Chapter 38
Pose Tracking in Augmented Reality of Cultural Heritage

Daniel Asmar
American University of Beirut, Lebanon

ABSTRACT

This paper briefly surveys pose tracking methods used for augmented reality applications in cultural heritage. The paper primarily benefits scholars and practitioners in the areas of electronic heritage. Pose tracking techniques are categorized as either being dependent or independent of their surrounding; accordingly, various solution methods in the literature are presented along with their advantages and disadvantages. I conclude the paper with a discussion on the open problems in pose tracking in cultural heritage and recommend future directions of research in this field.

INTRODUCTION

The past two decades have witnessed a boom in the field of e-heritage, in the form of many virtual reality (VR) and augmented reality (AR) applications, with cultural heritage as their main theme. These applications were born from a need to provide more entertaining techniques to promote our heritage, and their solutions became more tractable with the increase in processing power of computers and handheld devices. Traditional techniques of showcasing relics at exhibits used to do little more than place the object of interest under appealing lighting conditions; but with the emergence of augmented and virtual reality applications, it is possible to turn an old and dusty scene into a live and colorful stage.

There is a significant difference between VR and AR. In the former case, the entire scene, including existing and imagined, is constructed as a computer model, and the user views the complete scene through special goggles—such as the Oculus RIFT™—that block out the outside world and project a complete virtual 3D scene to user. In the latter case of augmented reality, the user views the real world with a virtual scene overlaid on top. To realize this, the user requires an AR device—such as the Google Glass™ or the Microsoft Hololens™.

AR is significantly more difficult than VR, due to several reasons, including the requirement for proper meshing of the existing and virtual scene at their intersection, as well as the need for accurate tracking

DOI: 10.4018/978-1-5225-5469-1.ch038

methods of the pose of the user's interface—such as an outward looking camera of a handheld device—in order to project the augmented scene on the interface from the correct perspective. In VR applications, pose tracking is not required since user motion is a control to the system, rather than a sensory input.

In this paper, I will focus on the pose-tracking problem in AR for cultural heritage. There are a number of surveys that treat the AR problem alone, but to my knowledge, none of them focus on AR for cultural heritage. For example, Bostanci, Kanwal & Clark (2015) put forward a survey covering pose estimation for AR applications, and group the various systems based on an indoor/outdoor taxonomy; but they do not restrict their survey to CH. The reason it is important to make the distinction is that CH sites are usually controlled by stringent rules, which do not allow the addition of infrastructure that might change the appearance of the sight in any fashion. The implication is that tracking systems relying on an infrastructure (e.g., Vicon®, markers, etc.) are generally not permitted, and the designer of the pose tracking system has to rely on the more difficult infrastructureless approach. Furthermore, although CH sites include indoor settings, many CH sites are located outdoors, in dynamically changing lighting conditions due varying sunlight conditions. As a result, the tracking systems that are used must be robust, and capable of dealing, to some degree, with these variations.

The remainder of this paper is structured as follows. In the `Background' section, I will first describe the different techniques for pose tracking that could be applied to any AR problem, then in the `Related Work' section, I will list and critically assess the solutions adopted by different people in the e-heritage community. Finally, in the `Discussion' and `Conclusions' sections, I will present my perspective on the state of pose estimation in cultural heritage, and the direction it should take, and then conclude the paper.

BACKGROUND

Pose estimation in AR applications for CH can be broadly categorized as either being dependent or independent of its surrounding environment. In the former case, the pose of the agent is calculated according to an interaction between sensors (e.g., cameras) and artificial beacons (e.g., markers), each placed either in the environment or on the user. In the latter case, the agent is equipped with sensors that are used for positioning, without any required infrastructure to support the localization process.

Dependent Pose Estimation

Dependent pose estimation is performed in either a user-centered, or environment-centered approach: in the first case the user is equipped with a sensor that can detect artificial landmarks or beacons, placed around the environment. In the latter case, the setup is reversed, where the environment is retrofitted with sensors and markers are placed on the agent; the basic idea is shown in Figure 1, where a number of cameras track salient features (small spherical balls) worn by the user, in order to triangulate the position of each feature. Typically, for an AR application, one of the spheres could be attached to the camera to be tracked.

Examples of such systems include the Vicon®, and Optitrack®; which operate in the infrared spectrum, where each camera is equipped with a ring light directed towards the target. By measuring the position of the target from several vantage points, it is possible to triangulate the location of the target. It is usually the practice to use at least six cameras in order to reduce the possibility of occlusion of any target as it moves around the scene. As the dimension of the scene grows, so does the number of required cameras.