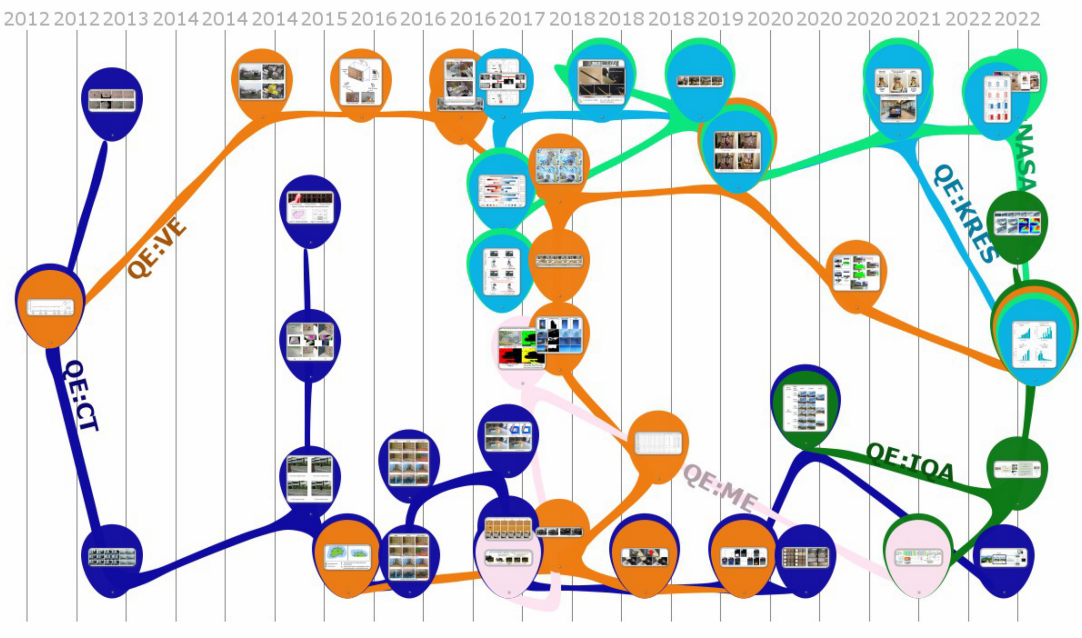

Fig. 1. We applied an interactive visualization tool named BalloonVis [29] to explore the relationship between the representative literature. Six categories of the survey visualized by the tool: the dark purple balloons represent "computation time" (CT) of quantitative evaluation (QE: CT), the green balloons indicate "IQA" of quantitative evaluation (QE: IQA), the orange balloons represent "visual evaluation" (VE) of quantitative evaluation (QE: VE), the pink balloons indicate "model evaluation" (ME) of quantitative evaluation (QE: ME), the bright green balloons represent "NASA-TLX evaluation standard" (NASA-TLX) of qualitative evaluation (QE: NASA-TLX), and the blue balloons indicate "evaluation standard proposed by Kruij and Riecke" [30] (KRES) of qualitative evaluation (QE: KRES).

A. Related Surveys

Mori et al. [3] conducted the initial survey of DR, categorizing the research based on underlying technologies, such as background observation, scene tracking, region of interest detection, hidden view generation, and composition in a five-step process. They also discussed displays and imaging devices in DR, and categorized DR works into three types, including see-through-based, projection-based, and tablet-based DR systems. Their survey is well-rounded and clearly classified, and it is more suitable for beginners in DR, aiming at educating those with less expertise in the field. Eskandari and Motamedi [33] provided a comprehensive overview of the definition, key processes, and supporting technologies of DR. Besides, they classified DR based on background recovery techniques into two categories: 1) inpainting-based diminished reality (IB-DR) and 2) observation-based diminished reality (OB-DR) [34]. OB-DR is further divided into prediminished reality (POB-DR) and real-time OB-DR (ROB-DR) [33].

Wang et al. [35] summarized the current research status of metaverse in their survey. They believed that there is still considerable scope for growth in the field of interaction design in the metaverse, as XR is in an early stage of development and has many limitations. Moreover, Wu et al. [36] summarized collaborative virtual reality (VR) techniques in the metaverse, breaking down the technical framework into three major categories: introducing the fundamental elements of creating the virtual world and the necessary approaches for constructing it. Then they discussed motion capture, XR, and brain–computer interaction, explaining how these three categories enable interaction between the virtual and real worlds. They believed that XR still has some challenges, including accurately perceiving spatial depth and improving XR devices for a more immersive experience. Finally, they elaborated on and concluded the discussion about data interaction. We discussed DR techniques for metaverse applications in this article based on the relevant surveys mentioned above.

B. Taxonomy of the Survey

We conducted a comprehensive academic literature review and searched conventional academic databases, including IEEE Xplore, ACM Digital Library, Web of Science, Springer, and Wiley, as well as Google Scholar for additional searches. We identified over 300 papers mentioning the concept of DR and other XR technologies. After excluding papers with limited relevance to DR, there were over a hundred articles. We statistically categorized the literature into two major types: quantitative (including four subcategories) and qualitative (including two subcategories) evaluation methods. For example, there are a series of representative papers that used peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), learned perceptual image patch similarity (LPIPS), real-time processing time, fillrate, and frame rate as key assessment metrics in their experiments. Other metrics, such as "subjects' visual acuity," "incoherence," and "mask ratio," were not selected as categorized metrics because they were cited rarely in these papers.

We employed the interactive tool named BalloonVis [29] (Fig. 1) to explore the literature in a focus+context exploration scheme. The general steps of using BalloonVis: 1) categorizing literature into distinct categories and subcategories;

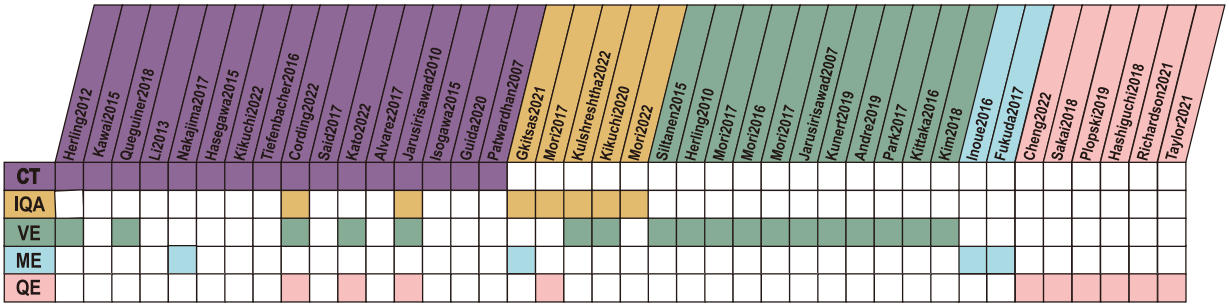

Fig. 2. Some representative papers selected from each category. The horizontal axis represents different papers, while the vertical axis means different evaluation criteria selected while evaluating the DR methods, including four quantitative evaluation categories: computation time (CT), IQA, visual evaluation (VE), model evaluation (ME), and qualitative evaluation (QE). Many papers herein share several analogous evaluation criteria across various categories. All papers are classified (with different colors) and sorted in columns according to the major property they presented.

2) identifying relationships between papers; and 3) creating visualizations that represent the subcategories along with the established relation information. Furthermore, we selected some representative papers from each category. As is displayed in Fig. 2, all papers are categorized based on their technical performance and user experience. The representative papers are separated from each category with a different color. This figure clearly demonstrates that many of these papers employ comparable evaluation methods across different categories.

II. Quantitative Evaluation

Quantitative evaluation is a numerical assessment method that utilizes statistical and computational techniques to analyze metrics of DR effectiveness. It primarily relies on metrics like accuracy, benchmarked against ground-truth comparisons, with the performance of DR methods evaluated individually for each function. There are standard data sets used for evaluating DR techniques. For example, Morozumi et al. [32] built a public data set, which serves as a fair basis for evaluating the research of DR work. This data set can be used in various scenarios and is divided into four categories: 1) objective evaluation in diverse indoor conditions; 2) assessment of outdoor daylight changes; 3) person removal scenario; and 4) dynamic scene with a miniature car. It supports the evaluation of DR methods in dynamic scenes, assessing the performance of DR methods, and comparing results in outdoor daylight changes.

Based on these four subcategories, we have created a table (see Table I) that tabulates the work conducted in each subcategory from a quantitative evaluation perspective. In addition, we counted the techniques and equipment used in DR applications mentioned in the relevant papers. In practice, we can select the quantitative evaluation methods that meet our needs according to specific situations, collect the appropriate data, conduct quantitative evaluations, and use the final results for decision making.

A. Computation Time

The general process of DR mainly includes five parts: 1) background observation; 2) sense tracking; 3) detection of the region of interest; 4) hidden view generation; and 5) composition [3]. DR techniques are often encapsulated in two stages: 1) preprocessing and 2) real-time processing. In the preprocessing stage, the raw image is mainly preprocessed to get the information required for masking or removing the specific targets. The preprocessed data is further processed and analyzed in the real-time processing stage, and the results are generated to achieve the final purpose. In practical applications, it is very difficult to get a completely accurate analysis result for the process events. However, the time period of it can be extended so that the realization time of different processes can be recorded by inpainting on the time metric, and this recorded calculation time can reflect the realization effect of DR well.

We summarized the issues in real-time DR implementations, as shown in Table II. There were 16 papers that mentioned real-time DR implementations. We described their deployment and architectures, challenges in real-time implementations, and techniques. In addition, we summarized their real-time performance according to the corresponding literature. In terms of performance, DR has some real-time challenges in the metaverse that combines the real world and virtual world. The solutions mentioned in the literature mainly focus on three aspects: 1) involving more preprocesses [5], [11], [15]; 2) utilizing high-performance real-time algorithms [16], [37]; and 3) designing better interface and interaction [53], [54]. In this survey, we summarized the preprocessing time and real-time processing time as computation time, considering computation time as a comprehensive representation of the effectiveness of DR. The processing time reflects the performance and behavior of the DR techniques at each stage.

1) Preprocessing Time: The typical implementation process of DR usually includes a preprocessing phase, although some DR techniques may not necessitate any preprocessing [4], [37], [55]. The preprocessing phase takes a photograph or video of the target structure as input, reconstructs the scene data, estimates the camera position and orientation of the target object, and produces a local feature structure as output. This output serves as the input for the real-time processing phase. For instance, Herling and Broll [11] proposed a pixel-based approach to image rendering and real-time DR, along with a real-time object selection and tracking algorithm [see Fig. 3(a)]. This method faced challenges in

TABLE I
CATEGORIES OF INDICATORS AND THEIR REPRESENTATIVE PAPERS. (IP: INPAINTING, RTOR: REAL-TIME OBJECT REMOVAL, AND BR: BACKGROUND RECOVERY)

Categories	Indicators	Representative papers	Techniques	Device
Computation time	Pre-processing time	Pixmix [11]	IP, RTOR	-
		Background geometry DR [15]	IP, RTOR	Logicool Qcam Pro 9000
		Mobile DR [5]	IP, RTOR	iPad Pro
	Real-time processing time	Internet photo collections [12]	BR, IP	-
		SLAM with viewpoint [37]	BR	RGB-D sensor
		Hand-held camera-based DR [38]	IP, RTOR	a consumer digital video camera
		GAN-based DR [16]	IP, RTOR	-
		Mono camera-based DR [39]	IP, RTOR	Acer Iconia tablet PC
		Structure propagation method [40]	IP, RTOR	-
		Landscape assessment [41]	IP, RTOR	iPad Pro
Image quality assessment	PSNR	PanoDR [42]	IP, RTOR	-
		Multi-View camera-based DR [43]	BR	Microsoft Kinect sensor, USB camera
		DR app for android smartphones [44]	IP, RTOR	RGB camera
		Automated furniture removal [10]	IP, RTOR	-
		Landscape assessment [41]	IP, RTOR	iPad Pro
		Plane-sweep algorithm [45]	RTOR	-
	SSIM	PanoDR [42]	IP, RTOR	-
		Multi-view camera-based DR [43]	BR	Microsoft Kinect sensor, USB camera
		DR app for android smartphones [44]	IP, RTOR	RGB camera
		Automated furniture removal [10]	IP, RTOR	-
		Landscape assessment [41]	IP, RTOR	iPad Pro
		Plane-sweep algorithm [45]	RTOR	-
	LPIPS	PanoDR [42]	IP, RTOR	-
		Automated furniture removal [10]	IP, RTOR	-
		Good keyframes [46]	IP	-
Visual evaluation	Video resolution	Unconstrained environments [4]	IP, RTOR	-
	RT res	Mobile DR [5]	IP, RTOR	iPad Pro
	Fillrate	PixMix [11]	IP, RTOR	-
		DR app for android smartphones [44]	IP, RTOR	RGB camera
	Frame rate	Diminished hand [47]	RTOR	RGB-D camera and RGB camera
		Interior design [48]	IP	-
		Mobile DR [5]	IP, RTOR	iPad Pro
		DLFR [49]	RTOR	camera
		Registration framework [50]	IP, BR, RTOR	6DoF sensor
		DR using real-time surface recon. [51]	IP, RTOR	iPad Pro
		3D-scanning [52]	IP, RTOR	iPad with Structure Sensor
		Landscape assessment [41]	IP, RTOR	iPad Pro
		Plane-sweep algorithm [45]	RTOR	-
Model evaluation	Global model	PanoDR [42]	IP, RTOR	-
		SLAM with viewpoint [37]	BR	RGB-D sensor
	Local model	PanoDR [42]	IP, RTOR	-

TABLE II
EVALUATION OF REAL-TIME IMPLEMENTATION OF DR. (CBF: CAMERA BLENDING FIELDS, IP: INPAINTING, RTOR: REAL-TIME OBJECT REMOVAL, AND BR: BACKGROUND RECOVERY)

Methods	Deployment and Architecture	Challenges in real time implementation	Techniques	Performance
Internet photo collections [12]	Photo collection database, 3D reconstruction system	Algorithm	BR, IP	Fair
SLAM with viewpoint [37]	SLAM, segmentation, and recognition framework	Performance and accuracy of SLAM	BR	Fair
Hand-held camera-based DR [38]	Synthesizing and overlaying the background	Complex environment, hardware limitations	IP, RTOR	Fair
GAN-based DR [16]	Combining semantic segmentation and GAN	Real time communication between devices	IP, RTOR	Good
Mono camera-based DR [39]	Multi-threaded algorithm design	Algorithm, hardware limitations	IP, RTOR	Fair
Structure propagation method [40]	Structure propagation that works near real-time	Computational efficiency, complex scenes	IP, RTOR	Fair
Landscape assessment [41]	Combining semantic segmentation and GAN	Complex scenes	IP, RTOR	Fair
Background geometry DR [15]	SLAM	Complex scenes, camera movement	IP, RTOR	Good
Pixmix [11]	DR-based on high-quality image inpainting	Complex scenes, hardware limitations	IP, RTOR	Fair
Mobile DR [5]	iPad Pro for 3D scanning	Hardware limitations	IP, RTOR	Fair
PanoDR [42]	360°DR Dataset, Structure-Disentangled DR Model	Sensitive to segmentation performance	IP, RTOR	Good
Multi-view camera-based DR [43]	Using CBF for DR	Sensitive to the application scope	BR	Fair
Automated furniture removal [10]	Automatic rectification	Artifacts	IP, RTOR	Fair
Plane-sweep algorithm [45]	Web cameras	Hardware limitations, GPU performance	RTOR	Good
Diminished hand [47]	View synthesis using re-designed CBF	Camera arrangement and algorithm	RTOR	Fair
Interior design [48]	A complete modular pipeline (8 steps)	Sensitive to environmental brightness	IP	Good

the target selection process, particularly when the target and background features had minimal differences, as reliable selection became more difficult. This also led to an increase in preprocessing time. Similarly, Kawai et al. [15] introduced a more comprehensive approach to scene DR compared to traditional methods. This approach involved using a combination of local planes to approximate the background geometry, correcting perspective distortion of textures, and constraining the search area, ultimately enhancing the quality of image restoration [see Fig. 3(b)]. This approach demonstrated that more naturalistic outcomes could be achieved without constraints on background geometries and keyframe management. Compared with other methods that require background processing, Queguiner et al. [5] proposed a DR application that did not require simplifying the background geometry [see Fig. 3(c)]. They utilized compute shaders on mobile devices to

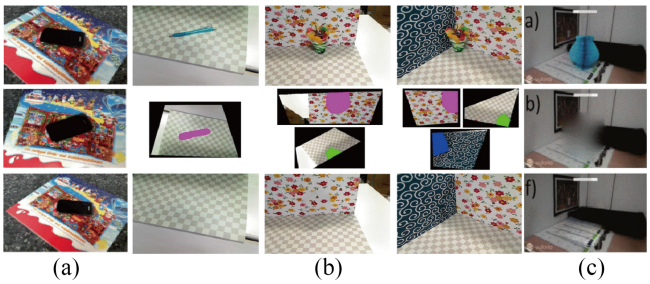

Fig. 3. Some approaches to evaluate DR using preprocessing time. (a) Even for DR on heterogenous backgrounds, the approach [11] produces a coherent video stream in real time. (b) Scenes for measuring computational time. Each row displays inputs, rectified images, and results of scenes [15]. (c) Different weighting functions to remove the blue objects [5].

address the issue of lighting variations in compensation areas, typically parts that need to be removed or covered in real scenes. These shaders could render and interpolate textures in real time on the graphics processing unit, compensating for lighting variations that might have occurred during the DR implementation process. Therefore, this method was capable of handling more complex 3-D scenes. It was important to note that this method was subject to hardware limitations, and required depth sensors and the iOS environment for scanning. Using other devices for scanning will in turn affected preprocessing time.

2) Real-Time Processing Time: Real-time performance is a critical performance metric for evaluating the efficiency of DR systems. In the field of DR technology, the term "real-time processing time" refers to the time required for an algorithm to process data in real time. The metric focuses on the time required from image or video capture to target processing completion, encompassing image analysis, object detection, and attenuation operations [56]. It is crucial to evaluate the effectiveness of DR technology. Short real-time processing times contribute to a seamless user experience, allowing developers to design more efficient DR algorithms and enhance overall system performance.

Li et al. [12] introduced a DR technique for camera positioning and real-time object removal using Internet photograph collections. This innovative approach, capitalizing on abundant online photograph data, enables a seamless transition in metaverse applications, establishing a connection between reality and the metaverse [see Fig. 4(a)]. However, this meant the program needed to efficiently analyze a substantial amount of image data for real-time DR, demanding significant computational resources. They assessed the processing time per frame in a test sequence, comparing it with other multiview methods. The results indicated satisfactory runtime performance.

A novel DR technique has been proposed by Nakajima et al. [37], which is capable of generating a 3-D model using SLAM, segmentation, and recognition frameworks without preprocessing. The technology eliminates unnecessary objects and has shown significant advantages on the UW RGB-D data set and scenes [see Fig. 4(b)]. Relying on SLAM for object alignment might have posed challenges in complex scenes, potentially leading to extended real-time processing.

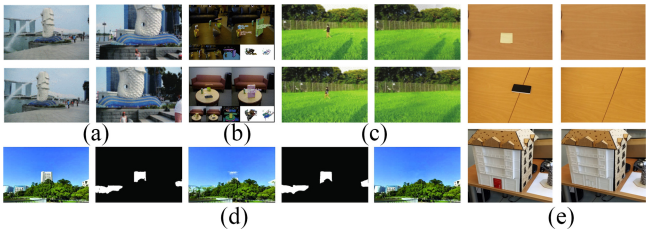

Fig. 4. Methods to evaluate DR using real-processing time. (a) Implementation of DR is achieved by leveraging the appearance and 3-D information provided by a large collection of photographs available on the Internet [12]. (b) It shows diminished results in each frame of the UW RGB-D data set when the category to be diminished is "Cereal Box" [37]. (c) Real-time processing effect of this approach [38] for grassland scenes. (d) This approach [16] detects the objects to be removed through semantic segmentation and uses GAN to complete the target area, achieving DR effect display. (e) Frame samples for the scaling sequence in simple (top), medium (middle), and complex (bottom) scenarios [40].

A method for DR using a handheld camera algorithm was proposed by Hasegawa and Saito [38]. The method utilizes a HOG-based human detector to extract pedestrians from each frame of the video sequence. Through masking and stitching background images, it effectively eliminates the masked pedestrian area, enhancing the representation of the real scene [see Fig. 4(c)]. To verify the effectiveness of the method, Hasegawa conducted experiments to calculate the fair, maximum, and minimum processing times for each frame in each scene. The experimental results showed that each scene could be processed in 0.28 s on fair, verifying that the method can achieve almost real-time processing. In the evaluation of image inpainting according to color vision [16], they proposed a DR method based on semantic segmentation and generative adversarial network (GAN). It was mainly applied to landscape assessment. The method used semantic segmentation to extraction unwanted objects/regions in landscape images and then employed GAN techniques for image restoration and reconstruction [see Fig. 4(d)]. The method effectively automated the extraction of undesirable elements in the image and restored the landscape, minimizing the need for manual intervention. Moreover, the GAN technique employed in this method preserved image details more effectively when compared to traditional image restoration methods, thus enhancing restoration quality. Nevertheless, the utilization of semantic segmentation and GAN techniques imposed high computational costs, potentially impacting real-time performance.

In order to be able to maintain the structure of the scene, Álvarez et al. [40] proposed a DR system that can run in real time. Optimizations in this system for still images included patch search reduction and image cropping and down sampling, while the global system for video used tracking techniques to maintain a stable reconstruction of the images in the video sequence [see Fig. 4(e)]. This method optimally reduces computational costs while maintaining acceptable visual quality. The utilization of parallel computing techniques is crucial for addressing future challenges. However, the method's emphasis on static scenes neglects translating and rotating objects, potentially leading to subpar target extraction and restoration.

Overall, real-time processing time is a very critical concept as it directly affects the practical feasibility and effectiveness of DR techniques. It is possible to effectively increase the value and usefulness of DR techniques by reducing real-time processing time, optimizing algorithms and utilizing large-scale data sets. In the pursuit of higher performance and better results, the continuous optimization of real-time processing time will bring a broader prospect for the development of this technology.

B. Image Quality Assessment

DR techniques must not only operate in real time and exhibit efficiency but also ensure that the diminished image meets visual requirements. Therefore, assessing the quality of image restoration is crucial to determine the effectiveness of a DR implementation. Image quality assessment (IQA) utilizes computational models to measure image quality, aiming to achieve results consistent with subjective quality evaluations. Hence, IQA can be employed to assess and optimize the quality of the diminished image, thereby enhancing reduction and achieving superior visual outcomes [57]. The primary objective of DR implementation is real-time removal, concealment, and attenuation of undesired objects in the perceptual environment [58]. The final presentation involves overlaying the effects of different frames of image restoration. Consequently, passing the IQA mechanism is of utmost importance for DR techniques. We summarized DR objective evaluation methods regarding IQA, including PSNR, LPIPS, and SSIM. These metrics are all full reference image assessment measurements.

1) PSNR: PSNR is a commonly used metric for IQA. In DR-related papers, PSNR is utilized as a metric to evaluate the quality of the image after a reduction effect has been applied. This is because the implementation of DR techniques can introduce issues, such as blurring and distortion to the image. To mitigate these problems, the diminished image can be assessed and optimized using PSNR. Researchers can quantitatively analyze the level of distortion in the image and make appropriate adjustments and improvements based on the evaluation results by measuring the PSNR. This iterative process aims to enhance the overall attenuated effect and ensure that the quality of the final rendered image meets the desired requirements.

PSNR is obtained by calculating the mean-square error (MSE) between the original image and the diminished image. A smaller MSE corresponds to a higher PSNR, indicating a better quality of the diminished image. Consequently, PSNR serves as a useful metric for evaluating the image quality of different DR techniques. It can be used to identify the best attenuation method and parameter settings, as well as to optimize existing DR techniques for improving the attenuation effect and overall visual experience.

A method called PanoDR, proposed by Gkitsas et al. [42], leverages panoramic images and stereoscopic geometry to achieve DR in indoor scenes. PanoDR automatically identified and masked elements that did not need to be displayed (such as rubbish bins and wires) with illusory objects, resulting in a tidier and safer scene for the user [see Fig. 5(a)]. They utilized

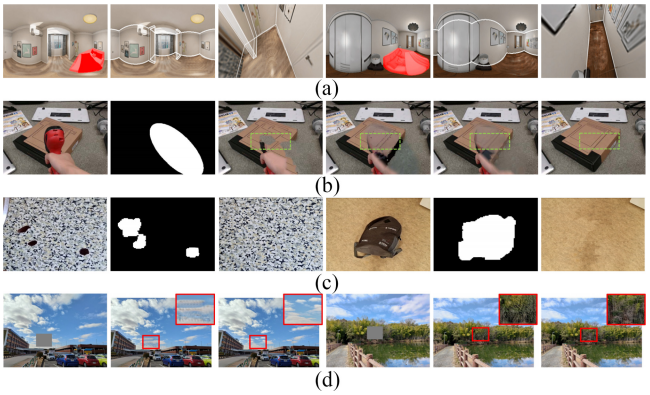

Fig. 5. Some approaches to evaluate DR using PSNR. (a) Results in this figure depict cases where RFR and PICNet provide reasonable structural coherency and aim at showcasing their model's fine-grained accuracy [42]. (b) Qualitative comparison of original view and DR methods provided by Photoshop Content-Aware Fill, surface splatting, and the proposed method [43]. (c) Application of the inpainting algorithm to various scenes [44]. (d) When the background is only one tree or only the sky, DR achieves the effect [41].

PSNR as an evaluation metric to evaluate the effectiveness of PanoDR. The quality of the DR images could be assessed by comparing the PSNR values of the method with other techniques. To measure the quality changes of images after DR processing, Mori et al. [43] also used PSNR as an evaluation tool. They employed multiple cameras to capture images from different angles of the workspace. They computationally corrected camera distortion and color bias to generate high-quality panoramic images. Subsequently, the DR technique was applied to eliminate unwanted elements and objects from the workspace, enhancing the user's visual experience and productivity in that area [see Fig. 5(b)]. The experimental results indicated that their method achieved relatively high PSNR values. As a result, it mitigated the tradeoff between work efficiency and reduction effect to some extent and provided better guidance for task completion.

To assess the restoration quality, Cording et al. [44] utilized PSNR as an evaluation metric and demonstrated the feasibility of their DR techniques by calculating PSNR values for five illustrative examples. They proposed a smartphone-based application for DR that automatically detected objects unnecessary for display in a real-time setting [see Fig. 5(c)]. In the virtual demolition method based on automatic DR [41], they proposed a virtual demolition method that utilized automatic DR techniques for landscape assessment. The method employed semantic segmentation and GAN to automate the virtual demolition process and selected PSNR as the metric for IQA to evaluate the demolished landscape [see Fig. 5(d)]. The PSNR was calculated to assess the overall demolition effect of the algorithm and verify the reliability of the evaluation results, thus guiding subsequent work.

In conclusion, PSNR, commonly employed as an IQA metric in DR research, plays a crucial role in evaluating the image quality of attenuation techniques. This metric is essential for optimizing the attenuation effect and enhancing the overall visual experience.

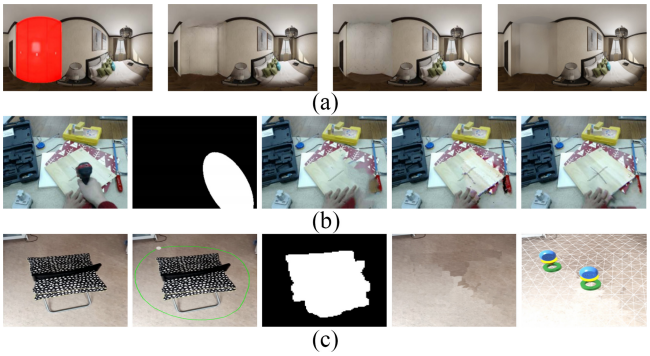

Fig. 6. Some approaches to evaluate DR using SSIM. (a) Qualitative results for diminishing objects from scenes on this test [42] set. From left to right: RFR, PICNet, and the proposed method (input image with the diminished area masked with transparent red). (b) DR result images of the similarity evaluations [43]. (c) Demonstration of the in-painting pipeline [44] running on an Android smartphone.

2) SSIM: SSIM stands for the SSIM, which is a commonly used indicator for IQA. In DR techniques, it is employed to evaluate the quality of diminished images. SSIM measures the structural similarity between two images by considering factors like brightness, contrast, and structure. The similarity between the original image and the diminished image is measured by SSIM, which yields a value ranging from 0 to 1, with 1 indicating that the two images are identical. SSIM provides a more precise measure of image distortion compared to PSNR. It can identify and quantify the loss of detailed information, thereby offering a more comprehensive evaluation of image quality.

In PanoDR [42], they utilized SSIM to assess the similarity of each pixel point and calculate the mean value, as shown in Fig. 6(a). SSIM took elements into account such as luminance, contrast, and structure, while also quantifying differences in the structure and content between the two images. In a similar vein, Mori et al. [43] employed two different methods to calculate SSIM values. They computed SSIM values for areas that were marked as masked or removed. Then, they performed SSIM calculations on the entire image following the masking process to evaluate the impact of the DR technique on the overall scene [see Fig. 6(b)]. Likewise, Cording [44] proposed a smartphone-based DR application and utilized SSIM as an evaluation metric. They compared the effectiveness of different algorithms in image removal and restoration [see Fig. 6(c)]. The efficacy of image processing could be objectively assessed, guiding algorithm optimization and improvement by calculating the SSIM value.

In summary, SSIM serves as an invaluable IQA index in the context of DR, providing crucial support and guidance for research and development in DR technology.

3) LPIPS: LPIPS is a deep-learning-based IQA metric that is widely used in DR papers. Compared to the common PSNR and SSIM metrics, LPIPS can more accurately model the features of human visual perception and can better calculate the quality difference between two images. This makes LPIPS suitable for measuring the quality of images after attenuation effects. Through a neural network trained in deep learning, LPIPS can learn a large number of image perception features

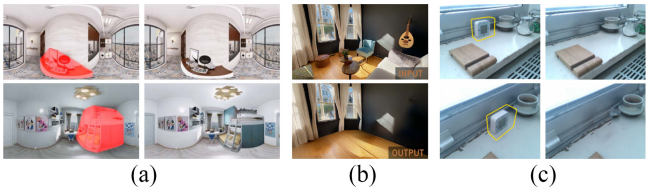

Fig. 7. Some approaches to evaluate DR using LPIPS. (a) Illustrative example of this process and the defects it solves is presented [42]. (b) System overview of this method [10], input data is provided through the input processor. (c) Two example keyframes in a scene and their good keyframes [46] to inpaint scores together with the inpainting results using the keyframes.

and measure the difference between images by calculating the Euclidean distance between image features. Compared to traditional evaluation metrics, LPIPS not only better reflects the features of human visual perception but also receives inspiration from the perceptual data during the evaluation process, which allows it to better address the issue of image distortion metrics.

PanoDR [42] also employs LPIPS for IQA to evaluate the effectiveness of their DR method [see Fig. 7(a)]. The quality of realistic reduction effects was evaluated by calculating the perceived similarity between the processed image and the original image, thus guiding the subsequent optimization and engineering development.

Similarly, Kulshreshtha et al. [10] proposed a layout-aware image restoration method for automatic furniture removal in indoor scenes [see Fig. 7(b)]. The method used a convolutional neural network that automatically learned the interrelationships between different furniture types and scene layouts, thus enabling accurate detection and removal of furniture elements from the image to be processed. They used LPIPS as a comparison metric to evaluate the performance of DR techniques, which is more suitable for evaluating the performance of large mask inpainting compared to PSNR or mean-squared deviation. A new video restoration method [46] was proposed by Mori et al. to use information from keyframes to help with missing frame restoration [see Fig. 7(c)]. In this method, a number of "good" keyframes were selected for the video restoration process. These keyframes were used together with the missing frames to train a deep-learning model. By mixing the information from the missing frames with the information from the good keyframes, the predictive power and repair effect of the deep learning model was improved. To mimic human perceptual behavior well, they chose LPIPS as a quality criterion for the DR effect and evaluated the quality stability of the method by comparing ground-truth images with patch images in multiple scenes and conditions to calculate LPIPS.

Using LPIPS to assess the image quality after DR operation allows a better measure of the differences between different techniques and parameter settings and provides an important reference for optimizing DR techniques.

C. Visual Evaluation

DR techniques can be employed to efficiently eliminate specific objects from a real-time video stream. Through the quantitative assessment of visual processing-related metrics,

Fig. 8. Demonstration of the image completion pipeline [4]; from left to right: original camera frame, original frame with active contour, object selection mask with bounding box, and resulting completion image.

the effectiveness and practical applicability of this technology can be ascertained. This type of evaluation contributes to the optimization and enhancement of DR techniques, thereby augmenting its utility. In our survey, the metrics related to visual processing encompass Video resolution, RenderTexture resolutions (RT Res), Pixel fill rate, and Frame rate.

1) Video Resolution: Video resolution refers to the number of pixels that can be displayed both horizontally and vertically, serving as a measure of the sharpness and level of detail in a video image. In DR technology, video resolution plays a crucial role, impacting the level of detail visible in the video image and the processing capabilities of the system. It provides insight into the capturing capacity of a camera or device and sets the processing resolution for the system. In the context of DR technology, higher video resolution generally translates to improve visualization. This enhanced resolution facilitates more accurate object extraction and elevates the perceived quality and viewing experience of the entire scene. While higher resolution necessitates increased computing power for image processing, potentially compromising the real-time nature of the system. Consequently, designers must strike a balance between video resolution and system processing speed to align with the available hardware and software conditions.

Video resolution holds significant importance in DR technology as it impacts various aspects, such as output quality, processing speed, and energy consumption. In their study, Herling and Broll [4] proposed an algorithm for self-enclosing object removal, enabling the DR technique in uncertain environments (see Fig. 8). This algorithm involved the extraction of high-resolution video images, which were subsequently compensated for within the scene using multiple layers of flexible templates and local polynomial regression. These techniques addressed a multitude of disturbances, including lighting variations, color variations, and motion blur. The algorithm achieved real-time application speed while ensuring accurate scene detail processing by optimizing the selection of video resolution. The choice of video resolution should be tailored to individual scenes, considering factors such as object size, complexity, and real-time requirements. This approach allowed the algorithm to adapt and cater to the diverse demands of different scenes.

Additionally, depending on the hardware configuration and technical requirements, it may be viable to opt for cameras with different video resolutions and scale down the resolution as necessary to boost the processing speed of the DR system. This tradeoff ensures that the system effectively meets performance requirements while efficiently executing object recognition, removal, and image restoration tasks.

2) RenderTexture Resolutions: RT res refers to the reduction of camera resolution to enable real-time repositioning in real-world technology. The chosen size of RT res is directly linked to the performance of DR techniques. Queguiner et al. [5] asserted that the size of RT res directly influences the real-time performance of real-world technologies. The real-time resolution will have a significant impact on metaverse applications. Specifically, a higher RT res can enhance the immersive and interactive experience within the metaverse. It allows for more detailed and realistic graphics, smoother animations, and responsive interactions. Users will be able to perceive the metaverse with greater clarity and precision, enhancing its life like and engaging feel.

In real-time systems, processing speed improves as the volume of data to be processed decreases. Therefore, minimizing the RT res value reduces the amount of data that requires processing, enhancing the real-time performance of DR technology. Simultaneously, a smaller camera resolution allows for faster data processing for operations such as color correction and brightness adjustment, thereby contributing to improving image quality. However, it is crucial to strike a balance when selecting RT res, as excessively small camera resolutions can result in blurred visual data and loss of detail, thereby compromising the weakened reality effect. Therefore, when choosing RT res, researchers need to consider the tradeoff between reducing real-world effects and optimizing real-time performance. Various factors, such as the size and complexity of objects to be removed, processing speed, and image quality, should be taken into account when determining appropriate RT res values for different scenarios.

3) Pixel Fill Rate (Fillrate): Fillrate is a vital technique employed to achieve object attenuation by replacing the area covered by an object with a background. Fillrate refers to the speed and efficiency at which a graphics processing unit (GPU) can draw and render images. It determines the number of pixels that the GPU can render on the screen. It plays a pivotal role in achieving DR effects, impacting both the quality of the effect and real-time performance. Herling and Broll [11] introduced PixMix, a real-time fillrate technology for video, that efficiently achieves high-quality DR effects.

The significance of fillrate in DR evaluation can be attributed to two main factors. First, at the pixel level, fillrate technology enhances object smoothing and reduces edge roughness by finely filling the background texture, effectively preventing noticeable visual defects. Second, the real-time performance of fillrate technology significantly influences the evaluation of DR. Employing fillrate technology enables the efficient realization of high-quality DR effects. Similarly, Cording [44] employed pixel fill rate as an indicator for evaluating the effectiveness of DR. Experimental results demonstrated that as the Fill Rate increases, the effectiveness of DR improves. In metaverse applications, a high fillrate is crucial because it determines the quality and smoothness of the images. If the fillrate is low, the rendered images may appear blurry, low-resolution, or exhibit noticeable pixelation, resulting in a poor visual experience for users. Therefore, when developing and designing metaverse applications, it is essential to optimize the pixel fill density while meeting the minimum

fill rate required for satisfactory effect quality. Balancing these factors ensures better real-time performance and enhances user experience.

4) Frame Rate: The frame rate significantly influences the quality of DR results. In general, higher video frame rates contribute to smoother and more natural DR effects, enhancing the user viewing experience. This is primarily because higher frame rates provide more information, leading to improved system recognition of video scene changes. Consequently, object segmentation and attenuation can be performed more accurately. Additionally, a higher frame rate enables faster image processing and pixel filling, reducing flickering or jittering in residual or foreign object areas and enhancing the quality of DR effects. Siltanen [48] proposed the application of AR in interior design to address the replacement of existing furniture. They emphasized the critical role of frame rate in achieving DR effects and user experience in AR environments, as it directly affected the smoothness and stability of displayed DR in AR scenes.

High frame rate images result in smoother and clearer video playback, reducing screen stuttering and faults, thus optimizing the display of DR techniques and enhancing user experience in metaverse applications. Furthermore, the frame rate directly impacts the real-time performance and responsiveness of the system. Developers must carefully select an appropriate frame rate that optimizes the effectiveness of the DR function while maintaining the balance between display quality and system performance requirements.

D. Model Evaluation

In DR techniques, model evaluation mainly involves evaluating different levels of models that describe a system or software application. Model evaluation includes the global model and the local model, and their difference lies in the scope and granularity of modeling. The global model provides a comprehensive view of the entire system, describing the structure, functionality, interactions, and behavior of the system. In contrast, the local model involves detailed modeling and analysis of a specific part or component within the system. It focuses on the specific details of the component's functionality, behavior, and interface. In the context of metaverse applications, global models and local models will play a crucial role in supporting the complex and diverse content and interactions within the metaverse.

1) Global Model Evaluation: In DR system, the global model encompasses a comprehensive understanding of both the real-world and virtual scenes. It provides a wealth of information including spatial data, image semantics, and virtual object positioning within the real scene. These functionalities help the system understand the interaction relationships among various elements and entities in the metaverse, facilitating the alignment between virtual objects and real-world scenes. Nakajima et al. [37] proposed a method based on SLAM and semantic object selection [see Fig. 9(a)]. In this approach, the global model contained two main layers: map construction and semantic object detection. The map construction layer entailed the utilization of SLAM technology to establish a real-time map that encompasses various data, such as spatial information and image features within the scene. The semantic object detection layer involves the identification of different categories of target objects (e.g., roads, sky, buildings, and pedestrians) based on the semantic segmentation information derived from the global model's scene. Based on semantic object detection, recognition techniques are employed to identify regions that resemble the objects requiring removal. The global model facilitates these techniques, which can further reduce the localization of the target area.

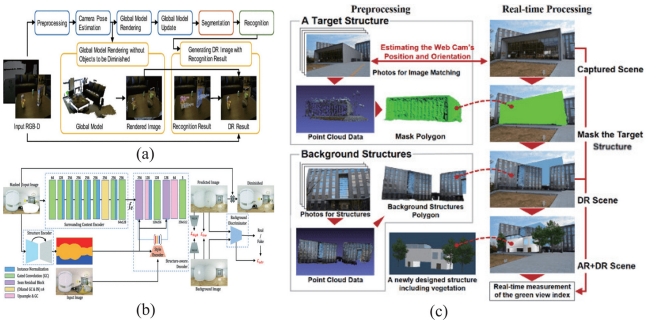

Fig. 9. Some approaches for evaluating DR using model evaluation. (a) Flow of the proposed method (depicted in blue: SLAM part, red: segmentation part, green: recognition part, and yellow: DR part) [37]. (b) Detailed design process of the PanoDR [42] work. (c) This system [59] enables simultaneous measurement of the green view index and simulation of building, urban, and planting designs.

Additionally, Gkitsas et al. [42] utilized a global model to achieve their DR techniques. They proposed an indoor DR method based on spherical panorama [see Fig. 9(b)], which enables natural, high-precision, and real-time DR operations. In this context, the global model refers to a model that predicted the position and size information of objects in the scene using the panoramic image from the indoor DR method based on spherical Panorama. Experimental results demonstrated that the global model accurately detected and segmented the object area that required reduction. It also adjusted the object position and size in the spherical panorama, ensuring visual consistency and preserving the naturalness of the scene. Hence, the global model plays a pivotal role in achieving DR techniques by providing highly accurate and comprehensive scene information.

2) Local Model Evaluation: Compared to the global model, the local model refers to a model that only describes a part or a specific component of the system. It provides a detailed modeling and analysis of a certain part or specific component in the system. The local model can help describe the details of specific areas or entities in the metaverse, including buildings, terrain, objects, characters, and more. Through the local model, specific characteristics and behaviors of individual objects in the metaverse can be presented. This allows for precise object positioning and size estimation, enhancing the accuracy and realism of object reduction while maintaining visual consistency and naturalness within the scene. The local model can also be used for real-time dynamic analysis. Fukuda et al. [59] employed a local model to divide the panoramic image into several smaller blocks. Real-time

detection was then performed on each block to achieve accurate object localization and size estimation. This information could be used to adjust the reduced object regions in the global model, thereby enhancing the accuracy and naturalness of the DR method [see Fig. 9(c)]. Consequently, the local model plays a vital role in DR techniques, contributing to improved effectiveness and realism. In other DR techniques, the local model can also be utilized to achieve precise positioning and size estimation of specific objects or regions, enhancing the real-time capabilities and accuracy of DR operations.

Overall, the global model and the local model complement each other in metaverse applications. The global model provides an overview and management of the overall situation, while the local model emphasizes individualized presentation and interactive experiences. If the two models work together, they enable DR techniques to accurately reduce or modify objects and features in real-world scenes, thus achieving a more realistic and immersive user experience in metaverse applications.

III. Qualitative Evaluation

The quantitative evaluation above may ignore some subjective metrics about the feedback from users. It may not have a comprehensive understanding of specific interaction requirements. However, qualitative evaluation provides a subjective perspective by focusing on the feelings and experiences of users. There are two qualitative evaluation metrics, i.e., the NASA-TLX and the metric [30] to evaluate system experience. The NASA-TLX is commonly used in recent studies related to measuring workload and assessing subjective workloads in various tasks, such as human–computer interaction (HCI), aviation, healthcare, and cognitive psychology. The second qualitative evaluation metric [30] has a wide range of applications. It is often used in research related to VR, AR, HCI, and user experience design. These perspectives are valuable for assessing and understanding the impact of immersive applications on users in various contexts.

A. NASA-TLX Evaluation Standard

1) Mental Demand: Mental demand refers to the level of cognitive and mental activity required to complete a task. It encompasses the user's concentration, the difficulty of the task, and the cognitive load and thought processes necessary to accomplish it. The mental demand experienced by a user is influenced by the task's complexity and difficulty, as well as the user's skills, knowledge, experience, and other individual factors. For instance, a proficient programmer may perceive a low mental demand when using a programming tool, as they have mastered its use and skills. On the other hand, a beginner with no programming experience may perceive a higher mental demand when using the same tool, as they need to allocate more cognitive and mental capacity to understand and utilize it. When designing HCI systems [60], it is crucial to understand and consider the mental demand experienced by the user. In the case of D-ball work [54], it allows players to customize difficulty levels based on their own capabilities. High cognitive load and emotional demands may lead to a poor user experience. If the mental demand is excessively high, users may experience fatigue, frustration, discomfort, and may even abandon the system or application altogether. Therefore, designers should strive to minimize the mental demand placed on users. This can be achieved by simplifying the interface, offering clear instructions and feedback, and reducing the amount of information that users must memorize. These design strategies enhance the user's experience and efficiency.

2) Physical Demand: The amount of physical activity required depends on both the quantity and difficulty of the task. Physical demand is a crucial factor to consider when evaluating a task or process, as it directly relates to the amount of physical activity required and the level of difficulty associated with completing the task. In the context of our survey on DR, we found that the experiments were not particularly physically or mentally demanding for the participants. During the DR process, participants were engaged in manipulating and interacting with virtual objects using a mobile device or computer interface. The physical actions required for this task mainly involved gestures, such as tapping, swiping, or using the rotation of the head to control the position of the perspective area, as demonstrated in the work by Lindemann and Rigoll [61]. These actions were relatively simple and did not impose significant physical strain or exhaustion on the participants.

3) Temporal Demand: Temporal demand refers to the urgency and constraints associated with completing a task within a specific timeframe. For instance, Hasegawa and Saito [38] implemented each scene with an fair processing time of 0.28 s, demonstrating that their proposed method can operate nearly in real time. In our quantitative evaluation, we primarily focused on computation time and will not delve further into this aspect in this section.

4) Performance: Performance workload is related to the accuracy and success achieved while performing the task. In interactive DR tasks, the completion success rate of the DR system is usually close to 100%, which meets the author's initial requirements. In addition, the performance workload also includes the user's evaluation of the overall perception of the system. In the research of Richardson et al. [62], subjective ratings are also used as the evaluation standard of task performance. These ratings consist of five composite subjective assessments: 1) perceived usefulness; 2) prominence; 3) perceivability; 4) effectiveness in attracting attention; and 5) intelligibility. These five comprehensive indicators can be applied to most interactive DR systems, helping researchers to better understand the needs for users so that they can improve the system performance.

5) Effort: Effort refers to the degree of mental and physical effort required by the user to perform the task. In the close quarters robotic telemanipulation work of Taylor et al. [63], effort refers to the extra effort and inconvenience for the user to change perspective. Additional physical and mental effort will have an impact on the accuracy of the user's manipulation of the robot. An efficient DR approach using real-time surface reconstruction [51] allows users to switch between reconstruction mode and DR mode at any time. Although their method would in principle allow both modes to

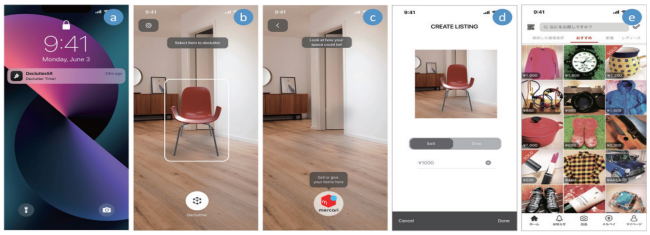

Fig. 10. DeclutterAR screens [53] design. DeclutterAR uses DR to remove real-world objects from AR scenes, allowing users to visualize their "cluttered space" as a "de-cluttered" space.

be activated simultaneously, the ability to switch at any time would be more convenient and less labor-intensive for the user.

6) Frustration: Frustration represents the degree to which individuals experience stress, discouragement, or dissatisfaction during the execution of a task [64]. This aspect assesses the frustration induced by various factors, such as task complexity, uncertainty, or issues related to the system's performance [65]. DeclutterAR [53] conducted interviews with users to gauge their frustration of using the app and their assessment of its feasibility and innovativeness, as shown in Fig. 10. Frustration can help identify issues in DR systems that may cause user frustration, such as image processing delays or DR effects that do not meet user expectations. By identifying these issues, targeted improvements can be made.

B. Evaluation Metric Proposed by Kruij and Riecke [30]

1) Cognition: Cognition refers to the cognitive perspective, which aims to assess whether the interactive system in question will impose cognitive difficulties or challenges on the user. The design of all DR systems takes into consideration the user's cognitive load. In the work InvisibleRobot [66], the authors proposed a DR method that superimposes background information onto the robot in the user's field of view through an optical perspective head-mounted display [67]. They presented two modes of InvisibleRobot visualization, where the user could choose to remove the robot's body from view completely or to remove the robot's interior while maintaining its silhouette. In their user study, participants were required to practice robot control for 5 min prior to the experiment and operate the InvisibleRobot separately from the traditional robot during the experiment, and they conducted several sets of experiments under different occlusion conditions. Based on the experimental results and user feedback, the authors assessed the user's ease of operation and psychological perceptual impressions from a cognitive perspective.

2) Game User Experience: Game user experience refers to the overall feeling and emotional response of users when interacting with a product or system. It encompasses their perception of ease of use, usability, satisfaction and happiness. User experience involves evaluating the intuitiveness, interactivity, response speed, stability, and interface design of the system. In the context of evaluating an interactive system, a generic game experience perspective is adopted. Commonly used approaches include: 1) the game experience questionnaire, which assesses factors, such as absorption, fluency, and immersion and 2) the immersive experience questionnaire, which evaluates aspects like emotion, awareness, departure from reality, challenge, and degree of control. Cording [44] and Guida and Sra [68] conducted statistical analyses of user perceptions. Lindemann and Rigoll [61] conducted a short post-test survey to evaluate user experience and obtain additional qualitative feedback of user's comfort in using the DR system.

3) Presence: Presence refers to the sensation of being immersed in a mixed reality environment, specifically, the extent to which users feel real and engaged when interacting with a virtual environment. It is influenced by various factors, including the realism of the virtual environment, the quality of interaction and equipment, among others. In DR systems, immersion and user experience are interconnected. Users' immersion heightens their interactive and sensory engagement within the virtual environment, which subsequently impacts their perception of the overall experience. Therefore, in order to improve the user experience of mixed reality system, it is necessary to consider the immersion and other factors, such as HCI, interface design, and system performance.

4) Vection: Vection pertains to self-motion, particularly the player's sense of movement within the virtual environment (locomotion), i.e., whether the player perceives themselves as moving or not. Our spatial and positional changes are determined by vision, hearing, the vestibular system, and somatosensory inputs (skin and subcutaneous perception, providing the sensation of "contact" and "force"). VR utilizes these elements to "trick" the player into experiencing virtual vection, which is a crucial factor influencing presence. In the work of Hashiguchi et al. [69], a system was developed to investigate whether users can perceive objects as being heavier or lighter than their actual weight by dynamically adjusting the visual representation of real objects using AR and digital reconstruction renderings. The system allowed users to interact with physical objects while observing them through an AR display.

IV. Discussion and Conclusion

DR is a novel method that can fastly reduce or eliminate objects in the real world, thus realizing the interaction effect between virtuality and reality. Our DR evaluation is mainly applied to reduce or modify indoor or outdoor scenes to help provide users with better interactive experiences and application services.

A. Limitation Discussion

1) Limitations of Evaluation Methods: We categorized DR methods into two types: 1) quantitative evaluation and 2) qualitative evaluation. There are some limitations and potential biases of DR evaluation, especially when researchers just use either qualitative or quantitative assessment. The essence of quantitative research lies in the use of digitized data and statistical analysis methods in order to quantify and measure phenomena to draw statistically generalized conclusions [70]. The results of the quantitative research are objective and precise, with a high degree of generalization. Qualitative

research is essentially the collection of nonquantitative data to understand and explain users' behavior, opinions, and experiences of experimental subjects [71]. Inevitably, using quantitative evaluation methods may ignore some subjective metrics about the feedback from users, and may not provide a comprehensive understanding of specific interaction requirements. In contrast, DR techniques evaluated solely through qualitative metrics may be susceptible to subjective feedback from participants, resulting in a lack of objectivity in the evaluation results. Researchers and developers can fully utilize the merits of the two types of evaluations.

2) Limitations of the Current DR Methods: We concluded with three challenges in the current DR methods.

1) *Challenges on Real-Time Processing:* DR techniques require very high real-time performance, requiring real-time processing of video streams with almost no latency. Despite the continuous improvement of computer processing power, it may still be limited by hardware [5] and algorithm [12], [56] limitations.
2) *Challenges in Handling Complex Scenes:* DR techniques may encounter difficulties in handling complex scenes. For example, accurately removing specific objects from the video stream may become complex, when multiple objects are overlapping or occluding in the scene [39], [41].
3) *Challenges on Algorithm Accuracy:* DR techniques need to accurately identify and segment objects in the video stream, remove them from the video sequences, and then recover the background. However, the accuracy of the algorithm may be affected by object recognition, tracking, and segmentation [58].

B. Potential Trend Discussion

We present four potential trend discussions.

1) Integrate More Advanced Computer Vision Methods Into DR Applications: Computer vision approaches, such as object detection and recognition, image generation, 3-D reconstruction, etc., can help DR improve user experience, increase efficiency, and provide more innovations and solutions for various application scenarios. One of the future directions is to integrate state-of-the-art computer vision techniques into DR applications.

2) Build DR Applications Based on Digital Twin Techniques: A digital twin like environment can be used to achieve in-context multiattribute comparisons of physical objects with text labels or textual information. For example, users can use additional attributes, such as ratings, publishers, comments, publication years, keywords, prices, etc., or a combination of these factors to regroup or rerank the virtual books, in the library case of DRCmpVis [72].

3) Apply Intelligent Modeling Methods or Artificial Intelligence Generated Content to Generate the 3-D Models of Physical Objects in DR Environment: The generated 3-D models can be used as virtual avatars to replace the physical objects. In interaction, artificial intelligence generated content (AIGC) may attempt to understand the user's perception. A DR method that can automatically generate or manually edit virtual objects can provide users with a more realistic drawing experience. AIGC-driven session interfaces can provide new opportunities for enriching virtual physical blended environments.

4) Embed LLM (Like ChatGPT) Into the Metaverse: LLM can better help annotate and illustrate the augmented information of physical objects. The GPT-like model will trigger content singularity, assisting human users in interacting with virtual objects in a DR environment.

C. Conclusion of This Article

Our survey is of significant importance to developers and researchers involved in the development of metaverse applications. First, it provides a perspective on the use of DR techniques, which is highly valuable for the improvement and development of metaverse applications. Second, our article includes evaluation methods and practical experiences regarding the use of DR techniques in the metaverse, which can provide useful guidance and inspiration for developers. Most importantly, as a nascent field, the metaverse requires ongoing research and innovation. This article offers an exploration of the use of DR techniques in metaverse applications, providing valuable reference and inspiration for researchers.

In this survey, we draw a summarization of how to evaluate DR-based methods and suggest possible future directions based on DR. We aim to help readers better understand the state-of-the-art of DR technologies in the metaverse and provide guidance for its further development and applications.

The implications of the survey findings for developers and researchers working on metaverse applications can be concluded into two aspects. First, our survey potentially provides a perspective on the evaluation of DR applications, which would be valuable for their improvement and development. DR allows operations, such as removal and visual restoration of the user's visual scene through an HMD device (e.g., Microsoft HoloLens, Meta Quest, and Google Glass). We believe DR applications are important supplements to AR [73] and MR applications because DR can provide flexible interaction performed on virtual avatars. Second, it provides evaluation choices on both technical performance and user experience. Various combinations of quantitative and qualitative evaluations can be recommended for users according to their requirements. They assist developers and researchers in optimizing the performance of metaverse applications.

DR, as one of the important XR technologies, is expected to show more application scenarios in various fields. From academic research to industrial applications [74], DR will continue to play an important role, bringing users a more realistic, smooth, and shocking experience. It is anticipated that future developments of DR in terms of hardware and software will be able to meet the evolving demands of the metaverse. Additionally, there tend to be more indicators established for evaluating DR. We think the evaluation methods may place greater emphasis on user perception and experience in the future. In addition, more application-specific evaluation methods will emerge according to different application areas of DR methods. As the metaverse develops, its application

may raise a series of ethical issues, such as privacy protection, data security [75], user safety, and behavioral norms in VR. For example, in DRCmpVis [72], when users wear HMDs in the library, positional deviations may lead to them bumping into wall bookshelves. It would be preferable that users are not psychologically burdened or have their physical health threatened in DR experience with their personal information kept confidential.

REFERENCES

[1] Y. Li et al., "DTBVis: An interactive visual comparison system for digital twin brain and human brain," *Vis. Inform.*, vol. 7, no. 2, pp. 41–53, 2023.

[2] I. F. Akyildiz, "Metaverse: Challenges for extended reality and holographic-type communication in the next decade," in *Proc. ITU Kaleidosc. Ext. Real. How Boost Qual. Exp. Interoper.*, 2022, pp. 1–2.

[3] S. Mori, S. Ikeda, and H. Saito, "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, p. 17, 2017.

[4] J. Herling and W. Broll, "Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2010, pp. 207–212.

[5] G. Queguiner, M. Fradet, and M. Rouhani, "Towards mobile diminished reality," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjun.*, 2018, pp. 226–231.

[6] S. Zokai, J. Esteve, Y. Genc, and N. Navab, "Multiview paraperspective projection model for diminished reality," in *Proc. IEEE ACM Int. Symp. Mixed Augment. Real.*, 2003, pp. 217–226.

[7] R. Liu et al., "Interactive extended reality techniques in information visualization," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 6, pp. 1338–1351, Dec. 2022.

[8] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer, "Towards understanding diminished reality," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2022, pp. 1–16.

[9] T. Kato, N. Isoyama, N. Kawai, H. Uchiyama, N. Sakata, and K. Kiyokawa, "Online adaptive integration of observation and inpainting for diminished reality with online surface reconstruction," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjunct*, 2022, pp. 308–314.

[10] P. Kulshreshtha, N. Lianos, B. Pugh, and S. Jiddi, "Layout aware inpainting for automated furniture removal in indoor scenes," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjunct*, 2022, pp. 839–844.

[11] J. Herling and W. Broll, "PixMix: A real-time approach to high-quality diminished reality," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2012, pp. 141–150.

[12] Z. Li, Y. Wang, J. Guo, L.-F. Cheong, and S. Z. Zhou, "Diminished reality using appearance and 3D geometry of Internet photo collections," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2013, pp. 11–19.

[13] S. Meerits and H. Saito, "Real-time diminished reality for dynamic scenes," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Workshops*, 2015, pp. 53–59.

[14] S. Mori, F. Shibata, A. Kimura, and H. Tamura, "Efficient use of textured 3D model for pre-observation-based diminished reality," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Workshops*, 2015, pp. 32–39.

[15] N. Kawai, T. Sato, and N. Yokoya, "Diminished reality based on image inpainting considering background geometry," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 3, pp. 1236–1247, Mar. 2016.

[16] T. Kikuchi, T. Fukuda, and N. Yabuki, "Diminished reality using semantic segmentation and generative adversarial network for landscape assessment: Evaluation of image inpainting according to colour vision," *J. Comput. Design Eng.*, vol. 9, no. 5, pp. 1633–1649, 2022.

[17] B.-K. Seo, M.-H. Lee, H. Park, and J.-I. Park, "Projection-based diminished reality system," in *Proc. Int. Symp. Ubiquitous Virtual Real.*, 2008, pp. 25–28.

[18] G. Albanis et al., "An AI-based system offering automatic DR-enhanced AR for indoor scenes," in *Proc. Adv. Intell. Virtual Real. Technol.*, 2023, pp. 187–199.

[19] R. Liu et al., "Narrative scientific data visualization in an immersive environment," *Bioinformatics*, vol. 37, no. 14, pp. 2033–2041, 2021.

[20] C. W. M. Leao, J. P. Lima, V. Teichrieb, E. S. Albuquerque, and J. Kelner, "Altered reality: Augmenting and diminishing reality in real time," in *Proc. IEEE Virtual Real. Conf.*, 2011, pp. 219–220.

[21] H. Tamura, H. Saito, F. Shibata, and Y. Kameda, "Diminished reality as challenging issue in mixed and augmented reality (IWDR2015) summary," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Workshops*, 2015, pp. 25–25.

[22] H. Okumoto, M. Yoshida, and K. Umemura, "Realizing half-diminished reality from video stream of manipulating objects," in *Proc. Int. Conf. Adv. Informat. Concepts, Theory Appl.*, 2016, pp. 1–5.

[23] N. Ienaga et al., "First deployment of diminished reality for anatomy education," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2016, pp. 294–296.

[24] S. Cacho-Elizondo, J.-D. Lázaro Álvarez, and V.-E. Garcia, "Assessing the opportunities for virtual, augmented, and diminished reality in the healthcare sector," in *The Digitization of Healthcare: New Challenges Opportunities*. London, U.K.: Palgrave Macmillan, 2017, pp. 323–344.

[25] T. Sawabe, Y. Okami, M. Kanbara, Y. Fujimoto, and H. Kato, "Diminished reality for sense of movement with XR mobility platform," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjunct*, 2022, pp. 348–351.

[26] I. Murph, K. Richardson, and A. McLaughlin, "Methods of training to overcome distraction via diminished reality," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 66, no. 1, pp. 1844–1848, 2022.

[27] D. S.-M. Liu and Y.-J. Chen, "Rain removal system for dynamic scene in diminished reality," *Signal, Image Video Process.*, vol. 14, pp. 945–953, Jan. 2020.

[28] T. Fukuda, Y. Kuwamuro, and N. Yabuki, "Optical integrity of diminished reality using deep learning," in *Proc. Int. Conf. Educ. Res. Comput. Aided Archit. Design Eur.*, 2017, pp. 241–250.

[29] X. Wang et al., "Hybrid line-based and region-based interactive set data visualization," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2021, pp. 1–7.

[30] E. Kruijff and B. E. Riecke, "Navigation interfaces for virtual reality and gaming: Theory and practice," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2018, pp. 1–4.

[31] S. Mori, Y. Eguchi, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, "Design and construction of data acquisition facilities for diminished reality research," *ITE Trans. Media Technol. Appl.*, vol. 4, no. 3, pp. 259–268, 2016.

[32] T. Morozumi, S. Mori, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, "[POSTER] design and implementation of a common dataset for comparison and evaluation of diminished reality methods," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2017, pp. 212–213.

[33] R. Eskandari and A. Motamedi, "Diminished reality in architectural and environmental design: Literature review of techniques, applications, and challenges," in *Proc. Int. Symp. Autom. Robot. Constr.*, 2021, pp. 995–1001.

[34] H. Matsuki, S. Mori, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, "Considerations on binocular mismatching in observation-based diminished reality," in *Proc. IEEE Symp. 3D User Interfaces*, 2016, pp. 261–262.

[35] H. Wang et al., "A survey on the Metaverse: The state-of-the-art, technologies, applications, and challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14671–14688, Aug. 2023.

[36] D. Wu, Z. Yang, P. Zhang, R. Wang, B. Yang, and X. Ma, "Virtual-reality interpromotion technology for metaverse: A survey," *IEEE Internet Things J.*, vol. 10, no. 18, pp. 15788–15809, Sep. 2023.

[37] Y. Nakajima, S. Mori, and H. Saito, "Semantic object selection and detection for diminished reality based on SLAM with viewpoint class," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2017, pp. 338–343.

[38] K. Hasegawa and H. Saito, "Diminished reality for hiding a pedestrian using hand-held camera," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Workshops*, 2015, pp. 47–52.

[39] P. Tiefenbacher, M. Sirch, and G. Rigoll, "Mono camera multi-view diminished reality," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2016, pp. 1–8.

[40] H. Álvarez, J. Arrieta, and D. Oyarzun, "Towards a diminished reality system that preserves structures and works in real-time," in *Proc. Int. Joint Conf. Comput. Vis., Imaging Comput. Graph. Theory Appl.*, 2017, pp. 334–343.

[41] T. Kikuchi, T. Fukuda, and N. Yabuki, "Automatic diminished reality-based virtual demolition method using semantic segmentation and generative adversarial network for landscape assessment," in *Proc. eCAADe Conf.*, 2021, pp. 529–538.

[42] V. Gkitsas, V. Sterzentsenko, N. Zioulis, G. Albanis, and D. Zarpalas, "PanoDR: Spherical panorama diminished reality for indoor scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 3711–3721.

[43] S. Mori, M. Maezawa, and H. Saito, "A work area Visualization by multi-view camera-based diminished reality," *Multimodal Technol. Interact.*, vol. 1, no. 3, p. 18, 2017.
[44] H. Cording, "Towards real-time object removal and inpainting through a diminished reality application for smartphones," B.S. thesis, Bachelor Comput. Sci. Eng., Delft Univ. Technol., Delft, The Netherlands, 2022.
[45] S. Jarusirisawad, T. Hosokawa, and H. Saito, "Diminished reality using plane-sweep algorithm with weakly-calibrated cameras," *Prog. Inform.*, vol. 7, no. 7, pp. 11–20, 2010.
[46] S. Mori, D. Schmalstieg, and D. Kalkofen, "Good keyframes to inpaint," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 9, pp. 3989–4000, Sep. 2023.
[47] S. Mori, M. Maezawa, N. Ienaga, and H. Saito, "Diminished hand: A diminished reality-based work area visualization," in *Proc. IEEE Virtual Real.*, 2017, pp. 443–444.
[48] S. Siltanen, "Diminished reality for augmented reality interior design," *Vis. Comput.*, vol. 33, no. 2, pp. 193–208, 2017.
[49] S. Mori, M. Maezawa, N. Ienaga, and H. Saito, "Detour light field rendering for diminished reality using unstructured multiple views," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2016, pp. 292–293.
[50] S. Mori, J. Qie, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, "[POSTER] background image registration as a post-processing technique in diminished reality rendering procedures," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2017, pp. 200–201.
[51] C. Kunert, T. Schwandt, and W. Broll, "An efficient diminished reality approach using real-time surface reconstruction," in *Proc. Int. Conf. Cyberworlds*, 2019, pp. 9–16.
[52] E. Andre and H. Hlavacs, "Diminished reality based on 3D-scanning," in *Entertainment Computing and Serious Games*. Arequipa, Peru, Springer, 2019, pp. 3–14.
[53] S. W. T. Chan, B. Ryskeldiev, and S. Nanayakkara, "DeclutterAR: Mobile diminished reality and augmented reality to location hoarding by motivating Decluttering and selling on online marketplace," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjunct*, 2022, pp. 870–874.
[54] S. Sakai, Y. Yanase, Y. Matayoshi, and M. Inami, "D-ball: Virtualized sports in diminished reality," in *Proc. 1st Superhuman Sports Design Chall. 1st Int. Symp. Amplif. Capabilities Compet. Mix. Real.*, 2018, pp. 1–6.
[55] C. Rolim and V. Teichrieb, "A viewpoint about diminished reality: Is it possible remove objects in real time from scenes?" in *Proc. 14th Symp. Virtual Augment. Real.*, 2012, pp. 141–146.
[56] N. Kawai, T. Sato, and N. Yokoya, "From image inpainting to diminished reality," in *Proc. 6th Int. Conf. Virtual, Augment. Mix. Real. Design. Develop. Virtual Augment. Environ.*, 2014, pp. 363–374.
[57] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, pp. 2378–2386, 2011.
[58] K. Takeda and R. Sakamoto, "Diminished reality for landscape video sequences with homographies," in *Proc. 14th Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2010, pp. 501–508.
[59] T. Fukuda, K. Inoue, and N. Yabuki, "PhotoAR + DR 2016—Integrating automatic estimation of green view index and augmented and diminished reality for architectural design simulation," in *Proc. 35th Int. Conf. Educ. Res. Comput. Aided Archit. Design Europe*, 2017, pp. 495–502, doi: 10.52842/conf.ecaade.2017.2.495.
[60] R. Liu, M. Gao, S. Ye, and J. Zhang, "IGScript: An interaction grammar for scientific data presentation," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2021, pp. 1–13.
[61] P. Lindemann and G. Rigoll, "A diminished reality simulation for driver-car interaction with transparent cockpits," in *Proc. IEEE Virtual Real.*, 2017, pp. 305–306.
[62] K. Richardson, A. C. McLaughlin, M. McDonald, and A. Crowson, "The effects of diminished reality on the detection of and response to notifications," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 65, no. 1, pp. 159–163, 2021.
[63] A. V. Taylor, A. Matsumoto, E. J. Carter, A. Plopski, and H. Admoni, "Diminished reality for close quarters robotic telemanipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 11531–11538.
[64] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, "NASA TLX: software for assessing subjective mental workload," *Behav. Res. Methods*, vol. 41, pp. 113–117, Mar. 2009.
[65] W. S. Helton, K. M. Jackson, K. Näswall, and B. Humphrey, "The national aviation and space agency task load index (NASA-TLX): Does it need updating?" *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 66, no. 1, 2022, pp. 1245–1249.
[66] A. Plopski, A. V. Taylor, E. J. Carter, and H. Admoni, "InvisibleRobot: Facilitating robot manipulation through diminished reality," in *Proc. IEEE Int. Symp. Mixed Augment. Real. Adjunct*, 2019, pp. 165–166.
[67] Y. Zhang, X. Hu, K. Kiyokawa, and X. Yang, "Add-on occlusion: Turning off-the-shelf optical see-through head-mounted displays occlusion-capable," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 5, pp. 2700–2709, May 2023.
[68] J. Guida and M. Sra, "Augmented reality world editor," in *Proc. ACM Symp. Virtual Real. Softw. Technol.*, 2020, pp. 1–2.
[69] S. Hashiguchi, S. Mori, M. Tanaka, F. Shibata, and A. Kimura, "Perceived weight of a rod under augmented and diminished reality visual effects," in *Proc. ACM Symp. Virtual Real. Softw. Technol.*, 2018, pp. 1–6.
[70] R. A. Croker, "An introduction to qualitative research," in *Qualitative Research in Applied Linguistics: A Practical Introduction*, J. Heigham and R. A. Croker, Eds., London, U.K.: Palgrave Macmillan, 2009, pp. 3–24.
[71] H. K. Mohajan, "Quantitative research: A successful investigation in natural and social sciences," *J. Econ. Develop., Environ. People*, vol. 9, no. 4, pp. 50–79, 2020.
[72] R. Liu, S. Ye, Z. Ding, G. Yang, S. Cheng, and K. Mueller, "DRCmpVis: Visual comparison of physical targets in mobile diminished and mixed reality," *IEEE Trans. Vis. Comput. Graph.*, early access, Jan. 25, 2024, doi: 10.1109/TVCG.2024.3358419.
[73] N. Kawai, M. Yamasaki, T. Sato, and N. Yokoya, "Diminished reality for AR marker hiding based on image inpainting with reflection of luminance changes," *ITE Trans. Media Technol. Appl.*, vol. 1, no. 4, pp. 343–353, 2013.
[74] Y. Li, B.-K. Seo, and K. Kim, "Exploring industrial uses of virtually altering the physical world," in *Proc. IEEE Conf. Virtual Real. 3D User Interfaces Abstr. Workshops*, 2023, pp. 434–437.
[75] K. Yagi, K. Hasegawa, and H. Saito, "Diminished reality for privacy protection by hiding pedestrians in motion image sequences using structure from motion," in *Proc. IEEE Int. Symp. Mixed Augment. Real.*, 2017, pp. 334–337.

Siru Chen received the bachelor's degree from Nanjing Normal University, Nanjing, China, in 2022, where she is currently pursuing the master's degree in computer science and technology.

Her current research focuses on extended reality (VR/AR/MR), human–computer interaction, and immersive computing and analytics.

Lingxin Yu is currently pursuing the master's degree in computer science and technology with Nanjing Normal University, Nanjing, China.

Her current research interests include extended reality (VR/AR/MR) + visualization, immersive computing and analytics, and HCI.

Yuxuan Liu is currently pursuing the bachelor's degree in artificial intelligence with Nanjing Normal University, Nanjing, China.

Her current major research interest is immersive computing and analytics.

Zhifei Ding is currently pursuing the master's degree with Nanjing Normal University, Nanjing, China.

He is the first author of the paper accepted in the IEEE TRANSACTIONS ON BIG DATA. His current research interests include big data visualization and immersive visualization.

Jiacheng Zhang is currently pursuing the master's degree in computer science and technology with Nanjing Normal University, Nanjing, China.

His current research focuses on extended reality (VR/AR/MR) + visualization, and immersive computing and analytics.

Xinyue Wang is currently pursuing the master's degree with Nanjing Normal University, Nanjing, China.

Her current research focuses on immersive computing and analytics, and 3-D reconstruction from a single image.

Jiahao Han is currently pursuing the master's degree with Nanjing Normal University, Nanjing, China.

His current research interests include extended reality (VR/AR/MR) + visualization and immersive computing and analytics.

Richen Liu received the Ph.D. degree from Peking University, Beijing, China.

He is an Associate Professor with Nanjing Normal University, Nanjing, China. He has published more than 40 conference/journal papers, including the conferences, such as ACM CHI, ACM MM, IEEE VIS, EuroVis, and PacificVis, and the journals, such as IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, IEEE TRANSACTIONS ON HUMAN–MACHINE SYSTEMS, IEEE TRANSACTIONS ON BIG DATA, *Bioinformatics*, and *SPE Journal*. His current research interests include XR + visualization, AI + metaverse, and HCI.

Dr. Liu served as multiple committees and reviewers.