WHEN MIXED REALITY MEETS INTERNET OF THINGS:

Toward the Realization of Ubiquitous Mixed Reality

From panoramic paintings and stereoscopic photos in the early 19th century, there has been a century-long effort to realize mixed reality, interweaving real and virtual worlds that interact with each other. Recently, over the past few years, we have witnessed the first wave of "affordable" mixed reality platforms, such as Oculus Rift and Microsoft HoloLens hitting the market. In particular, 2017 was the showcase year of mixed reality technologies: The Academy awarded its first Oscar to virtual reality storytelling[1]; AAA caliber virtual reality games started to hit the market with impacts[2]. Furthermore, major mobile operating systems, including Android and iOS, began to support augmented reality at the platform level (e.g., Android ARCore, Apple ARKit). Looking down the road, a recent forecast by Orbis Research projects over $40 billion mixed reality market worldwide by 2020[3].

Photo, istockphoto.com

Taiwoo Park and Mi Zhang *Michigan State University*
Youngki Lee *Singapore Management University*

Editor: Eduardo Cuervo

FIGURE 1. Still from *HYPER-REALITY* (2016) by Keiichi Matsuda.

Although mixed reality is one of the most intriguing technologies hitting the market today, there are still pessimistic voices against the rosy expectations of the current generation of mixed reality devices. Device users and researchers point out a series of challenges on social acceptance, digital fatigue, heavy and bulky headsets, and most importantly, lack of genuinely ubiquitous mixed reality experiences[4].

In this article, we focus on the vision of enabling genuinely ubiquitous mixed reality experiences. We argue that the next generation of mixed reality devices needs to incorporate the capability to closely interact with objects in a user's surrounding environment. As a potential approach to enable such capability, we propose to leverage the Internet of Things (IoT) technologies that integrate sensors and connectivity into everyday objects, such as toys, home appliances, and shopping carts. These sensors capture a user's interaction with the objects in her vicinity, expanding the capability of mixed reality platforms beyond tracking users' head and whole-body movements.

Given its unique capability of interacting with everyday objects, ubiquitous mixed reality can enable new applications. For example, surgeons can be trained interactively by connecting mixed reality headsets with sensor-rich surgical instruments [16]. Sports players would experience a new level of fun and efficiency as the sensors embedded in the sports equipment can track the players' movements and activities, and provide real-time feedback. *HYPER-REALITY* (2016) by Keiichi Matsuda [17] has proposed promising uses of mixed reality in various application domains, including shopping and advertising (see Figure 1).

In this article, we describe the challenges of realizing the vision of ubiquitous mixed reality. We then present our ongoing work on developing a virtual fitness coach and its supporting platform as a concrete example of ubiquitous mixed reality technology.

[1] https://www.polygon.com/virtual-reality/2017/10/27/16561624/virtual-reality-oscar-carne-y-arena
[2] https://www.vrandfun.com/master-list-of-all-of-the-upcoming-aaa-caliber-vr-games/
[3] https://www.reuters.com/brandfeatures/venture-capital/article?id=4975
[4] https://medium.com/the-mission/the-6-biggest-challenges-facing-augmented-reality-8d48c470286d

CHALLENGES OF REALIZING UBIQUITOUS MIXED REALITY

Infrastructure-Free Activity and Context Tracking

To achieve a high level of practicality, we envision that ubiquitous mixed reality should be enabled without any or only minimal infrastructure support. The infrastructure-free systems will quickly spread ubiquitous mixed reality by breaking the vicious cycle of limited infrastructure caused by limited users. Also, it will allow users to bring the systems anywhere to enjoy the mixed reality experience.

In the current mixed reality systems, the need of infrastructure support (e.g., lighthouses and cameras) acts as a major hurdle in implementing the vision of ubiquitous interaction platform due to the high initial installation cost and social barriers. For instance, many existing VR platforms fail to meet infrastructure-free interaction since they rely on additional sensing devices, such as lighthouses of HTC VIVE for laser beam spreading [11] and cameras of Oculus Rift to track infrared LED patterns of headsets and controllers [12].

Recently, Microsoft HoloLens and Windows mixed reality headsets implemented infrastructure-free, inside-out tracking enabled by a computer vision technology [13] that tracks environmental features with stereo cameras. Additionally, by employing similar technology to Oculus' constellation tracking, the Windows mixed reality headset provides an additional hand controller tracking feature [14]. However, its sensing performance is significantly affected by surrounding context, such as an environmental feature (e.g., the system would fail in an empty room) and low light. Also, it can only track objects with light markers (e.g., handheld controllers) within the camera's field-of-view. The latest review reported that infrastructure-free, inside-out tracking revealed a significant amount of compromise in terms of user experience while interacting with the surrounding environment [15]. Moreover, these mixed reality devices do not support context and activity monitoring to automatically provide timely mixed reality applications, as shown in the example of Figure 1.

Tight Accuracy and Latency Constraints

Recognizing surrounding context with high accuracy and low latency is another important requirement of ubiquitous mixed reality [3]. For typical examples of ubiquitous services, such as quantified self and smart environments, there are rooms of designing meaningful feedback with a medium level of context recognition accuracy and latency. For example, coarse-grained activity summary (e.g., amount and intensity levels of physical movement and/or workout) and energy usage summary for buildings could be provided based on coarse-grained user location and activity information. However, ubiquitous mixed reality should provide high accuracy with low latency in context monitoring, otherwise seriously degrading users' perceived interaction and fidelity [9].

In terms of object and motion tracking for mixed reality, there is a series of findings on the effect of visual-inertial lags and delays, including reduced realism [8], task performance [5] and cyber sickness [20]. As well, quick calculation examples of combined latency in regards to the expected head movement and rotation rate at 50 degrees per second suggests that 100 ms of latency causes 5 degrees of angular error [19]. Kennedy et al. have shared their findings from military simulator settings in [18], considering 5 to 50 ms of delay as "eliminated" cue asynchrony, which is the delay between visual and inertial stimuli, and their empirical delay limit of 40 ms so as not to get complaints from subjects. Regarding the precision, the recent measurement of the tracking performance of HTC VIVE reported its performance with accuracy and precision of RMS 1.5 mm and 1.9 mm respectively [6], with 22 ms latency in average [7], which will play a role of state-of-the-art mixed reality tracking technology in terms of tracking precision, accuracy, and latency.

JARVIS: A Ubiquitous Mixed Reality Platform for Interactive Fitness Coaching

As a concrete example of ubiquitous mixed reality technology, we have been developing JARVIS, a ubiquitous mixed reality platform for interactive fitness coaching, where the needs of real-time workout monitoring
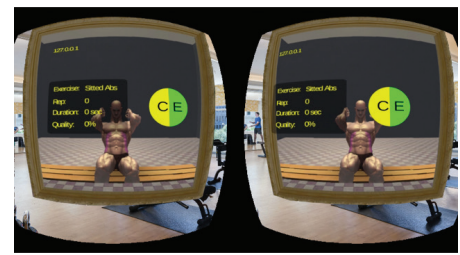
FIGURE 2. Example usage: (left) a user trying JARVIS (b) stereoscopic screen of the virtual coach. It supports either augmented or virtual display mode.

and interactive guidance converge. Today, working out in a gym has become an important part of people's modern lives. However, working out on the stationary exercise machines in the gym can quickly induce boredom. Moreover, without the help of a professional trainer, novice exercisers are hardly aware if they are using the right set of muscles or if their intensity of exercise are adequate. This prevents exercisers from making steady progress and eventually makes them lose their interest and motivation.

JARVIS fills this gap by equipping novice exercisers with the knowledge and skills to effectively build up their training routine. Specifically, JARVIS features dual-mode sensing mechanism: human-wearable and object-attachable. In its casual usage, a miniature IoT device is worn on a user's body as a form of smartwatch or cloth clip. When in need of precise motion tracking of a specific object of interest, the miniature IoT device can be attached to the object and works as a sensing device as part of a distributed interaction platform.

Here is an envisioned use scenario of JARVIS:

David has been thinking of building up his muscles using exercise machines in the gym. However, he failed to adhere to regular strength training schedules. He once worked with personal trainers, but they could not be with him all the time. Most of the existing fitness logging mobile apps require him to manually log exercise records. Although there are a few automatic fitness monitoring mobile apps, they could only summarize the exercises after the exercise session is finished, which was not very helpful for him during the training.
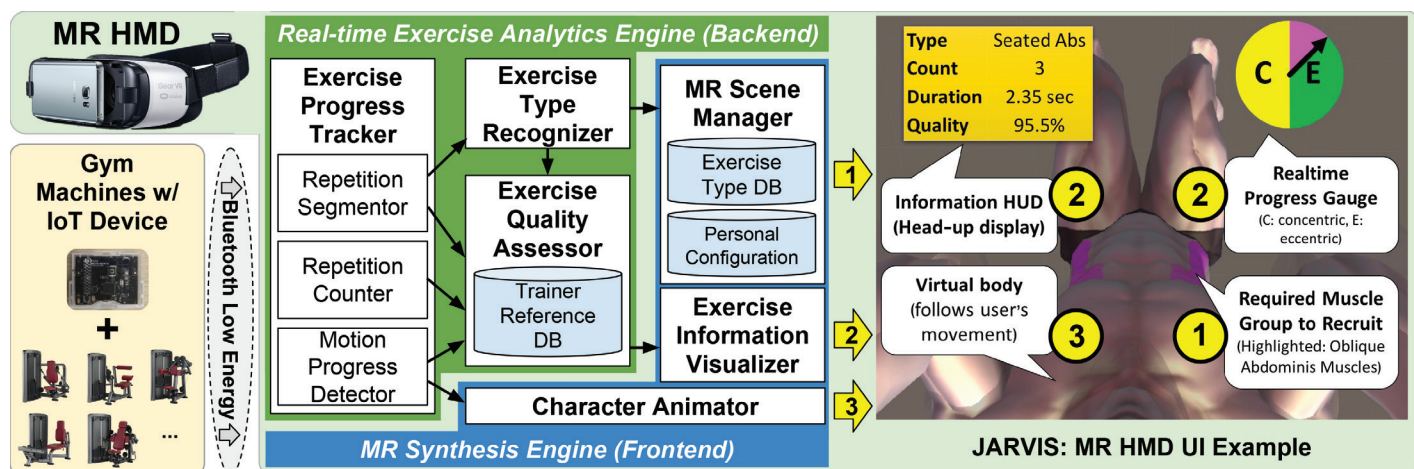
FIGURE 3. JARVIS system architecture for virtual fitness coach application.

David decides to try the JARVIS virtual fitness coach. He brings a mobile mixed reality head-mounted display (HMD) and a miniature IoT device with onboard sensors to any exercise machine at a gym or at a community park. He attaches the IoT device to the biceps curl machine and wears the HMD. After a few trial repetitions, JARVIS detects the type of exercise machine, and then displays a virtual exercise environment customized to the recognized biceps curl machine. JARVIS also provides an immersive and interactive mixed reality environment showing an "augmented David" following his exercise movements and provides real-time exercise progress information. Furthermore, JARVIS highlights the required muscle groups corresponding to the exercise being performed and David's personal goal of strength training. This helps David focus easily on the required muscle groups. Once he finishes the biceps curl training, JARVIS turns off the mixed reality exercise scene to let David navigate to the next exercise machine. With JARVIS, David can go to the gym with automatically generated training plans and an automated real-time immersive mixed reality exercise environment that makes him more confident in his strength training.

---

[5] https://www.theodysseyonline.com/mobile-vr-cant-take-heat

A number of unique technical and usability challenges arise to fully realize the envisioned virtual fitness coaching application. First, the application needs to accurately and quickly recognize the exercise type from the streaming sensor data, so that a virtual exercise environment customized to the current exercise type can be displayed in the HMD in real time to enable highly immersive and interactive user experiences. Second, the application needs to create smooth mixed reality scenes while preventing the HMD from being overheated. Otherwise, the overheated HMD can lead to negative user experiences[5]. Third, the application needs a user interface that can provide exercisers with immersive and informative coaching experiences.

JARVIS addresses these challenges in several ways. First, JARVIS utilizes miniature IoT devices that can be attached to exercise machines and use the onboard sensors to capture machine exercise information. Compared to the approach of putting sensors on the human body where non-exercise body movements are also captured, this machine-attachable approach only captures exercise-related movements. As a result, JARVIS is able to achieve high exercise-repetition segmentation and exercise-type recognition accuracies in real time, so as to meet the tight accuracy and latency constraints for ubiquitous mixed reality. Second, JARVIS employs a smart visual quality guide to adjust the trade-off

between visual quality and computation workload to reduce energy consumption and heat generated by HMD. Third, JARVIS incorporates an immersive and informative mixed reality user interface, which highlights the recommended muscle groups corresponding to the current exercise type to maximize the efficacy of the exercise (see Figure 3). In the current implementation, it achieves mixed reality through mixing a rendered avatar model with an external camera image of a mobile HMD (e.g., Samsung Gear VR).

Our evaluation results look promising: JARVIS segments the exercise repetitions in real time with an average accuracy of 97.96%. It recognizes exercise type with an average accuracy of 99.08% within 1.22 repetitions. This performance is comparable to or outperforms state-of-the-art mobile interaction [1, 2] and strength training monitoring platforms [10]. We have also conducted a field study with more than 10 exercisers. Our result shows that JARVIS can significantly enhance user enjoyment and perceived competence for strength training. Finally, with the aforementioned muscle group highlighting technique, we quantitatively analyze muscle activities of exercisers using high-fidelity clinical surface Electromyography (sEMG) sensors. Our result indicates that JARVIS provides a significant difference ($p < .01$; $p < .05$) in terms of sEMG muscle activation signals, implying JARVIS's significant potential of effective coaching for strength training [4].

CONCLUSION AND OPEN QUESTIONS

In this article, we present ubiquitous mixed reality, our envisioned next generation mixed reality technology that enables truly ubiquitous mixed reality experiences. Ideally, ubiquitous mixed reality provides infrastructure-free body and surrounding object tracking with high precision and low latency. It also provides quick, rich context and activity recognition to offer context-aware services in a timely manner. As a proof-of-concept design, we propose a combination of a mixed reality device and a dual-mode IoT device as an affordable set of devices, which can enable rich context sensing for high-fidelity mixed reality experiences, as well as the design of our interactive ubiquitous mixed reality platform, JARVIS. It demonstrated the performance of activity recognition and tracking in real time, as well as the potential of usability and efficacy of ubiquitous mixed reality application through a case of fitness coach application based on JARVIS.

Still, there is a long road to realizing ubiquitous mixed reality, and a number of questions arise. Would JARVIS be able to support significantly a large body of everyday life context? Will the proposed real-time activity recognition technology work for other situations without significant revision? We are currently working on other scenarios with JARVIS, such as training and education, to check whether its dual-mode IoT will be a feasible option.

We also acknowledge that our approach will not be the only solution to the envisioned ubiquitous mixed reality. Advances of computer vision with deep learning recently revealed a huge potential in recognizing human activities and actions with impressive recognition accuracy [21], and we believe that camera-based recognition and tracking technology will be a great complement to IoT in realizing ubiquitous mixed reality. In fact, who knows if we will wear mixed reality glasses with more than 10 cameras for body and surrounding object tracking in 2025? Privacy and social acceptance will still be the big hurdles. ∎

Taiwoo Park is an assistant professor at Michigan State University. He received his PhD, MS, and BS from Korea Advanced Institute of Science and Technology (KAIST). His research interest includes innovative, creative, and persuasive use of mobile, ubiquitous, and interactive technologies.

Mi Zhang is an an assistant professor at Michigan State University. His research is at the intersection of mobile computing, embedded sensor systems, and machine intelligence, with applications in health care and well-being. He received his PhD from the University of Southern California and BS from Peking University.

Youngki Lee is an assistant professor at Singapore Management University. He has broad research interests in building innovative software systems, which covers a wide design spectrum across systems, applications, and users. More specifically, his research interest lies in building mobile and sensor systems to enable accurate monitoring of human behaviour and contexts.

REFERENCES

[1] Liu, J., Zhong, L., Wickramasuriya, J., & Vasudevan, V. (2009). "uWave: Accelerometer-based personalized gesture recognition and its applications." *Pervasive and Mobile Computing*, 5(6), 657-675.

[2] Park, T., Lee, J., Hwang, I., Yoo, C., Nachman, L., & Song, J. (2011, November). "E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices." *In Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems* (pp. 260-273). ACM.

[3] MacIntyre, B., & Feiner, S. (1996). "Future multimedia user interfaces." *Multimedia Systems*, 4(5), 250-268.

[4] G. M. Karst and G. M. Willett. "Effects of specific exercise instructions on abdominal muscle activity during trunk curl exercises." *The Journal of Orthopaedic and Sports Physical Therapy*, 34(February 2001):4–12, 2004.

[5] Frank, L. H., Casali, J. G., & Wierwille, W. W. (1988). "Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator." *Human Factors*, 30(2), 201-217.

[6] "Analysis of Valve's 'Lighthouse' Tracking System Reveals Accuracy." https://www.roadtovr.com/analysis-of-valves-lighthouse-tracking-system-reveals-accuracy/

[7] Niehorster, D. C., Li, L., & Lappe, M. (2017). "The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research." *i-Perception*, 8(3), 2041669517708205.

[8] Ellis, S. R., Adelstein, B. D., Baumeler, S., Jense, G. J., & Jacoby, R. H. (1999, March). "Sensor spatial distortion, visual latency, and update rate effects on 3D tracking in virtual environments." In Virtual Reality, 1999. Proceedings., *IEEE* (pp. 218-221).

[9] Nielsen, J., & Levy, J. (1994). "Measuring usability: preference vs. performance." *Communications of the ACM*, 37(4), 66-75.

[10] Morris, D., Saponas, T. S., Guillory, A., & Kelner, I. (2014, April). "RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3225-3234). ACM.

[11] Niehorster, D. C., Li, L., & Lappe, M. (2017). "The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research." *i-Perception*, 8(3), 2041669517708205.

[12] "Community Download: Was Constellation Tracking a Long-Term Mistake for Oculus?" https://uploadvr.com/community-download-constellation-tracking/

[13] "Inside-Out Tracking" https://docs.microsoft.com/en-us/windows/mixed-reality/enthusiast-guide/tracking-system

[14] Motion controllers. https://docs.microsoft.com/en-us/windows/mixed-reality/enthusiast-guide/motion-controllers

[15] "Self-tracking headsets are 2017's big VR trend — but they might leave your head spinning." https://www.theverge.com/2017/1/12/14223416/vr-headset-inside-out-tracking-intel-qualcomm-microsoft-ces-2017

[16] Prada, J., Park, T., Lim, J., & Song, C.. "Exploring the potential of modifying visual stimuli in virtual reality to reduce hand tremor in micromanipulation tasks. *Current Optics and Photonics*. 1(6), 642-648.

[17] "Hyper-reality." http://hyper-reality.co

[18] Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., Dunlap, W. P., & Mulligan, B. E. (1987). "Guidelines for alleviation of simulator sickness symptomatology" (No. NAVTRASYSCEN-TR-87-007), Naval Training Systems Center, Orlando, FL.

[19] Azuma, R. (1993). "Tracking requirements for augmented reality." *Communications of the ACM*, 36(7), 50-51.

[20] Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). "Human factors issues in virtual environments: A review of the literature." *Presence*, 7(4), 327-351.

[21] Fang, B., Co, J., & Zhang, M. (2017, November). "DeepASL: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation." *In Proceedings of the 15th ACM Conference on Embedded Networked Sensor Systems*. ACM.