

Chapter 7

Multi-View RGB-D Synchronized Video Acquisition and Temporally Coherent 3D Animation Reconstruction Using Multiple Kinects

Naveed Ahmed
University of Sharjah, UAE

ABSTRACT

This chapter introduces a system for acquiring synchronized multi-view color and depth (RGB-D) video data using multiple off-the-shelf Microsoft Kinect and methods for reconstructing temporally coherent 3D animation from the multi-view RGB-D video data. The acquisition system is very cost-effective and provides a complete software-based synchronization of the camera system. It is shown that the data acquired by this framework can be registered in a global coordinate system and then can be used to reconstruct the 360-degree 3D animation of a dynamic scene. In addition, a number of algorithms to reconstruct a temporally-coherent representation of a 3D animation without using any template model or a-prior assumption about the underlying surface are also presented. It is shown that despite some limitations imposed by the hardware for the synchronous acquisition of the data, a reasonably accurate reconstruction of the animated 3D geometry can be obtained that can be used in a number of applications.

INTRODUCTION

Temporally coherent time-varying dynamic scene geometry has been employed in a number of applications. It can be used for 3D animation in digital entertainment productions, electronic games, 3D television, motion analysis, gesture recognition etc. First step in obtaining temporally coherent 3D video is to capture the shape, appearance and motion of a dynamic real-world object. One or more video cameras are employed for this acquisition, but unfortunately, data obtained by these video cameras has no temporal

DOI: 10.4018/978-1-5225-1025-3.ch007

Multi-View RGB-D Synchronized Video Acquisition

consistency, as there is no relationship between the consecutive frames of a video stream. In addition, for a multi-view video, all the cameras have to be synchronized to extract temporal correspondences at each frame of the video. This synchronization is typically achieved by means of a hardware-based camera trigger, which acts as an external synchronizer. From the acquired synchronized data, in order to reconstruct a temporally coherent 3D animation, a spatial structure between cameras has to be established along with the temporal matching over the complete video data.

In this chapter, a system for acquiring synchronized dynamic 3D data using multiple RGB-D cameras along with three new method for capturing spatio-temporal coherence between RGB-D images captured from multiple RGB-D video cameras are presented. Synchronized multi-view video (MVV) data is used in a number of applications, e.g. motion capture, dynamic scene reconstruction, free-viewpoint video etc. Traditionally, the MVV recordings are acquired using synchronized color (RGB) cameras, which are later processed for use in a number of applications (Aguiar et al., 2008; Carranza et al., 2003; Starck et al., 2007; Theobalt et al., 2007; Vlasic et al., 2008). The acquisition setups used for these earlier works comprised of a dedicated system for capturing synchronous high quality RGB MVV recordings, which were then used to reconstruct dynamic 3D scene representation.

One of the earlier works in this area was presented by Carranza et al. (2003), who used eight multi-view recordings to reconstruct the motion and shape of a moving subject and applied it in the area of free-viewpoint video reconstruction. Theobalt et al. (2007) extended this work so that in addition to capturing the shape and motion they also captured surface reflectance properties of a dynamic object. Starck et al. (2007) presented a high quality surface reconstruction method that could capture detailed moving geometry from multi-view video recordings. Later de Aguiar et al. (2008) and Vlasic et al. (2008) presented new method for reconstructing really high quality of dynamic scene using multi-view video recordings. Both of their methods first obtained the shape of the real world object using a laser scanner and then deformed the shape to reconstruct the 3D animation. Ahmed et al. (2008) presented a method of dynamic scene reconstruction with time coherent information without the use of any template geometry, but unlike one of the presented method they did not explicitly include multiple matching criteria for extracting time coherence in their method.

With the arrival of depth sensors, especially low cost Microsoft Kinect (Microsoft, 2010), there has been a wave of interest in incorporating them in a number of research areas including the reconstruction of dynamic scene geometry. For example, one or more depth sensors are employed in 3D shape scanning and dense 3D reconstruction of static objects; pose, motion and 3D shape estimation (Baak et al., 2011; Berger et al., 2011; Weiss et al, 2011). These works show that despite the limitation of depth sensors, i.e. low resolution and high noise, it is possible to employ them to get high quality results. Recently, two methods have been presented to capture data using multiple depth sensors. Kim et al. (2008) presented a system for the fusion and calibration of RGB and depth sensors. Their system uses a dedicated hardware setup for the synchronization of the color and depth cameras. Recently, Berger et al. (2011) presented a method to capture motion using four Kinects but without any active synchronization between the sensors. These methods do not try to extract any spatial or temporal coherence information from the acquired dynamic data. More recently, Ye et al. (2014), and Zollhöfer et al. (2014) presented probabilistic methods for shape and pose estimations of dynamic objects using a single RGB-D camera. Unlike one of the presented methods, both of these methods deform a template mesh to achieve 3D animation reconstruction. On the other hand, the presented methods do not rely on any a-priori model and is more general because the 3D geometry is directly acquired from the RGB-D video data.

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/multi-view-rgb-d-synchronized-video-acquisition-and-temporally-coherent-3d-animation-reconstruction-using-multiple-kinects/170217

Related Content

The Combination of Adaptive Filters to Improve the Quality of Medical Images in New Wavelet Domain

Nguyen Thanh Binhand Vo Thi Hong Tuyet (2016). *Classification and Clustering in Biomedical Signal Processing* (pp. 46-76).

www.irma-international.org/chapter/the-combination-of-adaptive-filters-to-improve-the-quality-of-medical-images-in-new-wavelet-domain/149382

Local Linear Regression on Hybrid Eigenfaces for Pose Invariant Face Recognition

Ajay Jaiswal, Nitin Kumarand R. K. Agrawal (2012). *International Journal of Computer Vision and Image Processing* (pp. 48-58).

www.irma-international.org/article/local-linear-regression-hybrid-eigenfaces/72315

Content Based Image Retrieval Using Active-Nets

David García Pérez, Antonio Mosquera, Stefano Berrettiand Alberto Del Bimbo (2009). *Artificial Intelligence for Maximizing Content Based Image Retrieval* (pp. 85-114).

www.irma-international.org/chapter/content-based-image-retrieval-using/4153

Supporting Spatial Cognition in Vascular Visualization

Timo Ropinski, Jennis Meyer-Spradow, Frank Steinickeand Klaus Hinrichs (2008). *User Centered Design for Medical Visualization* (pp. 114-135).

www.irma-international.org/chapter/supporting-spatial-cognition-vascular-visualization/30627

A Vision-Based Approach for Indian Sign Language Recognition

Karishma Dixitand Anand Singh Jalal (2012). *International Journal of Computer Vision and Image Processing* (pp. 25-36).

www.irma-international.org/article/vision-based-approach-indian-sign/75768