Precision Testing of a Single Depth Camera System for Skeletal Joint Detection and Tracking

Ketan Anand[1], Kathiresan Senthil[1], Nikhil George[2] and Viswanath Talasila[2]

[1] Dept. of Electronics and Instrumentation Engineering
[2] Dept. of Electronics and Telecommunication Engineering, Center for Imaging Technologies
M. S. Ramaiah Institute of Technology, Bangalore
anand_ketan@outlook.com
S_kathiresan@outlook.com

Abstract. Implementing a marker-less, camera-based system, to capture skeletal joint motion in a timely manner that is robust to variables such as lighting conditions, occlusion and changes in the subject's distance from the camera remains a challenge. This paper aims to test the feasibility of such a system using the Intel RealSense D435i depth camera to capture joint coordinates in the camera reference frame. The subject was placed in various lighting conditions to analyze the effect of glare on joint coordinate precision, a decimated filter and a spatial filter were applied to test which of these filters reduced noise in the image signal most effectively. To examine which setup is the most effective in measuring coordinates of the neck, elbows, wrists, knees and ankles precisely, variations were made in the subject's distance from the camera, position in the field of view and ambient lighting conditions. The empirical cumulative distribution function (ECDF) plots of the true distance obtained from the data collected were used to systematically eliminate undesirable recording conditions. The coordinates measured will be used to calculate gait parameters.

Keywords: Biomechanics, Gait analysis, Joint modelling, Medical Imaging, Motion Capture, Data Analysis, Depth camera.

1 Introduction

The emerging domain of tele-rehabilitation aims to eliminate bulky equipment traditionally used in the rehabilitation of those with motor impairment. The first stage of our proposed setup of a compact, simplified module for tele-rehabilitation involves a depth camera which detects the joint coordinates of the subject. This paper focuses on testing the feasibility of using the Intel RealSense D435i depth camera along with the *cubemos* Skeleton Tracking SDK to obtain the 3D coordinates of the subject's joints. Joint detection and tracking are vital for measuring gait parameters. The system proposed in this paper will subsequently be used to analyze the gait of children of the age of 6-12 years, affected with cerebral palsy. A marker-less setup to track skeletal joints is especially beneficial in the process of rehabilitating those with sensorimotor impairment such as cerebral palsy. This system is easily integrable with inertial measurement units

This work was supported by a BIRAC grant Ref. No. BT/AIR094S/PACE-19/19

(IMUs) to develop a comprehensive gait analysis tool. Gait analysis involves measurement of joint angles, as well as joint velocity and acceleration in 3D [1]. These gait measurements are used as an important diagnostic tool to decide the course of rehabilitation treatment in those with movement disorders.

Precision of the joint coordinates that are being measured is of importance since the data obtained is used in subsequent stages of the system proposed by us, to calculate diagnostic parameters such as gait cycles and base of support. Precision is an indicator of the reproducibility of data collected, over a given range. One metric used to determine the goodness of precision is the standard deviation, where a smaller standard deviation indicates a higher degree of precision.

The factors that were evaluated in testing the robustness of our depth camera-based system were: placing the subject in varied lighting, evaluating the performance of *cubemos* while the subject is at varied distances from the depth camera and testing various signal filters to reduce the standard deviation and improve the robustness of the system [2].

A thorough investigation of the precision of the Intel RealSense D3435i depth camera along with the *cubemos* Skeleton Tracking SDK was done by first performing camera calibration (Sec. 2.1) in indoor conditions. Further, carrying out an evaluation of the factors that influence the precision of joint coordinates (Sec. 3) helped eliminate unfavorable conditions, that would affect the robustness of our gait analysis system. The quantitative results for the measured neck, right wrist, and right ankle coordinates are presented there after (Sec. 4), followed by plotting the data in the form of an Empirical Cumulative Distribution Function (ECDF) to visualize the range covered by the measured data in various scenarios (Sec. 5).

2 Methodology

2.1 Camera Calibration

The process of calibrating the depth camera ensures that lens distortion does not worsen the precision of our setup, hence the reproducibility of the coordinates is maintained.

It was also noted that points in the corner of the field of view had negligible effect on the standard deviation of any of the coordinates. Since precision of coordinates was maintained in all regions of the field of view, calibration with the extrinsic parameters was not performed and hence, manipulating the camera coordinates with the intrinsic calibration parameters was sufficient.

Intrinsic Parameters. The intrinsic parameters of the camera are a transformation of the form $R^3 \rightarrow R^2$. This transformation maps 3D camera coordinates to 2D pixel coordinates [3]. The intrinsic parameter model of the Intel RealSense D435i camera is [4]:

$$Proj(x, y, z) = F.D_{Model}\left(\frac{x}{z} + \frac{y}{z}\right) + P \tag{1}$$

where, $D_{Model}: R^2 \to R^2$ is the lens distortion function, $P = (p_{x,} p_y)$ is the principal point indicating pixel offset from the left edge, and $F = (f_{x,} f_y)$ is the focal length in multiples of pixel size.

The intrinsic parameter model given above holds true for unfiltered data only. On application of signal filters such as the decimated filter or spatial edge-preserving filter, the intrinsic parameters are modified [5]. The description and controls of these filters are listed under Sec. 3.3.

Deprojection Model. The deprojection model uses the depth coordinate along with the pixel coordinates to reconstruct any point in focus, in camera coordinates. Obtaining these camera coordinates gives us the flexibility to estimate gait parameters in various biomechanical frames of reference. The deprojection model for the Intel RealSense D435i is given as [6]:

$$Deproj(i, j, d) = \left(d. U_{Model}\left(\frac{(i,j)-P}{F}\right), d\right) \tag{2}$$

where, $U_{Model}: R^2 \to R^2$ is the undo lens distortion function.

3 Evaluation

3.1 Effect of Lighting Conditions

Stereo cameras such as the Intel RealSense D435i utilize two infrared cameras and an onboard embedded D4 ASIC [7] to stream live depth data. A disparity map is generated where objects closer to the camera will have a greater horizontal shift between the right and left camera stream, compared to objects that are farther away. This shift along the epipolar line is known as disparity. The disparity is mapped to a particular depth value by a method known as *correspondence*:

$$Depth = \frac{Baseline \times Focal\ Length}{Disparity} \tag{3}$$

where, $Baseline$ indicates the midpoint between the left and right camera lens and $Disparity$ is the difference in pixel value of the feature under consideration.

Improper lighting conditions, that are overexposed, causes speckles in the IR pattern that the stereo cameras rely on to compute the depth coordinate. Speckles decrease depth image density, which results in incorrect correspondence. Further, if the loss of depth image density is at joints that *cubemos* is trying to detect, the *cubemos* model fails to recognize features of interest, resulting in inaccurate coordinate values [8].

3.2 Effect of Distance of the Subject from the Camera

Disparity is inversely proportional to distance of the subject from the camera. At distances with a depth component beyond 7 meters the disparity becomes negligible and insensitive to changes in depth [9]. From the quantitative results presented in Sec. 4, it is observed that a subject positioned at approximately 3 meters from the camera is best suited for precise depth measurement. Beyond 7 meters, the negligible depth value results in failure of correspondence.

Subjects that are too close to the camera (< 2 meters) and not visible from head-to-toe present an equally challenging problem to our setup. When *cubemos* fails to detect a certain joint, the coordinates logged are sparse in nature.

3.3 Effect of Various Signal Filters

Image signal filters are used to reduce noise in the image data. The two filters implemented in testing our single depth camera system were the Decimated filter and the Spatial Edge-Preserving filter.

Decimated Filter. The decimated filter reduces the complexity of the depth scene by reducing the sampling frequency by an integer value. The linear scaling factor provided by the Intel RealSense post processing toolkit runs between kernel sizes of 2x2 to 8x8. Scaling the image stream with a 3x3 kernel worked best for skeleton tracking as presented in Sec. 4.

Spatial Edge-Preserving Filter. The spatial edge-preserving filter exploits the fact that an RGB image is a 2D manifold in a 5D space. The spatial filter defines a domain transform from $R^2 \rightarrow R^5$ [10]. A 5x5 kernel is used in this transformed space to perform smoothing and edge-preserving operations. The controls of the spatial edge-preserving filter include a filter magnitude value which indicates the number of filter iterations, the smooth alpha value defines the weight of the current pixel and the smooth delta value defines the depth gradient below which smoothing will be performed. The controls of the spatial filter were set as follows:

Table 3.3.1. Filter controls of the Spatial Edge-Preserving filter

Parameter	Range	Value
Filter magnitude	[1-5]	5
Smooth alpha	[0.25-1.0]	1
Smooth delta	[1-50]	50

4 Quantitative Results

A thorough evaluation of the factors that influence the quality of the joint coordinates helped formulate a method of data collection that had the least standard deviation. The tabular columns below (Table 4.1 to 4.3) contain statistics of the neck, right wrist and right ankle coordinates. The joint coordinates that are being measured are in the format $(x,\ y,\ z, true\_distance)$ where the $true\_distance$ of the subject from the camera origin is a preliminary indicator of the combined standard deviation of the $x,\ y,\ z$ coordinates. Quantitative analysis was performed for the coordinates of the neck, elbows, wrists, knee and ankle, and a subset of the data is presented in the tables below.

Table 4.1. Statistics of neck coordinates

	Actual distance from the camera = 3m		Actual distance from the camera = 7m		Background with glare		Evenly lit, opaque background	
	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial
Mean	3.143m	2.928m	7.19m	6.5m	2.907m	3.107m	3.177m	2.807m
Standard Deviation	0.025m	0.016m	0.088m	0.077m	0.047m	0.033m	0.031m	0.024m
Range	2.834m to 3.276m	2.86m to 3.001m	6.857m to 7.522m	6.166m to 6.674m	2.540m to 2.966m	2.95m to 3.338m	2.305m to 3.398m	2.642m to 2.966m

Table 4.2. Statistics of right wrist coordinates

	Actual distance from the camera = 3m		Actual distance from the camera = 7m		Background with glare		Evenly lit, opaque background	
	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial
Mean	3.33m	3.072m	6.961m	6.914m	3.009m	3.196m	3.382m	3.009m
Standard deviation	0.045m	0.019m	0.059m	0.063m	0.056m	0.047m	0.027m	0.021m
Range	3.102m to 3.46m	3.010m to 3.144m	6.735m to 7.16m	6.582m to 7.208m	2.849m to 3.218m	3.028m to 3.414m	3.172m to 3.55m	2.849m to 3.218m

Table 4.3. Statistics of right ankle coordinates

	Actual distance from the camera = 3m		Actual distance from the camera = 7m		Background with glare		Evenly lit, opaque background	
	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial	Decimated	Spatial
Mean	3.24m	3.067m	7.019m	6.884m	2.932m	3.173m	3.305m	2.828m
Standard Deviation	0.043m	0.013m	0.071m	0.068m	0.045m	0.036m	0.015m	0.028m
Range	3.032m to 4.035m	3.027m to 3.156m	6.81m to 7.307m	6.595m to 7.131m	2.724m to 3.883m	3.081m to 3.317m	3.134m to 3.527m	2.583m to 2.909m

5 Visualizing the Results

Depth Maps. The depth maps of the subject, as seen by the camera create a colour coded image of pixels at the same distance. These depth maps give us useful insight into the effect of distance and glare on camera perception. The raw normalized depth map, with the joints in focus help log data more efficiently. The $true\_distance$ calculated from the $(x,\ y,\ z)$ joint coordinates are plotted as an Empirical Cumulative Distribution Function (ECDF). The empirical distribution function is computed as:

$$F_n(t)\ =\ \frac{1}{n}\sum_{j=1}^{n} I_{\{Z_j \le t\}} \qquad (4)$$

where, I is the indicator function which has a value equal to 1 if the distance for which frequency is being computed is found in the data and 0 otherwise.

Visualizing the ECDF Plots of our Data. The ECDF plots in Fig. 5.1 and Fig. 5.2 represents the distribution of the right ankle coordinate data while the subject was at distance of 3m and 7m from the camera, respectively. The ECDF plot makes visualizing the range, mean and standard deviation more intuitive and making inferences simpler. The inferences drawn help us select the appropriate conditions to implement our setup in. The ECDF plots for the neck, elbows, wrists, knees and ankles aid in making sound inferences regarding appropriate recording conditions, and a subset of these plots are presented below.

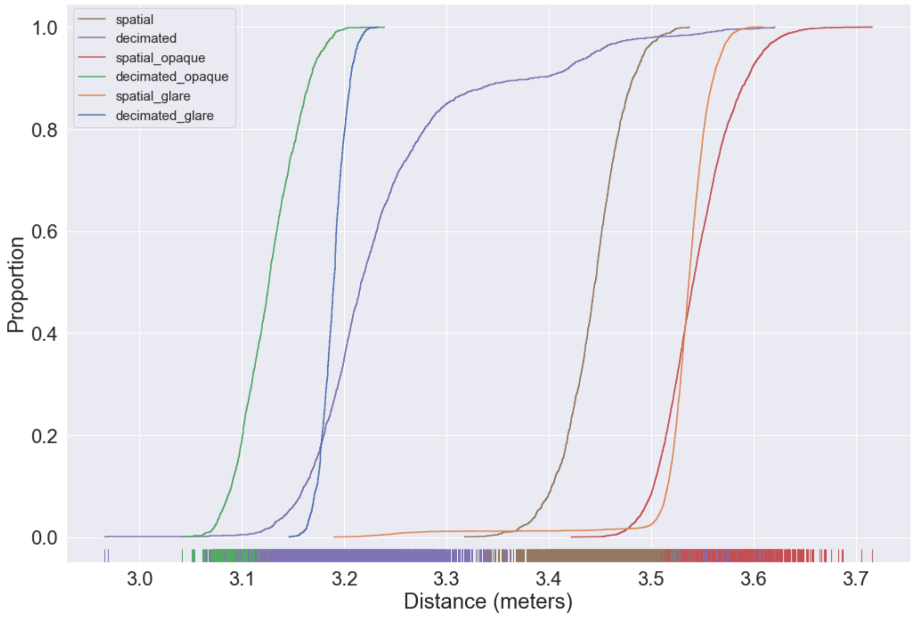

Fig. 5.1. ECDF plots of the right ankle coordinates logged at a distance of 3m from the camera

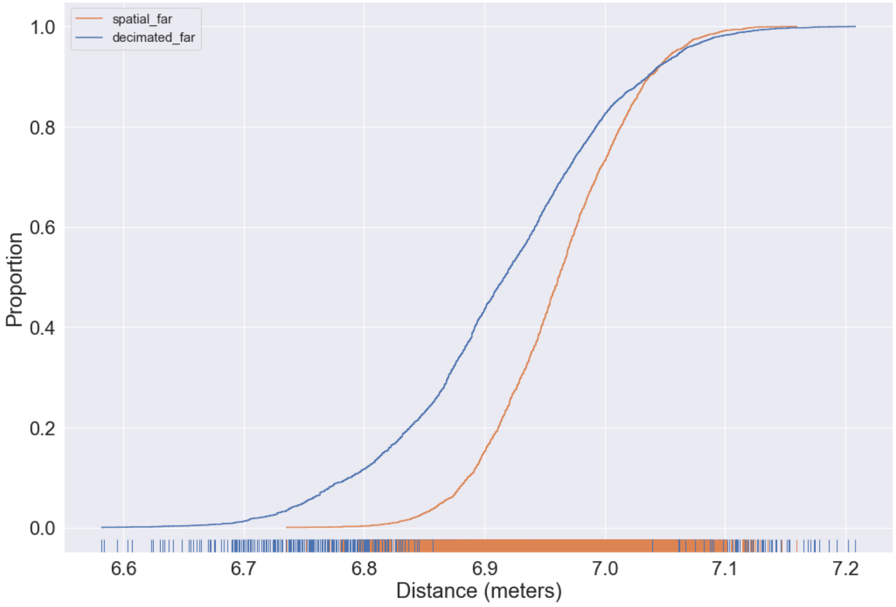

Fig. 5.2. ECDF plots of the right ankle coordinates logged at a distance of 7m from the camera

6 Conclusion

Rigorous analysis of the factors that could affect the performance of a single depth camera system for skeletal joint detection and tracking was performed. This yielded a set of conditions for joint coordinate measurement under which joint tracking was the most conducive. The ideal conditions for implementing our proposed setup are: no glare falling on the camera lens and the subject standing about 3m from the depth camera, while using a spatial edge-preserving filter to post process the raw depth image captured. The standard deviation of coordinates measured in favorable, controlled conditions was under 2cm. A high degree of precision ensures that the noise in the coordinate data is not amplified when the joint coordinates are used to calculate gait parameters, in our proposed tele-rehabilitation module. As observed from the ECDF plots and depth maps, that subjects placed at a distance greater than 7m from the camera and with overexposed lighting conditions, the precision decreases drastically. A controlled environment, that is obstacle free, with the subject clearly visible in the field of view of the camera provides the most suitable conditions to perform tele-rehabilitation. Precision testing of the Intel RealSense D435i depth camera with the *cubemos* Skeleton Tracking SDK completes the preliminary stage of our rehabilitation setup, using which data can be measured and calculation of vital gait parameters can be made [11]. In future work, we aim towards integrating this single depth camera system with IMUs to create a compact system that is capable of measuring data of a variety of gait parameters [12]. Further stages of the setup we are proposing would sense and model gait using this depth camera plus IMU based system, which would provide crucial metrics required for physical rehabilitation of gait [13].

References

1. Raghavendra P., Sachin M., Srinivas P.S., Talasila V. (2017) Design and Development of a Real-Time, Low-Cost IMU Based Human Motion Capture System. In: Vishwakarma H., Akashe S. (eds) Computing and Network Sustainability. Lecture Notes in Networks and Systems, vol 12. Springer, Singapore. https://doi.org/10.1007/978-981-10-3935-5_17.
2. L. C. Chin, S. N. Basah, S. Yaacob, M. Y. Din and Y. E. Juan, "Accuracy and reliability of optimum distance for high performance Kinect Sensor," 2015 2nd International Conference on Biomedical Engineering (ICoBE), 2015, pp. 1-7, doi: 10.1109/ICoBE.2015.7235927.
3. D. Herrera C., J. Kannala and J. Heikkilä, "Joint Depth and Color Camera Calibration with Distortion Correction," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 10, pp. 2058-2064, Oct. 2012, doi: 10.1109/TPAMI.2012.125.
4. Intel RealSense Documentation. https://dev.intelrealsense.com/docs/projection-texture-mapping-and-occlusion-with-intel-realsense-depth-cameras
5. Sander Schreven, Peter J. Beek, Jeroen B.J. Smeets, Optimising filtering parameters for a 3D motion analysis system, Journal of Electromyography and Kinesiology, Volume 25, Issue 5, 2015, Pages 808-814.
6. G. Balakrishnan, A. Dalca, A. Zhao, J. Guttag, F. Durand and W. Freeman, "Visual Deprojection: Probabilistic Recovery of Collapsed Dimensions," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 171-180, doi: 10.1109/ICCV.2019.00026.
7. Intel RealSense Documentation. https://dev.intelrealsense.com/docs/ post-processing-filters
8. Ling J., Tian L., Li C. (2016) 3D Human Activity Recognition Using Skeletal Data from RGBD Sensors. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2016. Lecture Notes in Computer Science, vol 10073. Springer, Cham. https://doi.org/10.1007/978-3-319-50832-0_14.
9. C. Høilund, T. B. Moeslund, C. B. Madsen and M. M. Trivedi, "Improving stereo camera depth measurements and benefiting from intermediate results," 2010 IEEE Intelligent Vehicles Symposium, 2010, pp. 935-940, doi: 10.1109/IVS.2010.5547978.
10. Gastal, Eduardo & Oliveira, Manuel. (2011). Domain Transform for Edge-Aware Image and Video Processing. ACM Trans. Graph. 30. 69. 10.1145/2010324.1964964.
11. S. Samprita, A. S. Koshy, V. N. Megharjun and V. Talasila, "LSTM-Based Analysis of a Hip-Hop Movement," 2020 6th International Conference on Control, Automation and Robotics (ICCAR), 2020, pp. 519-524, doi: 10.1109/ICCAR49639.2020.9108052.
12. K.R Vidyarani, Viswanath Talasila, N Megharjun, M Supriya, K.J Ravi Prasad, G.R Prashanth, An inertial sensing mechanism for measuring gait parameters and gait energy expenditure, Biomedical Signal Processing and Control, Volume 70, 2021, 103056, ISSN 1746-809, https://doi.org/10.1016/j.bspc.2021.103056.
13. V. N. Megharjun and V. Talasila, "A Kalman Filter based Full Body Gait Measurement System," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018, pp. 1-5, doi: 10.1109/CIMCA.2018.8739597.