Grab-Carry-Release:
Manipulating Physical Objects in a Real Scene through a Smart Phone

Kai-Yin Cheng* Yu-Hsiang Lin* Yu-Hsin Lin* Bing-Yu Chen* †Takeo Igarashi
*National Taiwan University †The University of Tokyo/JST ERATO

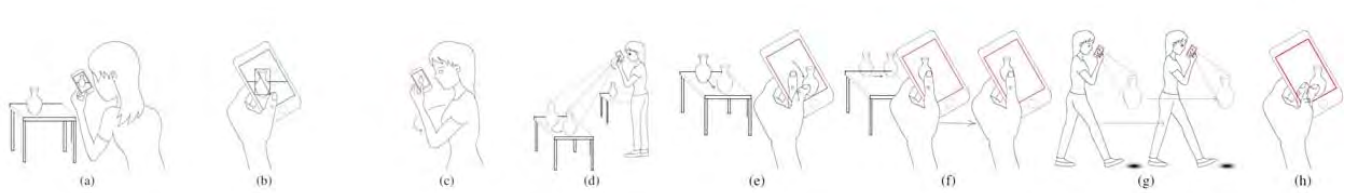

Figure 1. *Grab-Carry-Release* gestures. (a) Users aim at a real object that they want to move. (b) The *Grab* is performed to form a virtual 3D model of the real object. (c) Users can *Carry* the virtual model to the desired scene. (d) The ray-casting gesture allows users to place the virtual model at the target position along the look-at vector from the smart phone. The auto-snapping is enabled in this mode, and the model is placed on the horizontal or vertical surface where the users are aiming. (e) Through the vertical dragging gestures, users can move the virtual model nearer (inward gesture) or farther (outward gesture). (f) Users can move or carry the virtual model by pressing the model on the screen and moving the smart phone. (g) Since the distance from the virtual model to the users are maintained, while the users physically walk forward, the model is moved forward to maintain the distance. (h) To confirm the results, the *Release* gesture by double-tapping on the screen is performed.

1. Introduction

Individuals often discuss how to configure their real world space, such as switching the location of an object, designing furniture layouts or room decorations, shopping for suitable house appliances, or even commanding a robot to perform certain tasks. However, verbal communication is often insufficient to convey the imagined results for such spatial arrangements. Users must occasionally spend considerable efforts in either moving the real objects physically or using photos or videos to composite the mockup to simulate the layout. Nevertheless, those methods are too tedious for real-time communication.

This work presents a novel mobile AR approach, Grab-Carry-Release, to facilitate visual communication and experimentation with respect to real object manipulation. Users virtually grab a physical object on the see-through screen with their mobile devices, carry it to the target location, and release it. Through the grab gesture, a virtual model is formed for representing the physical object in a real scene. Users then carry the model to where they want to position it at by the developed two-mode gestures. Users utilize the ray-casting gesture to determine the rough global position and adopt the manual placement gestures to perform local adjustment. During positioning, a virtual shadow is generated to provide distance and spatial information. Confirming the target position, the user releases the object through the double-tap gesture. Following completion of the process, the user and other individuals can walk around the virtual model to verify the simulated results in a real scene.

Additionally, collocated users can also share the retrieved and formed model. For instance, several individuals can discuss the furniture layout in the same space by coordinating with the updated position status of a particular virtual object. Remote AR sharing can also be achieved because the captured depth information provides the rough size of a real object. For instance, users in an electric appliance store can capture a TV model and share it with their families at home. Due to the preserved rough size of the TV model, a family can simulate placing the model in a desired position to verify whether its appropriateness before purchasing it. All scenarios are prototyped and demonstrated in the user study to retrieve user feedback to further enhance interactions.

This paper examines the in-situ mobile AR interaction by attaching a depth camera to a smart phone. Kinect for Xbox 360 is used as a depth camera with iPhone 4 used as a smartphone. While Kinect provides 2D RGB and depth images (or so-called the 2D+Z images), the adapted GrabCutD [1] algorithm is adopted to curve out the desired physical object to form a virtual model. Visual feature points are also tracked by combining depth images from multiple view points to register the virtual model in the real scene.

Conventional AR systems usually show predefined information to users [2]. Either the environment is known previously, or the augmented information is predefined first. However, the proposed method does not require predefined information and allows users working with physical objects “in the wild” to create in-situ AR interaction by taking advantage of the readily available surrounding context in smart phone using scenarios.

References

VAIAPURY, K., AKSAY, A., AND IZQUIERDO, E. 2010. GrabCutD: improve grabcut using depth information. In *Proceedings of SMVC’10*, ACM, 57–62.

ZHOU, F., DUH, H. B.-L., AND BILLINGHURST, M. 2008. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings ofISMAR’08*, ACM/IEEE, 193–202.

.