Chapter 27
3D Camera Tracking for Mixed Reality using Multi-Sensors Technology

Fakhreddine Ababsa
University of Evry Val d'Essonne, France

Iman Maissa Zendjebil
University of Evry Val d'Essonne, France

Jean-Yves Didier
University of Evry Val d'Essonne, France

ABSTRACT

The concept of Mixed Reality (MR) aims at completing our perception of the real world, by adding fictitious elements that are not perceptible naturally such as: computer generated images, virtual objects, texts, symbols, graphics, sounds, smells, et cetera. One of the major challenges for efficient Mixed Reality system is to ensure the spatiotemporal coherence of the augmented scene between the virtual and the real objects. The quality of the Real/Virtual registration depends mainly on the accuracy of the 3D camera pose estimation. The goal of this chapter is to provide an overview on the recent multi-sensor fusion approaches used in Mixed Reality systems for the 3D camera tracking. We describe the main sensors used in those approaches and we detail the issues surrounding their use (calibration process, fusion strategies, etc.). We include the description of some Mixed Reality techniques developed these last years and which use multi-sensor technology. Finally, we highlight new directions and open problems in this research field.

DOI: 10.4018/978-1-61350-326-3.ch027

INTRODUCTION

In MR applications the vision-based approaches are often used to achieve the camera tracking. Vision-based techniques estimate the camera pose using only the visual information extracted from the acquired images. In most MR applications, camera tracking remains a difficult task which must be accurate and stable. It is known that a non-robust tracking or not enough accurate can generate a "jitter" effect on the Real/Virtual registration, and often leads to tracking failure. In order to deal with this problem, some MR systems use artificial markers, called also fiducials. The main idea consists in placing in the environment several markers among which the content, the size, the position and the orientation are known by the system. By using image processing methods, the MR systems can then extract and identify the markers and thus localize the camera. However, theses methods suffer generally from a lack of accuracy when the markers are occluded or in the case of blurring effect generating by abrupt motion of the camera. Other MR systems use Markerless tracking approaches in order to estimate the camera pose. The principle consists in using salient geometric features (points, edges, silhouettes) existing naturally in the scene. In this case, the registration between the Real and Virtual worlds is realized thanks to the alignment of the 2D information extracted from the images with the 3D model of the scene. These methods usually give a more precise solution than marker-based techniques. However, their main disadvantage lies in the reliability of the 2D-3D matching process. Indeed, an erroneous matching would engender false camera pose estimation. Furthermore, vision-based approaches remain very sensitive to working conditions. Their performances decrease significantly when they are used in uncontrolled environments where situations such as change in brightness, occlusions and sudden motion arise rather often. Multi-sensors techniques which combine various technologies and methods seem to open a new way to resolve the lack of robustness of vision-based methods. They generally fuse a vision-based tracking approach with measurements obtained from localization sensors (inertial, GPS, etc.) to compensate for the shortcomings of each technology when used alone.

The objective of this chapter is to present some original solutions which use multi-sensors technology in order to estimate the camera localization.

STATE OF THE ART

The idea of combining several kinds of sensors is not recent. The first multi-sensors system appeared with robotic applications where, for example, Vieville et al. (1993) proposed to combine a camera with an inertial sensor to automatically correct the path of an autonomous mobile robot. This idea has been exploited these last years by the community of Mixed Reality. Several works proposed to fuse vision and inertial data sensors, using a Kalman filter (You et al., 1999) (Ribo et al., 2002) (Hol et al., 2006) (Reitmayr & Drummond, 2006) (Bleser & Stricker, 2008) or a particular filter (Ababsa et al., 2003) (Ababsa & Mallem, 2007). The strategy consists in merging all data from all sensors to localize the camera following a prediction/correction model. The data provided by inertial sensors (gyroscopes, magnetometers, etc.) are generally used to predict the 3D motion of the camera which is then adjusted and refined using the vision-based techniques. The Kalman filter is generally implemented to perform the data fusion. Kalman filter is a recursive filter that estimates the state of a linear dynamic system from a series of noisy measurements. Recursive estimation means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. So, no history of observations and/or estimates is required.

You et al. (1999) developed a hybrid sensor combining a vision system with three gyroscopes