International Journal of Virtual and Augmented Reality
Volume 6 • Issue 1

Advanced Visual SLAM and Image Segmentation Techniques for Augmented Reality

Yirui Jiang, Cranfield University, UK*

Trung Hieu Tran, Cranfield University, UK

https://orcid.org/0000-0002-3989-4502

Leon Williams, Cranfield University, UK

ABSTRACT

Augmented reality can enhance human perception to experience a virtual-reality intertwined world by computer vision techniques. However, the basic techniques cannot handle complex large-scale scenes, tackle real-time occlusion, and render virtual objects in augmented reality. Therefore, this paper studies potential solutions, such as visual SLAM and image segmentation, that can address these challenges in the augmented reality visualizations. This paper provides a review of advanced visual SLAM and image segmentation techniques for augmented reality. In addition, applications of machine learning techniques for improving augmented reality are presented.

KEYWORDS:

Augmented reality, computer vision, image segmentation, machine learning, visual SLAM

1 INTRODUCTION

Nowadays, augmented reality (AR) has coexisted the real world with virtual objects. The technology has increased human experience in a virtual-reality intertwined world. It has grown in popularity over the last ten years, moving from laboratories into various real-life scenes (Van Krevelen & Poelman, 2010). However, there are numerous issues in AR, leading to the rise of innovations (Masood & Egger, 2019). Outdoor AR systems face challenges such as handling complex large-scale scenes, dealing with real-time occlusion, and rendering virtual objects. Although image segmentation could address the issues, it requires collaboration with other advanced techniques (Roxas et al., 2018). The collaboration is critical in the future trend of AR applications. Since traditional methods for sharing accurate spatial information are insufficient, various machine learning algorithms have been proposed to achieve low-cost and high-efficiency future collaboration systems (Zou et al., 2019). The volume of mobile and industrial AR applications is growing at an exponential rate; however, previous high

DOI: 10.4018/IJVAR.307063

*Corresponding Author

latency, low precision, and unfriendly user experience have hampered the widespread adoption of AR systems (Huang et al., 2013). To overcome these limitations, registration (Hoff et al., 1996), tracking (Runz et al., 2018), image segmentation (Kirillov et al., 2019) and occlusion (Tang et al., 2020) approaches have been proposed with machine learning-improved visualization techniques, i.e., visual SLAM (vSLAM) and image segmentation. This paper provides an overview of the main obstacles of the AR visualizations and their potential solutions. The most recent computer vision technologies are concentrated, particularly machine learning-enhanced innovative technologies.

To date, there exists a lack of comprehensive literature review on the topic of applying the latest machine learning-improved computer vision in AR systems. However, there are some literature reviews related to other research communities in the AR systems, which are shown in Table 1.

As part of the digital transformation of industry, AR improves industrial efficiency, safety, compliance, and costs. Ling et al. (2017) discussed commercial trends in industrial AR. Palmarini et al. (2018) conducted a systematic literature review to identify the most relevant industrial AR technical limitations. Li et al. (2018) examined various Virtual Reality (VR) / AR prototypes, products, and training evaluation paradigms. Gattullo et al. (2020) systematically reviewed the literature on visual assets. Egger et al. (2020) investigated the current challenges and future directions of AR manufacturing. Costa et al. (2022) provided an overview of the current state of the art in AR human-robot collaboration and future development trends. Notably, user research on AR cybersecurity applications is severely lacking. Alzahrani et al. (2022) identified, described, and synthesized research findings on the cybersecurity of the AR industry.

Visual tracking is a fundamental task in AR and has been an active research topic for many years. Kalkofen et al. (2011) pioneered the spatial integration of virtual objects in real-world settings. Rabbi et al. (2012) identified the difficulties of tracking an object in an unfamiliar environment. Billinghurst et al. (2015) investigated general tracking and displaying techniques. Li et al. (2018) reviewed the most recent deep learning-based tracking methods and divided them into three categories based on network structure, functionality, and training. The tracking methods used in AR-based robot maintenance are qualitatively evaluated (Koh et al., 2020). Jiao et al. (2021) reviewed the critical advances made by deep learning, including deep feature representations, network architecture, and four critical issues in visual tracking (e.g., spatiotemporal information integration, target-specific classification, target information update, and bounding box estimation). Zhu et al. (2022) provided an overview of AR visual object tracking on RGB-D videos.

Augmented reality provides users with an engaging, memorable, and impactful interactive experience, resulting in a digital world that closely resembles our physical world and offers a new perspective on reality. Rabbi et al. (2013) concentrated on AR hardware and user experience (UX). Irshad et al. (2014) summarized the UX of AR and identified areas where research was lacking. A thorough and detailed review was presented to assist AR developers in focusing on UX improvement by Irshad et al. (2017). The most influential AR user studies were presented by Dey et al. (2018). Recent AR systems have offered a range of device-specific interaction options and tailored solutions for delivering immersive experiences to users, but with an inherent lack of standardization across devices and applications. To address this issue, a systematic review and evaluation of explicit, task-based interaction methods in immersive environments are presented (Spittle et al., 2022).

The original clunky and poor user experience has been transformed by mobile AR, which you can take with you wherever you go, often on a smartphone device. The number of mobile AR users is increasing as hardware devices improve. AR network implications were highlighted by Westphal et al. (2017). Braud et al. (2017) investigated AR applications and their external infrastructures. Goh et al. (2019) provided context for AR 3D mobile interaction. Lee et al. (2022) reviewed the field of human interaction in connected cities, focusing on AR-driven interaction.

Most existing reviews of AR literature have paid less attention to visualization principles. This paper comprehensively reviews machine learning-enhanced AR visualization applications using vSLAM and image segmentation. The fundamental principles of technology are summarised, and