

CHAPTER

9

DEPTH SENSOR AND HANDHELD DISPLAY FUSION

Muhammad Anwar Ahmad and Norhaida Mohd Suaib

9.1 INTRODUCTION

Augmented reality (AR) is a technology that enhances the real world with digital content overlaid seamlessly through AR displays, such that it can be viewed and interacted with in real-time. These displays typically use sensors, cameras, and software to track the user's movements and position, and then project or overlay digital information onto the real-world view. Typical AR displays include head-mounted displays (HMDs), handheld displays, and projection-based displays. A mobile device, like a smartphone or tablet, can view a handheld AR experience. This kind of AR technology uses the camera and sensors in a mobile device to put digital content on top of the real world that the device is pointing at. This chapter presents a framework that leverages a high-fidelity stereo vision-based depth sensor to overcome the limitations of a handheld display device's processing power to compute depth estimation, particularly for indoor environments. A virtual object should be realistically positioned in the real world to make AR realistic. A key component of AR realism is tracking and registration, which entails the proper relative positioning of real objects and virtual objects to achieve realism. The virtual objects displayed should remain aligned

with the position and orientation of the real objects. Occlusion occurs after the tracking process, in which virtual objects should be correctly occluded by foreground objects. Thus, occlusion handling is performed to hide or remove the pixels of virtual objects that are obstructed by real objects. An AR tracking system should be independent and robust enough to adapt to the visual information in the real world, including occlusions. Figure 9.1 shows a simple simulation of how occlusion handling can improve the look of the AR environment. In Figure 9.1(a), the cube is expected to be behind the wireless mouse but appears to be on top of the mouse instead. After applying occlusion handling in Figure 9.1(b), parts of the cube that are obscured by the mouse are removed from the scene and look more accurate. This improves the overall look of the scene.



Figure 9.1 Simulation of occlusion handling in an AR environment
(a) Cube overlapping the mouse and (b) Obscured parts removed

AR occlusion handling methods can be classified as model-based and depth-based. This chapter focuses on the depth-based method, primarily using the spatial mapping method, involving the creation of a detailed and accurate representation of the physical environment in the digital space via three-dimensional (3D) meshes (Tian et al., 2019). Spatial mapping and occlusion handling on handheld AR are limited by the size of the devices and processing power, which is addressed by the presented framework, in which the backbone is the fusion of a handheld display with a depth sensor, via a fusion module.

9.2 SPATIAL MAPPING

Depth sensing enables devices to perceive and interpret the 3D structure of the physical environment, thereby allowing virtual content to be placed with greater precision and realism within a user's surroundings. Without a proper understanding of depth, virtual objects may appear incorrectly layered over real-world elements, reducing the overall realism and immersion. The integration of depth sensors into AR platforms also involves a range of technical considerations, including hardware compatibility, data processing, and system optimization. This section examines the technologies and approaches involved in depth sensing for AR, highlighting key methods and applications across spatial mapping, occlusion handling, and system-level integration.

9.2.1 Depth Sensing

Depth sensors are devices that can measure the distance and shape of objects by producing depth maps. Based on past work by Burger et al. (2023), there were three candidates for depth sensors to be utilized. The commercial product named ZED Mini by Stereolabs company is the best choice as it is suitable for fast, dynamic movements. Furthermore, it features spatial mapping capabilities. It also mentioned that ZED Mini's NEURAL mode is a good compromise between the Z-accuracy and fill rate, which is the fraction of pixels that contain valid measurements within a region of interest. The stereo vision principle behind the ZED Mini is shown in Figure 9.2 and discussed further below.

Two cameras observe a point P located at a distance z . The left and right camera planes display the optical projections p_l and p_r of P . Knowing the intrinsic and extrinsic parameters of the cameras, the depth z is inversely proportional to the horizontal parallax between projections p_l and p_r , as shown in the Equation 9.1 (Giancola et al., 2018). This principle basically mimics human binocular vision for depth perception, enabling effective depth sensing and spatial understanding.

$$z = Bf/(x_r - x_l) \quad (9.1)$$