ARToken: A Tangible Device for Dynamically Binding Real-world Objects with Virtual Representation

Hsuan-Yu Hsueh
sony0970610901@gmail.com
National Tsing Hua University
Hsinchu, Taiwan

Chien-Hua Chen
tim51072003@gmail.com
National Taiwan University of Science and Technology
Taipei, Taiwan

Irene Chen
cheniy@usc.edu
National Taiwan University of Science and Technology
Taipei, Taiwan

Chih-Yuan Yao
yuan.yao@csie.ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Hung-Kuo Chu
hkchu@cs.nthu.edu.tw
National Tsing Hua University
Hsinchu, Taiwan

ABSTRACT

Users are eager to interact with and control virtual proxies in the reconstructed virtual world using physical objects with haptic feedback since the development of AR/VR. Therefore, we propose ARToken, a device that virtualizes physical objects through AR tracking, and then links the relative positions of the Token device and virtual representations through its system, generating interactive virtual environments based on physical environments. In addition to tracking position through AR images, when the object exceeds the recognition range of AR Image, the IMU sensor can be used to detect rotation and displacement. ARToken sends this information through WiFi, allowing for real-time feedback of the state of the real object to the corresponding virtual representation. Also, this system allows for simultaneous tracking of multiple devices. To verify our design, we have produced two applications that exhibit how ARToken aids in generating virtual environments based on physical environments and links virtual and physical objects as one.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; Augmented Reality.

KEYWORDS

Augmented Reality; User interface toolkits; Immersive experience.

ACM Reference Format:
Hsuan-Yu Hsueh, Chien-Hua Chen, Irene Chen, Chih-Yuan Yao, and Hung-Kuo Chu. 2021. ARToken: A Tangible Device for Dynamically Binding Real-world Objects with Virtual Representation. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3460418.3480161

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3480161

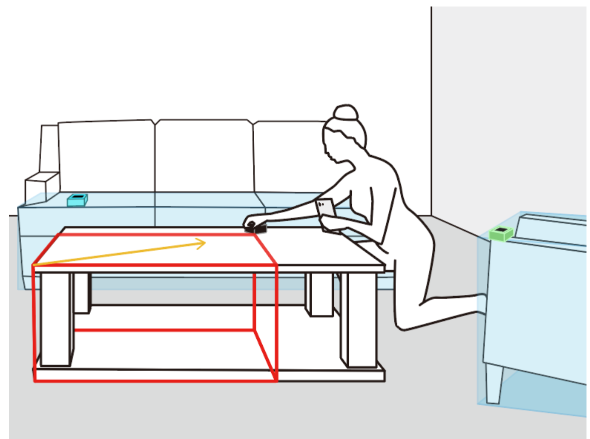

Figure 1: ARToken can reproduce physical environments in the virtual space, and track the translation and rotation of physical objects.

1 INTRODUCTION

Since the development of VR/AR technology, people have been craving a completely immersive experience. Providing users with haptic feedback effectively enhances this immersive experience and reduces the feeling of dissociation during usage.

The reconstruction of the real environment in the virtual space is one of the methods to realize the interactive experience of tactile feedback. Some studies reproduce indoor spaces and objects by taking 360-degree panoramic photos. Today's AR technology is even more powerful; it can directly look around a space and immediately reconstruct it. However, if some objects in the room moved, the positions captured by these methods will no longer be accurate, whether reconstructed through a panorama or AR. Therefore, conventional AR experience provides users with *one-way* interaction as any changes in real-world objects wouldn't feedback to the virtual world.

In order to support *two-way* interaction and tactile feedback, it is essential to link virtual representations to physical objects and control the movement of virtual representations by tracking the location and posture of the bound physical objects. In this work, we propose ARToken, a tool that can transform physical environments into virtual ones. We have adopted AR technology that can sync the coordinates of real and virtual spaces and objects in time. Using AR image detection and measurement, the token first detects the position of the object relative to the room, then measures the true size of the object through one's mobile device, so that both the position and size of the virtual representation are consistent.

In order to ensure that a physical object and its virtual counterpart move synchronously, after generating a virtual representation for a real object, we will place the ARToken somewhere on the real object, bind the token and object locations, then track the token's real-time position and rotation through the IMU sensor in the ARToken. Additionally, because there will likely be more than one movable object in a room, ARToken also incorporates IoT technology. First, each real object is individually attached to its token. Then, the token uses WiFi to connect the IMUs of all devices. The information detected by these sensors is then transmitted to the user in real-time.

This article also produced two practical applications to show the actual use of our device system. These applications demonstrate the creation of virtual environments that correspond to real spaces, the synchronized movement of real objects and their virtual counterparts, and the use of physical objects to advance in virtual games.

2 RELATED WORK

Our goal is to reproduce physical environments in the virtual space, track the movement and rotation of physical objects through several hardware devices, and provide physical feedback corresponding to users' actions to improve feelings of immersion. To set the stage for our work, we will describe work in three related areas: current AR technology, providing haptic feedback from virtual representations, and realistic space reconstruction.

2.1 Current AR Technology

Today's AR technology can detect patterns [8], 3D objects [2], and facial features to create virtual objects to interact with them. It can also scan environmental space measuring spatial distance [9], detecting horizontal and vertical planes [5], and determining depth [4, 13]. Recently, many mobile devices added the depth sensor to obtain more accurate real-world spatial information through its detection, which can even directly reconstruct a 3D model of the space.

2.2 Haptic Feedback from Virtual Objects

During AR/VR usage, if physical feedback can be provided, it can effectively enhance the user's immersive experience. Some attempts at achieving this have provided the user tactile feedback within a small range of motion (e.g. [1, 3, 6]), or used devices that can move autonomously to interact with the user (e.g. [12, 16]). These methods provide physical feedback for a virtual environment within

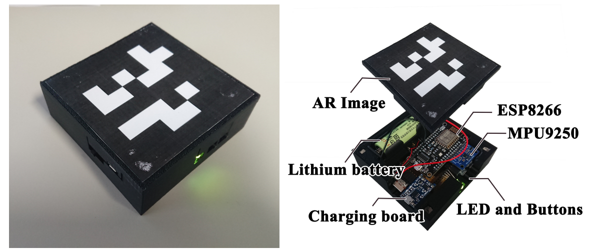

Figure 2: The hardware of an ARToken and components it contains.

a certain space by using reusable devices to change the physical environment to correspond to the virtual. On the other hand, if objects in physical environments can be reproduced in the virtual space, this would also be a viable method that could directly provide the user with haptic feedback during AR/VR use.

2.3 Realistic Space Reconstruction

Obtaining information on the size, shape, and location of objects in indoor spaces, or reconstructing real environments in the virtual environment, is a topic that has been researched and discussed greatly in the AR/VR industry. This technology can be used to reconstruct indoor spaces for digital viewing. In the AR/VR field, it can also be used to provide users with corresponding physical feedback in the virtual environment to enhance user immersion.

Currently, one of the methods to achieve this is to use 360-degree panoramic photos to reproduce the indoor environment. This method can not only detect the size and shape of the indoor space but also the corresponding position of indoor objects [15, 17, 18]. Another method that has been studied for a while is using a depth camera to scan and produce a virtual version of the environment [7, 10] by making calculations according to depth information obtained from different perspectives. Holoportation [11] demonstrates high-quality, real-time 3D reconstructions of an entire space, including people, furniture, and objects, using a set of new depth cameras. These 3D models can also be transmitted in real-time to remote users. ShareSpace [14] utilize multiple devices to tag the objects and people in the room, and track the position status of the devices in real-time, which is the most similar to our goal of obtaining object information in real-time.

3 DESIGN AND IMPLEMENTATION

ARToken is a device that can combine reality with virtuality. Users can utilize this device to reconstruct the real environment in the virtual world, and bind it to the reconstructed environment objects. If someone moves a bound object, the corresponding object will also immediately update its position in the reconstructed virtual world. In addition to providing real-time and accurate location information, it can also provide users with physical feedback corresponding to the virtual environment, or allow third parties to move objects in the current environment to interact with the user.

Each ARToken is specifically designed to be a size easily operable by the user in hand. It can combine the virtual and real-world through AR images and sensors, and track the translation and rotation of multiple objects in real-time. When the ARToken is within the recognition range of the user's AR camera, the device can be detected through AR. Once the user leaves the effective recognition range of AR, ARToken can be detected by the IMU sensor, and transmitted to the server via WiFi. The server can then relay the information to the user in real-time, achieving a connection between the real world and virtual space.

3.1 Design of ARToken Device

Because the user needs to have one hand available to control the phone when using ARToken, and each token needs to be placed in a certain position for tracking. We decided on making the device size 9cm x 9cm x 2.5cm.

In order to ensure that ARToken can be detected through AR, detect movement and rotation independently, and use WiFi to transmit information, we used the following parts:

- AR image: Used to obtain the coordinates in relation to the virtual and real environments.
- ESP8266 NodeMCU: A microcontroller that can receive sensor information and connect to WiFi for value transmission.
- MPU9250: A nine-axis IMU sensor that can obtain movement and rotation status through mathematical calculations.

We have also included various micro buttons and micro switches to receive more user input, and an RGB LED to display the status of the device.

3.2 Design of ARToken System

Figure 3 shows the architecture of the ARToken system, which includes a mobile phone capable of using AR functions, several ARToken devices, and a server responsible for receiving and transmitting data. The purpose of AR in our ARToken system is to form the connection between the real and virtual world, measure object sizes, and track virtual proxy. The IMU sensor in the ARToken will also track the virtual proxy and pass all of the information to the server, then the server will pass all the information it receives to the mobile phone. This system uses both AR and IMU sensors for tracking and switches the tracking method according to different situations.

3.3 Switch AR and IMU Tracking Data

The AR method for tracking objects is still more accurate than the IMU sensor's detection, but the recognition range is limited. On the other hand, the IMU sensor's detection is less accurate, but there is no range limitation. Therefore, in our system design for the ARToken tracking method, both AR and IMU sensors are used for tracking. When the ARToken is in the AR recognizable range, the system will use AR to track the object. Once it leaves the recognizable range, the system will automatically switch to tracking the device using the rotation and translation information detected by the IMU sensor.

4 INTERACTION

In this section, we will explain how our ARToken reconstructs virtual spaces that correspond to the real environment, and how to link real and virtual objects. We will also be explaining the entire reconstruction and the combination method and process in detail.

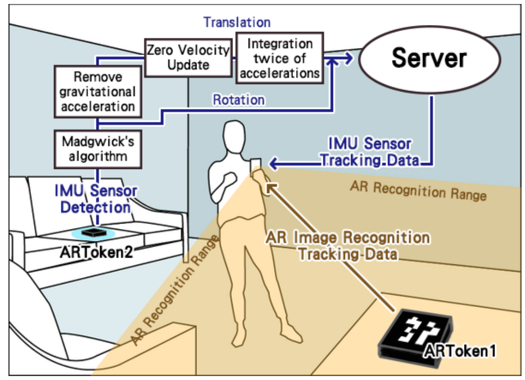

Figure 3: The ARToken system is comprised of IMU sensor and AR image recognition tracking data. Based on AR recognition range automatically switch the using tracking data.

4.1 Creation of virtual space

In order to use ARToken to reconstruct the real space, the user first needs to select the ARToken device to be used and connect with it by picking up the ARToken and recognizing the AR image with a mobile device (Figure 4 (A-1)). Once connected, users can choose to enter "Creation Mode" or "Attachment Mode." If there is no input after connecting for a period of time, the system will automatically disconnect from the selected device.

After selecting "Creation Mode," the user will first place the ARToken on a corner of the object, and then detect the ground plane on which the object is placed. In doing so, the height of the object is automatically obtained, and an initialized cube is generated (Figure 4 (A-2)). Then the user can measure the length and width of the object by moving the ARToken along its perimeter to create a virtual representation (Figure 4 (A-3)) whose position and size is consistent with those in the real environment (Figure 4 (A-4)).

4.2 Combination of virtual and reality

While the virtual representation has now been created, it has not yet been bound to the physical object(Figure 4 (B-1)). At this stage, if the real object is moved, this will not be reflected in the virtual representation. Therefore, we need to bind virtual representations and real objects through "Binding Mode" next. After the user selects the ARToken and enters Attachment Mode, one only needs to place the ARToken somewhere on the real object (Figure 4 (B-2)), then use AR image recognition to determine the corresponding location and bind (Figure 4 (B-3)). Afterward, if the real object is moved, the ARToken placed on the object will also translate the virtual representation in real-time, so as to achieve a true combination of reality and virtuality (Figure 4 (B-4)).

5 APPLICATIONS

ARToken offers three main features: the ability to measure objects and generate a virtual version of that object, the ability to track

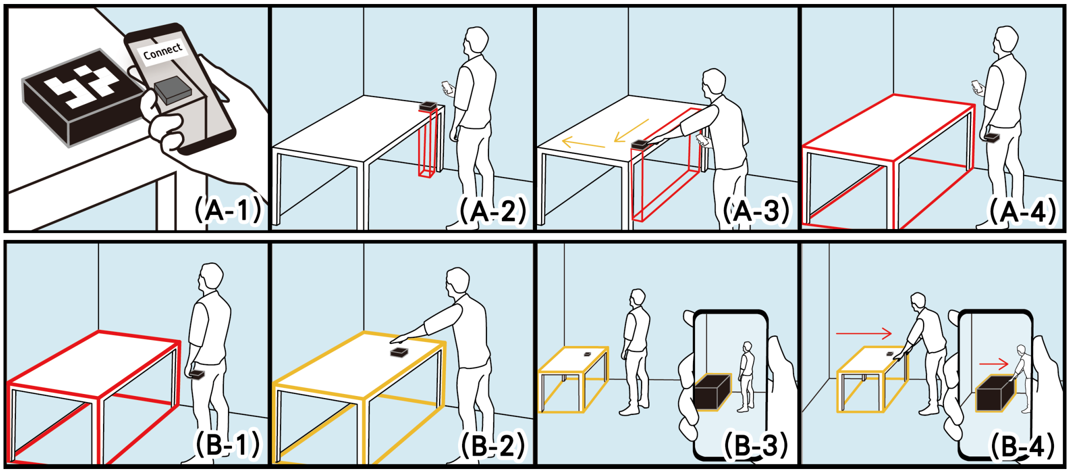

Figure 4: The user reconstructs the virtual environment process. (A)Creation of virtual space (B)Combination of the virtual and real object.

the translation of the object, and the ability to track rotation. With these capabilities, we are able to take the generated objects and transform them into virtual environments, and track them in the game world. In order to illustrate the variety of options ARToken offers, we have created two demo applications.

5.1 The Floor is Lava

"The Floor is Lava" is a simple game that demonstrates the defining feature of ARToken - using physical environments to generate virtual environments. The standard version of the game that our version is based on asks players to pretend that the floor is lava and furniture is landmasses to walk across. ARToken allows this fantasy to become a reality by utilizing its room/furniture measuring and tracking abilities. This application also allows for much more variation in the game in the form of special effects, visual feedback, and/or point tracking.

5.2 Escape the Ship

"Escape the Ship" is an escape-room-esque game where players must solve a series of puzzles to advance and win the game. This game also demonstrates the generation of virtual worlds based on information received from ARToken, and utilizes the translation and rotation tracking features of the tokens to clear the game. For example, use the rotation detection provided by the IMU sensor to place the objects in the room bound with ARToken to the corresponding position. Therefore, instead of tapping a screen, users can physically grab and rotate objects, and receive haptic and visual feedback. This product bridges the difference between seeing a virtual representation and squeezing some buttons to grab it, and seeing a virtual representation and reaching out with an empty hand to interact with it directly.

Figure 5: The ARToken application "The Floor is Lava"(left) and "Escape the Ship"(right). The former demonstrates how to avoid lava through the game in the virtual scene reproduced by the ARToken system, and the latter demonstrates that the user uses ARToken to place objects in the required state.

6 CONCLUSION

We propose a device that can reconstruct virtual spaces in relation to real environments. This device and system using AR technology to not only reproduce virtual representations from a physical environment with accurate placement but also reproduce the virtual objects in the correct size. More importantly, we are utilizing both AR image recognition and IMU sensor detection. Therefore, after the virtual and physical objects are bound, if the physical object is moved, this information can be reflected in the virtual world in real-time. Using IOT technology, we are also able to link multiple bound objects so that several ARTokens can transmit live information on each object's status to the users at the same time, achieving real-time tracking of various objects in the room. With this system, AR users can create a virtual space that corresponds to a real environment, control virtual objects with physical feedback, and allow others to control the user's environment through real objects. The increased variety and depth of user interaction in the virtual space effectively increases the user experience, immersion, and enjoyment.

ACKNOWLEDGMENTS

The project was funded in part by the Ministry of Science and Technology of Taiwan (109-2218-E-007-014- and 107-2221-E-007-088-MY3).

REFERENCES

[1] Bruno Araujo, Ricardo Jota, Varun Perumal, Jia Xian Yao, Karan Singh, and Daniel Wigdor. 2016. Snake Charmer: Physically enabling virtual objects. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. 218–226.

[2] Arpan Chakraborty, Ryan Gross, Shea McIntee, Kyung Wha Hong, Jae Yeol Lee, and Robert St. Amant. 2014. Captive: a cube with augmented physical tools. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 1315–1320.

[3] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D Wilson. 2017. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3718–3728.

[4] Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, et al. 2020. DepthLab: Real-Time 3D Interaction With Depth Maps for Mobile Augmented Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 829–843.

[5] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.

[6] Hsin-Yu Huang, Chih-Wei Ning, Po-Yao Wang, Jen-Hao Cheng, and Lung-Pan Cheng. 2020. Haptic-Go-Round: a surrounding platform for encounter-type haptics in virtual reality experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–10.

[7] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 559–568.

[8] Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE, 85–94.

[9] Johnny Lee. 2017. 4-1: Invited Paper: Mobile AR in Your Pocket with Google Tango. In *SID Symposium Digest of Technical Papers*, Vol. 48. Wiley Online Library, 17–18.

[10] Peter Ondrúška, Pushmeet Kohli, and Shahram Izadi. 2015. Mobilefusion: Real-time volumetric surface reconstruction and dense tracking on mobile phones. *IEEE transactions on visualization and computer graphics* 21, 11 (2015), 1251–1258.

[11] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*. 741–754.

[12] Ryo Suzuki, Hooman Hedayati, Clement Zheng, James L Bohn, Daniel Szafir, Ellen Yi-Luen Do, Mark D Gross, and Daniel Leithinger. 2020. Roomshift: Room-scale dynamic haptics for vr with furniture-moving swarm robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[13] Julien Valentin, Adarsh Kowdle, Jonathan T Barron, Neal Wadhwa, Max Dzitsiuk, Michael Schoenberg, Vivek Verma, Ambrus Csaszar, Eric Turner, Ivan Dryanovski, et al. 2018. Depth from motion for smartphone AR. *ACM Transactions on Graphics (ToG)* 37, 6 (2018), 1–19.

[14] Keng-Ta Yang, Chiu-Hsuan Wang, and Liwei Chan. 2018. Sharespace: Facilitating shared use of the physical space by both vr head-mounted display and external users. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 499–509.

[15] Shang-Ta Yang, Fu-En Wang, Chi-Han Peng, Peter Wonka, Min Sun, and Hung-Kuo Chu. 2019. Dula-net: A dual-projection network for estimating room layouts from a single rgb panorama. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3363–3372.

[16] Yan Yixian, Kazuki Takashima, Anthony Tang, Takayuki Tanno, Kazuyuki Fujita, and Yoshifumi Kitamura. 2020. ZoomWalls: Dynamic Walls that Simulate Haptic Infrastructure for Room-scale VR World. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 223–235.

[17] Yinda Zhang, Shuran Song, Ping Tan, and Jianxiong Xiao. 2014. Panocontext: A whole-room 3d context model for panoramic scene understanding. In *European conference on computer vision*. Springer, 668–686.

[18] Chuhang Zou, Alex Colburn, Qi Shan, and Derek Hoiem. 2018. LayoutNet: Reconstructing the 3D Room Layout From a Single RGB Image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.