Validation of a passive infrared marker 3D-tracking technique using the Microsoft Kinect™

[1,2]Sena, MP; [1]Lee, R; [1]Ray, NA; [2]Coughlin, DG; [2]Feeley, BT; [2]Lotz, JC
[1]University of California, Berkeley, Berkeley, CA; [2]University of California, San Francisco, San Francisco, CA,
mark.sena@ucsf.edu

Introduction

Marker-based motion capture (MoCap) systems track the 3D position of either active strobing infrared (IR) LEDs, or passive infrared reflectors (IRRs) placed on a subject's body. The clinical efficacy of gait analysis using MoCap is well-supported[1]. For example, orthopaedic surgeons utilizing gait reports made significantly different treatment decisions than surgeons without such reports (by 20%, p<0.01), in terms of choosing whether or not to operate on children with cerebral palsy[2].

MoCap researchers today face many of the same practical challenges they did more than twenty years ago[3], including manual marker identification routines and laboratory costs in the range of hundreds of thousands of dollars. These difficulties have limited the number of patients that could benefit from quantitative motion analysis.

In an effort to expand the setting for, and reduce the cost of optical MoCap, we have developed a passive IR marker tracking technique that utilizes the Microsoft Kinect™ (<$150) as the 3D optical measurement device. We combine depth and IR image data from the Kinect sensor in order to determine the 3D position of passive IRRs placed on a subject. Human "pose" estimates from the Kinect are then used to automatically identify each marker. Here, we determine the accuracy and precision of our technique using the Optotrak Certus® system ($<1^0$ and <1 mm resolution, NDI, Ontario, CA) as a standard for comparison.

Materials and Methods

A shared reference coordinate system was defined by placing three IRRs on a stationary object ~3m in front of the Optotrak and Kinect (Fig. 1B). IRR positions were registered using a digitizing probe (Optotrak) and a manual digitizing routine (Kinect). Thin plastic discs (Ø12cm) served as rigid bodies for attachment of an Optotrak "Smart Marker" and an IRR (1 pair per disc). IRR positions on each disc were manually registered at the start of the experiment using the Optotrak probe. Markers were identified automatically by the Kinect using an algorithm developed by the authors. Since both active and passive markers were fixed on each disc, the tracking performance of the same rigid bodies (origins=IRRs) could be compared between both systems.

A male subject facing ~45° away from the Kinect performed four squats within an ~6s time span. Discs were attached to a subject's right hip, knee, and ankle using Velcro straps such that each IRR was roughly aligned with the iliac crest, lateral epicondyle, and lateral malleolus, respectively. Data collected by the Optotrak at 100Hz and by the Kinect at 30Hz were temporally matched by manual alignment.

Error analysis was performed for the hip, knee, and ankle markers during the 4-squat experiment. Systematic error was defined as the raw difference between the Kinect and Optotrak 3D position data ($\boldsymbol{S} = \boldsymbol{p}_{\text{kin}} - \boldsymbol{p}_{\text{optk}}$). Random error was defined as the difference between Kinect and *transformed*-Optotrak data ($\boldsymbol{R} = \boldsymbol{p}_{\text{kin}} - \mathbf{G}\boldsymbol{p}_{\text{optk}}$), where **G** is the single constant 4-by-4 homogeneous transformation matrix that minimized $\boldsymbol{R}$ for the hip, knee, and ankle. The maximum absolute value of $\boldsymbol{S}$ (max[|$\boldsymbol{S}$|]) and root-mean-square of $\boldsymbol{R}$ (RMS[$\boldsymbol{R}$]) were calculated, and an ANOVA with multiple comparisons and a Bonferroni adjustment were used to identify differences between their *x, y,* and *z* (*x/y/z*) components. Overall accuracy and precision were defined as the norm, ||•||, or magnitude of $\boldsymbol{S}$ and $\boldsymbol{R}$, respectively, averaged over N time points for which data could be compared (hip: N=160, knee: N=96, ankle: N=63).

Results

Means and standard deviation across the hip, knee, and ankle for the following systematic and random error metrics are given in Table 1.

<u>*Systematic*:</u> The maximum absolute difference between *x/y/z* values recorded by the Optotrak and Kinect (max[|$\boldsymbol{S}$|]), ranged from 10.8-35.0mm; and the time-averaged distance between the apparent Optotrak and Kinect positions (time-avg.||$\boldsymbol{S}$||), ranged from 18.1-22.6mm. Figure 1C shows a representative plot of ||$\boldsymbol{S}$|| versus time for the knee.

<u>*Random*:</u> The RMS deviation of the Kinect *x/y/z* data from the *transformed*-Optotrak data (RMS[$\boldsymbol{R}$]) ranged from 2.7-11.5mm. RMS deviation in *z* was significantly greater than that of *x* and *y* (p<0.005). The time-averaged noise, or effective RMS distance between the Kinect and transformed Optotrak-data (time-avg.||$\boldsymbol{R}$||) ranged from 9.5-10.9mm.

Table 1: Error metrics (mm) averaged over the hip, knee, and ankle

component	Systematic Error		Random Error	
	max[|$\boldsymbol{S}$|]	time-avg. ||$\boldsymbol{S}$||	RMS[$\boldsymbol{R}$]	time-avg. ||$\boldsymbol{R}$||
X	20.0 ± 8.5		4.8 ± 0.44	
Y	23.1 ± 2.6	20.3 ± 2.3	4.0 ± 1.2	10.3 ± 0.7
Z	33.1 ± 3.0 *		10.4 ± 1.0†	

mean ± standard deviation *p > 0.05 †p < 0.005 (Bonferroni adj.)

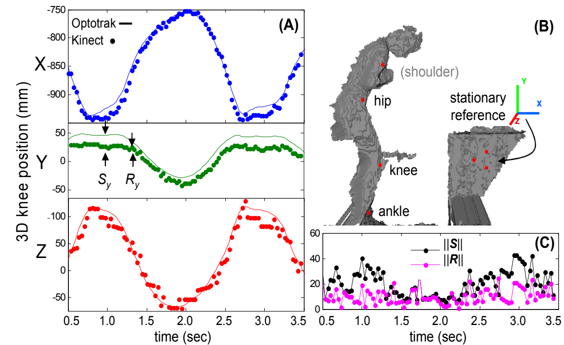

Figure 1: A) *x/y/z* components of the subject's knee position over the course of a two-legged squat (Optotrak—, Kinect●). B) 3D Kinect depth output with overlaid body and reference IR reflector positions (●). C) Knee: systematic (-●-) and random (-●-) error magnitudes versus time.

Discussion and Conclusion

These results demonstrate that by using our technique, it is possible to track multiple joint positions with an accuracy of 20±2.3 mm and a precision of 10.3±0.7mm. These values are larger than commercial MoCap errors; however, they are comparable to the magnitude of soft tissue artifact (up to 30mm)[4]. The finding that random noise in *z* was significantly greater (p<0.005) than that of *x*/*y* is consistent with the Kinect's lower depth resolution. Systematic error in *z* was also larger than that of *x*/*y*, but the difference was not significant (p=0.057).

Other low-cost human motion tracking systems have been developed for clinical applications using the Kinect[5] and the Nintendo Wii™[6]. Unlike our technique, however, these others require separate hardware, stationary camera positions, calibration, background subtraction, and data filtering before gait[5] and hand motion[6] analysis are possible. To the authors' knowledge, the technique presented here is the first that permits automatic 3D tracking and identification of multiple anatomical markers using only the Kinect, several passive IRRs, and a PC equipped with open-source Kinect libraries[7,8]. We acknowledge the relatively slow frame-capture rate (30Hz), 11-65% marker occlusion, and the lack of rigid body rotation data. Hardware improvements, the use of multiple Kinect units, and future studies will address these limitations.

The ability to perform 3D MoCap using the low-cost (<$150, as-sold) Microsoft Kinect™ may aid physicians and therapists in diagnosis, treatment, and monitoring of a range of musculoskeletal disorders in a clinical setting. Ergonomists and trainers, for example, might also use our technique for motion analysis in occupational and sports settings.

References

1. Wren, T. et al., *Gait Posture* **2011,** *34* (2), 149-153.
2. Wren, T. et al., *Gait Posture,* **2011**, 34(3), 364-369.
3. Davis, R. B., III, *IEEE Eng. Med. Biol. Mag.* **1988,** *7* (3), 35-40.
4. Peters, A. et al., *Gait Posture* **2010,** *31* (1), 1-8.
5. Stone, E. et al., Pervasive Health, Dublin, May 21-23, **2011**
6. Attygalle, S. et al., Virtual Rehab. 25-27 Aug. **2008**; pp 168-174.
7. http://www.openni.org/
8. Dirk-Jan Kroon, http://www.mathworks.com/matlabcentral/fileexchange/30242-kinect-matlab

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship (Sena, MP)