A Low-Cost Motion Capture System using Synchronized Azure Kinect Systems

Jongoh Jeong, Yue Wang, Mili Shah*
{jeong2,Yue.Wang,mili.shah}@cooper.edu
The Cooper Union
New York, New York

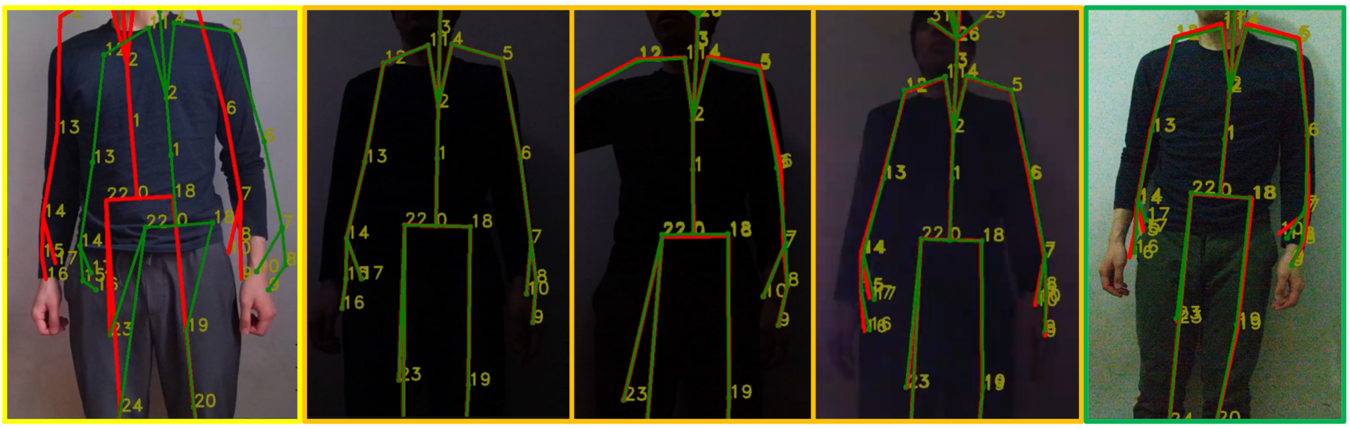

Figure 1: From left to right: (a) estimates from single (red) and weighted average (green) without transformation, (b) occlusion at subordinate #0, (c) occlusion at subordinate #1, (d) effect of direct light on master device, (e) final synchronized streams with offset due to parallax (red: single-device, green: 3-device system)

1 INTRODUCTION

Capturing three-dimensional body joint data is key to understanding human body movement and there are a great number of potential applications where motion capture (MOCAP) systems may be of effective use. Two general approaches to developing a motion capture system are (1) marker-based approach where active (LED-emitting) or passive (reflective) markers can be used, and (2) marker-less approach. Some camera-based professional MOCAP systems, such as Vicon and Qualisys, use passive markers that are visible in infrared (IR) cameras, while marker-less systems do not require additional materials besides cameras . For instance, the OpenPose algorithm from CMU[1] processes two-dimensional videos to for joint estimation, and the Azure Kinect uses RGB and IR cameras create a three-dimensional point in its camera space.

Some potential problems with professional MOCAP systems include cost and maintenance, and those with single marker-less MOCAP camera include occlusion and lighting effects that often disrupt the precision of joint estimation. To mitigate these issues, we developed a MOCAP system with higher precision than a single-camera system, while also having a lower cost and increased portability compared to a conventional MOCAP system by leveraging the body-tracking SDK on multiple synchronized Kinects.

2 APPROACH

We first configure the devices in a daisy-chain, or RS-232, method whereby we physically set the master and subordinate mode of each unit. Then we calibrate each pair of Kinects using the extrinsics computed by least-squares fitting of the two 3-D joint positions [1]. From this transformed subordinate joints positions, we select the data from devices that output with higher confidence, and take the weighted average of each joint's 3-D positions if there is more than one joint of that confidence level. We then perform perspective projection on the averaged homogeneous 3-D estimates to visualize in 2-D master camera view, using the master intrinsic parameters [2].

In order to verify the comparable functionality of our synced system to that of a single-device system, we tested in environments with occlusion and illumination variations. Maintaining the same posture with each camera either fully visible or blocked in view and also an additional light source directed at the master camera, we observed that there was no significant difference between the two settings. This implies that the system is selecting the joint data from each device adaptively by its confidence in estimation.

We hope to apply our low-cost synchronized MOCAP system to exoskeleton gait analysis, human body-controlled drone tracking, and graphic tool development for interactive avatars. Moreover, more precise joint angle estimation can be made, extending a single-device estimation by Islam *et al.* .

REFERENCES

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. 1987. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 5 (1987), 698–700.

[2] Wilhelm Burger. 2016. *Zhang's Camera Calibration Algorithm: In-Depth Tutorial and Implementation.* Technical Report. 10–11 pages.

*Advisor; Project repository: *https://github.com/andyj1/kinect*.
[1] https://github.com/CMU-Perceptual-Computing-Lab/openpose