Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/34/20230319

Mixed Reality: Present and future

Zhenyu Hu

Faculty of science and engineering, University of Liverpool, Liverpool, L69 3BX UK

sgzhu6@liverpool.ac.uk

Abstract. Mixed Reality (MR) is an emerging technology that can create a visual environment connecting the real world and the virtual environment. With this technology, the way information is presented has changed in a great number of domains. However, for common people, this technology still looks far away from life and seems unintelligible. Therefore, this essay reviews the development and recent research of MR, trying to help people without professional backgrounds understand this emerging technology and offer researchers another perspective on MR. By using the methods of text mining, this essay has reviewed more than twenty papers related to MR on different topics, including the definition, the principles, and the applications. The objective of this essay is to provide a comprehensive understanding of MR, which may promote wider applications and research.

Keywords: mixed reality, technologies of MR, current trends of MR, challenges of MR.

1. Introduction

With the unveiling of the latest Mixed Reality product, Vision Pro, by Tim Cook at the 2023 Worldwide Developers Conference, Mixed Reality (MR) technology has once again captured the world's attention on a grand scale. While MR has seen significant application in recent years, such as in driving systems [1] and healthcare, the precise definition of this innovative technology remains ambiguous, lacking a unified consensus in both industry and academia [2]. There are generally four definitions of MR [3], which can be summarized as the Reality-Virtuality Continuum, a synonym for augmented reality (AR), and collaboration or combination of virtual reality (VR) and AR. Nevertheless, it is undeniable that MR technology has its roots in both AR and VR. The technique of AR involves integrating 3D graphics into the real world by spatially registering them in the physical environment [3], offering users a visually enhanced experience based on the real physical world. On the other hand, VR merges virtual reality with the real world, enabling spatial interaction within a simulated environment.

Despite the significant impact of these three technologies across various industries, numerous challenges persist. Therefore, the purpose of this essay is to critically review and discuss the current development of MR technology, including the definition and technologies of MR, while also exploring future possibilities and challenges that lie ahead. By delving into these aspects, the aim is to gain a deeper understanding of the potential of MR and its implications for the future.

2. Definition of MR

In order to enhance comprehension and gain insights into this technology, this section will delve into the fundamental definition of MR and explore the distinctions.

2.1. Reality-Virtuality Continuum

The commonest definition is the Reality-Virtuality Continuum [4]. This conception was first brought up by Milgram et al. in A Taxonomy of Mixed Reality Visual Displays in 1974 and has been cited thousands of times in the next quarter century. The continuum utilizes computing techniques to create a scale model representing real-world classes. Figure 1 illustrates that on the left side is the real environment, which is the real world without any virtual objects. On the right side is the virtual environment, which is the environment created by computer graphical techniques. The mixed reality is everything between these two parts (not including the extrema). MR aims to merge real and virtual environments, enabling seamless coexistence and real-time interaction between physical and virtual objects in a vast interconnected space, catering to various user scenarios [5][6]. It serves as a class of simulators that combine both realms, resulting in a compelling hybrid experience that bridges the gap between the virtual and real worlds.

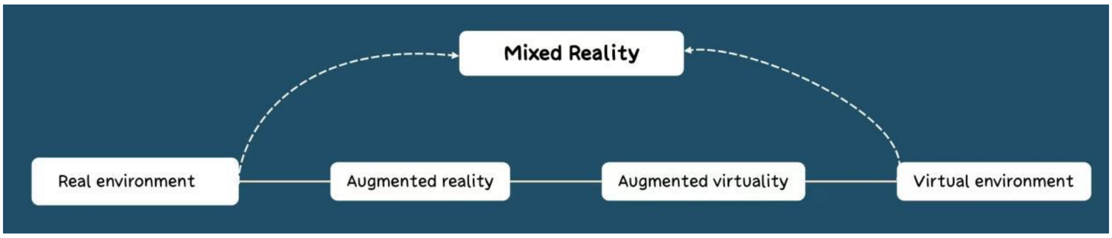

Figure 1. Reality - Virtuality continuum [5].

However, the Reality-Virtuality Continuum, while a valuable framework in the early 1990s, primarily centered on visual displays, thus limiting its ability to fully encompass novel developments like multi-user or multi-environment MR experiences. As technology has progressed, the scope of MR has expanded far beyond visual elements, incorporating diverse sensory inputs and interactive capabilities that go beyond what the early model could anticipate [3]. Therefore, relying solely on the Reality-Virtuality Continuum may hinder a comprehensive understanding of the full potential and complexity of modern MR applications.

2.2. Synonym for AR

In the second situation, it is commonly perceived that MR is an advanced iteration of AR, encompassing a broader range of functions and capabilities [3]. This view holds that despite the expanded functionalities, both MR and AR still share fundamental core techniques, such as the seamless integration of 3D graphics into the real world and spatial registration within the physical environment.

2.3. Collaboration and Combination

Another two notions of MR are the collaboration or combination of AR and VR [3]. The former refers to MR, which describes interactions between an AR user and a VR user who may be physically separated. This idea encompasses space mapping, wherein the environment of a local AR user is reconstructed in VR for a remote user's experience. The latter represents that MR is a system that combines separate AR and VR components, allowing these two parts to interact without strict integration, or an application capable of seamlessly switching between AR and VR based on requirements.

The common ground between these two concepts lies in the focus on intertwining real and virtual elements to create seamless and immersive experiences for users. However, disparity and controversy arise regarding whether VR should be considered part of MR. The former emphasizes the interaction between separate users in distinct environments, while the latter emphasizes the combination of distinct AR and VR elements within a single system.

2.4. Comparison of Definitions

This section has presented the four predominant definitions of MR. The diverse definitions can be attributed to the dynamic nature of this emergent technology, which, despite being in practical application for several decades, lacks a universally accepted consensus within the academic and

DOI: 10.54254/2755-2721/34/20230319

industrial domains. Moreover, as a derivative of AR and VR, in certain instances, it may be more pertinent to focus on researching the individual technologies of AR and VR rather than striving to establish a precise definition for MR itself.

Upon comparing the four definitions, it becomes evident that they all concur that MR originates from the integration of AR and VR. Additionally, AR, which exhibits stronger ties to the real world and does not entail a completely virtual environment, is recognized as a subset of MR by all definitions. However, the disparity and controversy arise regarding whether VR should be considered as part of MR. The first two definitions regard VR as a separate entity from MR, considering it solely as a method to enhance AR techniques. Conversely, the other two definitions view VR as an integral component of MR. Although a precise and standardized definition might aid in comprehending MR better, as mentioned above, in practical research and application, the distinction may not hold significant weight. Given that MR encompasses a fusion of techniques, the key lies in understanding how to effectively utilize it rather than solely focusing on a rigid definition.

3. Techniques and Development of MR

3.1. Principles and Techniques

At present, there exist two primary technologies for crafting MR spaces, namely optical see-through and video-see-through approaches. In the optical see-through technology, users perceive the real world directly through tools, such as transparent glass, facilitating the seamless integration of virtual elements into the real environment. Conversely, the video-see-through technology entails the simultaneous presentation of both virtual and real objects on an Liquid Crystal Display screen, allowing users to observe a blend of synthetic and real-world content [7].

The key distinction between these two technologies is how they deliver the MR experience to users. Optical see-through offers a more natural and immersive interaction by enabling users to view the real world unobstructed, with virtual elements seamlessly integrated into their visual field. This approach often involves utilizing specialized head-mounted displays or glasses to achieve the desired effect.

On the other hand, video-see-through technology leverages an LCD screen to superimpose virtual objects onto the real world as perceived through the display [8]. This method provides greater flexibility in terms of device portability and content rendering, as it does not rely on direct visual access to the real world. However, it may present some challenges in achieving perfect alignment and transparency of virtual objects in the user's view, which may potentially affecting the overall sense of immersion.

To establish a MR environment utilizing a see-through approach, several essential steps are involved, as depicted in Figure 2, including device recognition, space recognition, object detection and tracking, mapping recognition, visual recognition, and portable recognition.

Figure 2. MR environment creation steps [5].

3.2. Algorithms and Methods

In the process of establishing an MR environment, various crucial algorithms and methods are required at each step of the setup. Hence, this section will focus on discussing and summarizing different studies on these algorithms to introduce the development of the current MR.

3.2.1. Calibration. In MR, calibration technology refers to the process of establishing an accurate correspondence between physical locations in the real world and virtual objects or scenes in the virtual world [9], where virtual objects can be precisely aligned with the real world, providing users with a more realistic and immersive experience in MR.

In the current study, in the context of Head-Mounted Displays worn on the user's eyes, the calibration model addresses the instability between the handheld tool and the user's hand, enabling seamless interaction with virtual objects and leading to more accurate camera parameter estimation [5]. Additionally, another research team utilized a projection model with a matrix transformation formula to compute fish-eye camera parameters, which proved valuable in generating full-view spherical images for scene reconstruction [10].

3.2.2. Model of Space and Simulation. Space modelling is also significant to create an MR environment, which contains modelling large spaces, such as rooms, and incorporating real objects, such as tissues. Various methods exist for space modelling. For instance, geometry and VRML format was utilized for modelling MR room applications, employing descriptors to provide user information and location in the MR space [11]. Similarly, another research team proposed a toolkit for building large-space MR environments, using Voronoi spatial analysis to construct virtual scenes in 2D maps and 3D scenes through sweeping analysis [12]. The toolkit enables arranging virtual objects in real scenes with accurate matching using anchor data and a Voronoi-based structure for path management, object layer adjustments, and efficient rendering visibility preservation. Furthermore, a machine learning and Delaunay triangulation-based method was also used for creating an MR environment, which was generated using 3D data from scanning, mesh calculations, and region classification with specific rules, enabling precise placement of scene objects within classified regions [13].

3.2.3. Object Recognition and Tracking. Another two essential algorithms in MR are object recognition and tracking. These two technologies allow for the identification and tracking of real-world objects, facilitating the seamless integration of virtual objects into the real environment to create immersive MR experiences [5].

In the realm of object recognition, diverse methodologies have been investigated by researchers. One instance involved an image-processing-based strategy, utilizing probability-based segmentation, to identify mobile entities like humans from the foreground [14]. This enabled the introduction of virtual elements, such as homes, into real-world scenes. In contrast, a separate study proposed an image processing approach that employed semantic segmentation coupled with conditional random fields in recurrent neural networks (CRF-RNN) to establish a relevant context for presentation on a material-aware platform [15]. Moreover, spatial object recognition was embraced in another study to showcase pertinent information, like buildings, based on the user's field of vision [16]. This framework integrated distance-dependent filtering to determine the information to exhibit, contingent on the user's proximity to the subjects of interest.

Furthermore, MR tracking methods are typically classified into three categories: sensor-based, vision-based, and hybrid [17]. The hybrid method combines elements from both sensor-based and vision-based approaches to address their respective limitations effectively. Sensor-based tracking relies on various sensing devices, including GPS, visual markers, acoustical tracking systems, magnetic and inertial sensors, and hybrid systems, to estimate real object positions. On the other hand, vision-based tracking utilizes computer vision techniques and image processing, which can be either feature-based or model-based. Three mechanisms were proposed for self-inspection in the BIM model, which includes feature-based tracking using feature detection and SLAM (Simultaneous Localization and Mapping), model-based tracking involving contour detection, and spatial-based tracking using depth information [18]. These vision-based approaches allow for the accurate positioning of virtual objects relative to the real environment, enhancing the overall MR experience.

3.2.4. Registration and Mapping. Registration and mapping represent other essential technologies in the realm of Mixed Reality (MR), enabling the seamless integration of virtual and actual elements. This results in the precise interaction between virtual objects and the physical environment. As an illustration, in a specific study, the ICP algorithm was leveraged to superimpose a virtual representation of a historical building onto the physical space based on the user's location [19]. This technique tackled the complexity of mapping and registering extensive data by employing key user-selected objects from a scaled-down virtual model. Likewise, in another work, a novel method was introduced involving mathematical transformations such as perspective, affine, and isometric transformations, for aligning a virtual cube with the real world [20]. This approach enhanced the calibration accuracy with HoloLens, ultimately leading to improved precision.

3.2.5. Visualization and Rendering. The display of MR content necessitates the use of advanced analytical algorithms to tackle various challenges. These challenges encompass enriching video images by adding virtual files, achieving a close-to-reality representation of virtual objects with lighting and shadows, managing the level of detail concerning the user's position, addressing occlusion and screen off issues for invisible objects, employing standard schemas for faster display, handling bulky 3D data like point clouds, and synthesizing information from different display sources. For instance, in a specific study, the immersion of virtual objects into video images was achieved using chroma-keying with PCA, a technique useful for foreground identification through separating images into back and foreground components [21]. In another study, the application of spherical harmonics light probe estimation facilitated light calculation via deep learning, enabling processing of raw images without prior scene knowledge [22]. Furthermore, an innovative lighting model was introduced in a different work, incorporating a refractive object within the instant radiosity (DIR) model [23].

3.3. Current Development and Applications

3.3.1. Trends and Application. As an emerging technology that can create a virtual environment in real scenarios, MR has developed rapidly in recent decades and is widely used in various scenarios. A discernible increase in the volume of MR publications over recent decades has been evident, signifying a noteworthy surge in attention and interest towards this technology [23]. Additionally, it highlights that the realm of applications has experienced considerable expansion, followed by user and technical subjects, underscoring the extensive integration of MR across diverse domains [24].

For example, in the realm of application, researchers have explored and proposed various methods to enhance users' usage and interactive experiences in the current stage. For instance, they have created a small room equipped with multiple depth cameras and projectors to investigate various interaction and computation strategies associated with spaces inhabited by interactive displays. They introduced a novel prototype featuring a combination of multiple depth cameras and projectors, enabling ordinary walls and tables to become interactive surfaces, allowing users to interact within and between room-sized environments on interactive surfaces above them [24]. Additionally, smart glasses or head-mounted displays (HMDs) have been gaining popularity as the next-generation mainstream wearable devices [25]. Given the compact wearable platform of HMD systems, novel interface modalities are required. Multimodal fusion techniques empower users to interact with computers using various input modalities such as speech, gesture, and eye gaze [26].

In other realms, MR has also played an important role. In the field of education, researchers have explored and confirmed the usability of MR, especially in science centers and museums, where applying MR to educational exhibits garnered positive feedback [27]. Moreover, in classrooms, MR has transformed learning sessions into interactive and engaging "TV-show style games," significantly boosting student motivation [28]. In the medical field, MR offers advantages over haptic interfaces and robotic-assisted surgery, as it provides a touchless interface, ideal for the sterile environment of the operating room. With no physical contact required, surgeons can still maintain necessary control features [29].

4. Current Challenges and Future Studies

Despite significant advancements and widespread applications of MR technology in various fields, there are still some shortcomings that hinder its further development and extensive use.

In terms of algorithms, there is a need for improvements in large-space modelling, involving the consideration of sharing between different coordinate systems using standard schema-like ontologies. Additionally, an authoring system is required to enable seamless integration of raster and vector data in 2D and 3D, along with appropriate tracking systems to accurately determine the status of objects through sensor fusion [29]. Moreover, object recognition faces challenges with spatial object detection, leading to unwanted errors due to tracking tools like the global positioning system (GPS). To address this issue, image-processing techniques are necessary to extract contour data, followed by a comparison of the extracted data with maps to correct errors [16]. In the aspect of MR mobile apps, more research are needed on new computer graphic algorithms for automatic environment construction and realistic data visualization. Large-scale MR should include security mechanisms for diverse information levels (organizations, social network data, virtual objects, users, and environmental sensors). Enriching MR content with IoT demands new architectures for seamless integration. Future MR systems should prioritize interface automation, ensuring user adaptability [5].

Furthermore, the current stage of MR hardware is relatively large and expensive, posing another obstacle to its progress. To overcome this limitation, researchers are focusing on mobile devices and applications, which are currently the most popular trends in the MR research field. However, the computational power of mobile devices needs improvement, and there is a requirement for a better balance between performance and capacity to enhance the information provided [24].

5. Conclusion

This essay has preliminarily reviewed and discussed the four different primary definitions of MR, the main technologies and currently applied methods in MR, the trends of research, the challenges and research in the future.

Through our retrospective analysis, the evolution of MR technology has been presented, shedding light on its transformative journey. However, it is essential to acknowledge the limitations of this essay. Some topics may have lacked in-depth exploration, and the information presented may not fully represent the cutting-edge research being conducted today.

Despite these limitations, this essay serves as a valuable starting point for comprehending the multidimensional landscape of MR. By addressing challenges and opportunities, we hope to inspire future researchers to contribute innovative ideas and solutions to the advancement of MR technology.

In conclusion, MR is an ever-evolving field with vast potential, and continued exploration and collaboration will undoubtedly shape a future where immersive experiences become an integral part of our daily lives.

References

[1] Hashimoto, S.; Tan, J.K.; Ishikawa, S. (2007) A method of creating a transparent space based on mixed reality. Artif. Life Robot. 11: 57–60.

[2] Stretton, T.; Cochrane, T.; Narayan, V. (2018) Exploring mobile mixed reality in healthcare higher education: A systematic review. Res. Learn. Technol. 26: 21-31.

[3] Speicher, M., Hall, B.D., Nebeling, M. (2019) What is Mixed Reality? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 537, 1–15.

[4] Skarbez, R.; Smith, M.; Whitton, M. C. (2021) Revisiting Milgram and Kishino's Reality-Virtuality Continuum. Frontiers in Virtual Reality, 2: 10.3389

[5] Rokhsaritalemi, S., Sadeghi-Niaraki, A., Choi, S.-M. (2020) A Review on Mixed Reality: Current Trends, Challenges and Prospects. Applied Sciences, 10(2), 636.

[6] Chen, Z.; Wang, Y.; Sun, T.; Gao, X.; Chen, W.; Pan, Z.; Qu, H.; Wu, Y. (2017) Exploring the design space of immersive urban analytics. Vis. Inform., 1: 132–142.

[7] Potemin, I.S.; Zhdanov, A.; Bogdanov, N.; Zhdanov, D.; Livshits, I.; Wang, Y. (2018) Analysis of the visual perception conflicts in designing mixed reality systems. In: Proceedings of the Optical Design and Testing VIII, International Society for Optics and Photonics, Beijing, China, 5 November 2018.

[8] Claydon, M. (2015) Alternative Realities: From Augmented Reality to Mobile Mixed Reality. In: University of Tampere (Eds.), Master’s Thesis. Tampere, Finland.

[9] Zhang, Z.; Li, Y.; Guo, J.; Weng, D.; Liu, Y.; Wang, Y. (2018) Task-driven latent active correction for physics-inspired input method in near-field mixed reality applications. J. Soc. Inf. Disp., 26: 496–509.

[10] Nakano, M.; Li, S.; Chiba, N. (2007) Calibration of fish-eye camera for acquisition of spherical image. Syst. Comput. Jpn., 38: 10–20.

[11] Steed, A.; MacColl, J.; Randell, C.; Brown, B.; Chalmers, M.; Greenhalgh, C. (2004) Models of space in a mixed-reality system. In: Proceedings of the Eighth International Conference on Information Visualisation, IV 2004, London, UK, 14–16 July 2004; IEEE: Piscataway, NJ, USA.

[12] Xing, H.; Bao, X.; Zhang, F.; Gai, W.; Qi, M.; Liu, J.; Shi, Y.; De Melo, G.; Yang, C.; Meng, X. (2019) Rotbav: A Toolkit for Constructing Mixed Reality Apps with Real-Time Roaming in Large Indoor Physical Spaces. In: Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; IEEE: Piscataway,

NJ, USA.

[13] Nagata, K.; Masuko, S.; Hoshino, J. (2018) Construction of Mixed Reality Story Environment Based on Real Space Shape. In: International Conference on Entertainment Computing; Springer: Berlin/Heidelberg, Germany.

[14] Kakuta, T.; Vinh, L.B.; Kawakami, R.; Oishi, T.; Ikeuchi, K. (2008) Detection of moving objects and cast shadows using a spherical vision camera for outdoor mixed reality. In: Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology, New York, NY, USA, October 2008; pp. 219–222.

[15] Chen, L.; Tang, W.; John, N.; Wan, T.R.; Zhang, J.J. (2018) Context-Aware Mixed Reality: A Framework for Ubiquitous Interaction. arXiv:1803.05541.

[16] Kim, D.-G.; Jang, Y.-H.; Kwon, Y.-J. (2012) Ambient Browser: Mobile Mixed Reality System Based on Ambient Intelligence. Int. J. Future Comput. Commun., 1: 160.

[17] Costanza, E.; Kunz, A.; Fjeld, M. (2009) Mixed reality: A survey. In: Human Machine Interaction; Springer: Berlin/Heidelberg, Germany, pp. 47–68.

[18] Riexinger, G.; Kluth, A.; Olbrich, M.; Braun, J.-D.; Bauernhansl, T. (2018) Mixed Reality for on-site self-instruction and self-inspection with Building Information Models. Procedia CIRP, 72, 1124–1129.

[19] Brito, C.; Alves, N.; Magalhães, L.; Guevara, M. (2019) Bim mixed reality tool for the inspection of heritage buildings. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., 4, 25–29.

[20] Azimi, E.; Qian, L.; Navab, N.; Kazanzides, P. (2017) Alignment of the Virtual Scene to the Tracking Space of a Mixed Reality Head-Mounted Display. arXiv:1703.05834.

[21] Beato, N.; Zhang, Y.; Colbert, M.; Yamazawa, K.; Hughes, C.E. (2009) Interactive chroma keying for mixed reality. Comput. Anim. Virtual Worlds, 20, 405–415.

[22] Marques, B.A.D.; Clua, E.W.G.; Vasconcelos, C.N. (2018) Deep spherical harmonics light probe estimator for mixed reality games. Comput. Graph., 76, 96–106.

[23] Knecht, M.; Traxler, C.; Winklhofer, C.; Wimmer, M. (2013) Reflective and refractive objects for mixed reality. IEEE Trans. Vis. Comput. Graph., 19, 576–582.

[24] Chen, S.S.-C., Duh, H. (2019) Interface of mixed reality: from the past to the future. CCF Transactions on Pervasive Computing and Interaction, 1(1), 69–87. .

[25] Hong, I., et al. (2015) 18.1 A 2.71 nJ/pixel 3D-stacked gaze-activated object-recognition system for low-power mobile HMD applications. In: 2015 IEEE International Solid-State Circuits Conference-(ISSCC), pp. 1–3. IEEE.

[26] Ismail, A.W., Sunar, M.S. (2015) Multimodal fusion: gesture and speech input in augmented reality environment. In: Computational Intelligence in Information Systems, pp. 245–254. Springer.

[27] Woods, E., et al. (2004) Augmenting the science centre and museum experience. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 230–236. ACM.

[28] Freitas, R., Campos, P. (2008) SMART: a SysteM of Augmented Reality for Teaching 2nd grade students. In: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, vol. 2, pp. 27–30. BCS Learning and Development Ltd.

[29] Ruppert, G.C.S., Reis, L.O., Amorim, P.H.J., de Moraes, T.F., da Silva, J.V.L. (2012) Touchless gesture user interface for interactive image visualization in urological surgery. World J. Urol. 30(5), 687–691.