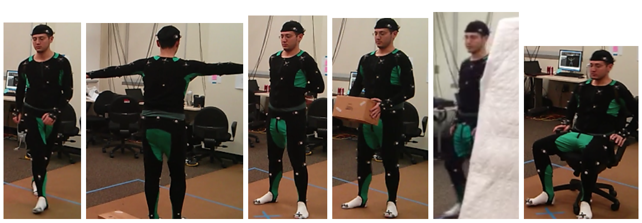
Accuracy Analysis of Skeleton Trackers for Safety in HRI

Akansel Cosgun, Martin Bünger and Henrik I. Christensen

***Abstract*— We present an evaluation study of two popular Kinect-based skeleton tracking software: OpenNI NITE and Microsoft Kinect SDK. We report the tracking robustness and joint error on the human skeleton for different activities such as walking, turning, sitting and holding an object.**

I. Introduction

With the recent introduction of RGB-D sensors such as the Kinect, skeleton tracking algorithms that provide the positions of important joints on the human body boosted person tracking research. Skeleton trackers facilitate several Human-Robot Interaction applications such as teleoperation, gesture recognition and object hand-overs. Two trackers have been popular in robotics community: OpenNI NITE [1] and Microsoft Kinect SDK [2]. While these trackers perform reasonably well at first glance, experiments are usually conducted in controlled environments (i.e. static sensor, human is facing the sensor). However, in real environments, human body can be in different poses depending on the activity, and that results in tracking errors. If the robot and the human are co-existing in the same space, the robot should have an understanding of how reliable its human perception system is, and plan accordingly to ensure safety. Robot assistants that help help humans will appear in factory floors in the future. Human workers are likely to be in different body configurations during work. We therefore evaluate both trackers in 6 different scenarios (Figure 1).



(a) (b) (c) (d) (e) (f)

Fig. 1. Scenarios that we have tested. a)**Walk:** Walking around in a circle while facing the Kinect for measuring performance in basic movement in the scene. b)**360:** Two full rotations around self with arms up and down, for tracking performance in different body orientations. c)**Hide:** Hiding one arm at the time behind the back and then hiding both at the same time, to explore how the two systems handles partial occlusions. d)**Box:** Picking up a box from the floor and extending it towards the sensor, for understanding how the system reacts when a object is introduced to the scene. e)**Occ:** Full body occlusion by hiding behind an obstacle and coming back, to measure if trackers can re-initialize. f)**Sit:** Sitting down for several seconds and then standing up, for tracking performance of sitting humans.

A. Cosgun and H. Christensen are with Center for Robotics and Intelligent Machines, Georgia Tech, Atlanta, GA, USA.

This work was funded by the BMW corporation.

Joint	System	Walk	360	Hide	Box	Occ	Sit
Head	MSSDK	15.8	17.2	16.1	13.2	**32.5**	14.8
	NITE	**10.6**	**11.8**	**13.3**	**12.2**	76.2	**11.0**
Neck	MSSDK	11.2	14.5	10.9	**8.5**	**31.8**	7.7
	NITE	**4.6**	**4.9**	**3.2**	10.5	76.6	**5.7**
Torso	MSSDK	**4.4**	**5.9**	**3.9**	**10.1**	**30.7**	**7.5**
	NITE	6.7	6.7	6.8	15.5	82.0	12.1
Shoulders	MSSDK	7.8	**16.8**	**5.8**	9.3	**34.6**	**7.9**
	NITE	**5.6**	18.6	7.1	**8.7**	82.4	9.2
Elbows	MSSDK	9.6	**28.6**	7.6	**6.4**	**42.5**	**8.7**
	NITE	**9.0**	32.0	**7.4**	9.1	78.5	11.1
Hands	MSSDK	14.8	**47.3**	15.6	**12.2**	**52.9**	**14.1**
	NITE	14.8	50.2	**11.0**	15.9	84.7	14.2

TABLE I

Joint tracking error in centimeters

Joint	System	Walk	360	Hide	Box	Occ	Sit
Head	MSSDK	100	100	100	100	66	100
	NITE	91	99	93	82	83	68
Neck	MSSDK	100	100	100	100	67	100
	NITE	100	100	100	100	83	99
Torso	MSSDK	100	100	100	100	67	100
	NITE	100	100	100	100	83	99
Shoulders	MSSDK	100	73	100	100	60	99
	NITE	100	100	100	100	83	99
Elbows	MSSDK	100	75	100	96	57	74
	NITE	99	90	100	71	59	95
Hands	MSSDK	100	76	58	48	47	86
	NITE	99	88	100	70	58	94

TABLE II

Reported percentage of tracked frames

II. Approach

We compare joint errors and robustness of OpenNI and MS-SDK. Both trackers report 3 joint states: tracked, inferred and not tracked. By robustness, we mean the ratio of frames that a joint has 'tracked' state to all frames. Ground truth data was extracted from Vicon motion capture system. Joint error is the Euclidean error between the ground truth to measured position.

III. Results

OpenNI performed better for head and neck, whereas MS-SDK was more accurate for other joints (Table I). Occlusions make tracking much worse and trackers confuse left/right hands when user rotates around self. Upper body joints for Box and Sitting scenarios had comparable accuracy to Walk scenario. Both trackers claimed that the joints are tracked well in most frames (Table II). However, given that average joints errors are usually more than 5 cm, we conclude both trackers are optimistic in its tracking capabilities. Therefore additional measures in addition to skeleton tracking should be considered for safe HRI.

References

[1] Openni nite 1.5, http://www.openni.org/files/nite, 2013.

[2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR*, 2011.