Kinematic Data Filtering with Unscented Kalman Filter

Application to Senior Fitness Tests Using the Kinect Sensor

Manuel Goulão
Instituto Superior Técnico, Universidade de Lisboa
manuel.goulao@tecnico.ulisboa.pt

ABSTRACT

Sedentary behavior and the lack of physical exercise are routines existing in a large segment of the population, specially in the elderly. This can lead to health complications and debilitating conditions, reducing the quality of life, as it drives to the inability to perform activities of daily living. The Microsoft Kinect sensor provides an accessible mechanism to implement automatic and assisted training systems, which can then be used, for example, for in-home training, or in care centers, either in a fixed platform or in a mobile assistive robot. However, the need to obtain movement parameters to perform fitness assessment requires data with a large degree of consistency. In this paper, we propose a method aimed at improving the accuracy of the skeleton data from the Kinect sensor, based on a kinematic filter approach using an unscented Kalman filter, which tries to achieve a higher coherence by combining the observations with previous knowledge on the movements being executed. In the end, it has been possible to extract information, such as angle and angular velocity corresponding to each degree of freedom of each joint, as well as to correct some of the errors present in the data.

KEYWORDS

Biomechanics, Kinect v2 SDK, Motion Capture, UKF.

1 INTRODUCTION

Nowadays, a sedentary lifestyle describe plenty of people's lives, with the lack of physical exercise being one of the key factors. This malicious behavior covers a large portion of the population, from multiple age groups and social backgrounds, reaching, in the USA, more than 50% of the population with ages over 45 [11]. Physical exercise takes a large part in the prevention of diseases and disabilities, as well as leading to a longer, and a better quality of life [8]. Additionally, this behavior may lead to crippling conditions, reducing the capabilities and functional independence of an elderly person, and making it impossible to perform normal Activities of Daily Living (ADLs) and to live a life with quality.

The present work was mostly motivated by the necessity to promote regular physical exercise in the elderly population, as it is an essential element in reducing the sedentary lifestyle that many people follow. For this sector of the population, functional fitness is essential, as the capability to perform ADLs is a major factor in the quality of life. Most ADLs are highly depend on specific fitness areas such as strength, aerobic endurance, flexibility, and agility [15]. Consequently, promoting the regular practice of physical exercise, as well as specifically training the areas responsible for the functional fitness of an elderly individual is of extreme importance.

The Microsoft Kinect sensor is an easily available and affordable device which, together with the available SDK that implements an easy-to-use back-end to interact with the it, grants the ability to a broad set of people to be able to design and implement their own systems. In particular, it is possible to acquire the position of the body joints of a user (skeleton features) that enable a wide range of applications related to exercise, and multiple applications in fitness using the Kinect have been described in the last years. However, it is recognized that skeletal features are sometimes noisy, particularly in cases of (self-)occlusion and clutter, which limits some applications that require the precise analysis of the users joints' trajectories over significant periods of time. In this work, the Kinect v2 and its available SDK were applied to create a real-time system, which aims at improving the accuracy of the joint position data provided by the SDK, by modeling the skeleton provided [17] as a kinematic chain, while using an approach based on a kinematic filter.

In the last years, elderly fitness training using automatic systems has been largely studied, with systems on multiple platforms being developed. For example, the use of the Nintendo Wii gaming console to study fitness training and its enjoyment [4], as well as balance training [21]. Mobile applications are another type of devices used [7], as are custom wearable devices [14]. On the other hand, the Kinect sensor allows the creation of systems without the requirement to wear, or hand held a device, which could be seen as an advantage over the other types of systems, as the manipulation of a device could be seen as difficult, particularly by older people [3]. Some recent applications employing the Kinect are, i.a., full body gait analysis and the extraction of stride information [2]; senior health monitoring and fall detection, by detecting abnormal gait using SVMs [13]; a mobile robot which follows a person and, by comparison of the skeleton positions against the ground's, performs real-time fall detection [9]; a mobile elderly companion robot to promote active living, which is capable of talking using text-to-speech methods, and which uses the Kinect sensor to detect and follow a user, detecting noise by setting thresholds based on past data [20]; clinical parameters extraction, revealing axial landmarks and upper body parameters to be valid, while ankle and feet positions to be unstable [12]. However, most works either do not provide enough information, or do not implement any type of pre-processing in the data before using it in their applications, which could lead to miscalculations in, for instance, clinical parameters which require a considerable degree of accuracy and consistency of the data. An exception is [19] that describes a Kinect physical therapy system, performing outlier removal using a mixture model, and filtering with a four-pass UKF considering a random-walk and constant segment lengths. However, that work presents a preliminary study which does not show many details on the implementation of the filter, preventing other researchers from replicating their results. Instead, in our paper, we present a complete implementation of a kinematic filter for Kinect skeleton features, and corresponding

filter parameters, that can be used by the research community to improve the signal quality of users' motion capture data.

Our system targets the fact that the SDK's acquisition algorithm [17] does not enforce any constraints on the data. As such, in this kinematic filter approach, both the continuity between frames is addressed by comparison of a prediction, from the equations of motion, with the data provided by the SDK, as well as the variation of the lengths between joints (bone, or link, lengths), which do not describe true features of the human movement, and are key characteristics in improving the accuracy of the data [10].

In this study, we establish a formulation and implementation of a system, which attempts to improve the accuracy and physical coherence of the data provided by the Kinect v2 SDK sensor. This data could then be translated as movement, or fitness features, which require a high degree of reliability. Moreover, these features could then be employed in an assessment model, and implemented in an assistant robot for the elderly, to automatically appraise the capabilities of an individual, as well as in a training system to promote the practice of regular physical exercise, or even connecting the two, creating a personalized system that would be able to identify and provide exercises adequate for each subject, which is a key factor in maintaining the interests of the participants [3].

In the subsequent sections, first, the methods employed to implement the kinematic filter will be described. Then, the experimental results will be presented, with multiple instances of different exercises, as well as results from different features, which the algorithm implements. Ultimately, the results will be analyzed and conclusions, as improvements or limitations of the implemented system, will be drawn.

2 METHODOLOGY

The data provided by the Kinect v2 SDK does not enforce any constrains on the subject's positions. Accordingly, an algorithm that is focused on specific issues is of interest to improve the accuracy of the data, as the addressing of the variable lengths of the links through time is expected to improve accuracy [10]. Also, the restriction of large and unexpected variations of the position in a short period of time, which could represent a mistake from the sensor, may improve the accuracy on the measurement of movements. An approach that is able to do so becomes even more appealing when the movement to be performed and captured by the Kinect is known, or there is some previous information about it. This is usually the case for training systems, and was also the case for the present work, where there is previous knowledge of which movement would be performed as the participants executed a predefined set of exercises.

In terms of the processing of the data, an approach using a kinematic filter was employed, as an unscented Kalman filter (UKF) [5] was used to model the skeleton data from the Kinect. In fact, the data was designed as two kinematic chains, considering their root as the abdomen of the model for both chains. While the first chain modeled the lower half of the body, from the abdomen to the left/right hip, knee, ankle, and foot; the second chain modeled the upper half of the body, from the abdomen to the spine, shoulder center, left/right shoulder, elbow, and wrist, and neck and head. By splitting the model in two, parallel processing of both parts is possible, reducing computation time.

The state for the UKF consisted on the spacial coordinates of the root in the universe coordinate system, and its orientation, as well as on the angles in the anatomical planes, in relation to the previous joint, which could be 1 (knees, and elbows), 2 (ankles), or 3 (neck, shoulders, shoulder center, spine, and hips). Finally, the lengths of each link connecting the joints were also considered. Then, starting at the root, and through successive transformations considering the rotations to follow the Euler angle convention, it was possible to compute the state of the system iteratively, employing an UKF. The lengths of the links were estimated at an initial calibration phase, at which the acquisition was performed with the subject still, with the lengths being averaged throughout this phase, and the values obtained were used to initialize the state.

The UKF model was employed using a constant velocity model, thus introducing variation through the acceleration. This filter, based on the Kalman filter [6], fits the problem at hand, as it is adequate for highly non-linear applications, using the unscented transform to approximate a non-linear function [5]. So, the system, with state $\mathbf{x}$ and observation vector $\mathbf{y}$, iteratively advances from the previous time step through a state-transition function F, to which there is an uncertainty associated $\mathbf{w}$ (described by mean 0, and covariance $\mathbf{Q}$), and the coordinates in the universe coordinate system may be obtained through an observation function H, to which there is also an uncertainty $\mathbf{v}$ (described by mean 0, and covariance $\mathbf{R}$).

$$\mathbf{x}_i = F(\mathbf{x}_{i-1}) + \mathbf{w}_i \tag{1}$$

$$\mathbf{y}_i = H(\mathbf{x}_i) + \mathbf{v}_i \tag{2}$$

The evolution of the state was set to follow the equations of motion, with the lengths set to the same as the previous state. And, the observation model was implemented according to the orientations computed, and the lengths of the links. This way, the system could not suddenly change, as it needed to follow these provided rules, while still adapting to the data at each iteration. Therefore, the mistakes in the data that violate the assumptions provided to the filter should most likely be attenuated, or even removed.

Therefore, the UKF works by computing, at each time step, a set of points considering that $\mathbf{x}$ has mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}$, called sigma points ($\mathcal{X}$), which have weights ($\mathbf{W}$) both for the mean reconstruction ($^{(m)}$) and for the covariance reconstruction ($^{(c)}$).

$$\mathcal{X} = \left[\bar{\mathbf{x}}, (\bar{\mathbf{x}} \pm (\sqrt{(L+\lambda)\mathbf{P}}))_i\right]^T \quad \text{with } i = 1, ..., L \tag{3}$$

$$\mathbf{W}_0^{(m)} = \frac{\lambda}{L+\lambda} \tag{4}$$

$$\mathbf{W}_0^{(c)} = \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta) \tag{5}$$

$$\mathbf{W}_i^{(m)} = \mathbf{W}_i^{(c)} = \frac{1}{2(L+\lambda)}, \quad i = 1, \ldots, 2L \tag{6}$$

and where L is the dimension of the state, $\lambda = \alpha^2(L + \kappa) - L$, with α, β, and κ, being parameters of the UKF. The square root of the covariance matrix $\mathbf{P}$ was computed using the Cholesky decomposition.

These points are then propagated though the non-linear function,

$$\mathcal{X}_i^* = F(\mathcal{X}_i) \tag{7}$$

and then reconstructed to obtain both the system and covariance matrix at the following iteration

$$\hat{\mathbf{x}}_i^- = \sum_{j=0}^{2L} \mathbf{W}_j^{(m)} \mathcal{X}_{i,j} \tag{8}$$

$$\mathbf{P}_i^- = \sum_{j=0}^{2L} \mathbf{W}_j^{(c)} (\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-)(\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-)^T + \mathbf{Q}_i \tag{9}$$

Then, the transformed sigma points are used in the observation model and the average is again reconstructed according to the weights, and the covariance matrix of the observations is computed, as well as a cross-covariance between both,

$$\mathcal{Y}_i = H(\mathcal{X}_i^*) \tag{10}$$

$$\hat{\mathbf{y}}_i = \sum_{j=0}^{2L} \mathbf{W}_j^{(c)} \mathcal{Y}_{i,j} \tag{11}$$

$$\mathbf{P}_{y_i} = \sum_{j=0}^{2L} W_j^{(c)} \left(\mathcal{Y}_{j,i} - \hat{\mathbf{y}}_i\right) \left(\mathcal{Y}_{j,i} - \hat{\mathbf{y}}_i\right)^T + \mathbf{R}_i \tag{12}$$

$$\mathbf{P}_{(x,y)_i} = \sum_{j=0}^{2L} W_j^{(c)} \left(\mathcal{X}_{i,j}^* - \hat{\mathbf{x}}_i^-\right) \left(\mathcal{Y}_{i,j} - \hat{\mathbf{y}}_i\right)^T \tag{13}$$

To combine the state with the observations, the Kalman gain attributes more or less weight to the observations from the Kinect in relation to the prediction, and is computed as,

$$\mathbf{K}_i = \mathbf{P}_{(x,y)_i} \mathbf{P}_{y_i}^{-1} \tag{14}$$

Finally, the new prediction combining the information from the state-transition and observation models, as well as the new covariance for this state is given by,

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i(\mathbf{y}_i - \hat{\mathbf{y}}_i) \tag{15}$$

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{K}_i \mathbf{P}_{y_i} \mathbf{K}_i^T \tag{16}$$

This process is performed twice, for the upper and lower body models, possibly in parallel. However, the mismatch between the root positions of both chains was not considered, since the main goal was to compute movement features, and most do not require this coherence. Nevertheless, multiple approaches such as the average between the two would suffice to correct any offset that could result.

When establishing the state-transition function, one may take advantage of previous knowledge on the movements being performed and acquired by the Kinect. In fact, multiple fast movements of a joint, compared to a slow or barely moving one, describe highly different systems. So, setting lower variations in the latter in relation to the former should make sense. Then, there would be a corresponding larger covariance value for the joints whose movements are larger in comparison with the other joints, and in comparison to other exercises requiring slower actions. Notwithstanding, as there was no ground truth available which could be used to select and validate the chosen covariance values, these were selected considering an empirical criteria. A smoothness of the movements was sought, while trying to maintain a considerable similarity to the original data, which was done by enabling the system to attain high enough variations which could be expected in the exercise, but constraining the movements corresponding to abnormal displacements of the joint's values.

For each iteration, it is possible to obtain the data from both the system of the UKF, i.e. the angles and angular velocities for each of the directions, and lengths of the links. Moreover, it is also possible to obtain the coordinates of each joint by taking the observation function at the computed state, for each iteration. This way, it was possible to analyze comparatively the original data from the Kinect with the data after being filtered using the UKF.

3 EXPERIMENTAL RESULTS

A dataset of 21 participants [1] was processed in this work, 11 young and 10 senior adults (average age of 61 years), performing four fitness assessment exercises based on the Senior Fitness Test (SFT) [15], which aim at different core fitness characteristics necessary to perform ADLs. This set of physical exercises establishes a standard, which can be used to study elderly fitness assessment, in a valid and reliable manner. Indeed, the employed tests, which focused on lower body training, were:

2-minute step: Test from the SFT, aiming at measuring the aerobic endurance of a subject, it consists on raising each knee in succession for the maximum number of repetitions achievable during the period of 2 minutes. This test is highly related with walking and climbing stairs, and also longer and more demanding activities requiring physical endurance. An example of the performance of this exercise may be seen in Figure 1.

30-second chair stand: Test from the SFT, which targets the measurement of lower-body strength, requiring the participant to stand up and sit down repeatedly, for the maximum repetitions possible during 30 seconds. Again, walking and climbing stairs are closely related to this exercise, as well as getting in or out of a car or tub, or other ADLs regarding lower-body strength. Figure 2 shows the execution of the 30-second chair stand test by an elderly participant.

8-foot up-and-go: Test from the SFT, focused on quantifying agility and dynamic balance of a participant, who starts seated down in a chair, and needs to stand up, walk 8 feet (2.4 meters), turn around, and return and sit down on the chair, with the time to complete the exercise working as the measure. ADLs requiring faster movements are closely related, like getting off the bus or cooking. The walking performed in this exercises may be seen in Figure 3, where 3 repetitions were captured.

Unipedal stance: The only of the four not belonging to the SFT, it focuses on assessing the static balance abilities. It requires raising and holding one leg off the ground, while keeping the arms crossed over the chest, for the longest possible, but for a maximum period of 30 seconds, at which the test is considered to be completed. Riding the bus or walking in a narrow sidewalk are relatable ADLs. The foot being raised above the ground during the full 30 seconds may be seen in Figure 4, where a correction is also performed by the filter right before the lifting.

The datasets consisted on a skeleton, acquired at 30 Hz, and including the timestamps of each frame, the orientation of the floor, the 3D positions of 23 joints, the orientation of each joint as a quaternion, and an indicator of the tracking status of each joint. In every test, the person would perform the exercise in front of

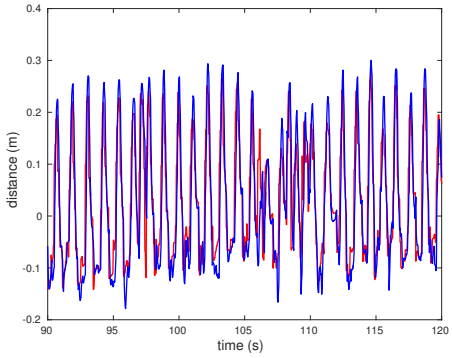

Figure 1: 2-minute step performed by elderly female. Right knee. Before (red) and after the kinematic filter (blue).

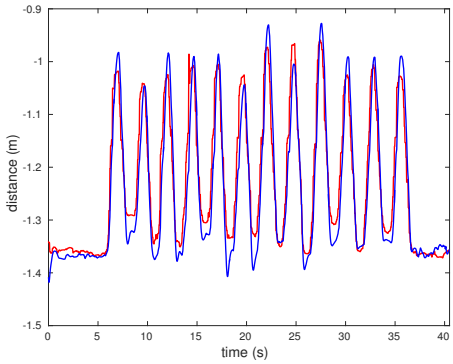

Figure 2: 30-second chair stand performed by elderly female. Right hip. Before (red) and after the kinematic filter (blue).

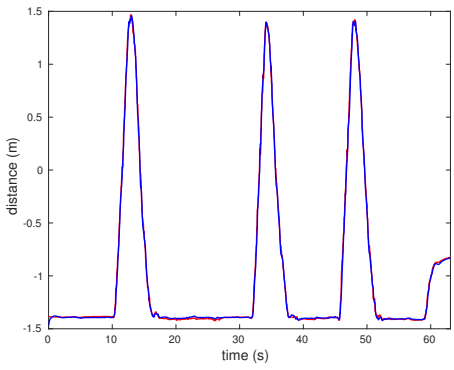

Figure 3: 8-foot up-and-go performed by elderly female. Right hip. Before (red) and after the kinematic filter (blue).

the Kinect camera, which captured their entire body. In some cases more than one person would be detected. In those cases only the main participant (which was always the closest to the camera) was considered while the others were discarded.

Next, for each test, the kinematic filter was applied with the covariance of the state being dependent on the movement being

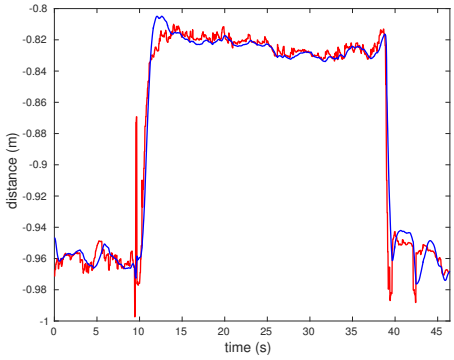

Figure 4: Unipedal stance performed by elderly female. Right foot. Before (red) and after the kinematic filter (blue).

executed. The system was initialized as a generic standing position for *2-minute step* and *Unipedal stance* tests, and as a generic seated position for *30-second chair stand* and *8-foot up-an-go* tests, since these were the most general positions that were expected at the beginning of each exercise. The velocities were initialized to zero, and the covariance of the state, $\mathbf{P}_0$, initialized to $\mathbf{P}_0 = 100 \cdot Q$. Nevertheless, since these do not correspond to the true initial position of the subject, there exist cases in which, at the start, it takes a small period of time until the system converges to the actual position, which may be seen at the beginning of the plot from Figure 8. Furthermore, regarding the lengths of the links, the first 30 frames (1 second of data) were averaged, since in most datasets, the participant remains still in the beginning of the video. In addition, the parameters of the UKF sigma points used were the most usual values found in the literature [18], with $\alpha = 1 \cdot 10^{-3}, \beta = 2, \kappa = 0$. Since an approach using the UKF was employed, it should be noticed that, as the computation of the square root of the covariance matrix was done with the Cholesky decomposition, the covariance matrix is required to be positive definite, and therefore no state variable must have covariance zero, thus there cannot be a constant variable, but only variables that have larger or smaller covariance than others. In this work, the covariances were considered as being diagonal, i.e., the variables started as independent from each other, and, regarding the state, joints which were expected to have larger variations in the exercise to be performed were set to have larger covariances than the rest, where considering the observations, it was set equal for the positions of every joint. Notwithstanding, all the values were picked by experimentation.

The selection of the specific values for the covariances of each exercise took advantage of the 21 available datasets, as a segmentation of the patients was employed while performing this operation. Accordingly, by using a few datasets for the estimation of these values, the possibility to introduce a bias on a specific patient should be reduced. Likewise, performing the testing on datasets which were not used to choose the covariances should yield less biased results, and demonstrate the generalization potential of the algorithm proposed in this work. In fact, the obtained results for the patients which were not included in the set used for the parameters selection exhibited similar behavior to the ones which had been used for that purpose, describing the movements performed by the

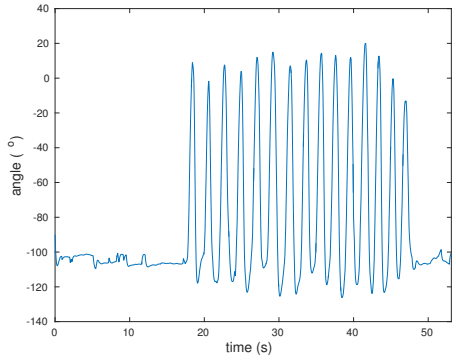

Figure 5: Left knee angle, during the 30-second chair stand exercise performed by an elderly male participant.

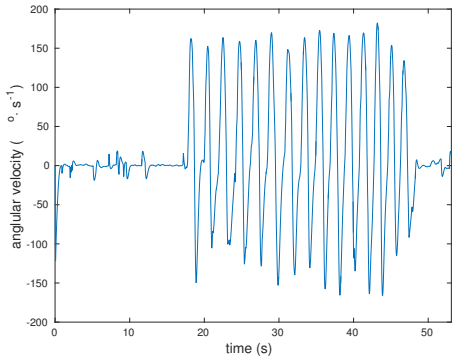

Figure 6: Left knee angular velocity, during the 30-second chair stand exercise performed by an elderly male.

participant in a smooth and clear manner. The values that were found to work best on the system are presented in Appendix A, where the detailed state and observation vectors are also described.

After the application of the kinematic filter, the state at each iteration described the angles and angular velocities, considering the degrees of freedom of each joint. In fact, Figures 5 and 6 show the plots of the knee rotation and velocity during the performance of the *30-second chair stand* exercise by an elderly male participant. In these plots, the system may be seen converging from the initialized state to the actual angle in the first iterations, and the exercise is easily identifiable in the successive flexion and extensions of the knee, as well as the zero velocities for maximum flexion and extension of the leg, and maximum velocities (in module) midway through the raise/lowering of the leg.

The acquired data presents a variety of mistakes, large variations in small amounts of time are likely not the true positions of the movement being captured. In fact, some errors in the data were corrected using the proposed method. For instance, in Figures 7 and 8, multiple mistakes are identifiable, which are then corrected by the filter. Accordingly, when the filter reaches these noisy samples, since they ought to be unlikely when considering the distribution of the state model, the observations will obtain a much lower gain (Equation 14) than the predictions, thus attenuating those mistakes.

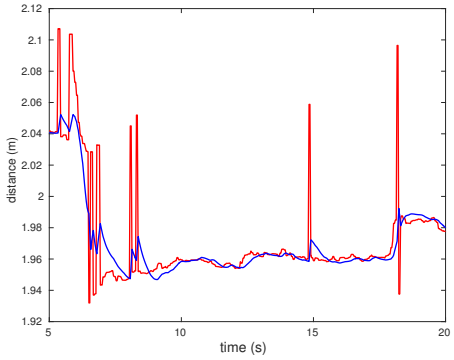

Figure 7: Segment of the Unipedal stance exercise from the left foot, before (red) and after kinematic filter (blue).

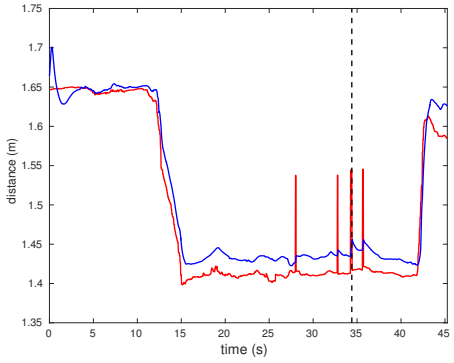

Figure 8: Left knee, Unipedal stance exercise by an elderly female, before (red) and after kinematic filtering (blue).

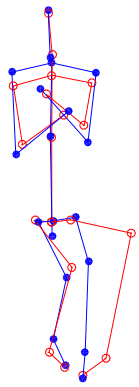

Figure 9: View of the model of an elderly female's Unipedal stance test. Before (red) and after the kinematic filter (blue).

One of these cases is more closely illustrated in Figure 9. This figure corresponds to the instant which is represented by a dashed line on Figure 8. Here, although being projected in the 2D coronal plane, it is possible to see the sudden displacement of the left leg, which is the support leg in the *Unipedal stance* test, being corrected by the kinematic filter.

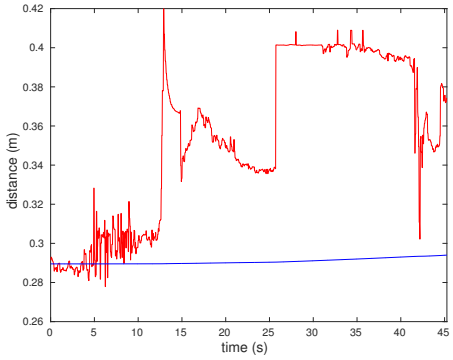

Figure 10: Unipedal stance, lower leg length of an elderly female. Before (red) and after the kinematic filter (blue).

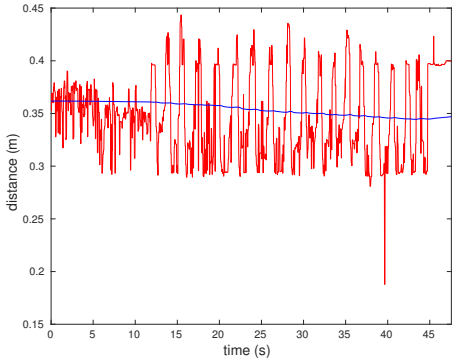

Figure 11: 30-second chair stand, lower leg length of an elderly female. Before (red) and after the filter (blue).

Also in Figure 8, one may identify an offset for the period in which the knee is flexed (approximately 15 to 43 seconds), this originates on the fact that the data provided by the Kinect SDK does not enforce the lengths of the links to be constant, varying by a large margin when compared to the little variation that is allowed by the achieved kinematic filter. As a matter of fact, Figure 10 shows the lower leg length varying through time of the same exercise, performed by the same participant, and being highly subjective on the position of the individual.

The length variations are clearly an erroneous feature processed by the Kinect SDK algorithm, as these phenomena do not describe the true positions of the subject in study. Again, this work tries to correct this issue, as seen in Figure 11, where a participant is performing the *30-second chair stand* exercise, and the lower leg length keeps oscillating, as the person is standing up or sitting down, in the data originated by the SDK. However, in the data from the kinematic filter, the length almost does not vary, following the initial value from the calibration phase.

4 CONCLUSIONS

We propose a real-time method aimed at the improvement of the accuracy of the data provided by the Kinect SDK, capable of being implemented in a diversity of applications, such as in an assistant robot which could focus its interactions in the physical training of its users. As this data does not enforce any kinematic constraints, by having previous knowledge on the movement being performed, it is possible to devise a kinematic filter model, based on the UKF, which does not only allow the prevention of certain wrong features of the skeleton data from the Kinect, but also provides information on the angles of each joint in either of the anatomical planes set to be movable by that joint, and also the corresponding angular velocity for the same directions. Therefore, the extraction of kinematic parameters could be enhanced by having a more cohesive anatomic model of the subject participating in the physical training. In this work, four standard physical exercise models to assess different fitness areas were employed, and the system was studied for each.

In fact, our system grants the ability to compute a set of kinematic parameters, as the computation of the angles and angular velocity of each joint in relation to the multiple directions are major components in the kinematic study of the human movement. The angle between two links may easily be computed in relation to a coordinate system which takes the links to be vectors and computes the angle by taking the scalar product of the two vectors. However, decomposing the movement in the angles of rotation in the anatomical movements of flexion and extension, adduction and abduction, and rotation, is not as easy, and using our approach these properties are computed by the filter and given as one of its outputs. Moreover, as seen above, the changing link lengths represent a flaw in the data obtained as they do not represent the true behavior of human movement, and this problem is highly attenuated by our model.

Furthermore, the segmentation of the dataset through the use of a subset of executions of each exercise to select the quantities to attribute to the covariances, and the evidence of the analogous results obtained while applying the filter to exercises of either set, provide a basis to validate the method's ability to generalize, as reproducible results are found for the whole dataset.

In opposition, the fact that there was no comparison to a standard, like a marker-based IR system with millimetric or sub-millimetric accuracy, reduces the capability to quantify the results from the algorithm, and would be an important testing mechanism. Nevertheless, this may be done in a future acquisition which could be eventually performed.

Additionally, a system such as the proposed kinematic filter attempts to solve an optimization problem, given some state information and the observed data, as it tries to find the state that better fits each case. However, this does not guarantee that the true state is found, and may lead to the computation of incorrect values and positions, which could lead to lower violations on the given constraints. Also, the fact that there is some previous knowledge about the movement being introduced hints that the system would perform worst in the acquisition of other movements, which limits the application to a set of specific use cases, and requires the creation of new profiles for new exercises introduced.

In the future, the system can be further developed, extending the state to include the orientation of each joint, provided as an orientation quaternion by the Kinect SDK. This way, one more indication of the actual position and orientation of the subject would be introduced, accomplishing a more meticulous system which can better model the kinematic chains, providing rotation

information in the observation model, adding to the one already in the state model.

Another interesting addition to the proposed model would be using the specific characteristics of the UKF and of the unscented transform. In fact, by using the predicted state, as its mean and uncertainty (covariance), to be able to identify large deviations from the prediction, and find a way to correct them, it could be possible to integrate a more precise outlier detection model. Also, if the distribution which describes the prediction could be able to pinpoint non-anatomical positions, e.g. bending the arm or leg backwards, these mistakes could also be found and then corrected.

Finally, testing the obtained results against other methods would be important and should be performed. This could be done by extracting certain movement parameters from the data and performing a comparison to the ones provided by specialized tests and health professionals, as well as comparing them to other state-of-the-art systems.

Despite the limitations mentioned above, we believe the method presented in this paper is already capable of improving, to a large extent, the signal-to-noise ratio of the Kinect skeleton data, and constitute an important asset for pre-processing the data for applications using this sensor for the measurement of biomechanical variables.

A UKF SYSTEM

A.1 Kinect skeleton from the Kinect SDK

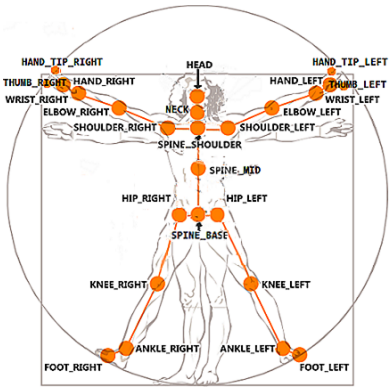

Figure 12: Set of skeleton joints provided by the the Kinect SDK. [16].

A.2 Observation vector

In the following vectors, the notation $\mathbf{y} = [\mathbf{x}]$ describes the three components of $\mathbf{x}$ as components of $\mathbf{y}$, i.e. $\mathbf{y} = [x_1, x_2, x_3]$.

Lower-body model:

$$\mathbf{y} = \big[\mathbf{x}_{root}, \mathbf{x}_{l-hip}, \mathbf{x}_{l-knee}, \mathbf{x}_{l-ankle}, \mathbf{x}_{l-foot}, \cdots \mathbf{x}_{r-hip}, \mathbf{x}_{r-knee}, \mathbf{x}_{r-ankle}, \mathbf{x}_{r-foot}\big]^T$$

Upper-body model:

$$\mathbf{y} = \big[\mathbf{x}_{root}, \mathbf{x}_{spine}, \mathbf{x}_{c-shoulder}, \mathbf{x}_{neck}, \mathbf{x}_{head}, \mathbf{x}_{l-shoulder}, \cdots \mathbf{x}_{l-elbow}, \mathbf{x}_{l-wrist}, \mathbf{x}_{r-shoulder}, \mathbf{x}_{r-elbow}, \mathbf{x}_{r-wrist}\big]^T$$

A.3 State vector

Lower-body model:

$$\begin{aligned}\mathbf{x} = [&x_{root}, y_{root}, z_{root}, vx_{root}, vy_{root}, vz_{root}, \theta x_{root}, \theta y_{root}, \cdots\\ &\theta z_{root}, \omega x_{root}, \omega y_{root}, \omega z_{root}, \theta x_{r-hip}, \theta y_{r-hip}, \theta z_{r-hip}, \cdots\\ &\omega x_{r-hip}, \omega y_{r-hip}, \omega z_{r-hip}, \theta z_{r-knee}, \omega z_{r-knee}, \theta y_{r-ankle}, \cdots\\ &\theta z_{r-ankle}, \omega y_{r-ankle}, \omega z_{r-ankle}, \theta x_{l-hip}, \theta y_{l-hip}, \theta z_{l-hip}, \cdots\\ &\omega x_{l-hip}, \omega y_{l-hip}, \omega z_{l-hip}, \theta z_{l-knee}, \omega z_{l-knee}, \theta y_{l-ankle}, \cdots\\ &\theta z_{l-ankle}, \omega y_{l-ankle}, \omega z_{l-ankle}, L_{r-root-hip}, L_{r-hip-knee}, \cdots\\ &L_{r-knee-ankle}, L_{r-ankle-foot}, L_{l-root-hip}, L_{l-hip-knee}, \cdots\\ &L_{l-knee-ankle}, L_{l-ankle-foot}]^T\end{aligned}$$

Upper-body model:

$$\begin{aligned}\mathbf{x} = [&x_{root}, y_{root}, z_{root}, vx_{root}, vy_{root} vz_{root}, \theta x_{root}, \theta y_{root}, \cdots\\ &\theta z_{root}, \omega x_{root}, \omega y_{root}, \omega z_{root}, \theta x_{spine}, \theta y_{spine}, \theta z_{spine}, \cdots\\ &\omega x_{spine}, \omega y_{spine}, \omega z_{spine}, \theta x_{l-shoulder-c}, \theta y_{l-shoulder-c}, \cdots\\ &\theta z_{l-shoulder-c}, \omega x_{l-shoulder-c}, \omega y_{l-shoulder-c}, \cdots\\ &\omega z_{l-shoulder-c}, \theta x_{l-shoulder}, \theta y_{l-shoulder}, \theta z_{l-shoulder}, \cdots\\ &\omega x_{l-shoulder}, \omega y_{l-shoulder}, \omega z_{l-shoulder}, \theta z_{l-elbow}, \cdots\\ &\omega z_{l-elbow}, \theta x_{r-shoulder-c}, \theta y_{r-shoulder-c}, \theta z_{r-shoulder-c}, \cdots\\ &\omega x_{r-shoulder-c}, \omega y_{r-shoulder-c}, \omega z_{r-shoulder-c}, \cdots\\ &\theta x_{r-shoulder}, \theta y_{r-shoulder}, \theta z_{r-shoulder}, \omega x_{r-shoulder}, \cdots\\ &\omega y_{r-shoulder}, \omega z_{r-shoulder}, \theta z_{r-elbow}, \omega z_{r-elbow}, \cdots\\ &\theta x_{up-shoulder-c}, \theta y_{up-shoulder-c}, \theta z_{up-shoulder-c}, \cdots\\ &\omega x_{up-shoulder-c}, \omega y_{up-shoulder-c}, \omega z_{up-shoulder-c}, \cdots\\ &\theta x_{neck}, \theta y_{neck}, \theta z_{neck}, \omega x_{neck}, \omega y_{neck}, \omega z_{neck}, \cdots\\ &L_{root-spine}, L_{spine-shouldercenter}, \cdots\\ &L_{l-shouldercenter-shoulder}, L_{l-shoulder-elbow}, \cdots\\ &L_{l-elbow-wrist}, L_{r-shouldercenter-shoulder}, \cdots\\ &L_{r-shoulder-elbow}, L_{r-elbow-wrist}, \cdots\\ &L_{shouldercenter-neck}, L_{neck-head}]^T\end{aligned}$$

A.4 Covariance values

In this section, the four exercises included in the dataset will be referred as:

- 2MS: *2-Minute Step* test
- 30SCS: *30-Second Chair Stand* test
- 8FUAG: *8-Foot Up-And-Go* test
- US: *Unipedal Stante* test

Table 1: Observation covariance values for the upper body model (per frame). ($\mathbb{I}$ represents the identity matrix.)

2MS	30SCS	8FUAG	US
$1 \cdot 10^{-4}\,\mathbb{I}$	$1 \cdot 10^{-4}\,\mathbb{I}$	$1 \cdot 10^{-3}\,\mathbb{I}$	$1 \cdot 10^{-4}\,\mathbb{I}$

Table 2: State covariance values for the lower body model (per frame). The vector entries of the table all have the same specified value in their components.

Variable	2MS	30SCS	8FUAG	US
$\mathbf{x}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\mathbf{v}_{root}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{root}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{hip}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{hip}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
θ_{knee}	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
ω_{knee}	$1 \cdot 10^{-1}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-5}$
$\boldsymbol{\theta}_{ankle}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{ankle}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{-5}$
link length	$1 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-11}$

Table 3: State covariance values for the upper body model (per frame). The vector entries of the table all have the same specified value in their components.

Variable	2MS	30SCS	8FUAG	US
$\mathbf{x}_{root}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\mathbf{v}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{root}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{root}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{spine}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{spine}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{neck}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{neck}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{shoulder-c}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{shoulder-c}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-6}$
$\boldsymbol{\theta}_{shoulder-l/r}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
$\boldsymbol{\omega}_{shoulder-l/r}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-5}$
θ_{elbow}	$1 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1 \cdot 10^{-8}$
ω_{elbow}	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$
link length	$1 \cdot 10^{-11}$	$1 \cdot 10^{-10}$	$1 \cdot 10^{-10}$	$1 \cdot 10^{-11}$

REFERENCES

[1] Alexandre Bernardino, Christian Vismara, Sergi Bermudez i Badia, Élvio Gouveia, Fátima Baptista, Filomena Carnide, Simão Oom, and Hugo Gamboa. 2016. A dataset for the automatic assessment of functional senior fitness tests using kinect and physiological sensors. In *International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*. 1–6.

[2] Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. 2012. Full body gait analysis with Kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. 1964–1967.

[3] Kathrin Gerling, Ian Livingston, Lennart Nacke, and Regan Mandryk. 2012. Full-body motion-based game interaction for older adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1873–1882.

[4] Lee EF Graves, Nicola D. Ridgers, Karen Williams, Gareth Stratton, Greg Atkinson, and Nigel T. Cable. 2010. The physiological cost and enjoyment of Wii Fit in adolescents, young adults, and older adults. *Journal of Physical Activity and Health* 7, 3 (2010), 393–401.

[5] Simon J. Julier and Jeffrey K. Uhlmann. 1997. A new extension of the Kalman filter to nonlinear systems. In *Int. symp. aerospace/defense sensing, simul. and controls*, Vol. 3. 182–193.

[6] Rudolph Emil Kalman et al. 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 82, 1 (1960), 35–45.

[7] Matthias Kranz, Andreas MöLler, Nils Hammerla, Stefan Diewald, Thomas PlöTz, Patrick Olivier, and Luis Roalter. 2013. The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices. *Pervasive and Mobile Computing* 9, 2 (2013), 203–215.

[8] I-Min Lee, Eric J. Shiroma, Felipe Lobelo, Pekka Puska, Steven N. Blair, Peter T. Katzmarzyk, Lancet Physical Activity Series Working Group, et al. 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet* 380, 9838 (2012), 219–229.

[9] Zaid A Mundher and Jiaofei Zhong. 2014. A real-time fall detection system in elderly care using mobile robot and kinect sensor. *International Journal of Materials, Mechanics and Manufacturing* 2, 2 (2014), 133–138.

[10] Štěpán Obdržálek, Gregorij Kurillo, Ferda Ofli, Ruzena Bajcsy, Edmund Seto, Holly Jimison, and Michael Pavel. 2012. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In *Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE*. 1188–1193.

[11] Department of Health and Human Services (US). 2008. *Physical Activity Guidelines for Americans*. Department of Health and Human Services (US).

[12] Karen Otte, Bastian Kayser, Sebastian Mansow-Model, Julius Verrel, Friedemann Paul, Alexander U. Brandt, and Tanja Schmitz-Hübsch. 2016. Accuracy and reliability of the kinect version 2 for clinical measurement of motor function. *PloS one* 11, 11 (2016), –0166532.

[13] Monish Parajuli, Dat Tran, Wanli Ma, and Dharmendra Sharma. 2012. Senior health monitoring using Kinect. In *2012 Fourth International Conference on Communications and Electronics (ICCE)*. 309–312.

[14] Paola Pierleoni, Alberto Belli, Lorenzo Palma, Marco Pellegrini, Luca Pernini, and Simone Valenti. 2015. A high reliability wearable device for elderly fall detection. *IEEE Sensors Journal* 15, 8 (2015), 4544–4553.

[15] Roberta E. Rikli and C. Jessie Jones. 2013. *Senior fitness test manual*. Human Kinetics.

[16] Microsoft Robotics. 2017. Kinect Sensor. (2017). Retrieved October 06, 2017 from https://msdn.microsoft.com/en-us/library/hh438998.aspx

[17] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM* 56, 1 (2013), 116–124.

[18] Eric A. Wan and Rudolph Van Der Merwe. 2000. The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*. 153–158.

[19] Qifei Wang, Gregorij Kurillo, Ferda Ofli, and Ruzena Bajcsy. 2015. Remote health coaching system and human motion data analysis for physical therapy with microsoft kinect. *arXiv preprint arXiv:1512.06492* (2015).

[20] Rong Wang, Huiguo Zhang, and Cyril Leung. 2015. Follow me: a personal robotic companion system for the elderly. *Int. J. Inf. Technol* 21, 1 (2015).

[21] William Young, Stuart Ferguson, Sebastien Brault, and Cathy Craig. 2011. Assessing and training standing balance in older adults: a novel approach using the 'Nintendo Wii'Balance Board. *Gait & posture* 33, 2 (2011), 303–305.