Remote Health Coaching System and Human Motion Data Analysis for Physical Therapy with Microsoft Kinect

Qifei Wang & Gregorij Kurillo
University of California, Berkeley
Berkeley, CA 94720, USA
{qifei.wang, gregorij}@eecs.berkeley.edu

Ferda Ofli
Qatar Computing Research Institute
Doha, Qatar
fofli@qf.org.qa

Ruzena Bajcsy
University of California, Berkeley
Berkeley, CA 94720, USA
bajcsy@eecs.berkeley.edu

Abstract

This paper summarizes the recent progress we have made for the computer vision technologies in physical therapy with the accessible and affordable devices. We first introduce the remote health coaching system we build with Microsoft Kinect. Since the motion data captured by Kinect is noisy, we investigate the data accuracy of Kinect with respect to the high accuracy motion capture system. We also propose an outlier data removal algorithm based on the data distribution. In order to generate the kinematic parameter from the noisy data captured by Kinect, we propose a kinematic filtering algorithm based on Unscented Kalman Filter and the kinematic model of human skeleton. The proposed algorithm can obtain smooth kinematic parameter with reduced noise compared to the kinematic parameter generated from the raw motion data from Kinect.

1. Introduction

In the last decade, we witness increasing interests for remote monitoring technologies in health care. Remote monitoring captures the activity and other anthropometric data from the subjects. It can perform online analysis of these data, provide health-care feedback to the subjects monitored, and also represent the data to the doctor for further analysis. One typical application is in the physical therapy, especially for the elderly subjects, the doctors need to monitor the subjects physical performance at home during the training sessions and provide further treatments based on the subjects performance.

Human performance capture systems have been widely studied and applied in many applications. However, traditional motion capture systems require particular system setup and markers attached to the subjects monitored which make it impractical for recording motion data at home. Other motion sensors, like the inertia sensors are sensitive to the noise make them also not suitable for motion capture in human daily activities. The emerging accessible and affordable sensing technologies, e.g. the depth sensor, such as Microsoft Kinect, make it possible to capture the humans motion data at home [1].

Recently, we built an interactive exercise coaching system based on Microsoft Kinect [2] intended to encourage elderly adults to perform physical therapy at home. Since this type of sensors were not designed for medical purpose, the acquired motion data are usually noisy with low fidelity for the motion analysis. For example, the joint position data from Microsoft Kinect usually exhibits significant jitters caused by low depth accuracy, occlusions, ambiguity, and loss of tracking. The body segment lengths are also vary during the motion. In order to perform online analysis of subjects performance and feedback to the subjects, we further studied the motion data accuracy from Microsoft Kinect. We also studied the kinematic parameter extraction from the motion data captured by Kinect.

This paper is organized as following: Section 2 introduces the interactive system we have built for the remote physical therapy; Section 3 introduces the work we have done on motion data accuracy evaluation of Microsoft Kinect; Section 4 describes our work on the human kinematic parameters extraction from Kinect motion data; Conclusions are summarized in Section 5.
2. Interactive Exercise Coaching System

This system is built for automated exercise coaching of elderly users that would motivate them and track their physical performance when they perform physical therapy at home. The system contains four main modules: data acquisition, data processing, performance evaluation, and feedback. The pipeline of this system is shown in Fig. 1.

In the data acquisition module, we use the Microsoft Kinect camera to record the subjects' motion. Microsoft Kinect captures both the texture and depth information of the scene and provides real-time human skeletal joint data. Based on the skeletal data estimated by Microsoft Kinect SDK, the data processing module removes the outlier data caused by loss tracking with the joint Gaussian and uniform distribution model introduced in Section 3. It further extracts the kinematic parameters from the skeletal motion data. The performance evaluation module does real-time motion analysis, including repetition counting, most moving joints, reachable range, and moving time, which can be further used to measure the flexibilities, endurance, strength, and balance of the subjects. The feedback module can illustrate the evaluation results in the user interface on the screen and also provide both the system automatically generated suggestions and doctors' suggestions to the subject. Moreover, the system stores the subjects' motion data in the database for further reference in the future.

3. Kinect Data Accuracy Evaluation

3.1. Data accuracy evaluation

In order to quantify the fidelity of the motion data captured by Microsoft Kinect, we evaluated its pose tracking with respect to the motion capture system. In this study, we evaluate both the first and second generations of Microsoft Kinect which are called Kinect 1 and Kinect 2, respectively, in the following of this section. The motion capture system we use in this evaluation is the PhaseSpace optical motion capture system which can provide high-precision motion capture data at 120 Hz. The Kinect 1 and Kinect 2 capture the motion data at 20 Hz. During the evaluation, all these three systems capture the human motion simultaneously. The motion data from these three devices are temporally synchronized using Network Time Protocol (NTP). The skeletal joint data from Kinect 1 and Kinect 2 are extracted based on the latest SDK of them. The skeletal joint data from motion capture system are obtained by the Recap2 software. We picked 20 common joints and 14 common segments among the three systems. The joints and segments picked in the evaluation are demonstrated in Fig. 2.

During the evaluation, we perform 12 different exercises that six of them are in the sitting pose which are designed by the physical therapist for the elderly people with physical problems. The other six are in the standing pose which is the most common motion that people may perform in their daily life. Each exercise is performed by the subjects five times and captured in three different view angles, including the frontal view, right side view with 30 degree and 60 degree. The degree here means the angle between the subjects' frontal direction and the optical axis of the Kinect cameras.

For the joint position, we evaluate the offset and variance of the joint positions obtained by Kinect 1 and Kinect 2 with respect to the corresponding joint position captured by motion capture system. All the joint positions from these three systems are transformed into a unified coordinate system by the camera calibration performed before the motion capture.

Tables 1 and 2 show the average joint offset and variance of the offset in the sitting and standing exercise, respectively.

We also evaluate the segment length based on the motion data. Tables 3 and 4 demonstrate the average segment length error and variance of the segment length in the sitting
Table 1. Joint position offsets in sitting exercises.

| Joint | Kinect 1 | | | Kinect 2 | | |
|-------|----------|---|---|----------|---|---|
|       | Mean (mm) | SD (mm) | Mean (mm) | SD (mm) |
| ROOT  | 256 262 263 | 25 20 25 | 100 102 106 | 17 18 16 |
| SPINE | 91 97 100 | 24 20 21 | 110 117 126 | 13 14 11 |
| NECK  | 79 65 62 | 25 21 23 | 84 78 73 | 14 16 14 |
| HEAD  | 74 70 67 | 26 21 21 | 50 51 50 | 13 15 12 |
| SHO_L | 90 89 97 | 26 24 33 | 76 76 82 | 16 24 29 |
| ELB_L | 81 86 98 | 27 29 35 | 87 103 114 | 17 28 25 |
| WRI_L | 76 90 118 | 33 48 55 | 59 84 115 | 25 53 44 |
| HAN_L | 85 106 134 | 41 60 68 | 64 95 125 | 31 60 53 |
| SHO_R | 78 74 69 | 28 24 23 | 80 83 78 | 17 20 17 |
| ELB_R | 95 93 89 | 28 38 34 | 88 77 70 | 21 25 19 |
| WRI_R | 64 93 110 | 30 53 65 | 61 64 71 | 25 28 24 |
| HAN_R | 83 113 130 | 35 65 80 | 74 71 75 | 24 28 26 |
| HIP_L | 188 200 215 | 25 20 22 | 115 122 139 | 16 17 15 |
| KNE_L | 96 95 93 | 24 24 27 | 76 95 114 | 16 18 25 |
| ANK_L | 54 77 81 | 17 25 32 | 93 103 113 | 16 18 29 |
| FOO_L | 66 74 86 | 19 26 38 | 93 105 119 | 18 25 39 |
| HIP_R | 207 210 211 | 25 21 24 | 133 128 132 | 15 18 16 |
| KNE_R | 123 118 128 | 23 23 34 | 120 117 119 | 15 17 24 |
| ANK_R | 67 75 91 | 17 21 27 | 112 122 132 | 15 19 27 |
| FOO_R | 67 74 84 | 17 20 26 | 126 134 139 | 19 20 28 |

Table 2. Joint position offsets in standing exercises.

| Joint | Kinect 1 | | | Kinect 2 | | |
|-------|----------|---|---|----------|---|---|
|       | Mean (mm) | SD (mm) | Mean (mm) | SD (mm) |
| ROOT  | 245 256 267 | 23 25 25 | 76 81 93 | 24 19 18 |
| SPINE | 79 89 102 | 22 22 24 | 112 129 144 | 17 16 17 |
| NECK  | 82 91 102 | 23 24 23 | 113 129 143 | 18 16 16 |
| HEAD  | 89 84 87 | 30 29 26 | 79 82 90 | 26 22 20 |
| SHO_L | 76 76 80 | 33 36 38 | 68 72 78 | 29 31 31 |
| ELB_L | 98 112 134 | 46 52 66 | 112 137 159 | 37 41 57 |
| WRI_L | 85 93 110 | 56 57 71 | 66 84 111 | 47 52 73 |
| HAN_L | 85 96 114 | 62 65 80 | 77 94 121 | 49 55 80 |
| SHO_R | 76 76 74 | 30 30 26 | 96 106 109 | 26 24 22 |
| ELB_R | 98 89 80 | 38 39 34 | 96 91 80 | 33 32 31 |
| WRI_R | 80 82 80 | 43 43 38 | 78 88 87 | 33 32 31 |
| HAN_R | 85 83 77 | 46 49 44 | 81 79 74 | 33 34 32 |
| HIP_L | 186 205 228 | 28 27 27 | 106 125 150 | 23 19 21 |
| KNE_L | 101 111 124 | 33 38 52 | 107 129 148 | 26 34 51 |
| ANK_L | 119 144 135 | 43 61 58 | 135 168 174 | 33 51 58 |
| FOO_L | 115 135 122 | 42 60 59 | 130 170 180 | 33 56 63 |
| HIP_R | 188 195 207 | 25 24 23 | 103 104 109 | 25 17 17 |
| KNE_R | 105 102 101 | 31 29 26 | 115 119 118 | 26 25 27 |
| ANK_R | 111 113 108 | 38 39 28 | 146 158 157 | 32 33 35 |
| FOO_R | 97 93 86 | 36 37 29 | 146 157 150 | 32 33 38 |
and standing exercises.

The results show that overall Microsoft Kinect 2 generates more stable skeletal joint tracking with less offset and variance as compared to Microsoft Kinect 1. The variance of the joint position and segment length are both increasing with the view angle increasing. Some joints usually have larger offset and variance than the other joints, e.g. hip and knee joints in Kinect 1, the ankle and foot joints in Kinect 2.

### 3.2. Outlier data removal

The joint position offsets and variance in general depend on various sources of error, such as systematic errors, noise from depth quantization, occlusions, and ambiguity, etc. Compared to the multi-view motion capture system, the Kinect camera captures the motion data from a single viewpoint [7]. The occlusions and ambiguity from a single viewpoint will cause loss of tracking. Although the Kinect can infer the joint position based on the related body part [3], it is not reasonable to perform motion data analysis based on the inferred motion data due to tracking loss. Therefore, we analyze the error distribution to discriminate between the random errors and the errors due to tracking loss.

Generally, the random systematic errors follow the Gaussian distribution while the errors due to tracking loss can

---

**Table 3. Mean and SD of the bone length difference, sitting pose**

| Joint    | Kinect 1 (Mean mm) | SD (mm) | Kinect 2 (Mean mm) | SD (mm) |
|----------|-------------------|--------|-------------------|--------|
| ARM_UP_L | -76 -74 -81       | 17 18 21 | -67 -62 -68       | 14 17 20 |
| ARM_LO_L | -46 -40 -32       | 17 22 27 | -39 -30 -24       | 14 19 23 |
| HAND_L   | -3 -4 -3          | 22 19 20 | -20 -17 -17       | 20 24 25 |
| ARM_UP_R | -68 -70 -69       | 17 17 18 | -67 -67 -69       | 15 14 13 |
| ARM_LO_R | -37 -30 -29       | 17 20 21 | -35 -28 -30       | 13 15 16 |
| HAND_R   | 2 3 3             | 21 24 24 | -15 -13 -9        | 20 19 17 |
| PELVIC_L | -28 -28 -36       | 6 5 8  | -54 -62 -75       | 3 3 4  |
| LEG_UP_L | 49 38 38          | 30 32 34 | 9 5 5            | 16 18 18 |
| LEG_LO_L | -90 -100 -81      | 23 28 31 | -80 -79 -75       | 17 18 21 |
| FOOT_L   | -13 -4 1          | 8 10 13 | 36 38 32          | 12 14 19 |
| PELVIC_R | -29 -33 -41       | 6 5 7  | -52 -47 -49       | 3 3 4  |
| LEG_UP_R | 44 46 53          | 31 28 31 | 1 9 18          | 16 18 18 |
| LEG_LO_R | -86 -88 -95       | 25 26 31 | -76 -82 -74       | 17 19 23 |
| FOOT_R   | -8 -2 8           | 7 9 11 | 41 44 40          | 12 11 12 |

---

**Table 4. Mean and SD of the bone length difference, standing pose**

| Joint    | Kinect 1 (Mean mm) | SD (mm) | Kinect 2 (Mean mm) | SD (mm) |
|----------|-------------------|--------|-------------------|--------|
| ARM_UP_L | -60 -62 -65       | 19 22 23 | -54 -54 -51       | 15 18 20 |
| ARM_LO_L | -30 -31 -34       | 16 18 20 | -34 -33 -32       | 17 17 18 |
| HAND_L   | -7 -5 0           | 21 21 23 | -32 -32 -27       | 20 21 23 |
| ARM_UP_R | -53 -58 -59       | 21 19 18 | -59 -66 -66       | 17 14 14 |
| ARM_LO_R | -22 -22 -22       | 17 18 20 | -30 -27 -25       | 18 18 17 |
| HAND_R   | -1 4 6            | 23 21 19 | -21 -16 -14       | 19 18 18 |
| PELVIC_L | -27 -30 -37       | 9 7 8  | -52 -63 -75       | 5 5 5  |
| LEG_UP_L | 113 119 137       | 28 33 39 | -13 -10 -4       | 19 21 25 |
| LEG_LO_L | -64 -66 -65       | 29 29 31 | -79 -85 -86       | 21 23 25 |
| FOOT_L   | -6 -1 2           | 9 12 12 | 43 42 34         | 9 12 14 |
| PELVIC_R | -28 -34 -41       | 7 6 8  | -47 -44 -44       | 7 4 5  |
| LEG_UP_R | 103 115 128       | 29 27 29 | -25 -18 -6       | 18 18 20 |
| LEG_LO_R | -61 -57 -58       | 26 25 25 | -79 -79 -77       | 20 20 20 |
| FOOT_R   | -2 1 4            | 8 9 10 | 48 45 32          | 9 11 16 |
be treated as outliers belonging to a uniform distribution. Therefore, we use a mixture model of a Gaussian distribution and a uniform distribution to approximate the distribution of the joint position offsets, as defined in equation (1).

\[ p(\theta) = \rho \times N(\mu, \sigma) + (1 - \rho) \times U(x_1, x_2). \]  

(1)

In equation (1), \( p \) denote the mixture distribution of joint position, \( \theta \), \( N(\mu, \sigma) \) and \( U(x_1, x_2) \) denote the Gaussian and uniform distribution respectively, \( \rho \) denotes the weighting parameter. Fig. 3 demonstrates the distribution fitting results for the right elbow in the exercise. The results show the mixture model of the Gaussian and uniform distributions overlaid on the data histograms.

By applying the mixture model and maximal likelihood estimation, we can remove the outlier data which satisfy the uniform distribution and make cleaner data for the further analysis. Table I demonstrates the joint offset of both standing and sitting exercises without the outlier data.

Table 5. Average joint position offsets without outliers.

|         | Mean (%) | SD (%) |
|---------|----------|--------|
|         | 0°       | 30°    | 60°    | 0°       | 30°    | 60°    |
| Kinect 1|          |        |        |          |        |        |
| Sitting | 5        | 8      | 6      | 35       | 38     | 35     |
| Standing| 8        | 7      | 8      | 39       | 36     | 33     |
| Kinect 2|          |        |        |          |        |        |
| Sitting | 2        | 5      | 3      | 23       | 36     | 23     |
| Standing| 8        | 5      | 6      | 49       | 46     | 43     |

4. Kinematic parameter estimation

The kinematic parameters, such as the segment length, joint quaternion, and positions are widely used in human motion analysis and human robotic interaction. For the high fidelity motion data, the kinematic parameters can be directly estimated from the skeletal joint position. However, generating the kinematic parameters from the noisy motion data captured by Kinect will propagate the noise into the following processing and analysis steps. In order to reduce the noise in the derived kinematic parameters from the Kinect noise motion data, we propose a kinematic filtering approach in this section.

4.1. Kinematic model

In motion analysis, the humans skeleton can be represented as a series joints which are connected by bones. Therefore, the each joint position can be derived by its parent joint position, the corresponding segment orientation and length. The humans skeleton can thus be represented as a kinematic chain. In this paper, we treat the root joint as the root of each kinematic chain. Based on the skeletal structure shown in Fig. 2, the SPINE, NECK, SHO_L, SHO_R, HIP_L, and HIP_R are modeled as the 3-DoF joints and their parent joints are ROOT, SPINE, NECK, NECK, ROOT, ROOT, respectively; The ELB_L, ELB_R, KNE_L, KNE_R are treated as the 1-DoF joints whose parent joints are SHO_L, SHO_R, HIP_L, HIP_R, respectively; The WRIST, WRIST, ARM_L, and ARM_R are the 2-DoF joints that their parent joints are ELB_L, ELB_R, KNE_L, KNE_R, respectively. For each joint, its position can be derived by its parent joint position, \( T_p \), the segment length of its corresponding segment, \( l_c \), the relative rotation matrix of its corresponding segment, \( R_p \), with respect to its parent segment, the rotation matrix of its parent segment, \( R_p \), as the following function \( F \):

\[ [T_c, 1]^T = F(T_p, l_c, R_t, R_p) =
\begin{bmatrix}
R_t \times R_p \\
0 \\
1
\end{bmatrix}
\times
\begin{bmatrix}
0 \\
0 \\
l_c \\
1
\end{bmatrix} \]  

(2)

4.2. Kinematic filtering

With the kinematic model presented in the last section, we propose a kinematic filtering algorithm based on Kalman filtering to generate the kinematic parameter from the Kinect motion data. The input state vector is constituted by the root joint position, the segment length, and the segment rotation quaternion. The observation vector is constituted by the rest joint positions. The state transition is modeled by a random-walk process. The observation measuring function is a set of the kinematic model function as defined by equation (2). Since the state transition and the observation measuring functions are nonlinear, we apply the Unscented Kalman Filter (UKF) for the filtering. Since the segment lengths for a certain person are constants, we propose a four-pass UKF to meet with this constraints. In the first forward pass, the UKF is defined by the following functions

\[ x(t) = x(t - 1) + \delta(t) \]  

(3)

\[ y(t) = F(x(t)) + \varepsilon(t) \]  

(4)
In equations 3 and 4, the $x(t)$ and $x(t + 1)$ denote the state vector at time $t$ and $t + 1$, respectively. $\delta$ denote the random term in the random walk process, $y$ denote the observation vector, $\epsilon$ denotes the observation noise. In the first backward pass, the observation measuring function is the same as equation 4. The state transition function is defined as

$$x(t) = x(t + 1) + \delta(t) \quad (5)$$

After the first two pass, the length of each segment are assigned as a constant in the following forward and backward passes. The covariance and the random terms in the random walk process of the segment lengths are all set to zeros.

Fig. 4 demonstrates the filtering results of UKF for the motion data from Kinect 1. The segment quaternion curves after filtering are much smoother than the input noisy data and also demonstrate strong periodic pattern which is in accordance to the input motion. Therefore, the UKF based on the kinematic model is a practical solution to extract kinematic parameters.

5. Conclusions

This paper summarize the recent progress in the motion analysis for physical therapy with the affordable and accessible device like Microsoft Kinect. Based on preliminary studies, our remote health coaching system receive positive feedback from the users for its accessibility and interaction. Since the motion data captured by Microsoft Kinect is noisy, we further studied the data fidelity of each joint position and segment length. We also propose a mixture model to distinguish the outlier data caused by loss of tracking. It can helps to reduce the noise in the further motion data analysis. Finally, the kinematic filtering based on UKF provide a solution to extract smooth kinematic parameters from the noisy motion data.

References

[1] H. M. Hondori and M. Khademi. A review on technical and clinical impact of microsoft Kinect on physical therapy and rehabilitation. Journal of Medical Engineering, 2014:16, 2014.
[2] F. Ofli, G. Kurillo, S. Obdrzalek, R. Bajcsy, H. Jimison, and M. Pavel. Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned. IEEE Journal of Biomedical and Health Informatics, page 15, Epub ahead of print. 2015.
[3] 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 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1297–1304, 2011.
[4] E. Wan, R. Van Der Merwe, et al. The unscented kalman filter for nonlinear estimation. In Proceedings of IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC), pages 153–158, 2000.
[5] Q. Wang, G. Kurillo, F. Ofli, and R. Bajcsy. Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. arXiv preprint arXiv:1512.04134, 2015.
[6] Q. Wang, G. Kurillo, F. Ofli, and R. Bajcsy. Unsupervised temporal segmentation of repetitive human actions based on kinematic modeling and frequency analysis. In International Conference on 3D Vision (3DV), pages 562–570, 2015.
[7] Z. Zhang. Microsoft Kinect sensor and its effect. MultiMedia, IEEE, 19(2):4–10, Feb 2012.