Validation of Accuracy of the *Super Pop* VR^{TM} Kinematic Assessment Methodology using Markerless versus Marker-Based Motion Capture Systems

Sergio García-Vergara[1], Paul Robinette [1], Yu-Ping Chen[2], and Ayanna M. Howard[1]

[1]Georgia Institute of Technology, Electrical and Computer Engineering Department, 85 5th Street NW, Atlanta, GA 30332
sergio.garcia@gatech.edu, paul.robinette@gtri.gatech.edu, ayanna.howard@ece.gatech.edu

[2]Georgia State University, College of Health & Human Science, Division of Physical Therapy, Atlanta, GA 30302 USA
ypchen@gsu.edu

Abstract. Therapists and clinicians have been combining virtual reality (VR) systems for rehabilitation purposes with motion capture systems to accurately keep track of the users' movements and better analyze their kinematic performance. The current state-of-the-art motion capture technology is limited to the clinical setting due to its cost, the necessity for a controlled environment, requirement of additional equipment, among others. Given the benefits of home-based rehabilitation protocols, more portable and cost-effective technology is being coupled with the VR systems. In this work, we focus on validating the accuracy of the Kinect™ camera from Microsoft. We compare its performance to a current state-of-the-art motion capture system. Namely, we 1) analyze the difference between the outcome metrics computed with data collected with the Kinect™ camera and the outcome metrics computed with data collected with the motion capture system, and 2) compare the spatial trajectories generated by both systems for the hand, elbow, and shoulder joints. Data were collected from ten able-bodied adults to quantify these comparisons. In general, results from both analyzes support the validity and feasibility of using the Kinect™ camera for home-based rehabilitation purposes.

Keywords: Technological Rehabilitation · *Super Pop* VR^{TM} · Physical Therapy · Kinect™ vs. OptiTrack · Home-based Rehabilitation

1 Introduction

In the spirit of using virtual reality (VR) systems for rehabilitation purposes, therapists and clinicians have been combining these VR systems with motion capture systems to keep track of the user's movements. Currently, the most precise technology are marker-based systems, like the Vicon and OptiTrack systems. These use optical sensors to track reflective markers attached to the user's body and determine the three-dimensional (3D) coordinates of the markers, thus being able to locate the position of the user's body and limbs. Although shown to be highly accurate, these marker-based systems are limited when applied to the home setting. Some limitations include: cost, the necessity for a controlled environment, and the time required for marker placement. As such, marker-based motion capture systems have generally been limited to the clinical setting.

This is one of the main reasons we decided to use the Kinect$^{\mathrm{TM}}$ camera from Microsoft as the 3D depth capturing system for our VR system. In this work we quantify the accuracy of the Kinect camera as a motion capture system. We validate the precision of the device and study the feasibility of using it for rehabilitation purposes by comparing it to a marker-based motion capture system: the OptiTrack system. Previous studies have shown that the Kinect can achieve competitive motion tracking performance as the OptiTrack [1]. Although positive results were achieved, these studies focus only on spatial comparisons. In this work we further investigate the Kinect's capabilities specific to our proposed system and kinematic parameters of interest, namely for the assessment of an individual's upper-body kinematic progress in a home environment.

2 System Description

We focus this work on upper-body motor skills, of which the most dominant form is reaching movements which are correlated to various rehabilitation scenarios. The ability to reach is critical for most, if not all, activities of daily living such as feeding, grooming, and dressing [2]. In general, a reaching movement or exercise requires an individual to move their arm from an initial position to a target position. While interacting with the proposed system, it evaluates the user's kinematic performance and provides corrective feedback with respect to a set of defined reaching tasks. For this work we define the reaching exercises such that they are correlated to the virtual platform as described in the following section.

2.1 Virtual Environment

In order to enable a non-biased data collection process for randomized trials, we employed a platform called *Super Pop VR*^{TM} [3, 4], a motivating virtual reality game used to track upper-body movements using a three-dimensional depth camera - we make use of the Kinect camera from Microsoft. During game play, the user moves their arms to complete a set of reaching tasks. The movements are mapped into the virtual environment which is displayed on a computer or projecting screen. The objective of the game is for the user to move their arms to 'pop' the virtual bubbles that appear on the screen. More details and images about the setup and the system can be found in [3–5].

For this work we define a reaching task as the arm displacement generated while moving one's hand from 'popping' one bubble towards 'popping' another in the *Super Pop VR*^{TM} environment. To map the required reaching exercises to the game's virtual environment, the game employs a special type of virtual bubbles called 'Super Bubbles' (SBs). At various instances during game play, a set of SBs is displayed on screen prompting the user to move their dominant hand and complete the reaching task. An example of a reaching task in the Super Pop environment is shown in Figure 1. Namely, the reaching task requires users to raise their dominant hand to 'pop' the START bubble directly above their dominant shoulder, then make a downward movement to 'pop' the second SB, and finally reach the TARGET bubble located to their side, thus creating a 90° movement.

The positions of the three SBs are the same for all users, but the distance between them is a function of the length of the user's dominant arm. Before each game, the users are asked to raise their arms as high as they can to calibrate the game settings. The Kinect camera captures the three-dimensional Cartesian coordinates of the user's dominant hand and shoulder, and positions the SBs on the circumference of the circle centered at the user's shoulder which radius equals the length of the user's dominant arm as shown in Figure 1.

Fig. 1: Example of a 90° trajectory created by the position of the three Super Bubbles.

3 Hypothesis

We determine the Kinect camera's feasibility by comparing its performance to a state-of-the-art marker-based system, the OptiTrack system. Our objective is to evaluate our system's ability, using the Kinect, to yield results similar to those yielded by the OptiTrack system. In this manner, we can thus support our claim that the Kinect camera can be used as a motion capture system for home-based rehabilitation as well as identify any limitations with its use. The hypotheses of this study are:

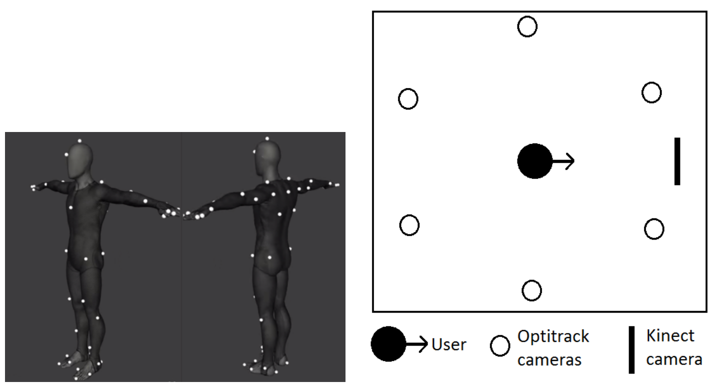
1. The differences between the outcome metrics computed with data collected with the Kinect camera and the outcome metrics computed with data collected with the OptiTrack system are negligible. This is to say that the

differences are small enough such that we consider the results from the two systems to be equal to each other.

2. The trajectories generated with Kinect and OptiTrack data are spatially similar to each other for the hand, elbow, and shoulder joints.

4 Experimental Design

Ten able-bodied adults were recruited to interact with the *Super Pop VR*^{TM} system. Six females and four males ranging in age between 24 and 31 years old played the game. All participants were asked to complete a 90° reaching task (as described in Figure 1) ten times for each arm. Their interactions were recorded with both the Kinect camera and the OptiTrack system. The participants were asked to wear a non-infrared reflective suit with passive infrared markers attached to it for the extraction of the OptiTrack data. The reflective suit had 37 markers in total (Figure 2a). Our OptiTrack setup consists of six *Flex 3* cameras. The layout of the testing environment is shown in Figure 2b. Details about how to calibrate the OptiTrack cameras can be found in [6].



(a) Marker Configuration

(b) Layout for Testing Environment

Fig. 2: (a) IR marker configuration, and (b) layout showing placement of the user and the Kinect and OptiTrack cameras.

To validate the first hypothesis, we focus on the kinematic parameters that depend only on the tracking systems: path length (PL), and elbow and shoulder range of motion (EROM and SROM respectively). We do not consider the parameters that depend on additional external variables because these can potentially introduce errors that are not derived from the two systems directly. For example, the movement time parameter depends only on the computer's system clock, thus it is independent of the tracking systems' capabilities. For each participant, we computed a percent error difference between the outcome metrics computed with Kinect data and the outcome metrics computed with OptiTrack data. We computed a percent error value for each movement task the participant completed and averaged the values per arm. The final percent error difference per participant is computed using (1).

$$\mathrm{PE}^{p,\alpha,\beta} = \frac{1}{n} \sum_{t=1}^{n} \frac{\left| V_K^{p,\alpha,\beta}(t) - V_O^{p,\alpha,\beta}(t) \right|}{\mathrm{nf}^{\beta}} \, 100\% \tag{1}$$

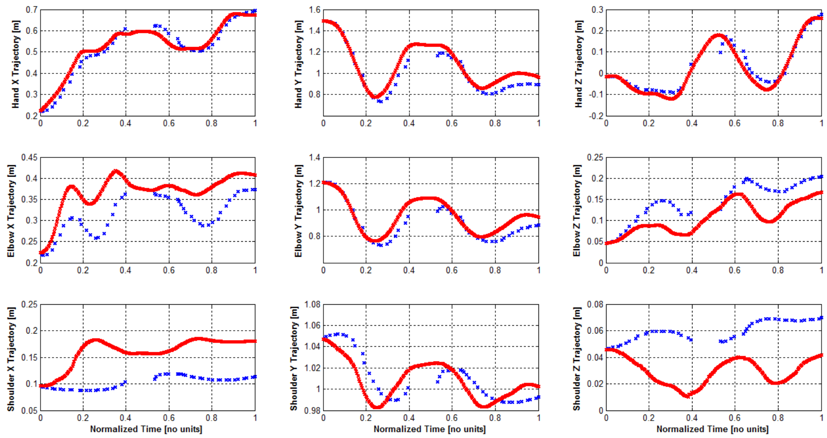
where $\mathrm{PE}^{p,\alpha,\beta}$ is participant p's average percent error difference of parameter β for n completed movements tasks with arm α, $V_K^{p,\alpha,\beta}(t)$ and $V_O^{p,\alpha,\beta}(t)$ are the outcome measures computed with Kinect and OptiTrack data respectively for participant p, parameter β and trial t, and nf^{β} is the normalization factor for parameter β. Each parameter has its own normalization factor. The EROM and SROM parameters are normalized with respect to their maximum allowed ROM (150° and 180° respectively [7]). Given that there is no maximum value allowed for the PL parameter (i.e. any trajectory can have an infinite length in theory), its percent error is normalized with respect to the value computed with OptiTrack data since we consider it to be the ground truth value.

To validate the second hypothesis, we use the deviation from path (DfP) parameter to quantify the similarity between the trajectories generated with Kinect and OptiTrack data for the hand, elbow, and shoulder joints. For a given participant, we compute the area between the Kinect and OptiTrack curves for each completed movement task, each joint, and each arm. The final comparison between the two trajectories is the average of the areas from the movement tasks a given participant completed.

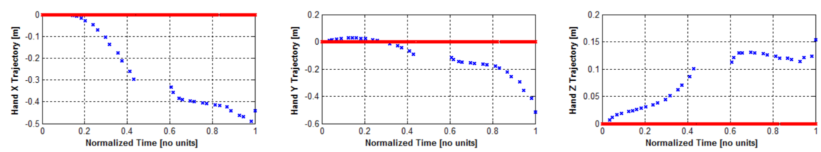
Before performing the computations, we first eliminate the trials with corrupt data. Corrupt data occurs when one or both of the motion capture systems loses track of the user's movements. For the OptiTrack, this happens when the cameras lose track of one or more of the suit's IR markers. For the Kinect, this happens when the camera loses track of one or more of the user's joints. When one or both of these events occur there is an incorrect estimate of the user's position, thus inaccurate joint coordinates are stored. Examples of trials without and with corrupt data are shown in Figure 3.

5 Experimental Results

The average percent error differences between the outcome measures computed with Kinect data and the outcome measures computed with OptiTrack data per participant for the PL, EROM, and SROM parameters are shown in Table 1. The average area values between the trajectories generated with Kinect and OptiTrack data per participant for the hand, elbow, and shoulder joints are shown in Table 2.



(a) Trajectories without corrupt data.



(b) Hand trajectory with corrupt OptiTrack data.

Fig. 3: Examples of Kinect (blue) and OptiTrack (red) trajectories without (top) and with (bottom) corrupt data.

6 Discussion and Conclusion

The percent errors averaged from all participants for the PL, EROM, and SROM parameters are relatively low (Table 1). The percent error differences for the EROM and SROM parameters range between 5% and 7% for both arms, while the percent error differences for the PL parameter range between 10.5% and 12.5% for both arms. These low values suggest that the differences between the outcome measures computed with Kinect data and the outcome measures computed with OptiTrack data are small enough such that we can conclude that the two capturing systems yield similar outcome measures. This supports our first hypothesis and our claim that we can substitute the marker-based motion tracking system with a more portable and cost-effective option. Namely, these results suggest that the Kinect is a viable option for home-based rehabilitation because it provides similar accuracy as the current state-of-the-art marker-based motion capture systems within 5-7% for the EROM and SROM parameters and 10.5-12.5% for the PL parameter.

The areas between the trajectories generated with Kinect and OptiTrack data for the hand, elbow, and shoulder joints averaged from all participants are

Table 1: Average percent errors for each parameter per participant.

	Right Arm PE [%]			**Left Arm PE [%]**		
Participants	**PL**	**EROM**	**SROM**	**PL**	**EROM**	**SROM**
1	15.1	1.9	2.0	12.9	7.1	0.8
2	-	-	-	-	-	-
3	8.3	3.3	3.0	10.7	3.2	4.4
4	19.4	5.0	9.0	34.5	5.1	9.5
5	12.3	3.7	2.2	10.2	6.5	1.0
6	4.3	2.8	3.4	3.7	2.2	1.4
7	6.6	9.7	7.6	-	-	-
8	5.2	13.1	20.1	5.9	5.9	7.6
9	4.2	8.8	13.3	6.3	14.4	11.8
10	19.0	6.1	3.0	15.7	5.9	7.5
AVG	10.5	6.0	7.1	12.5	6.3	5.5
STD	6.1	3.8	6.2	9.7	3.7	4.2

*PE: percent errors, PL: path length, EROM and SROM: elbow and shoulder range of motion.
**Missing values are due to all trials having corrupt data.

Table 2: Average areas as computed by the DfP parameter for the hand, elbow, and shoulder joints per participant.

	Right Arm DfP [10^{-3} m^2]			**Left Arm DfP** [10^{-3} m^2]		
Participants	**Hand**	**Elbow**	**Shoulder**	**Hand**	**Elbow**	**Shoulder**
1	62.2	48.4	38.9	48.1	46.8	19.8
2	-	-	-	-	-	-
3	68.2	34.3	41.2	79.2	47.1	39.7
4	111.4	56.1	36.0	167.8	62.9	25.2
5	95.1	57.3	42.4	132.5	62.2	52.1
6	80.5	53.7	53.1	76.5	49.1	41.7
7	58.6	94.5	45.3	-	-	-
8	108.4	38.5	37.9	116.1	50.9	34.3
9	102.8	91.5	93.6	73.7	88.0	63.7
10	116.1	50.1	37.0	175.9	120.2	41.6
AVG	89.3	58.3	47.3	108.7	65.9	39.8
STD	22.3	21.1	18.1	47.0	25.9	14.0

*DfP: deviation from path, PL: path length, EROM and SROM: elbow and shoulder range of motion.
**Missing values are due to all trials having corrupt data.

relatively low (Table 2). The results are in the order of less than $0.15m^2$. These low values support our second hypothesis that the trajectories generated by both tracking systems are spatially similar to each other. This further supports the validity and feasibility of using the Kinect camera for home-based rehabilitation purposes.

Acknowledgments. This work was supported in part by the NSF Grant 1208287. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A. Sawchuk, A. Rizzo, *et al.*, "Towards pervasive physical rehabilitation using microsoft kinect," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, pp. 159–162, IEEE, 2012.
2. Y.-P. Chen, L.-J. Kang, T.-Y. Chuang, J.-L. Doong, S.-J. Lee, M.-W. Tsai, S.-F. Jeng, and W.-H. Sung, "Use of virtual reality to improve upper-extremity control in children with cerebral palsy: a single-subject design," *Physical therapy*, vol. 87, no. 11, pp. 1441–1457, 2007.
3. S. García-Vergara, Y.-P. Chen, and A. M. Howard, "Super pop vrtm: An adaptable virtual reality game for upper-body rehabilitation," in *Virtual, Augmented and Mixed Reality. Systems and Applications*, pp. 40–49, Springer, 2013.
4. S. García-Vergara, L. Brown, H. W. Park, and A. M. Howard, "Engaging children in play therapy: the coupling of virtual reality games with social robotics," in *Technologies of Inclusive Well-Being*, pp. 139–163, Springer, 2014.
5. L. Brown, S. Garcia-Vergara, and A. M. Howard, "Evaluating the effect of robot feedback on motor skill performance in therapy games," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pp. 1060–1065, IEEE, 2015.
6. NaturalPoint, "Optitrack calibration," 2015. [Online; accessed 13-September-2015].
7. M. Sardelli, R. Z. Tashjian, and B. A. MacWilliams, "Functional elbow range of motion for contemporary tasks," *The Journal of Bone & Joint Surgery*, vol. 93, no. 5, pp. 471–477, 2011.