Observation-based diminished reality: a systematic literature review

Roghieh Eskandari[1] · Ali Motamedi[1]

Abstract

Diminished reality (DR) is a set of techniques for visually removing unwanted objects within an environment in real time. This study attempts to carry out a systematic literature review to investigate the potential benefits, challenges, and efficacy of DR in various applications. To investigate the relevant studies, a classification framework with six aspects, namely, paper type, DR type, processing workflow, background data type, display device type, and DR environment, is provided. The relevant papers were mainly sourced from the Scopus academic database. From an initial number of 1284 papers published from 2000 to 2024, 67 were selected as key articles for analysis. Based on the findings, this study offers recommendations for implementing observation-based DR-supported functionality for AR applications going forward.

Keywords Diminished reality · Augmented reality · Mixed reality · See-through vision · Systematic literature review

1 Introduction

The evolution of digital environments has been significantly shaped by the development of technologies that alter our perception of reality. The concept of the Reality-Virtuality Continuum, first articulated by Milgram et al. (1995), serves as a crucial framework for understanding the range of experiences created by these technologies. The continuum is often visualised as a one-dimensional axis (X-axis), with the real world on one end and the virtual world on the other. Virtual Reality (VR), which is situated at the virtual end of the spectrum, immerses users in entirely computer-generated environments, while Augmented Reality (AR) overlays digital information onto the real world, enriching the user's perception. Lying between these two extremes is Mixed Reality (MR), where real and virtual elements interact seamlessly. Mann et al. (2018) introduced the concept of Mediality, which represents the degree to which reality is modified. This adds a second axis (Y-axis) to the Reality-Virtuality Continuum, creating what the authors call the

Mediated Reality (X, Y) Continuum. Adding the Mediality axis, enables Mediated Reality to expand the original one-dimensional continuum into a two-dimensional space, allowing for a more nuanced understanding of how reality can be altered by technology.

Diminished Reality (DR), a subset of Mediated Reality, involves the selective removal of elements from a user's perception of the real world, thereby modifying their experience by eliminating rather than augmenting information (Mann 1999).

Since its introduction, DR has been conceptualised and defined in various ways in the literature, as evidenced by its multifaceted applications and theoretical underpinnings. Mori et al. (2017) further refined the concept of DR, describing it as a set of methodologies for visually diminishing, replacing, inpainting, and seeing through objects in a perceived environment in real time. Using these methodologies, DR can be categorised into two primary types: inpainting-based DR and observation-based DR (Ob-DR). Inpainting-based DR techniques recover the hidden background by estimating pixels according to sounding pixels, while Ob-DR methods rely on direct observations to replace visual information.

This Systematic Literature Review (SLR) aims to provide a comprehensive overview of the current state and future challenges of Ob-DR. The focus on Ob-DR herein is due to its unique approach, which involves substituting real-world elements with observations that originate directly from the

✉ Ali Motamedi
　ali.motamedi@etsmtl.ca

　Roghieh Eskandari
　roghieh.eskandari-toushmanlou.1@ens.etsmtl.ca

1　Department of construction, École de technologie supérieure (ÉTS), Montreal, Canada

existing reality, thereby preserving the integrity of the real-time experience. Although the review by Mori et al. (2017) provided valuable insights into the implementation of DR technology, our study seeks to delve deeper into the theoretical foundations of OB-DR, explore its extensive range of potential applications, and discuss its latest developments and case studies.

2 Research methodology

The review approach adopted in the present study is based on the SLR method, a well-known strategy used to find, assess, and understand related research for a particular subject (Kitchenham 2004). According to the method outlined by Kitchenham (2004), an SLR protocol should include a clearly defined background outlining the study rationale, specific research questions, a comprehensive search strategy detailing search terms and databases, as well as study selection criteria. Additionally, it involves developing quality assessment checklists and procedures, defining a data extraction strategy, and outlining a synthesis strategy for the extracted data. Accordingly, in conducting the SLR, this study used a three-step procedure comprised of (1) the planning step, (2) the conducting step, and (3) the reporting step.

Table 1 Research questions, exclusion criteria, and filtering criteria for this SLR study

|	
Research questions	RQ1: What types of articles are found in the DR literature? RQ2: Which evaluation methods do researchers use to assess the usability of DR? RQ3: What DR types do researchers introduce in the literature? RQ4: Which data sources do researchers use to generate the background scene in DR? RQ5: What display devices do researchers use to display DR results? RQ6: What development pipelines do researchers follow in the DR literature? RQ7: What DR environments do researchers use to implement DR use cases?
Exclusion criteria	EC1: Studies published prior to 2000 are excluded. EC2: Only papers published in scientific journals or conference proceedings. EC3: Papers not written in English are excluded.
Filtering criteria	FC1: Papers that use inpainting techniques to remove objects from a scene rather than observation-based DR techniques for recovering hidden backgrounds are excluded. FC2: Papers that use DR techniques in offline data, such as video frames are excluded, as this review focuses on real-time applications. FC3: Studies that employ non-visual sensors, such as ground-penetrating radars or seismic sensors (e.g., Wei et al. 2019), to see through objects are excluded.

2.1 Planning step

This step focuses on the review development protocol. Table 1 presents the scope of interest in this study, consisting of the Research Questions (RQs), Exclusion Criteria (EC), and Filtering Criteria (FC). The Scopus academic database was selected for the planning step. Data was extracted from this database and analysed using Microsoft Excel.

2.2 Conducting step

A lot of thought was put into the selection of keywords for this study to ensure a comprehensive exploration of the concept of DR and its associated techniques. DR is not only defined by the process of visually diminishing or removing objects from a user's perception of the real world, but is also closely linked with techniques such as see-through vision, AR X-ray vision, and ghosted view. These techniques provide a semi-transparent representation of the scene, allowing users to perceive both the virtual and real worlds simultaneously (Mori et al. 2017). Therefore, keywords such as See-through vision, AR X-ray vision, AR X-ray system, seeing-through, and Ghosted view were included in the search strategy to ensure all relevant literature was captured. These terms were used with the OR operator to extract articles that included them in their abstracts, titles, and keywords. We conducted two separate search runs, one in November 2022, and the other in April 2023, obtaining 1090 papers. The collected literature underwent three assessment phases, namely, screening, eligibility, and inclusion, as outlined in Fig. 1. This approach enabled us to gather a diverse range of articles and ensure a thorough examination of the relevant literature on DR.

During the screening phase, we first removed duplicate papers and then applied exclusion criteria (EC1:EC3), which considered the publication year, the paper type, and the language. We identified approximately 300 papers as non-relevant and subsequently discarded them, leaving a total of 712 papers for eligibility assessment.

During the eligibility phase, the set of filtering criteria (FC1:FC3) was applied to exclude papers that were not relevant. After the papers were carefully reviewed and the above filters applied, a total of 385 papers were eliminated. To ensure that all key articles were covered in full, backward and forward snowballing techniques were utilised; these involved examining the citations and reference lists (as depicted in Fig. 1). Finally, 67 papers remained for a detailed analysis in the inclusion phase.
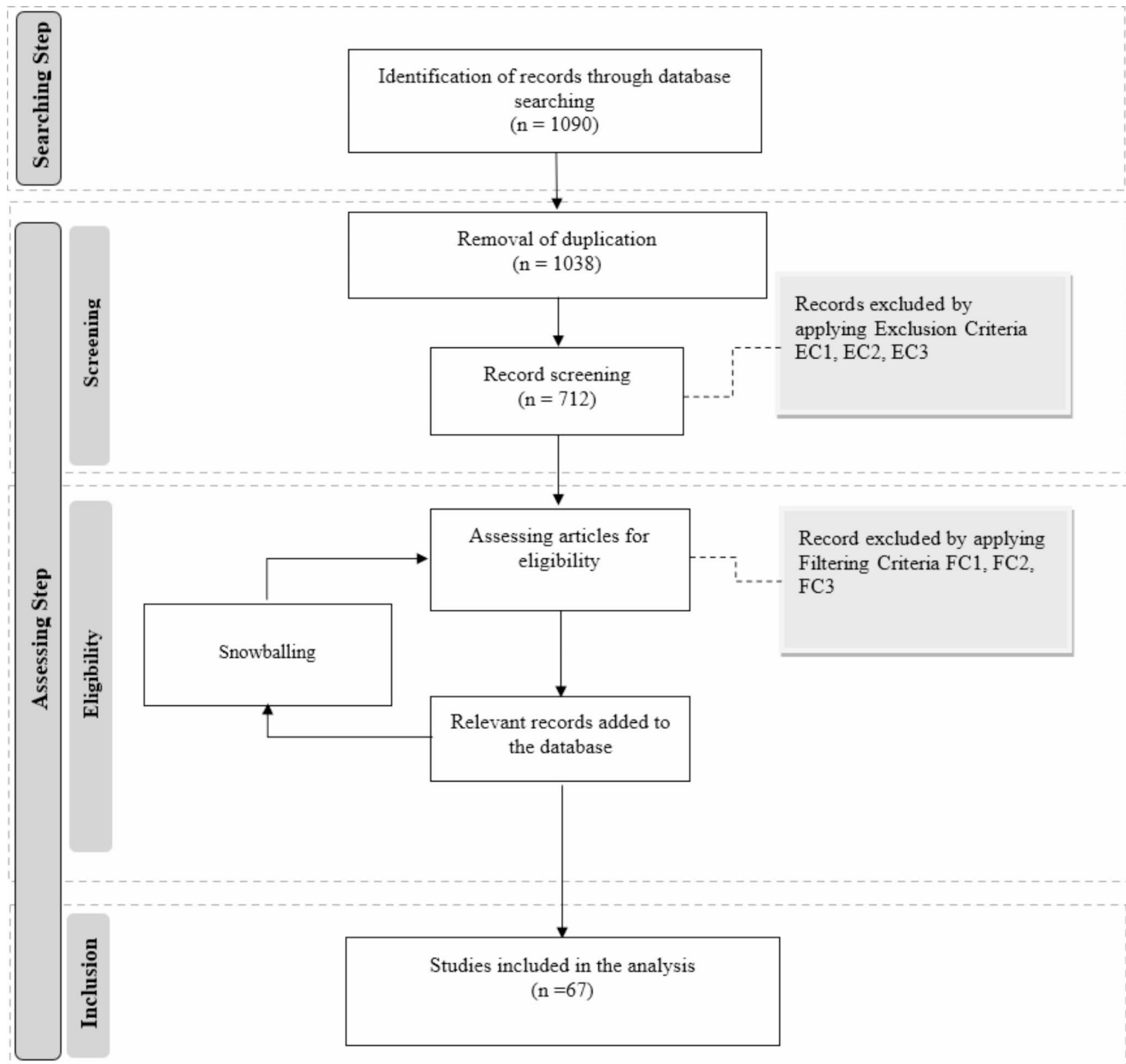
Fig. 1 Search results following review protocol (n = number of papers)

2.2.1 Data extraction and synthesis

A classification framework was stablished based on research questions prior to the reporting step. Following an in-depth analysis of the content of the collected papers, six dimensions were extracted, namely, paper type, diminished reality type, background data type, display device type, processing workflow, and DR environment (as shown in Fig. 2). This framework was chosen because it covers the relevant aspects of DR and enables a standardised comparison of the techniques in the literature. Each dimension and its options will be discussed in detail in the following sections.

2.2.1.1 Paper type Papers can be divided into three main types: (1) Technical Method Development: papers that explore new technical methods to implement DR, (2) Application Development: papers that concentrate on developing applications using DR techniques, and (3) Evaluation: papers that assess the quality and effectiveness of previously developed DR methods.

2.2.1.2 Diminished reality type Depending on the background recovery techniques employed, DR techniques are categorised into two main groups: image inpainting-based
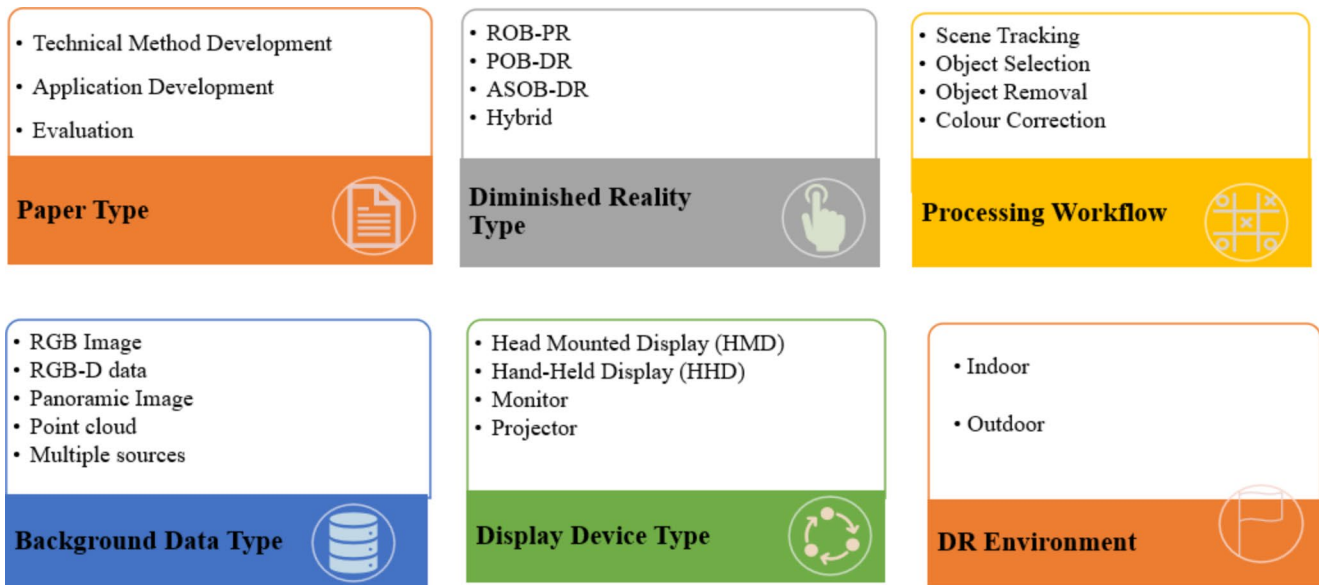
Fig. 2 Classification taxonomy

diminished reality (IB-DR) and observation-based diminished reality (OB-DR) groups.

Inpainting-based diminished reality (IB-DR)

This technique involves filling the Region of Interest (ROI) with plausible visual information instead of showing the real scene from the hidden background. Therefore, this category does not need an additional camera to observe the background from different views or pre-recorded observations, such as 3D scanning of the scene (Mori et al. 2016). Early IB-DR techniques focus on estimating the visual information from pixels or image patches surrounding the ROI within the same image (e.g., Kawai et al. (2015) and Gkitsas et al. (2021). However, given the limitations of these techniques in achieving reliable results in large ROIs or complex scenes, IB-DR techniques based on data-driven approaches that learn from a large database of images have recently been favoured for study (e.g., Pintore et al. 2022; Kikuchi et al. 2022; Kari et al. 2021).

Observation-based diminished reality (OB-DR)

In this category, the hidden view is recovered from background observation results. Therefore, OB-DR methods provide more convincing outputs than do IB-DR methods. According to Mori et al. (2017), OB-DR can be classified into four sub-categories: Pre-Observation-Based Diminished Reality (POB-DR), Real-time Observation-Based Diminished Reality (ROB-DR), Active Self-Observation-Based Diminished Reality (ASOB-DR), and the combination of inpainting-based methods with OB-DR. Each of these sub-categories refers to how the background data is captured or observed, which directly influences the effectiveness of the DR technique.

In the POB-DR category, the background observation occurs beforehand during the pre-processing step, either in the absence of the objects to be removed or from different viewpoints of the hidden background. On the other hand, ROB-DR methods utilise real-time background data collected using additional sensors. In the ASOB-DR category, the background is observed with a time difference by moving either the object or the camera. Similar to ROB-DR, this category lacks the background information dataset. The background information is retrieved from preceding frames of a video sequence, where the object to be diminished was not yet present in the scene.

2.2.1.3 Background data type The background information must be recovered to eliminate an object from a perceived scene. The background data type refers to the data source used for hidden background generation. Four background data sources are observed in the DR literature: RGB image, depth data, panoramic image, and point cloud. Some studies use a combination of these types to generate the background.

2.2.1.4 Display device type The display device serves as the hardware interface for users to interact with DR applications, and comes in various forms, such as (1) a Head Mounted Display (HMD), which is a device mounted on the user's head or placed on a helmet; (2) a Hand-Held Display (HHD), which is any portable device, such as a smartphone of a tablet computer; (3) a Monitor, which is a static computer screen; and (4) a Projector, which for its part is a

device that allows to project visual annotations onto real-world items.

2.2.1.5 Processing workflow An effective implementation of DR techniques requires the adoption of an appropriate and efficient technical approach. A review of the literature on observation-based DR reveals that most implemented methods integrate four main steps: scene tracking, object selection, object removal, and colour correction.

The first, the scene tracking step, involves estimating the position of physical objects relative to the cameras, facilitating 3D scene recovery. Subsequently, the object selection step identifies the object or ROI to be eliminated within the scene. Following this, the object removal step focuses on recovering the background information of the selected object from the user's viewpoint. Finally, colour correction is conducted to address any colour differences that remain at the boundaries of the recovered background, ensuring visual coherence and seamlessness in the final output.

2.2.1.6 DR environment The DR environment refers to the specific location where the DR case studies were implemented, including indoor and outdoor settings.

3 Results

Collected and organised literature, using the classification framework mentioned in Sect. 2.2.1, is available in Additional file 1. This supplementary file contains a list of all the studies reviewed in this paper. This data is used in the reporting step for various types of analysis.

3.1 Preliminary analyses

The distribution of DR publications at conferences and in journal proceedings provides insights into the current state of the field and its maturity level. The DR papers selected for the present work have been published in various journals and conference proceedings. As shown in Fig. 3a, about 75% of the publications (50 papers) were published in conference proceedings, and about 25% were published in journal proceedings (17 papers).

Figure 3b shows the distribution of DR publications by venue. The International Symposium on Augmented Reality (ISMAR) is the most popular conference in the domain
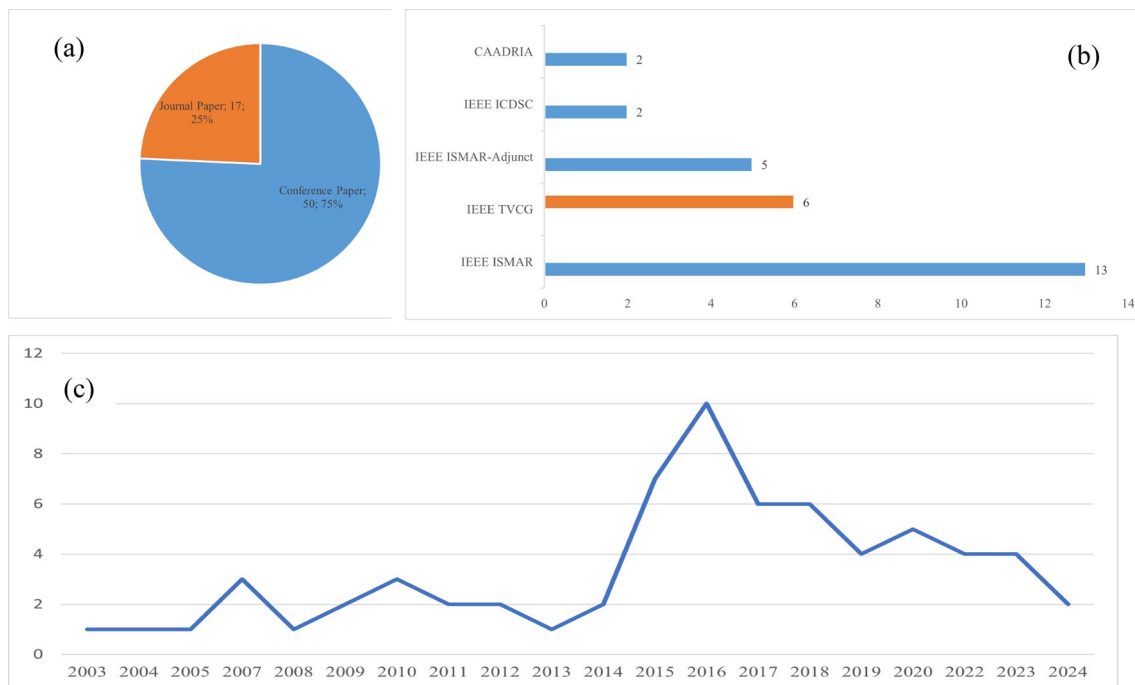
Fig. 3 a Distribution of DR publications in journals and conference proceedings, **b** distribution of publications by venue (venues with less than two publications have been removed from this chart: ICDSC, international conference on distributed smart cameras; ISMAR, inter-national symposium on mixed and augmented reality; TVCG, transactions on visualization and computer graphics)), **c** distribution of studies according to publication year

of diminished reality (13 papers). Regarding the journal papers, IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) is the top journal, with six publications. IEEE is the most-cited publisher in this domain (38 papers).

Journal papers offer advantages over conference papers due to their rigorous peer review process and higher credibility. That notwithstanding, most DR research is currently published in conference proceedings. Encouraging researchers to prioritise journal publications could enhance the impact and credibility of DR research, shaping future directions in the field.

Figure 3c illustrates the distribution of DR papers by publication year, revealing fluctuations over time. Publications were at a peak in 2016 (11 papers), suggesting notable research activity within the field at the time, possibly influenced by events such as the International Workshop on Diminished Reality (IWDR 2016). Fluctuations in publication frequency may reflect evolving trends, technological advancements, or shifts in research focus within the DR domain.

3.2 Distribution analysis

In this section, we explore the distribution patterns of the introduced factors across selected papers.

3.2.1 Paper types

3.2.1.1 Technical method development papers As illustrated in Fig. 4, the technical method development paper type is dominant, with 33 studies (50% of papers). This prevalence may be attributed to the novelty of DR visualisation, which prompts researchers to explore new techniques for its implementation. Furthermore, given the advancements in supporting technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), 5G communication, and sensors, the potential applications of AR in various industries seem to be promising (Siriwardhana et al. 2021). These advancements can also contribute significantly to progress in DR technology, enabling more sophisticated and seamless experiences. Consequently, we anticipate a rise in the number of publications focusing on the technical method development in this area.

3.2.1.2 Application development papers As shown in Fig. 4 and 27 studies focus on application development. DR is employed in various domains, such as Architecture/Engineering/ Construction-Facility Management (AEC-FM), the automobile industry, medicine, privacy protection, robotics, visuo-haptic systems, drone control, workplace productivity, and sports. As can be seen, the AEC-FM and automo-
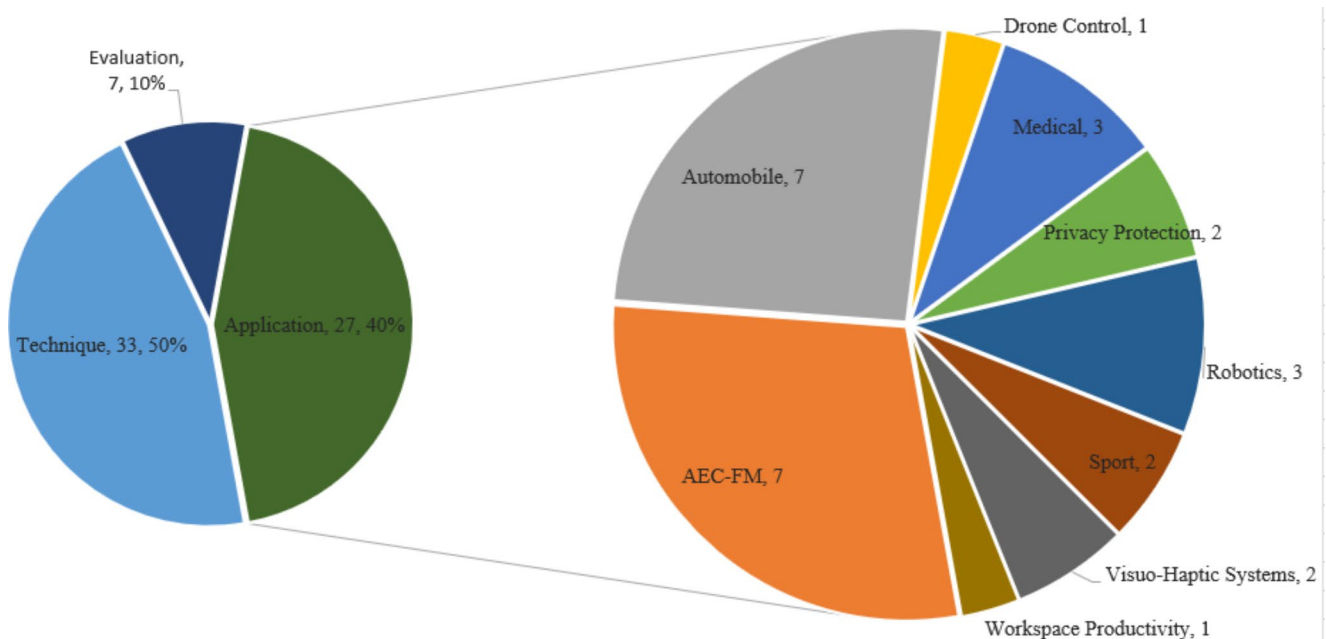
Fig. 4 Distribution of application paper types

bile industries are the most researched application domains, each with a total of 7 occurrences.

In the AEC-FM Industry, DR can improve communications among participants during the design process. In this case, the renovation plan can be changed by designers, and the occupants can observe the DR results. This process allows occupants to fully grasp the designer's ideas and clearly understand the renovation outcomes. For example, DR was employed to display interior renovation plans by Zhu et al. (2019). They subsequently used a collaborative design system that allows multiple people to simultaneously participate in the same environment during the design phase (Zhu et al. 2020).

Another category of common use cases of DR is outdoor landscape simulation. While AR can be employed to assess a future landscape by superimposing a 3D design model onto the actual buildings, DR can be used to visually eliminate buildings. For example, Kido et al. (2020) introduced a DR system to create a realistic simulation of the environment during redevelopment projects by visually eliminating objects in real time. Inoue et al. (2018) developed an AR/DR system in which the green view index was measured simultaneously with the DR simulation in an urban design application.

In the automobile industry, significant technological advances, such as the Advanced Driving Assistance System (ADAS), have facilitated the use of DR. To provide a better visualisation of road signs, DR techniques can be applied to eliminate obstacles, such as buildings and other cars, from the driver's field of view (Lindemann and Rigoll 2017). For example, Rameau et al. (2016) used a stereo-vision system that enables the driver to see through cars driving in front, while Rameau et al. (2016) used a Robot Operating System (ROS)-based DR module to see through the different objects on the road.

The application of DR technology in the automobile industry also has the potential to enhance passenger comfort and safety during self-driving experiences. For example, Sasai et al. (2015) developed an MR system to visualise the road surface by overlaying the wheel trajectories on the surface of the car dashboard. The results showed that the proposed MR system tended to reduce anxiety in passengers of autonomous vehicles in some situations.
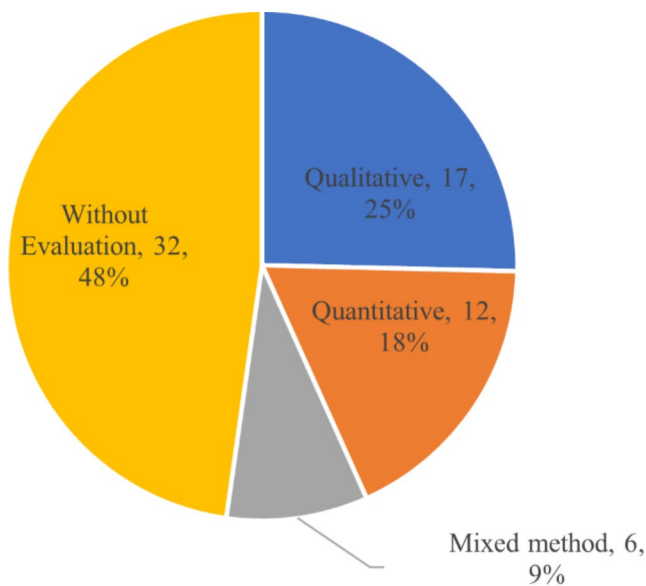
Furthermore, DR has been employed in a variety of other applications. For example, Hashiguchi et al. (2018) presented a system that can modify the visual representation of real objects using AR and DR renderings. The aim was to ascertain whether users perceive objects as heavier or lighter than their actual weight, similar to the Size Weight Illusion (SWI) effect. In a reduction scenario, the user can perceive an object as lighter than it is as part of the object is visually removed using a DR technique. In human-robot collaboration scenarios, DR can facilitate communication and collaboration between humans and robots by providing a mediated view of the environment. DR can help human operators better understand the robot's actions and intentions by selectively hiding or highlighting certain workspace elements (Weidner et al. 2023) or example, DR enhances the user's visibility in controlling the robotic arm in telemanipulation tasks by generating see-through images (Kittaka et al. 2016; Taylor et al. 2020). In industrial environments, DR can improve productivity by providing workers with augmented views of their surroundings, highlighting important information and removing distractions. Hasegawa and Saito (2015) explored the potential of DR for privacy protection. They presented a DR system to protect pedestrians' privacy by hiding them in video frames. Erat et al. (2018) explored DR for drone navigation in narrow or constrained environments. Yokoro et al. (2023) proposed a DR system to minimise visual distractions that can disrupt concentration in a workspace.

3.2.1.3 Evaluation papers As shown in Fig. 4, the evaluation paper type is less dominant in the DR literature (seven papers). In this area, Morozumi et al. (2017) attempted to generate a standard dataset to assess the DR approach in a static environment, using miniature sets. They focused on observation-based DR methods and evaluated their performance using ground truth data. Similarly, Peereboom et al. (2023) utilised simulation-based approaches to evaluate the effectiveness of DR/AR applications in pedestrian safety. In another study by Jütte, Poschke, and Overmeyer (2023) a simulation-based approach was used to evaluate DR performance. They employed synthetic data generated in Unity to replicate a real see-through application environment. These initiatives contribute to the advancement of the field by providing researchers with standardised methods for evaluating and comparing different DR approaches, ultimately leading to the development of more robust and user-friendly DR solutions.

In addition to evaluation-focused papers, some studies assess their DR approaches within application development or technical development papers. Figure 5 illustrates the distribution of all of the papers by evaluation method. As shown, 25% of the studies consulted tried to evaluate their presented methods using qualitative methods (17 papers) and 18% of publications employed quantitative evaluation methods (12 papers). However, 48% of the studies did not present any evaluation process, and 9% used mixed methods, i.e., combined quantitative and qualitative methods.

Quantitative evaluation methods employ ground truth data to assess the quality of DR results using similarity measures, such as Learned Perceptual Image Patch Similarity (LPIPS), Mean Absolute Error (MAE), Peak Signal-to-Noise
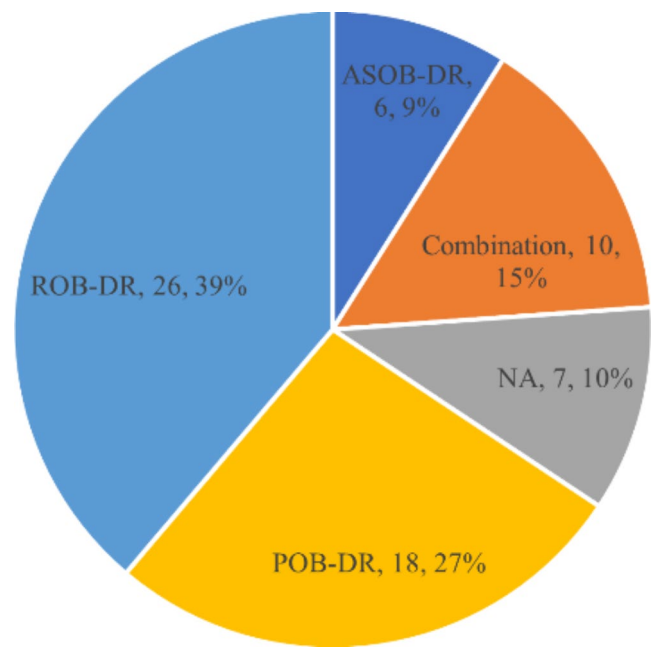
Fig. 5 Distribution of publications according to the evaluation method

Fig. 6 Distribution of DR types. NA (not available) indicates studies that did not mention the DR type

Ratio (PSNR), Normalized Cross Correlation (NCC), and Structural Similarity Index Measure (SSIM). For example, Morozumi et al. (2017) utilised a miniature set in a simulated environment to quantitatively evaluate observation-based DR methods using ground truth data. Namboku and Takahashi (2020) performed an accuracy assessment by comparing their methods with ground truth data manually generated from scene images. Gomes et al. (2012) used the PSNR similarity measure as a simple analytical approach for assessing the quality of results.

Obtaining ground truth data can be challenging, especially in outdoor environments due to lighting changes, which may affect the accuracy of the quantitative evaluation. Thus, as also recommended in Morozumi et al. (2017), most practical solutions can be evaluated by creating the dataset using a miniature set under simulated illumination conditions in indoor environments.

Regarding qualitative evaluation methods, Image Quality Assessment (IQA) techniques can be employed to evaluate the performance of DR methods without the need for ground truth data. For instance, Avery et al. (2008) validated a mobile-based see-through vision system in an outdoor environment using hypothesis testing conducted for participants. Yue et al. (2017b) utilised the Simulator Sickness Questionnaire (SSQ) (Kennedy et al. 1993) to evaluate the comfort level of participants wearing MR headsets with the DR function enabled. Interviews are also a common approach for qualitatively evaluating participants' overall performance in DR studies. For example, semi-structured interviews were utilised in a study conducted by Erat et al. (2018) to gather qualitative insights into participant experiences and perceptions.

In addition, some studies have used a mixed methods approach. For instance, Lee and Kim (2024) employed a combination of quantitative cognitive performance measures and qualitative post-study interviews to comprehensively evaluate the effectiveness of DR interventions in enhancing cognitive environments.

### 3.2.2 Diminished reality types

As shown in Fig. 6 and 26 papers focused on ROB-DR (39% of publications), followed by POB-DR (18 papers, 27%), combination type (10 papers, 15%), and ASOB-DR (six papers, 9%).

3.2.3 Background data type

As shown in Fig. 7, RGB images represent the most dominant data source for background generation in DR studies (58%, 39 papers). RGB images can be collected using surveillance cameras (e.g., Kameda et al. 2004; Mei et al. 2011), hand-held RGB cameras (e.g., Kido et al. 2020), or RGB cameras mounted on a drone (e.g., Erat et al. 2018). For example, Jarusirisawad (2007) used multiple RGB cameras to synthesise a free-viewpoint image without undesired objects. Mori et al. (2015) recovered hidden backgrounds using multi-view perspective RGB images that were captured in advance. Li et al. (2013) used an Internet-based photo collection for the purpose of removing people in a video sequence. RGB images are sometimes used to generate a 3D model of the background. For example, Kido et al. (2020) presented a

Fig. 7 Distribution of the background data type. NA (not available) indicates evaluation studies that did not mention the background data type
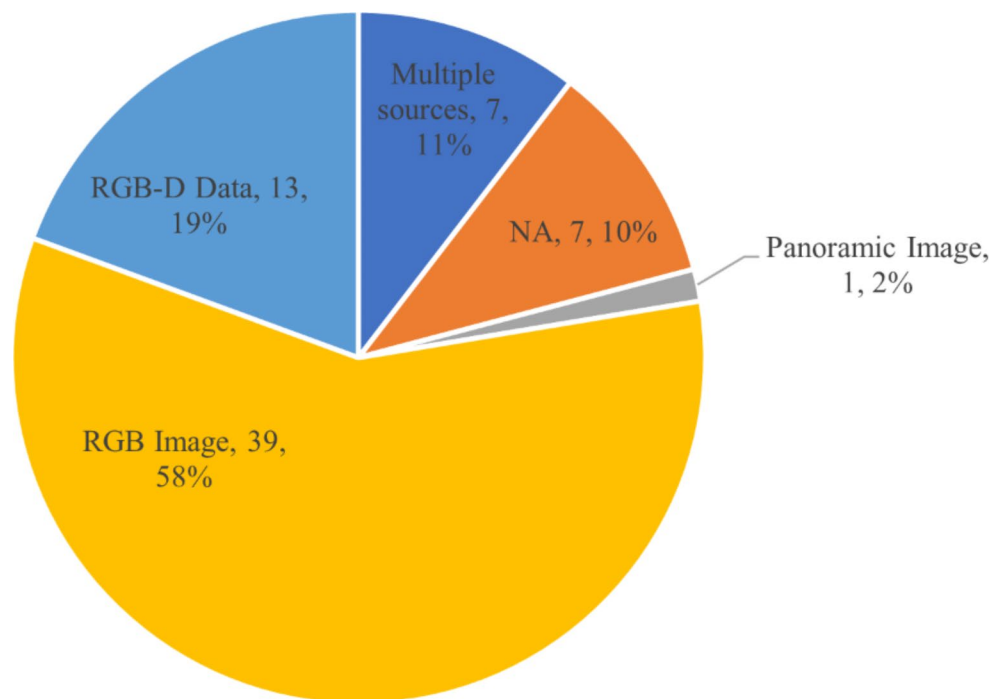
method to virtually diminish existing landscape objects using 3D models generated by SfM modelling software.

RGB-D data is targeted by 19% of the studies (13 papers). For example, Andre and Hlavacs (2019) used g depth data to recreate the structure of objects and colour data to reconstruct their appearance. In another study conducted by Qiaozhi (2016), a Kinect 3D sensor is employed to create a 3D model of the scene. This model is captured prior to adding any objects to the scene, and then a DR technique is used to remove unwanted objects. Meerits and Saito (2015) proposed a framework for generating mesh using an RGB-D camera. The 3D reconstruction results in their approach contain many missing pieces that appeared as black holes due to the limitations of using one RGB-D camera. As indicated by Habert et al. (2017), a dual RGB-D camera system can circumvent this problem.

Panoramic images obtained from 360-degree cameras were used in 1% of the surveyed DR studies. Using panoramic data for background scene reconstruction has the advantage of providing a wide field of view that instantly scans the entire background scene.

13% of DR studies utilised multiple sources, i.e., a combination of various data sources. For example, Zhu et al. (2020) used a combination of point clouds and BIM data for 3D background reconstruction to overcome the separate limitations of each method. Point clouds are useful in creating a large-scale mesh of the environment; however, they may not be used to reconstruct complex objects in detail. On the other hand, BIM data provides a detailed 3D model of complex physical objects; however, creating them is time-consuming

and resource-intensive. Therefore, combining point clouds and BIM data can provide a more comprehensive and accurate 3D model of the background environment in DR.

3.2.4 Display device type

As shown in Fig. 8, monitors are the dominant interaction device for displaying DR results (48%, 32 papers). HHDs, consisting of mobile phones and tablet PCs, are the second most used devices (22%, 15 papers).

The selection of display device is closely linked to the required processing power and the nature of the application. Conducting research on a PC is more advantageous than using HMDs and HHD thanks to the superior processing capabilities of the former in handling complex computations and storing large datasets. For instance, the method presented by Lin and Popescu (2022) demonstrated a fast performance on a workstation using only the system's Central Processing Unit (CPU), allowing it to keep up with video feeds. In addition, DR technologies are often still at the experimental or prototyping stage, and are undergoing iterative development and refinement. During this stage, researchers usually utilise PCs due to their flexibility and ease of use. HHDs, despite their portability and convenience, often lack high computational power. However, recent advancements in mobile technology have led to notable improvements in the computational capabilities of tablets and phones.

18% of studies employed HMD devices to display DR results. Video See-Through Head-Mounted Displays (VST-HMDs) use video cameras installed on the headset
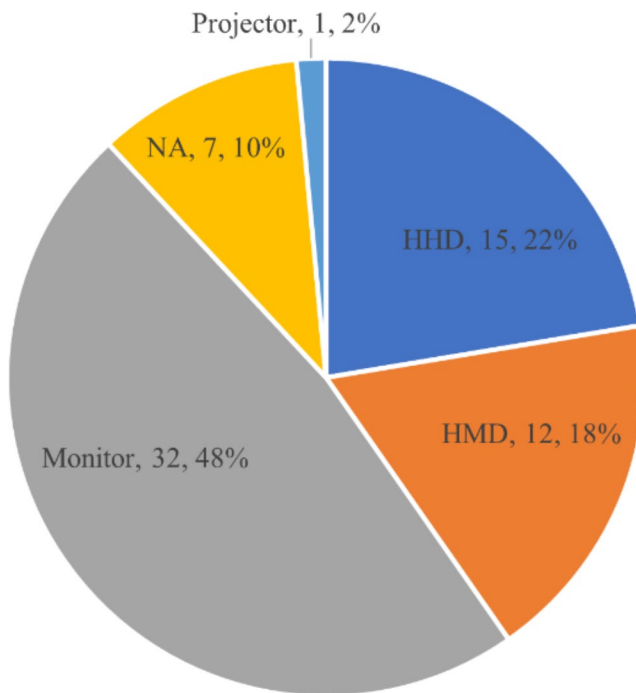
Fig. 8 Distribution of display device type. NA (not available) indicates evaluation studies that did not mention the display device type

to capture the real-world environment and display it on the screen in front of the user's eyes. Subsequently, virtual content is superimposed onto the real-world view captured by the cameras. This process thus creates an opaque display, where the user cannot view the real world directly, but only through the video display. For example, Chan et al. (2022) employed a VST-HMD to display a live video feed of the environment captured by its front cameras and superimposed object models onto it in real time. On the other hand, Optical See-Through Head-Mounted Displays (OST-HMDs), such as Google Glass and Microsoft HoloLens, provide users with a semi-transparent representation of the real world provided by mirrors installed on the headset. The user sees the virtual content and the real-world environment simultaneously using OST-HMDs. The semi-transparent nature of OST-HMD devices has shown efficiency in DR applications, where the occlusion of a target object from the user's view is a common challenge. One example is the work by Taylor et al. (2020), in which an OST-HMD is utilised to observe the occluded areas behind the robot. Nevertheless, OST-HMDs may present holographic content with low contrast, potentially resulting in non-photorealistic DR outcomes.

One of the challenges in utilising HMDs is managing latency effectively. According to Overmeyer et al. (2023), the measured latencies shed light on various stages of the system's operation, including sensor data processing, scene rendering, and display on the HMD. The observed average

latency for displaying frames on the HMD underscores the consistency of the system's performance in delivering visual feedback to the user. However, the increase on latency with the addition of cameras highlights potential scalability challenges, particularly in maintaining real-time responsiveness under heavy computational load. Moreover, the discrepancy in processing time between frame production and display on the HMD emphasises the importance of optimising rendering pipelines to minimise latency and ensure the timely delivery of critical information to the user.

Using an HMD device imposes higher demands on the fidelity and precision of the rendered visual information than is the case with fixed displays (Overmeyer et al. 2023. This heightened requirement for quality stems from the integration of virtual content with the user's real-world environment, necessitating seamless alignment and minimal discrepancy between the two. Spatial and temporal errors in the rendering process, as highlighted by the observed end-to-end latency in Overmeyer et al. (2023), can compromise the immersive experience and overall acceptance of the system.

Projectors are less dominant in DR applications (1%, one paper). Projectors can be used to project a see-through image obtained from a camera, capturing the background onto the surface of objects to create the illusion that the objects have been visually removed from the scene. For example, Sasai et al. (2015) presented a DR system that projects a see-through image on the dashboard of a car.

3.2.5 Processing workflow

Reviewed studies showed that most implemented methods consist of a combination of four main steps: scene tracking, object selection, object removal, and colour correction.

3.2.5.1 Scene tracking It is essential to track the position of objects relative to the cameras across multiple frames. The tracking method is highly dependent on whether the camera and the target object in the scene are fixed or moving.

As shown in Figs. 9 and 10 papers used a pre-calibration method to estimate the camera pose in the scene. While not essential, pre-calibration serves as a beneficial precursor to scene tracking, laying the groundwork for accurate and consistent tracking. By determining the intrinsic and extrinsic parameters of the camera, pre-calibration ensures alignment between virtual content and the user's perspective, enhancing the reliability and precision of subsequent tracking processes (Ono et al. 2023).

In the case of a moving camera, a six degrees of freedom (6 DoF) camera pose estimation, including three elements for the position and three elements for the orientation relative to the object, is performed. Tracking approaches in the
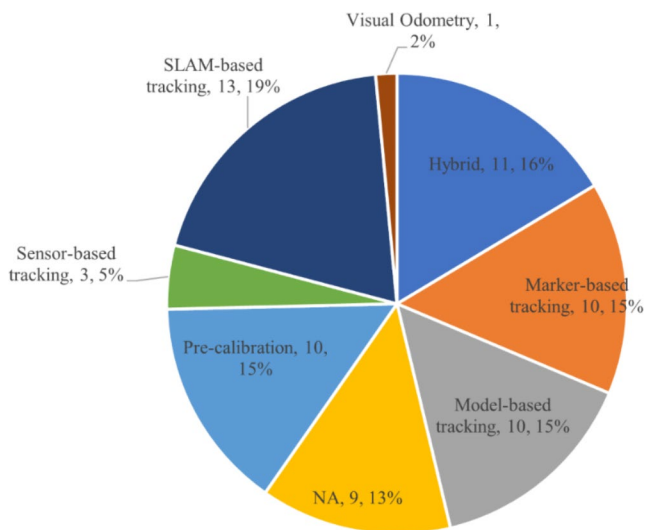
Fig. 9 Distribution of scene tracking methods. NA (not available) indicates studies that did not mention their tracking method
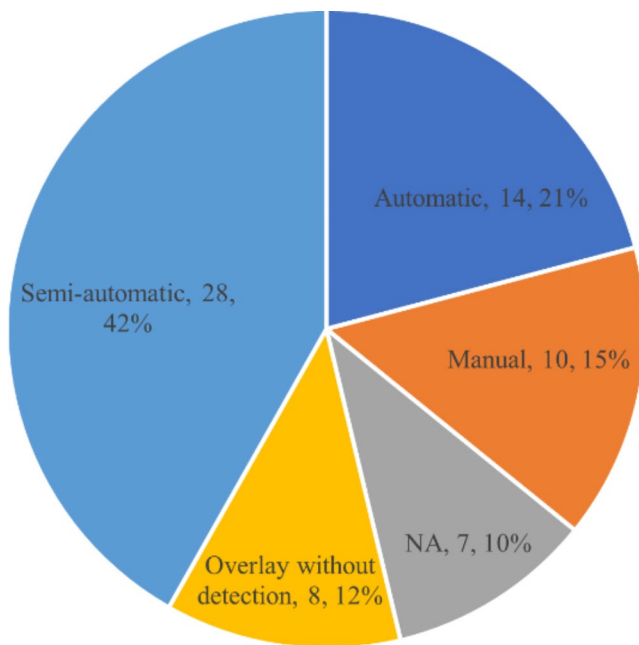
Fig. 10 Distribution of object selection methods. NA (not available) indicates evaluationstudies that did not mention their object selection method

DR literature can be classified into sensor-based and vision-based methods. In sensor-based methods, positioning algorithms using sensor data are used to determine the position and orientation of the camera. As illustrated in Fig. 9, three papers used a sensor-based tracking method. In these studies, sensors such as the gyroscope, ultrasonic, and GPS were utilised to identify the coordinates and the camera's orientation. Positioning technologies, such as GPS and inertial sensors, can be jointly used to obtain positioning and orientation information outdoors. However, this solution has

some drawbacks: it cannot be used continuously indoors, and expensive and heavy hardware is required in locations without GPS availability (Rolland et al. 2001). Additionally, while offering better accuracy, high-precision GPS systems are costly, and low-precision GPS systems require expensive hardware to compensate for their lower accuracy. For instance, Differential GPS (DGPS) systems and Real-Time Kinematic (RTK) GPS systems demand additional infrastructure and specialised equipment, further increasing costs (Radočaj et al. 2022).

Alternatively, vision-based methods use computer vision and image processing techniques to estimate the camera pose and track objects in the scene. While these methods are more accurate and reliable, they are computationally more expensive than sensor-based methods. Three types of vision-based methods are recognised in DR studies: (1) Model-based tracking methods, (2) self-localisation methods (such as Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey, 2006) and Visual Odometry (VO) (Nistér et al. 2004), and (3) marker-based methods. Model-based approaches rely on a 3D model of the surroundings for camera pose estimation and tracking. As shown in Fig. 9, model-based tracking is a dominant approach in the DR literature (10 papers). In these methods, 3D models can be generated using images and the SfM technique. The Perspective n Point (PnP) problem method (Fischler and Bolles 1981) is then utilised to estimate the camera pose, using corresponding features (e.g., Inoue et al. 2018; Oishi et al. 2017).

Another field of research has sought to achieve self-localisation. The SLAM-based approach and VO are vision-based methods used frequently in DR studies for indoor environments (as shown in Fig. 9). These methods use feature correspondences in consequent frames to estimate the camera's location in indoor environments. SLAM generates a map of the scene using RGB-D images and estimates the camera positions on the map. Visual-SLAM (Davison 2003) can overcome the problem of requiring a depth sensor by determining the depth data only from RGB images. For example, Mei et al. (2011) created a map of the environment using visual-SLAM and tracked objects within the map. Visual-SLAM and VO can cause a drift in the estimated trajectory of the camera, originating from the accumulated error of localisation (Khoshelham and Ramezani 2017).
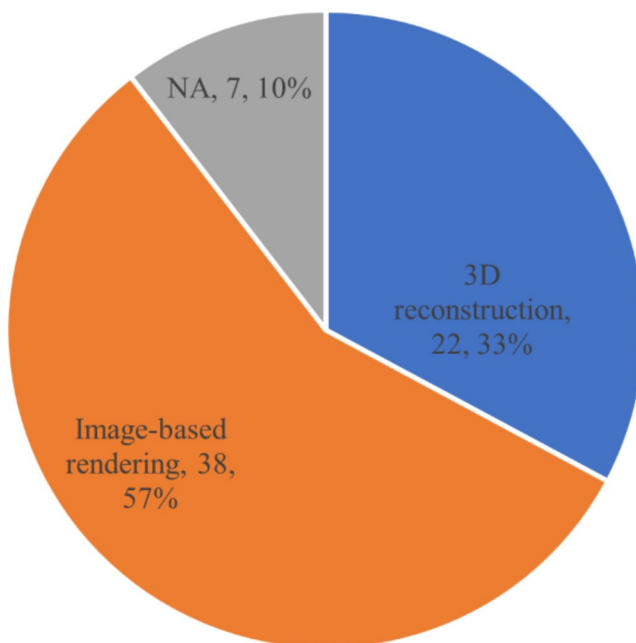
Furthermore, these methods are prone to error in weakly textured indoor environments, such as a hallway, because of the lack of features to be detected in the images. To overcome these issues, SLAM requires loop closure to eliminate the accumulating errors, which is impossible in many use cases. On the other hand, VO requires an initial location to start tracking. However, reliance on an initial location often

leads to relative positioning results, which is a limitation of VO methods (Acharya et al. 2019).

Marker-based methods have been used in 10 papers in the DR literature. These methods rely on explicit image patterns or markers placed in the scene to estimate the camera's pose. Technologies such as OptiTrack and Vuforia are common tracking technologies in this category. Taylor et al. (2020), for example, use OptiTrack to track the scene and the Vuforia tracking library is used in Queguiner et al. (2018) for camera pose estimation. Although these methods are robust, fast, and low-cost, they have drawbacks, such as requiring uniform lighting conditions and recognisable markers that strongly contrast with the environment. Furthermore, they may not be suitable for environments such as construction sites, where markers may be obstructed by workers, equipment, and machines (Palmarini et al. 2018).

Hybrid methods use a combination of tracking methods. For example, the approximate pose estimation of the GPS data combined with the matching algorithms to provide accurate camera pose estimation results presented in Habert et al. (2017). In another study, the hybrid tracking method utilised in Avery, Piekarski, and Thomas (2007) combines marker-based tracking, geometric information from CAD models, and non-vision sensors to achieve accurate scene tracking and camera registration in outdoor environments.

3.2.5.2 Object selection   Three types of object selection methods have been identified in the DR literature: (1) manual, (2) automatic or semi-automatic, and (3) overlay without detection. As shown in Figs. 10 and 15% of stud-

ies used the manual selection method (10 papers). In these methods, users manually select the ROI to find the object. For example, Maezawa et al. (2018) placed an AR Magic Lens (Baričević et al. 2012) as a virtual loupe to see through the occluded object. The ROI is determined by a user sweeping to find a good focus on the background object. Semi-automatic methods need the user to input a bounding box or a circle around the object in the first frame. The object will then be automatically detected in the next frames. As can be seen in Figs. 10 and 42% of studies (28 papers) attempted to use semi-automatic methods for object or ROI selection. For example, Queguiner, Fradet, and Rouhani (2018) used a semi-automatic method to select an ROI in the pre-processing step. Then, the selected region is tracked automatically during the run-time process.

Figure 10 shows that 21% of studies used the automatic object selection method (14 papers). In these methods, various techniques, including deep learning approaches such as semantic segmentation, are proposed to detect objects. Object detection based on deep learning is effective for Identifying ROIs due to its high performance in real-time applications (e.g., MobileNetSSD in Kido et al. (2020), CNN-based object recognition in Thompson et al. (2018), HOV-SVM in Hasegawa and Saito (2015), Graph cut algorithm segmentation in Hashimoto, Uematsu, and Saito (2010), and FCHarD-Net with a HarDNet network in Kikuchi et al. (2022).

The last category includes studies that overlay images without detecting specific objects (eight papers, 12%). For example, Sugimoto et al. (2014) used superimposing frames by time synchronisation techniques to give an impression of object removal from the scene. Although this method reduces the ROI detection computation cost, it may result in unexpected artifacts surrounding the overlayed region.

3.2.5.3 Object removal   Figure 11 shows the distribution of various object removal techniques. Two main methods were identified, namely, 3D reconstruction-based methods (22 papers, 33%) and image-based rendering methods (38 papers, 57%).

In cases where a 3D model of the scene is available, DR results can be generated using the reconstructed model, which provides depth information about the surrounding environment. Common methods such as SfM can be used to generate the geometric information of the scene. For example, Inoue et al. (2018) used a reconstructed 3D model of the scene generated using photogrammetry software to recover the hidden background.

Although methods based on 3D reconstruction effectively represent the scene's geometric information, they suffer from some limitations, including long processing times

Fig. 11 Distribution of object removal techniques. NA (not available) indicates studies that did not mention their object removal method

and high computational costs. Additionally, photo-realistic rendering using SfM algorithms for 3D reconstruction of the scene may be prone to errors such as projection errors, which occur due to the misregistration of projected images with the 3D model.

Dense reconstruction of the scene generated from sensors, such as stereo cameras (e.g., Rameau et al. 2016), and structure-from-light sensors (e.g., Kunert et al. 2019) can overcome these problems. As an example in this regard, Rameau et al. (2016) presented a method in which a depth map of the scene is generated using stereo cameras to warp a colour image to the user's view.

Image-based rendering approaches can directly generate scenes using collected images without needing 3D reconstruction. These approaches create new views by transferring the pixel values from the input images to their corresponding positions in the new views (Chang and Guo-Ping 2019). Popular image-based rendering techniques include light field rendering (Levoy and Hanrahan, 1996), Unstructured Lumigraph Rendering (ULR) (Buehler et al. 2001), and View-Dependent Texture Mapping (VDTM) (Debevec, Taylor, and Malik, 1996). For example, Mori et al. (2017a, b) used ULR to recover the background for work area visualisation. In this method, images from different viewpoint are created by assigning weights to each camera's image based on the geometric relationships between the calibrated cameras and a simplified 3D model of the scene, such as a polygon mesh.

3.2.5.4 Colour correction   Colour correction is a post-processing step designed to enhance the DR result and minimise discrepancies between the recovered ROI and the rest of the image in the current user's view. An advanced method is necessary because a naive copy of the rendered background information in the user's view could produce undesirable colour discrepancies due to illumination changes between the reconstructed model and the main view. In this process, colour differences are corrected at the boundary of the ROI and then interpolated within the ROI. This procedure is also known as *composition* or *seamless blending* in some studies (e.g., Queguiner et al. (2018).

Alpha blending and Poisson-based blending techniques are common in the DR literature for colour correction. The alpha blending technique creates new blended pixels by combining weighted background pixels and foreground pixels. Studies indicate that alpha blending is computationally inexpensive and produces efficient solutions in dynamic scenes (Mori et al. 2015; Kido et al. 2020).

The Poisson-based blending technique is another colour correction approach utilised in some studies (e.g., Kawai et al. 2016; Meerits and Saito, 2015). It solves the Poisson's equation to seamlessly integrate a source image region into a target image, ensuring smooth transitions and consistent gradients. This method handles differences in illumination and colour between the source and target images, producing more natural and accurate results than do simpler blending techniques. Although computationally expensive, Poisson-based rendering provides high-quality and visually consistent outputs (Kawai et al. 2013).
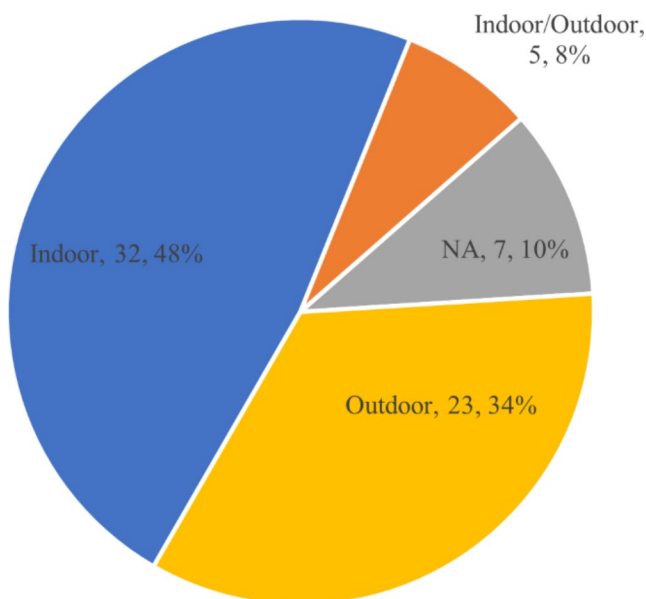
3.2.6 DR environment

Figure 12 presents the distribution of DR studies by environment. As illustrated, indoor settings are the most dominant, with 32 papers accounting for 48% of the publications. This is followed by outdoor settings, with 23 papers accounting for 34% of the publications.

### 3.3 Multi-dimensional relationship analysis

The analysis of the alluvial chart in Fig. 13 reveals insights into the relationships between DR types and background data types. One significant finding is the strong association between ROB-DR and RGB Image data, indicating the prevalent use of the RGB Image in real-time observation-based DR techniques. Additionally, the substantial presence of RGB-D data in POB-DR types underscores its importance in enhancing the depth perception and accuracy of DR results.

Figure 14 provides a visual representation of the relationships between different scene tracking methods and their prevalence across various DR types.
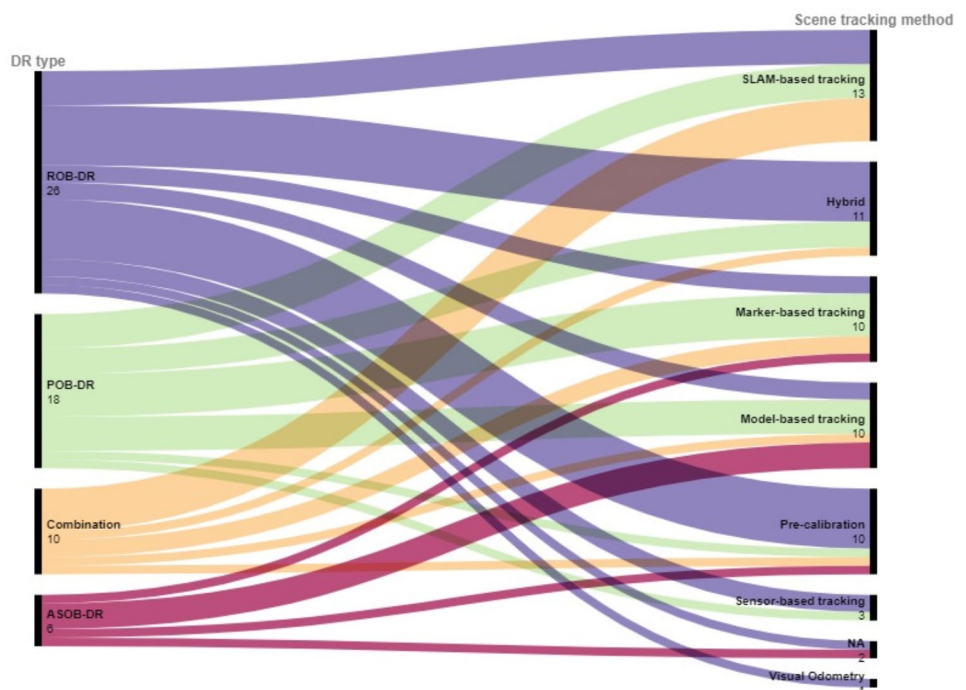

Fig. 12 Distribution of DR environment. NA (not available) indicates evaluation studies that did not mention DR environment

Fig. 13 Relationships between different DR types and background data types
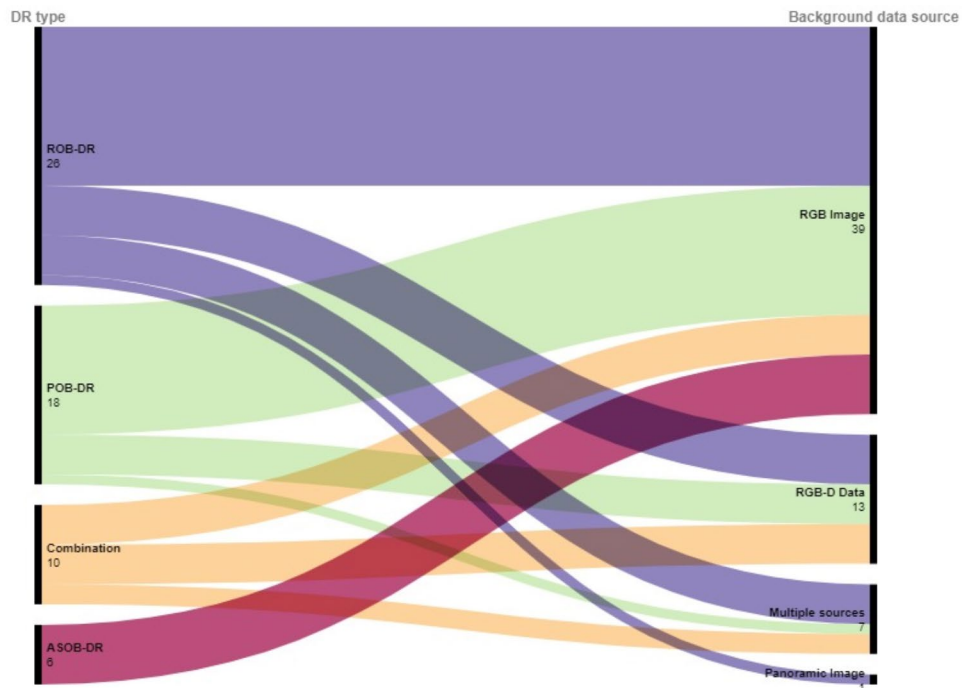
Fig. 14 Relationships between DR types and scene tracking methods

The prevalence of pre-calibration in ROB-DR indicates its critical role in enabling real-time object removal within dynamic environments. ROB-DR techniques typically operate in scenarios where immediate and precise tracking of objects is essential for the seamless integration of virtual and real-world elements. Pre-calibration ensures that the tracking system is accurately calibrated beforehand, allowing for efficient and reliable tracking of objects in real time.

The prevalence of model-based and marker-based tracking methods for POB-DR can be attributed to their compatibility with the nature of POB-DR, in which a pre-existing 3D model allows for accurate and reliable tracking

of objects without the need for extensive real-time environment mapping.

Figure 15 illustrates the distribution of DR types by application domain. The ROB-DR type is more commonly utilised in the automobile sector, whereas the AEC-FM industry uses POB-DR methods predominantly.

The prevalence of ROB-DR in the automobile industry can be attributed to the industry's need for real-time data to ensure safety, performance, and efficiency. On the other hand, the dominance of POB-DR in the AEC-FM industry is mainly due to the effectiveness of 3D modelling technologies, especially those based on 3D laser scanning techniques, which are known for their ability to create highly accurate and detailed digital representations of built environments (Tang et al. 2010). This data collection accuracy enables the effective application of POB-DR techniques in the AEC-FM industry.

Figure 16 illustrates the relationships between DR types and the DR environment. The ROB-DR and ASOB-DR categories that require real-time data of the background are predominantly employed in outdoor environments. In contrast, the POB-DR category is mainly utilised in indoor environments.

The prevalence of POB-DR methods in indoor environments can be attributed to several factors. The indoor environment provides a more controlled lighting environment, which is crucial in minimising the impact of illumination inconsistencies. Such inconsistencies arise from differences in the illumination between the run-time processing and the 3D virtual model or images of the background environment

that are collected in advance. However, the choice of any one among POB-DR, ASOB-DR, and ROB-DR methods is contingent upon the requirements of the application. While POB-DR methods may perform optimally in indoor environments, they might call for the adoption of techniques tailored to outdoor settings to ensure precise background recovery.

4 Discussion, recommendations, and future directions

The literature review presented in the preceding sections reported several findings. Based on it, we identified key challenges associated with observation-based DR, and these challenges are detailed in this section to help in developing practical DR systems.
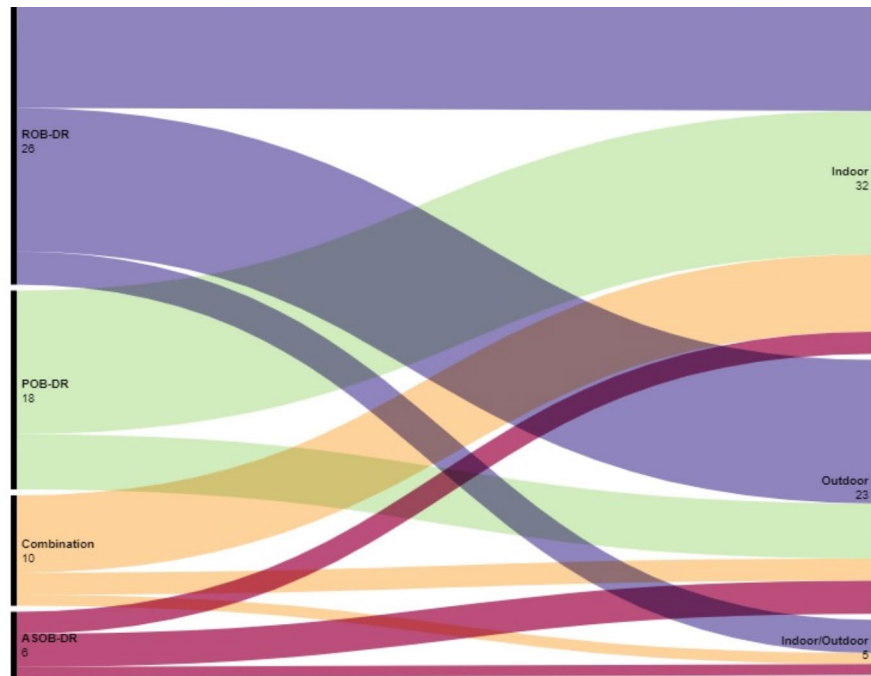
4.1 Real-time processing

One of the main challenges with observation-based DR is the requirement to process large amounts of data in real time. This calls for high accuracy and speed to ensure that data, such as video streams, are updated in real time without delay and with minimal latency. Real-time processing requires a combination of hardware and software optimisation, including powerful GPUs, specialised algorithms, and efficient data management techniques (Mohamed et al. 2023). Additionally, the application must be able to handle various types of input sources, such as different types

Fig. 15 Relationships between DR types and application domains
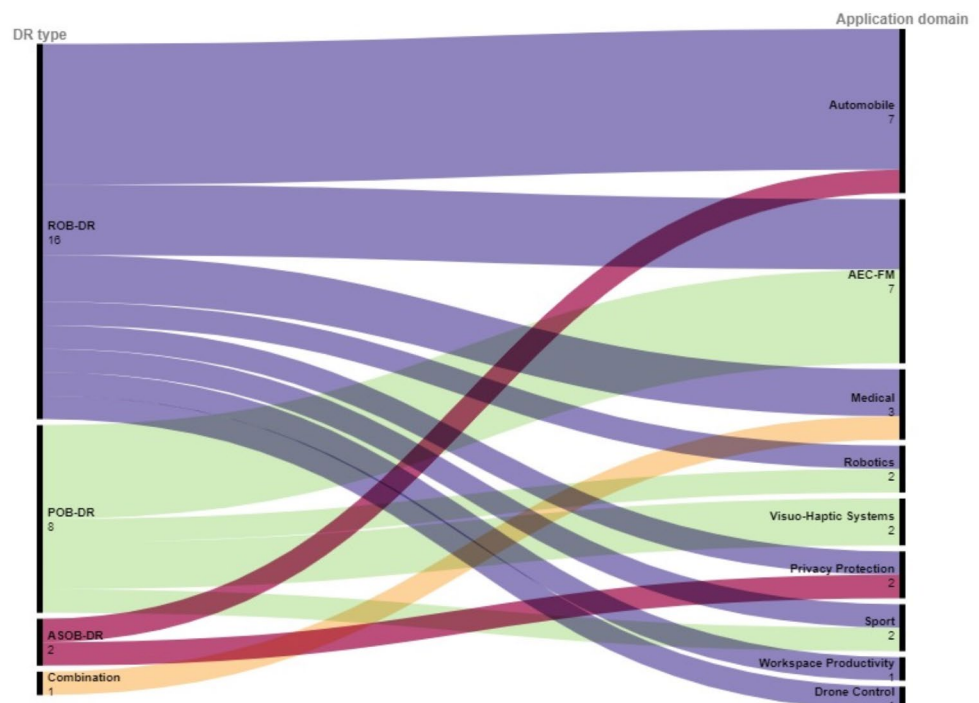
Fig. 16 Relationships between DR types and DR environment

of cameras, which can add complexity to the processing pipeline. The challenge is to achieve DR in real time while maintaining a high level of accuracy and visual quality. This requires a careful balance between processing speed and visual fidelity.

4.2 Object detection and tracking

The efficacy of DR hinges significantly on its ability to achieve precise object detection and tracking. This requires the use of advanced computer vision techniques, including deep learning-based methods. These techniques require large amounts of data for training and sophisticated algorithms for real-time processing. However, the accuracy and reliability of these algorithms can be affected by factors such as lighting conditions, occlusions, and environmental changes (Mirani et al. 2022). In some cases, the application may need to track multiple objects simultaneously, inserting an additional level of complexity to the tracking process (Luo et al. 2021).

To address these challenges, DR researchers can explore new techniques for object detection and tracking, including the use of advanced machine learning algorithms (e.g., Ren et al. 2015), multi-sensor data fusion (e.g., Senel et al. 2023), and efficient data processing techniques. Additionally, the use of edge computing-based object detection architectures (e.g., Ren et al. 2018) can help improve the accuracy and speed of object detection and tracking in DR applications.

4.3 Evaluation

Given the diverse nature of DR applications and technical approaches, there is a growing need for researchers to develop standardised evaluation methods to ensure consistent and reliable assessments of DR application performance and usability across different contexts (Morozumi et al. 2017) In response to this need, simulation-based approaches offer a cost-effective, controlled, and safe environment for the development and evaluation of DR applications. By replicating realistic scenarios without the need for expensive equipment or setups, simulations enable researchers and developers to precisely control environmental variables such as lighting conditions, object interactions, and user movements.

4.4 User experience

The challenge with the user experience in DR applications lies in the need to balance the functionality and effectiveness of the application with the user's comfort and convenience (Peereboom et al. 2023). DR applications can be particularly challenging in this regard, especially in cases that require the user to wear a device, such as a headset or glasses, in order to see the modified view. This can be uncomfortable for some users, particularly over extended periods of time, and can lead to issues such as eye strain or fatigue (Ariansyah et al. 2022).

To address these challenges, developers of DR applications must consider the user experience, starting from the design phase through to deployment, by conducting an

interview or qualitative evaluation. The evaluation includes ensuring that the user interface is intuitive and easy to use, that the device is comfortable to wear, and that the DR result is visually appealing and easy to understand. For example, the study by Peereboom et al. (2023) employed quantitative evaluation methods to assess user comfort and satisfaction with different DR/AR designs in a pedestrian crossing scenario, providing valuable insights into user preferences and experiences.

4.5 Cost

The cost challenge in DR applications can represent a significant barrier to the widespread adoption of the technology. There are several factors that contribute to the cost challenge, including hardware and software development costs.

Hardware is one of the primary cost drivers for DR applications. The devices used to display the DR view, such as computer tablets, smart glasses, and HMDs, can be expensive. Additionally, these devices may require supplementary components, such as sensors and cameras. For instance, the Structure IO sensor mounted on an iPad for a DR application, as described in the study by Andre and Hlavacs (2019), can further increase the overall cost.

Software development costs can also pose a significant challenge for DR applications. These applications require advanced computer vision and machine learning algorithms to process live video or images in real time, and these can be complex and time-consuming to develop. Additionally, developing the user interface and user experience can also be costly, as they require significant resources and expertise.

To address the cost challenge in DR applications, developers can explore alternative approaches such as leveraging existing hardware or software components or utilising open-source software libraries, such as OpenCV for Unity asset. Additionally, developers can work to optimise the performance of their software to reduce the hardware requirements and lower the cost of the devices used to display DR results.

Furthermore, developers can work to identify new use cases and industries where the benefits of DR applications outweigh the cost. For example, DR applications may be particularly valuable in fields such as manufacturing, where the technology can be used to improve worker safety or increase productivity (Maezawa et al. 2018).

In the coming years, the development of DR technology is expected to prioritise several key areas. Firstly, there will be a strong focus on improving the accuracy and effectiveness of the technology. This will involve the implementation of new technical approaches and evaluation methods to achieve better results. Secondly, the DR technology will expand its application base beyond its current use cases to fields, such as manufacturing or healthcare, where the benefits of the technology can have a significant impact. Thirdly, enhancing the user experience will be a key priority, as DR technology is integrated into other emerging technologies, such as AI and the IoT, to create more seamless and immersive experiences for users. Finally, embracing advancements in display device technology could lead to improved accuracy and quality in DR techniques, fostering further advancements in the field.

5 Conclusion

In this work, a systematic literature review of recent studies related to observation-based DR techniques was conducted. Relevant keywords in the literature were searched using the Scopus search engine. 67 studies meeting the study criteria were selected as key articles. After an in-depth review of these articles, the results were discussed to help answer the research questions.

The findings of this study revealed the potential of the DR function in many applications, such as interior re-design simulations and outdoor landscape simulations in the AEC-FM industry; efforts to increase driver safety and assist drivers in the automobile industry; helping surgeons in the medical field; and contribution to many other fields, such as visuo-haptic systems, robotics, sports, and drone navigation. The majority of the papers examined were related to technical method development, which concentrated on presenting new algorithms or improving existing techniques. Monitors were the most common interaction devices employed for displaying DR results. In addition, SLAM-based tracking was the most dominant method for scene tracking in these studies. Most of the studies employed a semi-automatic method for selecting the objects to be removed in the scene. For object removal and colour correction, IBR and alpha blending were respectively the most commonly used approaches.

This systematic literature review revealed that the development and implementation of DR applications face several challenges. These challenges include real-time object detection and tracking, evaluation method establishment, functionality balancing, user experience, and cost. To address these challenges, developers must work to optimise the performance of their software and develop intuitive user interfaces. Additionally, they can explore alternative approaches such as leveraging existing hardware or software components or utilising open-source software libraries. Tackling these challenges through new techniques and technologies can help researchers and developers overcome obstacles and unlock the full potential of DR in various applications.

References

Acharya D, Khoshelham K, and Stephan Winter (2019) BIM-PoseNet: indoor camera localisation using a 3D indoor model and deep learning from synthetic images. ISPRS J Photogrammetry Remote Sens 150:245–258

Andre E, Hlavacs H (2019) Diminished Reality Based on 3D-Scanning. In: Joint international conference on entertainment computing and serious games, pp 3–14. Springer

Ariansyah D, Erkoyuncu JA, Eimontaite I, Johnson T, Oostveen A-M, Fletcher S, Sharples S (2022) A head mounted augmented reality design practice for maintenance assembly: toward meeting perceptual and cognitive needs of AR users. Appl Ergon 98:103597

Avery B, Piekarski W, Bruce H (2007) Visualizing occluded physical objects in unfamiliar outdoor augmented reality environments. In: 2007 6th IEEE and ACM International symposium on mixed and augmented reality, pp 285–286. IEEE

Avery B, Thomas BH (2008) User evaluation of see-through vision for mobile outdoor augmented reality. In: 2008 7th IEEE/ACM international symposium on mixed and augmented reality, pp 69–72. IEEE

Baričević D, Lee C, Turk M, Höllerer T, Bowman DA (2012) A handheld ar magic lens with user-perspective rendering. In: 2012 IEEE international symposium on mixed and augmented reality (ISMAR), pp 197–206. IEEE

Buehler C, Bosse M, McMillan L, Gortler S (2001) Unstructured Lumigraph Rendering. In: Proceedings of the 28th annual conference on computer graphics and interactive techniques, pp 425–432

Chan SWT, Bektur R, Nanayakkara S (2022) DeclutterAR: mobile diminished reality and augmented reality to address hoarding by motivating decluttering and selling on online marketplace. In: 2022 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), pp 870–874. IEEE

Chang Y, Guo-Ping W (2019) A review on image-based rendering. Virtual Real Intell Hardw 1(1):39–54

Davison AJ (2003) Real-time simultaneous localisation and mapping with a single camera. In: Computer vision, IEEE international conference on, vol 3, pp 1403–1403. IEEE Computer Society

Debevec PE, Taylor CJ (1996) . Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In: Proceedings of the 23rd Annual conference on computer graphics and interactive techniques, pp 11–20

Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping: part I. IEEE Rob Autom Magazine 13(2):99–110

Erat O, Isop WA, Kalkofen D, Schmalstieg D (2018) Drone-augmented human vision: exocentric control for drones exploring hidden areas. IEEE Trans Vis Comput Graph 24(4):1437–1446

Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395

Gkitsas V, Sterzentsenko V, Zioulis N, Albanis G, Zarpalas D (2021) PanoDR: spherical panorama diminished reality for indoor scenes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3716–3726

Gomes P, Vieira F (2012) and Ferreira M. The See-through system: from implementation to test-drive. In: 2012 IEEE vehicular networking conference (VNC), pp 40–47. IEEE

Habert S, Ma M, Fallavollita P (2017) Multi-Layer visualization for medical mixed reality

Hasegawa K, Saito H (2015) Diminished reality for hiding a pedestrian using hand-held camera. In: 2015 IEEE international symposium on mixed and augmented reality workshops, pp 47–52. IEEE

Hashiguchi S, Mori S, Tanaka M, Shibata F (2018) Perceived weight of a rod under augmented and diminished reality visual effects. In: Proceedings of the 24th ACM symposium on virtual reality software and technology, pp 1–6

Hashimoto T, Uematsu Y, Saito H (2010) Generation of See-through Baseball Movie from Multi-Camera Views. In: 2010 IEEE International Workshop on Multimedia Signal Processing, pp 432–437. IEEE

Inoue KA, Fukuda TO, Cao R, Yabuki N (2018) Tracking robustness and green view index estimation of augmented and diminished reality for environmental design. In: Proceedings of CAADRIA 2018, pp 339–48

Jarusirisawad S, Saito H (2007) Diminished reality via multiple handheld CAMERAS. In: 2007 First ACM/IEEE international conference on distributed smart cameras, pp 251–258. https://doi.org/10.1109/ICDSC.2007.4357531

Kameda Y, Takemasa T, Ohta Y (2004) Outdoor see-through vision utilizing surveillance cameras. In: Third IEEE and ACM international symposium on mixed and augmented reality, pp 151–160. IEEE

Kari M, Grosse-Puppendahl T, Coelho LF, Fender AR, Bethge D, Schütte R, Holz C (2021) Transformr: Pose-aware object substitution for composing alternate mixed realities. In: 2021 IEEE international symposium on mixed and augmented reality (ISMAR), pp 69–79. IEEE. https://ieeexplore.ieee.org/abstract/document/9583783/

Kawai N, Yamasaki M, Sato T, Yokoya N (2013) Diminished reality for AR marker hiding based on image inpainting with reflection of luminance changes. ITE Trans Media Technol Appl 1(4):343–353

Kawai N, Sato T, Yokoya N (2015) Diminished reality based on image inpainting considering background geometry. IEEE Trans Vis Comput Graph 22(3):1236–1247

Kawai N, Sato T, Nakashima Y, Yokoya N (2016) Augmented reality marker hiding with texture deformation. IEEE Trans Vis Comput Graph 23(10):2288–2300

Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3(3):203–220

Khoshelham K, Ramezani M (2017) Vehicle positioning in the absence of GNSS signals: potential of visual-inertial odometry. In: 2017 joint urban remote sensing event (JURSE), pp 1–4. IEEE

Kido D, Fukuda T, Yabuki N (2020) Diminished reality system with real-time object detection using deep learning for onsite landscape simulation during redevelopment. Environ Model Softw 131:104759

Kikuchi T, Fukuda T, and Nobuyoshi Yabuki (2022) Diminished reality using semantic segmentation and generative adversarial network for landscape assessment: evaluation of image Inpainting according to Colour Vision. J Comput Des Eng 9(5):1633–1649

Kitchenham B (2004) Procedures for performing systematic reviews. Keele, UK, Keele University. vol 33, pp 1–26

Kittaka T, Fujii H, Yamashita A, Asama H (2016) Creating see-through image using two RGB-D sensors for remote control robot. In: 2016 11th France-Japan 9th Europe-Asia congress on mechatronics (MECATRONICS) /17th International conference on research and education in mechatronics (REM), pp 086–091. https://doi.org/10.1109/MECATRONICS.2016.7547121

Kunert C, Schwandt T, Broll W (2019) An Efficient Diminished Reality Approach Using Real-Time Surface Reconstruction. In: 2019 international conference on cyberworlds (CW), pp 9–16. IEEE

Lee JH, Kim LH (2024) Augmenting reality to diminish distractions for cognitive enhancement. arXiv. http://arxiv.org/abs/2403.03875

Levoy M, Hanrahan P (1996) Light Field Rendering. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques, pp 31–42

Li Z, Wang Y, Guo J, Cheong LF, Zhou SZ (2013) Diminished reality using appearance and 3D geometry of internet photo collections. In: 2013 IEEE international symposium on mixed and augmented reality (ISMAR), pp 11–19. IEEE

Lin C, Popescu V (2022) Fast intra-frame video splicing for occlusion removal in diminished reality. In: Raya MA, Bourdot P, Marchal M, Stefanucci J, Yang X, Zachmann G (eds) Virtual reality and mixed reality. vol 3484, pp 111–134. Lecture Notes in Computer Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-16234-3_7

Lindemann P, Rigoll G (2017) A diminished reality simulation for driver-car interaction with transparent cockpits. In: 2017 IEEE virtual reality (VR), pp 305–306. IEEE

Luo W, Xing J, Milan A, Zhang X, Liu W, Kim T-K (2021) Multiple object tracking: a literature review. Artif Intell 293:103448

Maezawa M, Mori S, Saito H (2018) A refocus-interface for diminished reality work area visualization. Electron Imaging 2018(4):112–1

Mann S (1999) Mediated Reality. Linux J 1999(59es):5-es

Mann S, Havens JC, Iorio J, Yuan Y, Furness T (2018) All reality: values, taxonomy, and continuum, for virtual, augmented, extended/mixed (X), mediated (X, Y), and multimediated reality/intelligence. AWE 2018. http://wearcomp.org/all.pdf

Meerits S, Saito H (2015) Real-time diminished reality for dynamic scenes. In: 2015 IEEE international symposium on mixed and augmented reality workshops. IEEE. pp 53–59

Mei C, Sommerlade E, Sibley G, Newman P (2011) Hidden view synthesis using real-time visual SLAM for simplifying video surveillance analysis. In 2011 IEEE international conference on robotics and automation. IEEE. pp 4240–45

Milgram P, Takemura H, Utsumi A, Kishino F (1995) Augmented reality: a class of displays on the reality-virtuality continuum. In: Telemanipulator and telepresence technologies, vol 2351. International Society for Optics and Photonics. pp 282–292

Mirani IK, Tianhua C, Khan MA, Aamir SM, Menhaj W (2022) Object recognition in different lighting conditions at various angles by deep learning method. arXiv http://arxiv.org/abs/2210.09618

Mohamed I, Elhenawy I, Salah A (2023) A survey on GPU-Based visual trackers. In: Recent advances in computer vision applications using parallel Processing. Springer, pp 71–85

Mori S, Shibata F, Kimura A, Tamura H (2015) Efficient use of textured 3D model for pre-observation-based diminished reality. In: 2015 IEEE international symposium on mixed and augmented reality workshops. IEEE. pp 32–3

Mori S, Eguchi Y, Ikeda S, Shibata F, Kimura A, Tamura H (2016) Design and construction of data acquisition facilities for diminished reality research. ITE Trans Media Technol Appl 4(3):259–268

Mori S, Ikeda S, Saito H (2017a) A survey of diminished reality: techniques for visually concealing, eliminating, and seeing through real objects. IPSJ Trans Comput Vis Appl 9(1):1–14

Mori S, Maezawa M, Saito H (2017b) A work area visualization by multi-view camera-based diminished reality. Multimodal Technol Interact 1(3):18

Morozumi T, Mori S, Ikeda S, Shibata F, Kimura A, Tamura H (2017) [POSTER] Design and Implementation of a common dataset for comparison and evaluation of diminished reality methods. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct). IEEE. pp 212–213

Namboku Y, Takahashi H (2020) Diminished reality in textureless scenes. In: International workshop on advanced imaging technology (IWAIT) 2020, vol 11515. International Society for Optics and Photonics. pp 1151522

Nistér D, Naroditsky O, Bergen J (2004) Visual odometry. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR.

Oishi K, Mori S, Saito H (2017) An instant see-through vision system using a wide field-of-view camera and a 3D-Lidar. In: 2017 IEEE International symposium on mixed and augmented reality (ISMAR-adjunct). pp 344–347. https://doi.org/10.1109/ISMAR-Adjunct.2017.99

Ono T, Ueda K, Ishii H, Shimoda H (2023) Development of a Calibration Method of Hidden Background Observer Cameras for Diminished Reality Using Radio Direction Finding. In: 2023 IEEE international conference on systems, man, and cybernetics (SMC). IEEE. pp 4265–4270. https://ieeexplore.ieee.org/abstract/document/10394605/

Overmeyer L, Jütte L, Poschke A (2023) A real-time augmented reality system to see through Forklift components. CIRP Ann 72(1):409–412

Palmarini R, Erkoyuncu JA, Roy R, Torabmostaedi H (2018) A systematic review of augmented reality applications in maintenance. Robot Comput Integr Manuf 49(February):215–228. https://doi.org/10.1016/j.rcim.2017.06.002

Peereboom J, Tabone W, Dodou D, de Winter J (2023) Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios. ResearchGate

Pintore G, Agus M, Almansa E, Gobbetti E (2022) Instant automatic emptying of panoramic indoor scenes. IEEE Trans Vis Comput Graph

Qiaozhi L (2016) Diminished reality based on kinect fusion

Queguiner G, Fradet M, Rouhani M (2018) Towards mobile diminished reality. In: 2018 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-adjunct). IEEE. pp 226–231

Radočaj D, Plaščak I, Heffer G, Jurišić M (2022) A low-cost global navigation satellite system positioning accuracy assessment method for agricultural machinery. Appl Sci 12(2):693

Rameau F, Ha H, Joo K, Choi J, Park K, Kweon IS (2016) A real-time augmented reality system to see-through cars. IEEE Trans Vis Comput Graph 22(11):2395–2404

Ren S, He K, Girshick R, and Jian Sun (2015) Faster R-Cnn: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28

Ren J, Guo Y, Zhang D, Liu Q, Zhang Y (2018) Distributed and efficient object detection in edge computing: challenges and solutions. IEEE Netw 32(6):137–143

Rolland JP, Larry D, Davis, and Yohan Baillot (2001) A survey of Tracking technologies for virtual environments. Fundamentals of Wearable computers and augmented reality. CRC, pp 83–128

Sasai S, Kitahara I, Kameda Y, Ohta Y, Kanbara M, Morales Y, Ukita N, Hagita N, Ikeda T, Shinozawa K (2015) MR visualization of wheel trajectories of driving vehicle by seeing-through dashboard. In: 2015 IEEE international symposium on mixed and augmented reality workshops. IEEE. pp 40–46

Senel N, Kefferpütz K, Doycheva K (2023) Multi-sensor Data fusion for real-time multi-object tracking. Processes 11(2):501

Siriwardhana Y, Porambage P, Liyanage M, Ylianttila M (2021) A survey on mobile augmented reality with 5G mobile edge computing: architectures, applications, and technical aspects. IEEE Commun Surv Tutorials 23(2):1160–1192

Sugimoto K, Fujii H, Yamashita A, Asama H (2014) Half-diminished reality image using three Rgb-d sensors for remote control robots. In: 2014 IEEE international symposium on safety, security, and rescue robotics. IEEE. pp 1–6

Taylor AV, Matsumoto A, Carter EJ, Plopski A, Admoni H (2020) Diminished reality for close quarters robotic telemanipulation

Thompson RL, Zhen HU, Cho J, Stovall J, Sartipi M (2018) Enhancing driver awareness using see-through technology. SAE Technical Paper

Wei L, Khan M, Mehmood O, Dou Q, Bateman C, Magee DR (2019) Web-based visualisation for look-ahead ground imaging in tunnel boring machines. Autom Constr 105:102830

Yokoro K, Perusquia-Hernandez M, Isoyama N, Uchiyama H, Kiyokawa K (2023) DecluttAR: an interactive visual clutter dimming system to help focus on work. In: Proceedings of the augmented humans international conference 2023. pp 159–170

Yue Y-T, Yang Y-L, Ren G, Wang W (2017) SceneCtrl: mixed reality enhancement via efficient scene editing. In: Proceedings of the 30th annual ACM symposium on user interface software and technology. pp 427–436

Zhu Y, Fukuda T, Yabuki N (2019) Synthesizing 360-degree live streaming for an erased background to study renovation using mixed reality

Zhu Y, Fukuda T, Yabuki N (2020) Integrated co-designing using building information modeling and mixed reality with erased backgrounds for stock renovation