

PAPER • OPEN ACCESS

A review on multimodal interaction in Mixed Reality Environment

To cite this article: Aladin Mohamad Yahya Fekri and Ismail Ajune Wanis 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012049

View the [article online](#) for updates and enhancements.

Recent citations

- [Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review](#)
Joo Chan Kim *et al*
- [MR-Deco: Mixed Reality Application for Interior Planning and Designing](#)
F E Fadzli *et al*
- [3D telepresence for remote collaboration in extended reality \(xR\) application](#)
F E Fadzli *et al*



240th ECS Meeting ORLANDO, FL

Orange County Convention Center **Oct 10-14, 2021**

Abstract submission due: April 9

SUBMIT NOW

A review on multimodal interaction in Mixed Reality Environment

Aladin Mohamad Yahya Fekri^{1,2}, and **Ismail Ajune Wanis**^{1,2}

¹Mixed and Virtual Environment Research Lab (mivielab), Vicubelab, Universiti Teknologi Malaysia, 81310 Johore, Malaysia

²School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

E-mail: yahyafekri@gmail.com, ajune@utm.my

Abstract. Mixed Reality (MR) environment proposes to blend together virtual and real environments where both world coexist and can respond to each other. As MR has the understanding of real environment, it is possible for user to use their surrounding as a means to interact with the virtual content. The user interaction in Augmented Reality (AR) and Virtual Reality (VR) are widely explored however very less in MR. Interaction must be intuitive as possible to produce natural user interaction. Multimodal interaction is an advanced interaction techniques when involve the combination of more than one input modalities such as speech with gesture, gaze and gesture, and gaze with speech. The inputs or sensory however are not restricted to just two modalities. Therefore this paper presents a review on multimodal interaction in MR environments.

1. Introduction

In the rapid advancements in computer vision, display technology, input systems and graphics processing, more researchers and large companies had their interest in Mixed Reality (MR). The term MR was originally discovered by Paul Milgram and Fumio Kishino [1], as a virtuality continuum where it extends from real world to completely virtual environments with Augmented Reality (AR) and Augmented Virtuality (AV) lies in between. AR lies close to real world continuum and it is an experience that overlay virtual imaginary onto the real world [2]. Virtual Reality (VR) on the hand is an experience that totally immerse users inside virtual imaginary world and cut off user's view from real world.

The emerging of many upgraded and new input systems invites more urges to make a natural user interaction. Interaction in AR usually obeys the VR principle. In MR however it looks different. For example, user can defend themselves from virtual enemy attack by using a physical book or bottle to act as a shield and use the same physical book to smash and damage the enemy. Multimodal interaction by combining several input modalities in order to achieve more natural user interaction in their application. As Bunt [3] and Quek [4] agreed that the interaction between this world and human is naturally multimodal. As to apply multimodal interaction in MR, Ismail [5] stated that it is a solution for more advanced interaction technique using gesture with speech inputs in AR. MR devices such as Microsoft HoloLens and Magic Leap did not cut users from the physical world, unlike immersive devices such as Oculus Rift and HTC Vive where these devices uses an opaque headset that



blocks out the physical environment from user's view. Despite the differences, these available headset however offer user with full 6 Degree of Freedom (DOF), but both in rotational and translation movement within their respective environment. The user interaction in AR and VR are widely explored but very less in MR. Therefore, this paper presents the previous works on tangible interaction and multimodal interaction. The paper gives a review on multimodal interaction in MR environment.

2. User Interaction Techniques

There are many interaction techniques available and some of them are (1) Tangible User Interface (TUIs) and (2) Multimodal Interaction.

2.1 Tangible Interaction Techniques

Tangible interface invites different interaction techniques with real tangible objects are acting as a tool to perform 3D object manipulation. There several examples to present tangible interaction methods such as the Magic-Lens [6], Cubical Interface (CUI) [7], and the Magic-Cup [8]. As in Fig. 1 (a), the Magic Lens has used a physical handle to portray as a virtual lens. The interaction by moving the both hands to move the lens so it works to zoom or scaling. While the Magic-Cup (as in Fig. 1 (b)), uses the method of "covering" virtual objects with a real shape of cup, when the shape can actually hold a virtual object. A cube-based tangible AR interface (as in Fig. 1 (c)) when a user uses their both hands to manipulate and compose virtual objects. Fig. 1 (d) shows the EditDunia research project on Ancient Malacca [9] which has executed the paddle-based interaction method. The paddle is a single marker attached as a tool for user to perform virtual object manipulation and move the object around the ground marker.

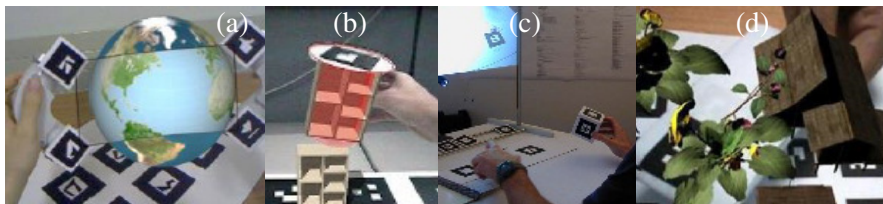


Figure 1. (a) CUI, (b) MagicLens, (c) MagicCup and (d) EditDunia, are the cases of tangible interface.

2.2 Multimodal Interaction Techniques

Multimodal interaction is an advanced interaction techniques when involve the combination of more than one input modalities such as speech with gesture, gaze and gesture, and gaze with speech. Multimodal interaction in AR has a great potential to enhance the user interaction and it is promising when more than input modalities have been invoked in the AR interface. For instance, G-SIAR [10] has introduced a natural interaction technique by two-handed free-hand gesture for object manipulation. It invites gesture and speech to execute a direct manipulation interaction technique to dynamically manipulate virtual content. As proposed by [9], the extended version of Ancient Malacca project [5], they also have created an advanced method to improve the user interaction. They have employed inputs modality, gesture and speech input to perform 6DOF virtual object manipulation. Compare to G-SIAR, [9] have advanced an intuitive and natural interaction by utilizing the sensor-based hand tracking when the motion data drawn as skeleton hands to represent the gesture input.

3. Multimodal Interactions in MR

Interaction methods in MR are based on the methods people have implemented in AR environment, researchers and experts are considering AR subsets to MR technology. Therefore, the methods in AR have been thoroughly studied recently the experts have begun to explore multimodal interaction in

MR. Table 1 shows the previous works with regards to multimodal interaction in MR spectrum. It describes and gives an overview of what the system is all about and the input modalities that are used.

Table 1. Previous works multimodal interaction in MR Spectrum

| Year | Title | Input modalities |
|------|--|---|
| 2014 | Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in AR [10] | Speech, Gesture |
| 2015 | Vision-Based Technique and Issues for Multimodal Interaction in AR [11] | Facial Recognition, Gesture, Speech |
| 2015 | Multimodal Fusion: Gesture and Speech Input in AR Environment [9] | Speech, Gesture |
| 2015 | ARZombie: A Mobile AR Game with Multimodal Interaction [12] | Facial Recognition, Touchscreen |
| 2016 | Simplified Industrial Robot Programming: Effects of Errors on Multimodal Interaction in WoZ experiment [13] | Touch-Sensitive Table, Gesture, 6D Pointing Device, Touchscreen, Direct Robot Arm |
| 2016 | An Intelligent Multimodal Mixed Reality Real-Time Strategy Game [14] | Speech, Gesture |
| 2016 | Hybrid Team Interaction in the MR Continuum [15] | Eye, Head, Hand, Gesture |
| 2017 | Multimodal Interaction in AR [16] | Speech, Gesture |
| 2018 | Wall mounted level: a cooperative MR game about reconciliation [17] | Digital Input Device, Physical Touch |
| 2018 | Designing an Augmented Reality Multimodal Interface for 6DOF Manipulation Techniques [5] | Speech, Gesture |

As presented in Table 1, in 2014 G-SIAR [10] has introduced natural interaction techniques which combines speech and gesture. These techniques support 6DOF movement, uniform scaling, as well as physics-based interaction. It be seen in 2018, [5] has proposed new multimodal interaction using gesture and speech that invites 6DOF with sensor-based interaction. In 2015, [12] has developed an AR game that uses facial recognition and detection to recognise players face and superimposed a 3D zombie face on top of player's face. Next in 2016, [14] has produced a role-playing MR table top game and the players give instructions via speech-gesture interaction. The game was to combine interaction styles with gameplay mechanics in order to control the character. MR game [17] was developed in 2018 which is a cooperative game that powers multimodal interactions to support its chronicle of 'reconciliation'. However in 2016, [13] found that gesture and 6D pointing are the fastest modalities in many interaction metaphors for robot-human. A robot-human system has been proposed by [15], it enable collaboration between a robot, a human, and a remote human where the remote human are connected via VR. This system leverage eyes, head, hand and gesture tracking as a mean to interact and communicate with and within the system.

4. Conclusion

Across the MR spectrum, each of the technology whether be AR, VR or MR, all of them has their own respective advantages and roles in enhancing our physical reality and making this world a better place to live in. In today's rapid technological advancement, more new devices will be announced and will push the limits and expand the range in MR spectrum. Interaction also need to be as intuitive as possible and natural that permits users to interact naturally and multimodality without the needs to wear the devices. It can also be seen that researchers nowadays are making a collaborative system to

invites more users to use the system simultaneously whether it is currently present user or a remote user or even with a robot. Multimodal interaction is an advanced interaction techniques when involve the combination of more than one input modalities such as speech with gesture, gaze and gesture, and gaze with speech. The inputs or sensory however are not restricted to just two modalities. Interaction is one of the biggest issues as interaction is the component that allows user to engage and express themselves with the system. Lacking in interaction metaphor can make the system to be less intuitive and need more improvements. Hence this paper presents the works in multimodal interaction for MR environments. Based on this chronological timeline the experts and researchers are started to explore the remote collaborative works in MR while significantly grasping the multimodal interaction for MR.

Acknowledgments

We would like to express our appreciation to Mixed and Virtual Reality Laboratory (miviellab) at Universiti Teknologi Malaysia (UTM). This work was funded by UTM-HIR High Impact Research Grants Scheme (Q.J130000.2451.04G77).

References

- [1] Milgram P *et al* 1994 *IEICE TRANS. Info. and Sys.* **77** 1321.
- [2] Akçayır M *et al* 2017 *Edu. Res. Rev.* **20** 1.
- [3] Bunt H *et al* 1998 *Springer Science & Bus. Media* **1374**.
- [4] Quek F *et al* 2002 *ACM Trans. on Cop. Hum. Inter. (TOCHI)* **9** 171.
- [5] Ismail A W *et al* 2018 *In Pro. SAI Intel.Sys. Conf.* 309.
- [6] Lee H *et al* 2011 *Virtual Reality* **15** 133..
- [7] Looser J *et al* 2004 *ACM 2nd Int. Conf. Comp. Grap. & Inter. Tech.* 204.
- [8] Kato H *et al* 2003 *IEEE Int. In Augm. Rea. Toolkit Work.* 75.
- [9] Ismail A W & Sunar M S 2015 *Spr. In Comp. Intel. Inf. Sys.* 245.
- [10] Piumsomboon T *et al* 2014 *IEEE Int. Sym. Mix. & Augm. Rea. (ISMAR)* 73.
- [11] Ismail A W *et al* 2015 *ACM The 8th Int. Sym. Vis. Inf. Comm. & Inter.* 75.
- [12] Cordeiro D *et al* 2015 *IEEE In 7th Int. Conf. on Inter. Entert. (INTETAIN)* 22.
- [13] Materna Z *et al* 2016 *In RO-MAN* 200.
- [14] Link S *et al* 2016 *IEEE In Vir. Real.* 223.
- [15] Moniri M M *et al* 2016 *ACM 22nd Conf. on Vir. Real.Soft. & Tec.* 335.
- [16] Chen Z *et al* 2017 *IEEE Int. Conf. on Sys. Man. & Cyb.* 206.
- [17] Swearingen K L *et al* 2018 *ACM SIGGRAPH Posters* 20.