# Resolving Occlusion for 3D Object Manipulation with Hands in Mixed Reality

Qi Feng
Waseda University
Okubo 3-4-1
Tokyo 1690072, Japan

Hubert P. H. Shum
Northumbria University
Newcastle upon Tyne NE1 8ST
United Kingdom

Shigeo Morishima
Waseda University
Waseda Research Institute for Science and Engineering
Tokyo 1690072, Japan

## ABSTRACT

Due to the need to interact with virtual objects, the hand-object interaction has become an important element in mixed reality (MR) applications. In this paper, we propose a novel approach to handle the occlusion of augmented 3D object manipulation with hands by exploiting the nature of hand poses combined with tracking-based and model-based methods, to achieve a complete mixed reality experience without necessities of heavy computations, complex manual segmentation processes or wearing special gloves. The experimental results show a frame rate faster than real-time and a great accuracy of rendered virtual appearances, and a user study verifies a more immersive experience compared to past approaches. We believe that the proposed method can improve a wide range of mixed reality applications that involve hand-object interactions.

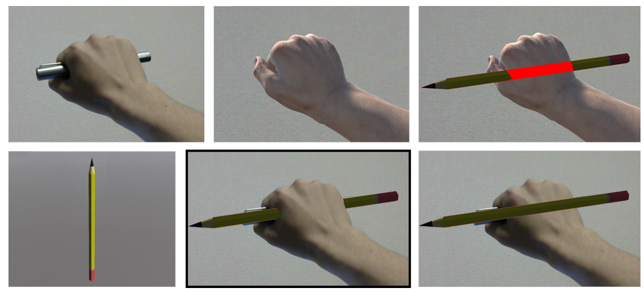

**Figure 1: The overview of the approach: top row: (1) real scene; (2) fitted hand model based on captured finger joints; (3) estimated occlusion mask; bottom row: (4) the model of virtual object; (5) rendered mixed reality scene with occlusion; (6) rendered mixed reality scene without occlusion.**

## CCS CONCEPTS

• **Computing methodologies → Mixed / augmented reality**;

## KEYWORDS

Mixed Reality, Occlusion, Hand tracking

**ACM Reference format:**
Qi Feng, Hubert P. H. Shum, and Shigeo Morishima. 2018. Resolving Occlusion for 3D Object Manipulation with Hands in Mixed Reality. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology, Tokyo, Japan, November 2018 (VRST '18),* 2 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Recent advance of displaying technologies combined with the progress in tracking techniques has pushed mixed reality to the verge of being a commodity as well as a feasible tool for assisting 3D tasks [Krichenbauer et al. 2018]. To provide an intuitive way to interact with augmented contents, physical interactions are equally important as virtual imagery [Bach et al. 2018].

Using different input devices to interact with virtual objects can lead to disparity in completing the same task [Krichenbauer et al. 2018]. As the most common way to interact with objects, the free-hand interaction with virtual objects is more preferable over other mentioned ways [Bach et al. 2018].

The quality of overlays is of great importance since incorrect occlusions and other unrealistic representations can easily break the experience [Azuma 2017]. To correctly composite augmented objects with realistic images, semi-automatic approaches with manual inputs have been proposed [Tian et al. 2010]. One way to tackle occlusions without the necessity of assigning the occluding object during usage is using depth sensors to determine the relative positions between in-scene objects [Newcombe et al. 2011]. However, the noises and computational cost are still problematic in MR applications. Using predefined 3D models to perform the fitting task via tracking is more viable for real-time usage. Tian [Tian et al. 2015] tried to reconstruct the scene and 3D objects based on a 3D time-of-flight camera. However, the performance greatly depends on the tracking robustness and is usually weak of deformable objects. Due to complex articulations of hands with intense self-occlusions, these solutions are incapable of handling such cases.

As an increasing interest in hand-objects interactions appears with the availability of commercial depth cameras to capture gestures, we focus only on hands and utilize the nature of grabbing poses to fulfill the necessity of depth relations between in-scene objects to develop a system that can aid egocentric scenes of hand-holding virtual 3D objects by delivering the correct occlusion. By combining a precise hand tracking approach, the customized models of the hand and an extra calibration step, our system generate better results than model-based and tracking based solutions. A comparative user study also verifies a more immersive experience.


*VRST '18, November 2018, Tokyo, Japan*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 2 METHODOLOGY

The architecture of our system consists of two phases: the preparation phase and the usage phase. During the preparation phase, our system calibrates the parameters of the user's hand by an RGB camera combined with a piece of marked paper to customize a hand model for further usage, as shown in Figure 2. Meanwhile, we chose a wide range of objects that are usually used in daily life, e.g. a pen, to construct a virtual object database to display during usage. During the run-time, we use a low-cost depth sensing device, e.g. Leap Motion [Weichert et al. 2013], to track the user's hand and obtain the coordination of finger joints and the direction of palms. Based on the relative positions of perspective, fingers and the direction of the palm, we label each pixel of the virtual object within the contour of the user's hand with 'visible'and 'invisible'. The label is decided by comparing the current 3D displacement of the virtual object and the customized hand model displacement which is determined by the captured fingers joints and the palm direction. The system can then acquire the mask of occluded portion. Based on the occluding mask, a correct representation of virtual objects is then rendered on the head-mount display to build the illusion that the user is holding something else with their hands.

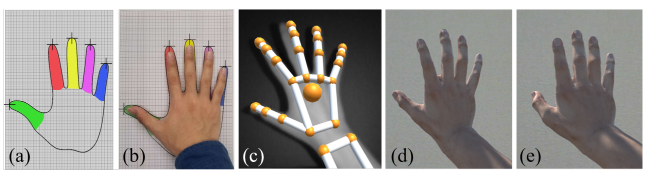


**Figure 2: An example of the process of customizing hand model: (a) marked sheet; (b) calibrating hand through RGB camera; (c) tracked joints during usage; (d) slender hand; and (e) tubby hand after customization.**

## 3 RESULT AND DISCUSSION

During the experiment, in order to verify the quality of the mixed reality experience of the proposed method, we implemented a complete table-top application to showcase the idea and study the immersitivity using the Unity3D. The configuration and results of the system are shown in Fig. 3. The top-left figure shows the real scene without any overlay; The top-right figure shows the result without any occlusion handling; The bottom-left figure shows the result when applying the approach proposed by Liang et al. [Liang et al. 2015]. The transparency of the virtual can was adjusted to 70% according to the direction of the palm in this case; The bottom-right figure shows the result of our approach.

To evaluate the concept and investigate whether this approach can improve the MR experience or not, we designed a participant-based cooperative qualitative evaluation. From both the results of the questionnaire (as shown in Fig. 4) as well as the comments in the subsequent open discussions, we confirmed the usefulness, usability and a positive impression of our implementation.

## 4 FUTURE WORK AND CONCLUSION

We have presented a robust approach to handle the occlusion during interaction between hands and object happens. We combined

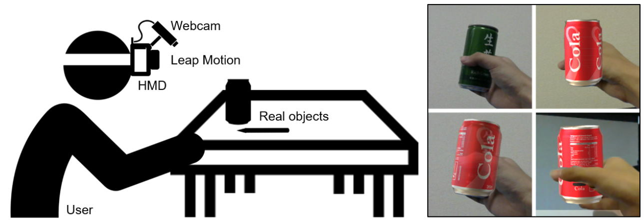


**Figure 3: The real-world configuration of the implemented application (left) and results of three different representation schemes for augmenting the original can with a virtual Cola can (right).**

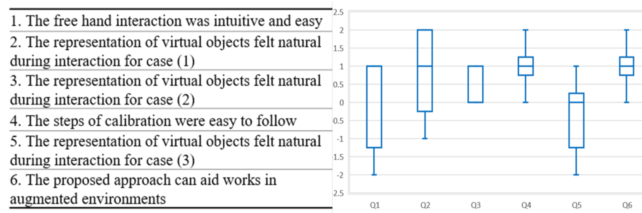


**Figure 4: Questionnaire results from the immersivity study. Each statemented is scored from -2 as "highly disagree" to +2 as "Highly agree" with a step of 1.**

the tracking-based method with the model-based method for estimating occluding mask. The implementation of the approach can be used to solve problems for both hands interacting with arbitrary rigid objects with a good accuracy and speed with the potential to be further adapted to other applications. As a future topic, we envision using improved neural networks to solve occlusions from a single RGB image with a high precision and a fast speed.

## REFERENCES

Ronald T. Azuma. 2017. Making Augmented Reality a Reality, In Imaging and Applied Optics 2017. *Imaging and Applied Optics 2017*, JTu1F.1. http://www.osapublishing.org/abstract.cfm?URI=AIO-2017-JTu1F.1

Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, and Hanspeter Pfister. 2018. The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *IEEE Transactions on Visualization & Computer Graphics* 1 (2018), 1–1.

Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2018. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *IEEE transactions on visualization and computer graphics* 24, 2 (2018), 1038–1048.

Hui Liang, Junsong Yuan, Daniel Thalmann, and Nadia Magnenat Thalmann. 2015. Ar in hand: Egocentric palm pose tracking and gesture recognition for augmented reality applications. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 743–744.

Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*. IEEE, 127–136.

Yuan Tian, Tao Guan, and Cheng Wang. 2010. Real-time occlusion handling in augmented reality based on an object tracking approach. *Sensors* 10, 4 (2010), 2885–2900.

Yuan Tian, Yan Long, Dan Xia, Huang Yao, and Jincheng Zhang. 2015. Handling occlusions in augmented reality based on 3D reconstruction method. *Neurocomputing* 156 (2015), 96–104.

Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. 2013. Analysis of the accuracy and robustness of the leap motion controller. *Sensors* 13, 5 (2013), 6380–6393.