https://doi.org/10.2352/EI.2023.35.12.ERVR-214

Tangible Extended Reality with Sensor Fusion

Yang Cai, University of California, San Diego, California, USA
Mel Siegel, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Abstract

Many extended reality systems use controllers, e.g. near-infrared motion trackers or magnetic coil-based hand-tracking devices, for users to interact with virtual objects. These interfaces lack tangible sensation, especially during walking, running, crawling, and manipulating objects. Special devices such as the Teslasuit and omnidirectional treadmills can improve tangible interaction. However, they are not flexible enough for broader applications. They are bulky and expensive. In this study, we developed a configurable multi-modal sensor fusion interface for extended reality applications. The system includes wearable IMU motion sensors, gait classification, gesture tracking, and AR/VR systems data streaming interfaces. This system has several advantages: First, it is reconfigurable for multiple dynamic tangible interactions such as walking, running, crawling, and operating with an actual physical object without any controllers. Second, it fuses multi-modal sensor data from the IMU and sensors on the AR/VR headset, e.g., for floor detection. And third, it is more affordable than many existing solutions. We have prototyped tangible extended reality in several applications, including medical helicopter preflight "walking around" checkups, firefighter search and rescue training, and tool tracking for airway intubation training with haptic interaction with a physical mannequin. From our preliminary experiments, we found that realistic, tangible VR simulation is useful for users to discover hidden problems. We found that realistic and tangible VR simulators not only can help with training, but also can help to improve existing designs and interfaces at relatively low-cost.

Introduction

All realities are interactions. Our senses convey a picture of reality that narrows our understanding of its fullness. Virtual reality (VR) and augmented reality (AR) systems have been designed to interact with virtual and real worlds through human-computer interfaces.

Handheld controllers are typical interfaces to sense hand and head movement through near-infrared motion trackers or magnetic-coil-based hand-tracking devices for users to interact with virtual objects, such as HTC VIVE, Magic Leap, and Oculus Quest [1-3]. Some systems track hand gestures with cameras, e.g., HoloLens [4]. Unfortunately, these interfaces lack tangible sensation, especially in terms of walking, running, crawling, and manipulating an object. *It is obviously unnatural to simulate walking by pushing a joystick. Using a handheld controller to simulate holding a heavy object such as a firehose is also unnatural.*

There are some special devices such as the Teslasuit [5] and omnidirectional treadmills [6] that can be used for improving tangible interaction. However, they are not flexible, and they are bulky, and they are too expensive for broad application.

Our study aims to explore a sensor fusion solution that can track user motion without a handheld controller. Specifically, the proposed technology has the following features: It can capture the user's movements, such as walking, running, crawling, standing, and walking up and down stairs. It is embeddable on the surface of a physical body or a physical object. And it can communicate with the VR engine through Wi-Fi or Bluetooth.

Fusion of Motion Tracking for Leg and Head

We start with the simulation of walking with an IMU (inertial measurement unit) sensor. A basic 6-DOF IMU with accelerometers and gyroscopes is sufficient to sense walking dynamics. Figure 1 shows the minimal configuration of the 6-DOF IMU. Ideally, a 9-DOF IMU with magnetic field sensing is preferred because of drifting yaw values in 6-DOF sensors. In addition, an altitude rate sensor provides valuable information about walking upstairs or downstairs.

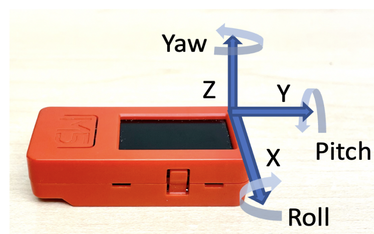

Figure 1. The 6-DOF IMU sensors with accelerometers and gyroscopes

Fortunately, many existing VR headsets have head-tracking sensors already. While a basic 6-DOF IMU can be a pedometer for step counting, head tracking can provide the user's heading data. By fusing these two datasets in real-time, we can simulate the walking motions by stepping at the same location so that the user can feel the locomotion without using a treadmill and walking for a distance outside the boundary.

Figure 2 illustrates the basic modules for the fusion of the wearable IMU sensor and the VR head-tracking sensors for walking detection and tracking, including standing, walking speed, and heading. Many VR headsets also have user hand-tracking functions such as pointing, pulling, pushing, rolling, grabbing, and holding. This would enable tangible extended reality with affordable wearable sensors, without using VR hand controllers.

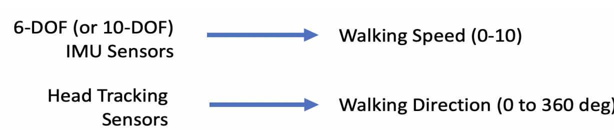

Figure 2. Fusion of wearable Inertial Measurement Unit and VR head-tracking sensors for walking detection and tracking.

The IMU data can be wirelessly streamed to the VR headset via WiFi or Bluetooth signals. We prefer Bluetooth to Wi-Fi because it is easier to pair and set up. Figure 3 shows the motion data format.

In our initial experiments we implemented the motion classification on the headset, which can integrate the foot motion and head motion data conveniently. However, sending raw IMU data from an embedded system to a headset is not efficient in terms of data communication and computation on the headset.

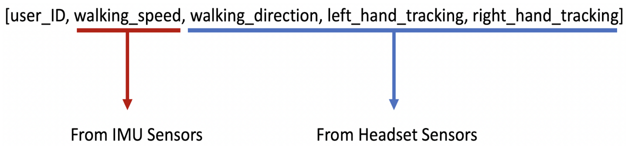

Figure 3. A fused data string from the IMU sensor and the headset

The fusion of motion-tracking data from the leg and head enables the user to experience VR simulation naturally in a lab, an office, or a home environment without using a handheld VR controller. Figure 4 shows a user who wore the motion-tracking sensor on his right foot to experience virtual medical helicopter preflight checkup activities such as walking around the helicopter, closing a door that is ajar, etc.

Figure 4. Walking, heading, and hand tracking in action in a VR environment.

Where to wear the IMU sensor and how to recognize the wearer's gaits are critical to the tangible extended reality experience. We started with the "sensor-on-kneecap" position naturally because it generates significant motion signals. However, considering the need to accommodate crawling and kneeling gaits, we ended up putting the IMU sensor on the toe of one of the shoes.

The 6-DOF IMU sensor contains accelerometers in 3-DOF and angular sensors in 3-DOF (pitch, yaw, and roll). Figure 5 shows the accelerometers moving in parallel to the ground along the *X-axis* and the gait moving forward, where the red line indicates significant motion along the *X*-axis and the green line indicates a subtle motion along the *Z*-axis. This pattern is typical when the user steps repeatedly on the same location. In the VR simulation system, we can translate a detected stepping motion into a scalable horizontal walking gait.

If we move the accelerometers perpendicularly to the ground, then the sensors are moving up and down along the *Z*-axis. Figure 6 shows the acceleration along the *Z*-axis (green line) and the acceleration along the *X*-axis (red line).

The movements in Figures 5 and 6 are in ideal conditions. Actual walking is much more complicated, as it includes pitch, roll, yaw, altitude changes, etc. These additional variables should be considered in advanced gait classification models.

In this study, it is challenging to simulate walking backward while trying to step in the same place. We designed a simple way to distinguish walking forward and backward: see Figure 7. We let the user move the foot that has the IMU sensor backwards about 15 cm and use the tip of the shoe to tap the floor vertically to simulate walking backward. The pseudocode of the heuristic rules is illustrated here:

```
If Pitch is greater than α then
  If acc_X-axis is greater than β
     Then Walk_backwards
  Else
     If acc_Z-axis is greater than β
     Then Walk_forward
```

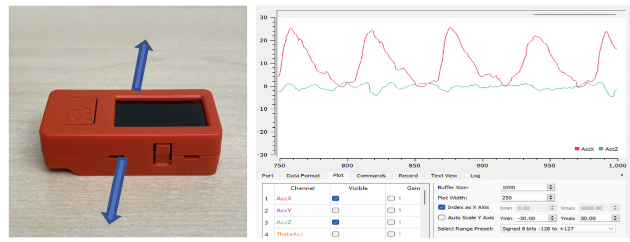

Figure 5. Acceleration is parallel to the ground along the X-axis (in red).

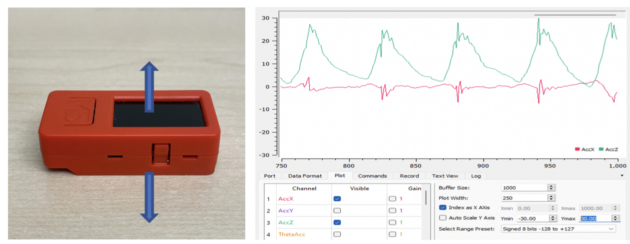

Figure 6. Acceleration is perpendicular to the ground along the Z-axis (in green)

Walking Forward Walking Backward

Figure 7. Gestures signifying walking forward and walking backward

Fusion of Distance and IMU Measurements

The advantages of fusion of the 6-DOF IMU sensors and the motion sensors on the headset include simplicity, light weight, and low cost. However, it is still challenging to simulate complex human movements in extreme conditions, for example, firefighters' crawling, "duck walk", and "shuffle walk." Modern 3D motion capture methods certainly can classify broader locomotions with acceptable accuracy. However, those methods may take much more space and expense to implement.

We have further explored the advanced sensor fusion of a distance sensor and a 9-DOF IMU, which are lightweight, wearable, and inexpensive. The sensor system can be integrated into professional equipment such as a firefighter's helmet.

The simplest distance sensor is LIDAR (Light Detection and Ranging), which has been widely used in autonomous driving vehicles. In our case, we just use a much simpler form: a one-directional LIDAR without rotation parts or multiple scan lines. The range of the LIDAR sensor we selected approaches 40 meters. In addition, the magnetic sensors and altitude sensors enable the onboard processor to determine the heading and stair walking gaits.

In this study we used advanced sensory fusion to detect seven gaits: crawl, walk, up-stairs, down-stairs, run, stand still, and duck walk. Figures 8 through 10 show data for pitch, and for acceleration along Y- and X-axes, for each of the seven gaits.

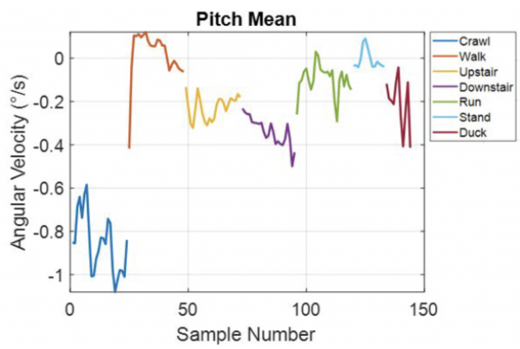

Figure 8. Dynamic pitch data for the seven gaits

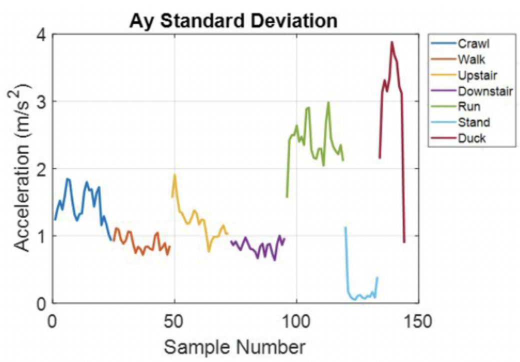

Figure 9. Acceleration data along Y-axis for the seven gaits

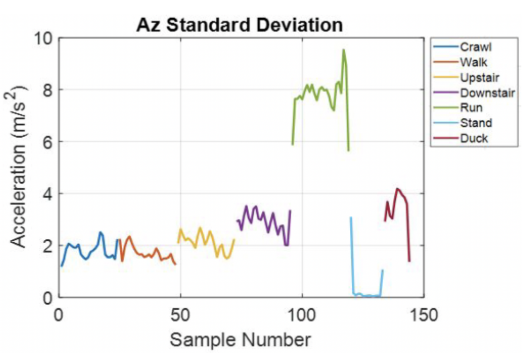

Figure 10. Acceleration data along Z-axis data of the seven gaits

Note that the Ay and Az data for standing still are indicative of the accelerometers' noise levels in their operating environment.

We then use a decision tree to fuse the sensory data and classify the seven gaits. We selected this model because it is simple, computationally lightweight, trainable, and explainable. The model is so small that we can implement the learning and classification model in an embedded system. Furthermore, a decision tree is also a visualization tool that enables human-computer interaction in a semantically intuitive way. Figure 11 shows an example of the output from the decision tree. There are more details about the firefighting case study in our previous published papers [7] and [8].

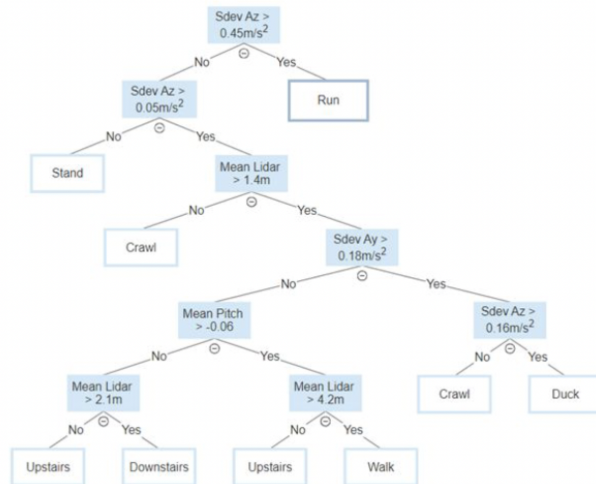

Figure 11. Decision Tree for classification of the 7 gaits

Our experiments show that the wearable system can recognize seven gaits with an accuracy of 86%. Figure 12 shows the confusion matrix of the classification results.

Actual Class

Output Class	Crawl	Downstair	Duckwalk	Run	Stand	Upstair	Walk
Crawl	90.8%	0.0%	3.9%	0.0%	0.0%	0.0%	0.0%
Downstair	5.0%	83.6%	0.0%	0.6%	0.0%	8.0%	0.8%
Duckwalk	0.0%	0.0%	80.6%	0.0%	0.0%	2.3%	0.0%
Run	0.0%	0.0%	0.0%	93.6%	0.0%	0.0%	0.0%
Stand	0.0%	0.0%	0.0%	0.0%	91.4%	0.0%	0.0%
Upstair	3.3%	15.8%	2.9%	5.8%	5.9%	81.7%	13.9%
Walk	0.8%	0.6%	12.6%	0.0%	2.6%	8.0%	85.2%

Overall Accuracy: 86.6%

Figure 12. Confusion matrix of the classifier for the seven gaits

Application Case Studies

Tangible applications are vital to the development of Extended Reality technologies. In our studies, real-world applications are always our driving forces for developing motion capture, classification, and interaction. Here we would like to introduce a few examples of applications as case studies, including virtual reality training for medical helicopter preflight checkups and firefighting and mixed reality training for intubation procedures.

Medical helicopters are often used for emergency response. The training of helicopter pilots, on-board emergency medical doctors, and operators in the communication control center is time-consuming and intensive. In this case, we developed VR-based exercises as a rehearsal preceding field-based training. The exercises contain two scenarios: 1) preflight checkups for objects left on the launch pad, e.g., bags, bottles, and tools, as well as unclosed doors, etc.; and 2) fire on the launchpad, including escaping and turning on the water cannons on the launchpad. In both scenarios the trainee needs to walk and run on the helicopter launchpad. Therefore we developed a simple wearable 6-DOF IMU motion sensor for sensing walking and running simulation, along with the IMU sensors on the user's headset for heading and hand tracking.

We scanned the actual helicopter launchpad and the medical helicopter and converted them into 3D VR models. We also extracted the actual background landscape from databases and recorded videos around the actual launchpad. The photorealistic VR models and real-time gait recognition and tracking, and dynamic scene rendering enable tangible perception and response in the extended reality environment.

Furthermore, we developed haptic interfaces for gloves and a helmet so that users can receive tangible tactile perception about directional guidance and force feedback when the user shuts an unclosed helicopter door, or the user's head hits the helicopter tail, resulting in a head injury incident.

From our preliminary lab experiments, we found that realistic, tangible VR simulation is useful for users to discover hidden problems beyond what is covered in existing training courses. For example, for helicopter landing safety reasons, there is no guide rail at the edges of the launchpad. From a user point of view, it is difficult to find the escape staircases ascending from the launchpad because they are below his line-of-sight. We reported these findings to the communication and control center. We also found that realistic and tangible VR simulators can not only help with training: they can also help to improve existing design and operation at nearly zero cost. Figure 13 and 14 are screenshots of the tangible VR system.

Figure 13. Helicopter preflight checkups and firefighting scenarios with fusion of the 6-DOF IMU and head tracking sensors on the VR headset.

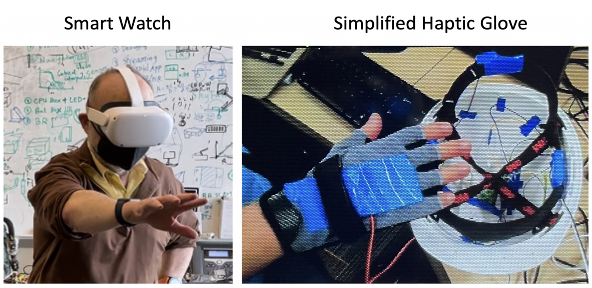

Figure 14. Smartwatches and haptic gloves enable tactile feedback.

Our second application case is mixed reality training for an intubation (airway management) procedure. The intubation process requires visual and haptic perception. To reflect the critical perceptions, we overlay a high-fidelity digital human model on top of a physical manikin. The mixed-reality approach creates more tangible extended reality experiences to trainees than do individual simulators separately. The user can simulate the intubation with the actual intubation blade, including LED lighting and physical force feedback. The digital models can be extended to extreme scenarios such as bleeding, wound care, and a realistic noise environment. Figure 15 and 16 show the training system in operation, the physical manikin, and the overlaid digital human on the 3D augmented reality glasses.

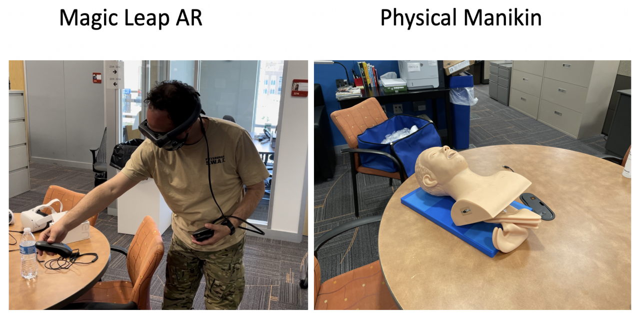

Figure 15. Integrating an AR headset with a physical manikin enables VR object overlay and haptic feedback.

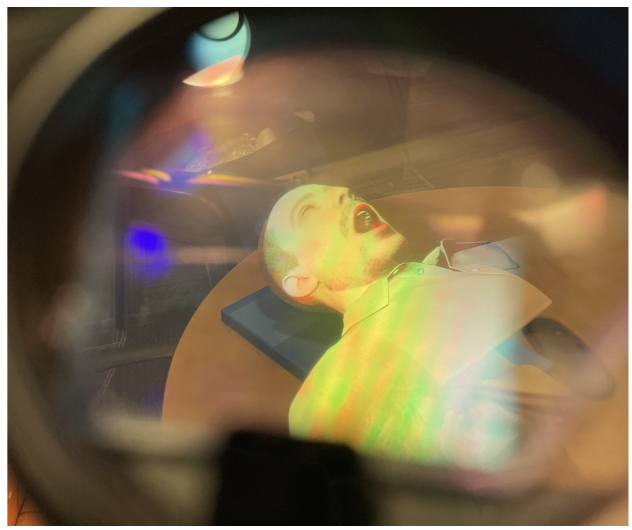

Figure 16. Screenshot of a high-fidelity digital human model overlaid on a physical manikin for airway intubation training.

Conclusions

Many extended reality systems use controllers, e.g., near-infrared motion trackers or magnetic coil-based hand-tracking devices for users to interact with virtual objects. These interfaces lack tangible sensation, especially during walking, running, crawling, and manipulating objects. Special devices such as the Teslasuit and omnidirectional treadmills can improve tangible interaction. However, they are not flexible enough for broad application. They are bulky and expensive. In this study, we developed a configurable multi-modal sensor fusion interface for extended reality applications. The system includes wearable IMU motion sensors, gait classification, gesture tracking, and AR/VR system data streaming interfaces. Our system has several advantages: First, it is reconfigurable for multiple dynamic tangible interactions such as walking, running, crawling, and operating with an actual physical object without any conventional controllers. Second, it fuses multi-modal sensor data from the IMU and other sensors on the AR/VR headset, e.g., for floor detection. And third, it is more affordable than many existing solutions. We have prototyped tangible extended reality in several applications, including medical helicopter preflight "walking around" checkups, firefighter search and rescue training, and tool tracking for airway intubation training with haptic interaction using a realistic human physical mannequin.

From our preliminary experiments, we found that realistic, tangible VR simulation is useful for users to discover hidden problems. We found that realistic tangible VR simulators not only can help with training: they can also help improve existing designs and operations at near-zero cost.

References

[1] VIVE: https://www.vive.com/us/

[2] Magic Leap: https://www.magicleap.com/en-us/

[3] Oculus: https://www.meta.com/quest/products/quest-2/

[4] HoloLens 2: https://www.microsoft.com/en-us/hololens/hardware

[5] Teslasuit: https://teslasuit.io/

[6] Omnidirectional treadmill: https://www.virtuix.com/

[7] Y. Cai, S. Hackett, B. Graham, F. Albert, M. Siegel, "Heads-Up Lidar Imaging with Sensor Fusion" in *Proc. IS&T Int'l. Symposium on Electronic Imaging: The Engineering Reality of Virtual Reality*, 2020, pp 338-1 - 338-7, https://doi.org/10.2352/ISSN.2470-1173.2020.13.ERVR-338

[8] S. Hackett, Y. Cai, and M. Siegel, "Activity Recognition from Sensor Fusion on Fireman's Helmet," *2019 12th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI)*, Suzhou, China, 2019, pp. 1-6, DOI: 10.1109/CISP-BMEI48845.2019.8966005.

[9] Y. Cai, Instinctive Computing, Springer-London, 2017

Author Biography

Yang Cai is the founder and head of Visual Intelligence Studio and a Research Scientist at Calit2, UC San Diego, CA, USA. He is the author of six AI books, including the monograph "Instinctive Computing (Springer-London, 2017)

Mel Siegel is Professor Emeritus at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA and an IEEE Life Fellow. He is an expert in sensing and sensors.

Acknowledgment

This project is in part sponsored by the NIST Public Safety Communication Research program. The authors would like to thank Scott Ledgerwood of NIST, and Professor Lenny Weiss of the University of Pittsburgh and UPMC for their support, as well as our many students and interns for system implementation.