# An integrated platform for live 3D human reconstruction and motion capturing

Dimitrios S. Alexiadis, Anargyros Chatzitoﬁs, Nikolaos Zioulis, Olga Zoidi, Georgios Louizis, Dimitrios Zarpalas and Petros Daras, Senior Member, IEEE

### Abstract

A robust, fast reconstruction method from multiple RGB-D streams is then proposed, based on an enhanced variation of the volumetric Fourier Transform based method, parallelized on the GPU, accompanied with an appropriate texture mapping algorithm. A novel framework is proposed for the quantitative evaluation of real-time 3D reconstruction systems. Finally, a generic, multiple depth streams-based method for accurate real-time human skeleton tracking is proposed. Detailed experimental results with multiple Kinect v2s datasets verify the validity of the arguments and the effectiveness of the proposed system and methodologies.

### Introduction

This paper describes the main elements of an integrated platform, including capturing and fast 3D reconstruction of human 3D shape/appearance and skeleton-based motion tracking, which targets TI and future 3D applications. The elements of the continuously being developed platform has already allowed the realization of a number of relevant applications, as in http://vcl.iti.gr/3dTI/, i) ski competition among users spread around Europe [3], ii) 3D hang-out communications [4], iii) multi-player networked 3D games (“SpaceWars” and “Castle in the Forest”), where users participate via their on the-ﬂy reconstructed 3D “replicants”, and iv) athletes’ training via professionals’ performance capturing and reconstruction for “quick-post” 4D media. Additionally, the paper describes a novel framework for the objective evaluation of the 3D reconstruction process, where the 3D ground-truth model is not available, as in real-time reconstruction applications.

### Capturing and Calibration

##### Capturing system

The devices are placed on a circle radius of 2m to 4m all pointing to the center of the captured area. K = 4 Kinect v2 sensors were used for optimality conditions of complexity and coverage area.

The Kinect v2 only operates with a computer per each sensor in master – slave configuration. System has 2 operation modes: Real – time and quick – post. First one continuously polls the slave nodes with the half of the sensor’s clock interval. The latter, gathers the data to the master after the recording sessions is closed. These modes enable either i) on-line 3D reconstruction, thus making it suitable for TI applications when combined with real-time efﬁcient data compression, or ii) temporally complete and higher quality results, exploiting all the recorded data in a quick-post processing step.

##### Calibration

Spatial (external) calibration of the sensors is achieved through a novel registration method, utilizing an easy-to-build calibration structure that serves as a registration “anchor”. The registration is performed separately for each sensor, with respect to that “anchor”, using an exact digital replica of the calibration object. The approach is based on the Scale Invariant Feature Transform (SIFT) and Procrustes Analysis.

The calibration structure is realized with 4 standardized IKEA package boxes, of size 56 x 33 x 41 cm3 as well as 32 unique Quick Response (QR) markers of area 13 x 13 cm2, placed at the corners of the boxes’ side faces. It is positioned at approximately the center of the capturing space, so that a) the full object is in the ﬁeld-of-view of all sensors and b) in the case of Kinect v1, to eliminate any bias, as in this case the noise increases with the measured surface’s distance.

##### Quick – post synchronization

To synchronize the data recorded during the quickpost operation mode, a post-synchronization procedure is employed. Each sensor continuously acquires pairs of timestamped depth and color images. While it is not the exact case, the depth and color components are synchronized (in practice they are synchronized up to 16msec) and therefore the depth timestamps are used. Each Kinect generates timestamps according to its local time-line Tk. An audio synchronization scheme is used to place the local timelines onto a global one. Audio signals of speciﬁc duration are simultaneously recorded from each sensor. Let the audio signal from the k-th Kinect be denoted as Ak(t). Its delay with respect to the reference Kinect k0 is calculated by ˆ dk = argmax(Rk,k0(d)), where Rk,k0(d) is the cross correlation of audio signals Ak(t) and Ak0(t). From these delays, the audio timestamp offsets ˇ Tk are obtained, which are used to place the local time-lines Tk onto the reference one T.