Robotics & Automation Engineering Journal
ISSN: 2577-2899

Juniper Publishers
key to the Researchers

Research Article
Volume 1 Issue 2 - September 2017
DOI: 10.19080/RAEJ.2017.01.555558

Robot Autom Eng J

A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming

Schlagenhauf F, Sahoo PP and Singhose W*

Department of Mechanical Engineering, Georgia Institute of Technology, Georgia

Submission: July 29, 2017; **Published:** September 08, 2017

***Corresponding author:** Singhose W, Department of Mechanical Engineering, Georgia Institute of Technology, Georgia, Email: Singhose@gatech.edu

Abstract

Motion capture systems can acquire human body motion for use in programming robotic applications that mimic human motion. Activities such as traffic control and aircraft marshalling could be automated using robots that replicate realistic human motions. In order to develop such robotic systems, natural human motion must be studied and parameterized. Microsoft Kinect V2 is a non-invasive, low-cost camera primarily used in the video gaming industry which can be used for human motion analysis. This study evaluates the joint-tracking abilities of a dual-Kinect system and compares it to a Vicon 3D Motion Capture system. The dual-Kinect system provides robust and accurate tracking of complex upper-body motions.

Keywords: Motion Capture; Kalman filter; Upper-Body motion; Kinect

Introduction

Figure 1: Robot Directing Traffic and Pedestrians (Adapted from [12]).

The robotics community is increasingly making use of human motion capture and joint tracking analysis. This study aims to assess the potential of employing a camera system to study upper-body joint motions by comparing the joint tracking abilities of a dual-Kinect system and a Vicon 3D Motion Capture system. Such motion data can be used to train a robot with data collected from humans performing representative motions. An example of such robotic applications is shown in Figure 1, which shows a robot being used in place of a police officer to direct pedestrians and traffic. Note that work related musculoskeletal disorders (WRMSDs) are a major issue plaguing traffic policemen and factory workers [1,2]. Muscular fatigue is induced due to long working hours, as well as incorrect or sub-optimal motion techniques [3].

Vicon systems use high definition cameras which, while being accurate, are expensive and have mobility constraints. On the other hand, the Microsoft Kinect V2 is a non-invasive, low-cost camera primarily used in the video game industry which can be used for human-body motion analysis. A primary differentiating factor between Kinect and Vicon is the necessity of retro-reflective markers in the Vicon system. However, the Kinect does not require markers for human-body tracking because a proprietary Microsoft Software Development Kit (SDK) possesses the ability to track human body joints. Furthermore, the Vicon system is composed of multiple cameras set up around the perimeter of the measurement workspace. On the other hand, the Kinect is a single sensor package that only sees what is directly in front of it.

Numerous studies evaluated the accuracy of skeleton and joint tracking using the first version of the Kinect V1 sensor (Kinect for Xbox 360) [4-6]. Many of these studies focused on gait analysis and, therefore, limit their scope to lower-body movements [7]. The Kinect V1 was compared to a Vicon 3D Motion Capture system to establish optical motion capture methods with respect to applications in ergonomics,

rehabilitation, and postural control [8-10]. Overall, these studies found that the Kinect V1's precision is less than the optical motion capture system, yet the Kinect has various advantages such as portability, markerless motion capture, and price.

To improve the Kinect V1's motion capture precision, some approaches used additional wearable inertial sensors [11]. With this approach, they were able to obtain more accurate joint angle measurements. This paper describes a method of improved markerless tracking of human upper-body motions using a dual-Kinect system. It establishes the accuracy of Kinect tracking by comparing tracked marker trajectories with end effector trajectories of a robot arm. Experiments evaluate the dual-Kinect's joint tracking abilities for human upper-body motions by closely studying the wrist trajectories [12].

The Kinect V2 baseline performance is evaluated in Section 4, 5 describes a novel dual-Kinect motion tracking system using sensor fusion and Kalman filtering and applies it to human upper-body motions. Finally, the paper Present's conclusion in Section 6.

Kinect V2 Baseline Performance Evaluation

This section evaluates the performance of the Kinect V2 system by testing its ability to track a known motion generated by a 3-degree-of-freedom robot. The end-effector trajectory was recorded using the Microsoft Kinect V2 for XBox One sampling at 30Hz.

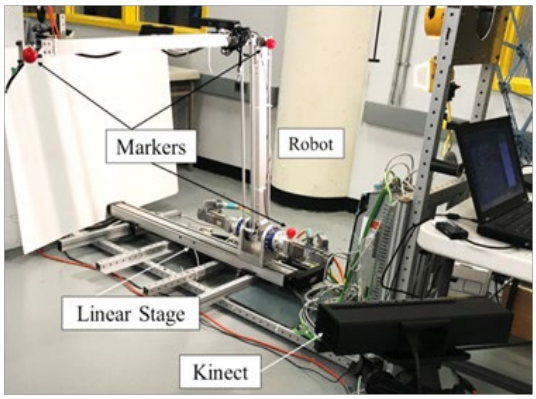

Figure 2: 3-DoF Robot Arm and Experimental Setup.

Figure 2 shows the robot arm and Kinect sensor. The system is controlled using Siemens Simotion drives and PLCs. A MATLAB code is used to generate the trajectory that the robot arm executes [13]. The Kinect is placed at an angle facing the robot arm. The robot arm has three red markers placed on the base, on the elbow, and on the end effector.

By default, the Kinect is able to track joints once it identifies a human body in its field of view. However, this experiment tracks the position of red markers. Given the absence of a human silhouette, the Kinect's built-in skeleton tracking algorithm could not be used. Therefore, depth data was collected from the Kinect sensor and processed using marker detection methods.

Red markers were chosen because red components can easily be detected in real time based on images acquired by the Kinect. Marker tracking data was captured in MATLAB using Kin2, a Kinect 2 toolbox [14]. A red color filter is first applied to the RGB frame, followed by a conversion to binary and circle detection. This process accurately locates the markers in the Kinect's field of view. Once the X and Y coordinates of the pixel that coincides with the centroid of the tracked marker are obtained, a built-in function provided by the Kinect for Windows SDK can be used to map between locations on the color image and their corresponding locations on the depth image. This function is used to obtain the X, Y, and Z coordinates of the tracked marker in 3D space.

The data was transformed so that the Kinect's coordinate system matched that of the robot arm. A tracked position of the end effector $P_{EE}^{K} = [xyz1]^{T}$, given in the Kinect's coordinate system, was transformed into the robot's coordinate system to yield P_{EE}^{R} using a 4×4 homogeneous transformation matrixA_R:

$$P_{EE}^{R} = A_R P_{EE}^{K} (1)$$

The robot's end effector trajectory was compared to the data collected by the Kinect. Three trials were conducted tracking a baseline trajectory that was developed for painting and sandblasting operations [13]. The test trajectory required the robot arm to move along the linear stage, while both the shoulder and the elbow joints were moved.

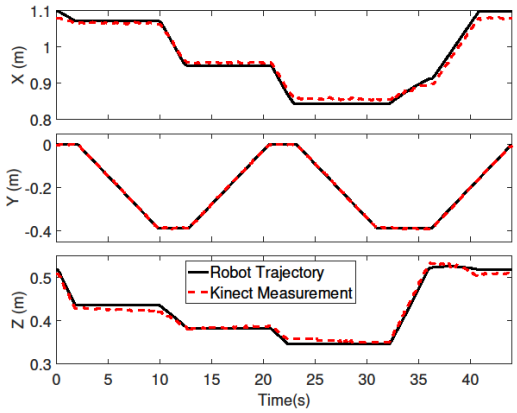

Figure 3: XYZ Coordinates of End Effector during Trial 1.

Figure 3 shows the XYZ coordinates of the end effector and their respective measured trajectories for trial 1. The Kinect tracks the Y coordinate almost perfectly. However, the X and Z coordinate show some deviation from the trajectory. This deviation is caused by the robot bending under the force of gravity, or by error in the alignment of the two coordinate systems.

How to cite this article: Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558. DOI: 10.19080/RAEJ.2017.01.555558.

The mean and maximum absolute errors between the Kinect measurements and the programmed robot trajectory were calculated for the three separate trials and are listed in Table 1. Mean absolute errors were smaller than 10mm for all trials, except the X component of trial 1, where the mean absolute error was 10.5mm. Generally, the robot trajectory could be tracked especially well in the Y direction. The tracking performance was worst in the X direction. Maximum absolute errors in the tracked position were smaller than 27mm for all trials.

Table 1: Mean absolute error (MAE), max. Absolute error, mean absolute deviation (MAD) and max. absolute deviation during tracking experiments. All values in mm.

	MAE			Max. abs. error		
	X	Y	Z	X	Y	Z
Trial 1	10.51	1.63	7.48	23.77	7.47	15.91
Trial 2	9.86	5.61	7.89	22.45	15.55	19.55
Trial 3	9.49	6.93	8.72	23.17	20.42	26.67
	MAD			**Max. abs. deviation**		
	X	Y	Z	X	Y	Z
Trial 1	1.67	4.36	1.76	9.47	13.57	15.03
Trial 2	1.42	2.46	0.87	8.24	10.23	5.33
Trial 3	1.55	3.42	1.75	7.95	11.3	12.13

The average end effector trajectory from the three trials was calculated. Figure 4 shows the comparison between the XYZ coordinates of the programmed end effector trajectory and the averaged Kinect tracking data.

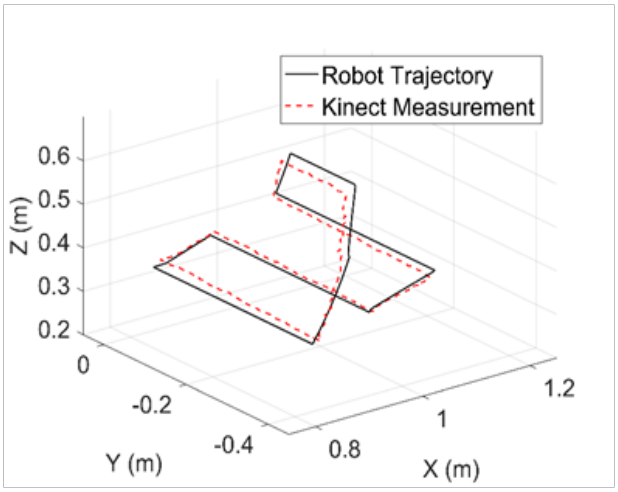

Figure 4: End Effector Trajectory in 3D Space.

To investigate the repeatability between the three different trials, the mean and maximum absolute deviation from the average Kinect measurement was calculated for each trial. The values are listed in the bottom portion of Table 2. The mean absolute deviations were smaller than 2mm for the X and Z component of the motion for all trials, and smaller than 4.4mm for tracking motion in the Y direction. The maximum absolute deviation was smallest for tracking the X component of the trajectory, and did not exceed 15.1mm for all trials. Thus, the tracking of the end effector position with Kinect was repeatable in the conducted experiments.

Table 2: Test Motion Definitions.

Number	Name	Description
1	Two-Handed Wave	Starting from T-Pose; Lift both arms vertically until hands are above the head
2	"Slow Down" Motion	With arms extended to the front; Lift both arms towards the shoulders while bending the elbows
3	Torso Twist	Rotate the torso about 90° to the left and right while lifting the hands to shoulder height with elbows bent

Dual-Kinect Tracking

Complex human motion can include rotational movement of the limbs and torso. Simple motions such as a single hand wave facing the Kinect can be tracked fairly accurately using a single-Kinect sensor; however, the more complex motions that are non-planar and involve rotation of the spine cannot be tracked as accurately. Data is lost when joints move out of the field of view or are occluded.

To improve tracking of complex motion, a second Kinect sensor was integrated into the system. In the dual-Kinect system, joint position data is acquired simultaneously with two Kinect sensors placed at a 90° angle to each other. Figure 5 shows this configuration and indicates the position of the test subject.

Tracking data of human motion was recorded at the Indoor Flight Facility at Georgia Tech. The room has 15 Vicon MX3+ cameras sampling at 100Hz. Data was concurrently obtained using the Microsoft Kinect V2. The software used to process tracking data was Vicon Nexus 2.5, Vicon Body Builder 3.6.4, and the Microsoft SDK provided to Kinect developers.

How to cite this article: Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558. DOI: 10.19080/RAEJ.2017.01.555558.

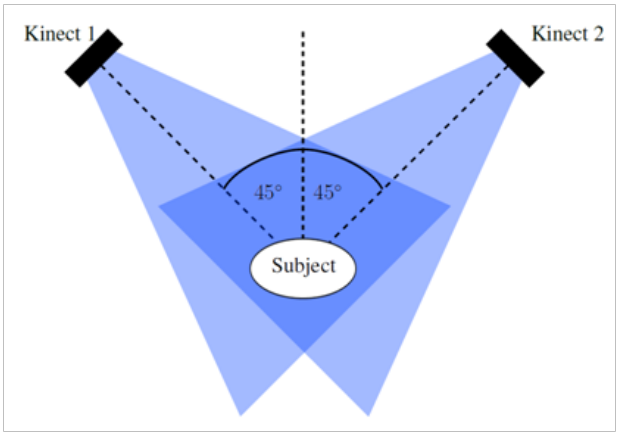

Figure 5: Dual-Kinect Configuration.

The subject for this study was a female university graduate research assistant. Thirty Nine (39) markers were placed on the test subject according to the Plug-In Gait Model for Vicon tracking and the subject stood at a distance of 2 meters from the Kinect sensors. A coordinate transformation was used to transform the captured data into a common coordinate system. The joint position data was subsequently fused and filtered using a linear Kalman filter.

Coordinate transformation

Rapid, real-time calibration of the two Kinects, without any additional calibration objects, can be accomplished using the initial 3D position estimates of the joints. To ensure no joint occlusion, the test subject is required to stand with straight legs and both arms fully extended, pointing sideways in a T-Pose for less than two seconds, during which 50 frames are acquired by both Kinect sensors. Then, the joint position estimates are averaged and fed into the calibration algorithm, which is based on an approach similar to the multiple Kinect Calibration described by Córdova-Esparza et al. [15]. The coordinate transformation is calculated via Corresponding Point Set Registration [16].

Considering two sets of 3D points, Set_A and Set_B, with Set_A given in coordinate frame 1 and Set_B given in coordinate frame 2, solving for R and t from:

$$Set_A = R \bullet Set_B + t \quad (2)$$

Yields the rotation matrix R and translation vector t needed to transform the points from coordinate frame 2 into coordinate frame 1. The problem of finding the optimal rigid transformation matrix can be divided into the following steps:

- Find the centroids of both datasets.
- Bring both datasets to the origin.
- Find the optimal rotation R.
- Find the translation vector t.

The rotation matrix R is found using Singular Value Decomposition (SVD). Given N Points P_A and P_B from dataset P_A and P_B respectively, with P=[x y z]^T, the centroids of both datasets are calculated using:

$$centroid_A = \frac{1}{N}\sum_{i=1}^{N} P_A^i \quad (3)$$

$$centroid_B = \frac{1}{N}\sum_{i=1}^{N} P_B^i \quad (4)$$

The rotation matrix R is:

$$R = VU^T \quad (5)$$

$$H = \sum_{i=1}^{N}\left(P_A^i - centroid_A\right)\left(P_B^i - centroid_B\right) \quad (6)$$

$$[U,S,V] = SVD(H) \quad (7)$$

The translation vector t can then be found using:

$$t = -R \bullet centroid_B + centroid_A \quad (8)$$

Using the rotation matrix and translation vector, thejoint position data from Kinect 2 can be transformed intothe coordinate system of Kinect 1.

Data Fusion

The joint positions collected from both Kinects are used to calculate a weighted fused measurement. In addition to the 3D coordinates of the joints, the Kinect sensor assigns a tracking state to each of the joints, with 0 = 'Not Tracked',1 = 'Inferred', 2 = 'Tracked'. This information is used to intelligently fuse the data collected by both Kinects. If the tracking state of a joint is 'Tracked' by both Kinects, or the tracking state of the joint is 'Inferred' in both Kinects, then the average position is taken. If a joint is 'Tracked' by one Kinect, but 'Inferred' or 'Not Tracked' by the other, then the fused position only uses data from the 'Tracked' joint.

The fused position p_{fused} of each joint can becalculated using the position estimates p_1 from Kinect 1 and p_2 from Kinect 2:

$$p_{fused} = w_1 p_1 + w_2 p_2 \quad (9)$$

The weighting factors w_1 and w_2 are assigned using the trackingstate information for each joint obtained from both Kinects:

$$w_1 = \frac{TrackingState1}{TrackingState1 + TrackingState2} \quad (10)$$

$$w_2 = \frac{TrackingState2}{TrackingState1 + TrackingState2} \quad (11)$$

After completing the coordinate transformation and sensor fusion steps described in the previous sections, the fused joint position is fed into a Kalman filter as a measurement.

If it can be assumed that the tracked joints are executing linear motion, then the linear Kalman filter can be used to estimate the states. A commonly used example of discrete time state space modeling of an object in 3D space is presented in

How to cite this article: Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558. DOI: 10.19080/RAEJ.2017.01.555558.

[17]. For the sake of simplicity, the equations are derived to track a single joint's position.

When the velocity of the joint is zero, the statevector for a problem with three spatial dimensions is given by $s = [xyz]^T$ and the state space model is:

$$s_{k+1} = As_k + w_k \quad (12)$$

$$z_k = Cs_k + v_k \quad (13)$$

w_k Is the process noise, and v_k is the measurement noise, and the state transition matrix is given by:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (14)$$

It is assumed that the process and measurement noises are zero-mean, Gaussian noise vectors. The observation matrix C takes into account the observed coordinates of the joint position and is given by:

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (15)$$

Figure 6 shows the left wrist position during a set of three test motions: a two-handed wave motion (from t=0s to t=10s), a "slow down" motion (from t=10s to t=22s), and a torso twist (from t=22s to t=42s). Each motion was performed five times by the test subject before moving on to the next test motion. Table 2 fully describes the test motions performed in this study.

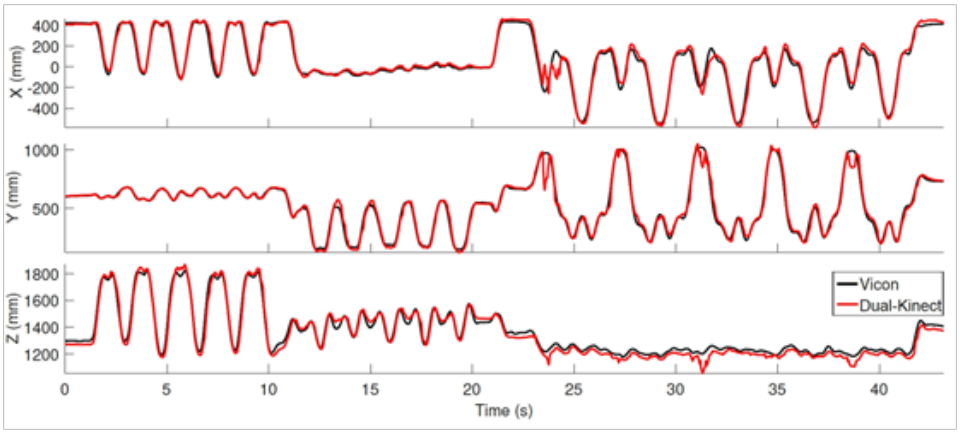

Figure 6: Left Wrist Joint Trajectory from Dual-Kinect compared to Vicon measurements.

The left wrist position was closely tracked for the wave motion and the "slow down" motion. During the torso twist motion, however, there was some discrepancy between Kinect and Vicon tracking data for extreme positions, when the wrist moved out of the field of view of both Kinect sensors. Generally, the wrist position could be tracked well by the dual-Kinect system for the majority of the test motions.

Conclusion

A baseline performance evaluation of the Kinect's depth tracking abilities showed that, in general, the Kinect can track motions in a reliable manner, and with acceptable accuracy for capturing data to program a robot through motions mimicking simple human upper-body motion. However, due to reliance on only one viewpoint, occlusion can lead to problems while tracking complex human motions. In order to overcome this problem and further improve joint tracking, a dual-Kinect system was developed. The proposed setup offers a low-cost, markerless, and portable alternative to marker-based motion tracking. It also eliminates the disadvantage of tedious marker setup and subject preparation time of a marker-based system.

Acknowledgment

We would like to thank Dr. Mark Costello and Seth Burdette for their assistance in the use of the Indoor Flight Facility at Georgia Tech for Vicon experiments.

References

1. Phadke SSD, Revati R, Iqbal R (2015) Work Related Musculoskeletal Symptoms among Traffic Police: Cross Sectional Survey Using Nordic Musculoskeletal Questionnaire.
2. Ha C, Roquelaure Y, Leclerc A, Touranchet A, Goldberg M, et al. (2009) The French Musculoskeletal Disorders Surveillance Program: Pays de la Loire network. Occup Environ Med 66(7): 471-479.
3. Ramsey JG, Musolin CPEK, Mueller C (2014) Evaluation of carpal tunnel syndrome and other musculoskeletal disorders among employees at a poultry processing plant. National Institute for Occupational Safety and Health, Health Hazard Evaluation, Report 2014-0040, p. 3232.
4. Mobini A, Behzadipour S, Foumani MS (2014) Accuracy of Kinect's skeleton tracking for upper body rehabilitation applications. Disabil Rehabil Assist Technol 9(4): 344-352.
5. Schmitz A, Ye M, Shapiro R, Yang R, Noehren B (2014) Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. J Biomech 47(2): 587- 591.

How to cite this article: Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558. DOI: 10.19080/RAEJ.2017.01.555558.

6. Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, et al. (2014) Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture 39(4): 1062-1068.
7. Pfister A, West AM, Bronner S, Noah JA (2014) Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. Journal of Medical Engineering & Technology 38(5): 274-280.
8. Fernández-Baena A, Susín A, Lligadas X (2012) Biomechanical Validation of Upper-Body and Lower-Body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments. 4Th International Conference on Intelligent Networking and Collaborative Systems, pp. 656-661.
9. Clark RA, Pua YH, Fortin K, Ritchie C, Webster KE, et al. (2012) Validity of the Microsoft Kinect for assessment of postural control. Gait Posture 36(3): 372- 377.
10. Martin CC (2012) A real-time ergonomic monitoring system using the Microsoft Kinect. IEEE Systems and Information Engineering Design Symposium, p. 50-55.
11. Destelle F (2014) Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors. in 22nd European Signal Processing Conference (EUSIPCO) pp. 371-375.
12. Sokol B (2014) Kinshasa's traffic robots: 'I thought it was some kind of joke- in pictures. ed.
13. Harber JA (2016) A Dual Hoist Robot Crane for Large Area sensing. Georgia Institute of Technology.
14. Terven JR, Córdova-Esparza DM (2016) Kin2 A Kinect 2 toolbox for MATLAB. Science of Computer Programming 130: 97-106.
15. Córdova-Esparza DM, Terven JR, Jiménez-Hernández H, Vázquez-Cervantes A, Herrera-Navarro AM, et al. (2016) Multiple Kinect V2 Calibration. Automatika. Journal for Control, Measurement, Electronics Computing and Communications 57(3): 2016.
16. Besl PJ, McKay ND (1992) Method for registration of 3-D shapes. Ed pp. 586-606.
17. Stohne V (2014) Real-time filtering for human pose estimation using multiple Kinects. KTH, School of Computer Science and Communication (CSC).

DOI: 10.19080/RAEJ.2017.01.555558

How to cite this article: Schlagenhauf F, Sahoo P P, Singhose W. A Comparison of Dual-Kinect and Vicon Tracking of Human Motion for Use in Robotic Motion Programming. Robot Autom Eng J. 2017; 1(2): 555558. DOI: 10.19080/RAEJ.2017.01.555558.