# Human Pose Estimation from Imperfect Sensor Data via the Extended Kalman Filter

Vlad Joukov, Rollen D'Souza, and Dana Kulić[(✉)]

Department of Electrical and Computer Engineering, University of Waterloo,
Waterloo, Canada
{vjoukov,rs2dsouz,dana.kulic}@uwaterloo.ca

**Abstract.** Accurate human pose estimation is of vital importance for a variety of human-robot interaction applications, including cooperative task execution, imitation learning and robot-assisted rehabilitation. As robots move from controlled indoor environments to unstructured and outdoor environments, the ability to accurately measure human pose without fixed sensors becomes increasingly important. In this paper, we present a general framework for accurately estimating human pose from a variety of sensors, including body-worn inertial measurement units and cameras, that can be used in indoor and outdoor environments to accurately estimate human pose during arbitrary 3D movements. Using a kinematic model of the human body, the sensor data is fused to estimate the body joint angles and velocities using a constrained Extended Kalman Filter which automatically incorporates feasible joint limits. For periodic movement such as gait, performance can be further improved via online learning of the gait model, individualized to the user. The proposed approach can deal with intermittent data availability and measurement errors during highly dynamic movements.

**Keywords:** Human pose estimation · Motion capture · Extended Kalman Filter

## 1   Introduction

Accurate human pose estimation is of vital importance for a variety of human-robot interaction applications, including cooperative task execution, imitation learning and robot-assisted rehabilitation [1]. The gold standard for human motion data collection is marker-based motion capture, based on either optical or magnetic marker technology. A number of commercial solutions are available, e.g. [2,3]. With optical motion capture systems, small reflective markers are attached to body landmarks, and observed with a set of cameras. The marker images and exact knowledge of the relative placements of the cameras are used to extract accurate 3D positions of each marker, achieving positioning sub millimeter positioning accuracy [2]. The marker positions coupled with a kinematic

---

model of the human body and an association between each marker and the corresponding link in the model are then used to estimate the body pose. However, even with the high accuracy of marker positioning reported by the system manufacturers, the quality of pose estimation is frequently corrupted by three common measurement issues: (1) temporary marker occlusions, (2) unlabeled markers, and (3) mis-labeled markers.

Temporary marker occlusions occur when no camera has a direct line of sight to the marker, and can be caused by either occluding objects in the scene or by self occlusion. Unlabeled markers appear when a marker is measured by one or more cameras in the capture scene, but cannot be associated with any of the known markers in the kinematic model. Mis-labeled markers, also known as *marker swapping*, occur when markers are incorrectly associated to known markers in the kinematic model. This issue occurs frequently when many markers are in a small volume, for example during close contact between demonstrators or fine hand/finger movements when the hands and fingers are brought together.

These issues can be systematically addressed during off-line post processing, but for interactive applications, on-line pose estimation is required. Aristidou and Lasenby [4] proposed an approach for missing marker handling via marker position prediction, based on previously known marker positions and rigid body assumptions. Meyer et al. [5] proposed an approach for on-line marker labeling for full-body motion capture based on a custom initialization and a probabilistic iterative estimation procedure. Maycock et al. [6] propose an approach for hand movement tracking using unlabeled markers, based on a Global Nearest Neighbor (GNN) approach. Marker occlusions are handled using interpolation. Mandery et al. [7] propose an approach for pose estimation using unlabeled marker measurements using the smart sampling Kalman filter.

While marker-based systems provide high accuracy in indoor settings, they require extensive camera setup and calibration, as well as line-of-sight visibility between the cameras and the markers, implying a restricted capture space. For many practical applications, these requirements are too restrictive. Recently, alternatives to camera-based motion capture based on body-worn Inertial Measurement Units (IMUs) have been proposed [8]. While IMUs enable capture in arbitrary environments, they suffer from gyro drift and poorer pose estimation accuracy than marker-based systems. Kalman filter approaches coupled with a kinematic model of the body have frequently been applied for IMU-based pose estimation, to fuse the accelerometer and gyroscope measurements and deal with gyro drift [9,10].

In this paper, we propose a general framework for human pose estimation based on the Extended Kalman Filter (EKF) and a skeleton model of the body. Using the motion model and state and observation covariances estimated by the EKF, we automatically eliminate missing or incorrectly measured markers, perform marker matching and pose estimation, handle joint limits and sensor measurement noise and bias. The proposed approach has been extensively validated with a variety of dynamic movements and demographics.

## 2   Proposed Approach

### 2.1   Problem Statement

We model the human body with an articulated rigid body skeleton, where adjacent bones are articulated via a set of $N$ *joint angles* $q_i, i \in 1...N$. The joint vector additionally includes the 6 dimensional pose of the body relative to the inertial frame. For the case of marker-based motion capture, our measurement consists of $M_k$ marker Cartesian positions and velocities, the number of which may be different at each time step $k$.

$$\mathbf{z}_k^m = \begin{bmatrix} \mathbf{y_{M_k}} \\ \mathbf{\dot{y}_{M_k}} \end{bmatrix}$$

where $\mathbf{y_{M_k}}$ is the vector of Cartesian marker positions, and $\mathbf{\dot{y}_{M_k}}$ is the vector of Cartesian marker velocities.

For IMU based motion capture, IMUs are attached to a set of body limbs, and our measurement consists of angular velocities and linear accelerations measured at each IMU:

$$\mathbf{z}_k^i = \begin{bmatrix} \omega_{\mathbf{K}} \\ \mathbf{a_K} \end{bmatrix}$$

where $K$ is the number of IMUs, $\omega_{\mathbf{K}}$ is the vector of angular velocity and $\mathbf{a_K}$ is the vector of linear acceleration measurements in the local IMU frame.

Given the observations at frame $k$, our objective is to estimate the pose $q_k$. Additionally, in the case of marker measurements, to deal with missing and incorrectly labeled or unlabeled markers, we must first associate the observed markers with the corresponding skeleton location, and discard incorrect markers.

### 2.2   The Extended Kalman Filter Formulation

The EKF state consists of the joint positions, velocities and accelerations, $\mathbf{x} = [\mathbf{q} \, \mathbf{\dot{q}} \, \mathbf{\ddot{q}}]^T$. We assume a constant acceleration state evolution model, such that

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & dt & dt^2/2 \\ 0 & 1 & dt \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k$$

where $\mathbf{x}_k$ is the vector of joint angles, velocities and accelerations at time $k$, $dt$ is the sampling time interval and $\mathbf{w}$ is the process noise, assumed to be zero-mean Gaussian noise with covariance $Q_k$.

The observations are related to the state via the non-linear forward kinematics, $\mathbf{z} = h(\mathbf{x})$. To perform state estimation, the forward kinematics is linearized about the current operating point.

$$\mathbf{z}_k = \begin{bmatrix} J \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k$$

where $J$ is the Jacobian of the measurement equation with respect to the state $\mathbf{x}$, and $\mathbf{v}$ is the observation noise, assumed to be zero-mean Gaussian noise with covariance $R_k$.

The unconstrained EKF formulation described above can lead to joint angle estimates which are not physically feasible. To ensure that the estimated joint angles remain physically feasible, the joint angles are constrained to remain within joint limits, by restricting the Kalman gain to ensure that the updated state remains within the constraints [11,12].

In IMU based estimation, another significant source of error is the gyroscope drift, particularly about rotation axes parallel with gravity, i.e., the yaw rotation in the world frame. With some apriori knowledge of the motion, we can estimate the mean yaw angle of links with respect to the torso or the world frame. Placing a virtual yaw sensor on a link and assuming it always takes this mean as measurement effectively prevents drift from accumulating [13]. The measurement noise covariance of the virtual sensor can be used as a tuning parameter to allow for accurate yaw motion estimation while reducing the effect of drift. Figure 1 shows the virtual yaw sensor drift correction for a gyroscope with $0.01\,rad/s$ bias experiencing sinusoidal motion about the world $z$ axis.

For the case of periodic movement, the assumption of constant acceleration can be removed by learning an individualized model of the periodic movement over time, to more accurately model the acceleration. This approach explicitly models the state as a parameterized sum of sinusoids, and learns the model parameters during online observation of the movement. The learned model enables drift free integration of the velocity and acceleration measurements, and improves pose estimation during periodic movement such as gait [13].

At each time step, the EKF estimates the measurement covariance $P_k$, based on the Kalman filter update equations [14].
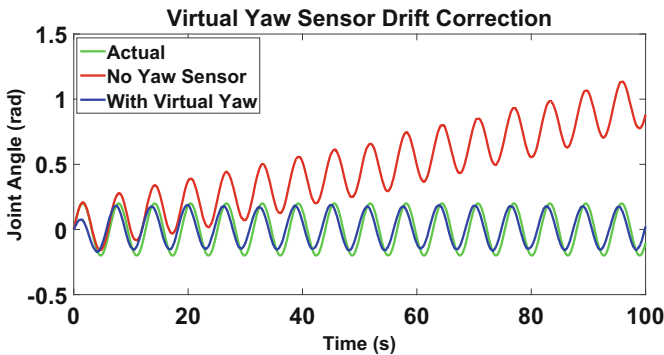


**Fig. 1.** Tracking of a single joint model with an IMU sensor experiencing sinusoidal motion about the world $z$ axis. EKF continues to integrate the bias ($0.01\,rad/s$) in the gyroscope measurement accumulating error in the joint angle estimate. A virtual yaw sensor is added with a mean measurement of 0, it does not allow the bias error to accumulate and maintains accurate motion estimation.

## 2.3  Incorrect and Missing Marker Detection

To allow for incorrect and missing markers during online measurement, at each time step, the probability distribution of the location of each model marker is predicted using the current state estimate, the observation model, and the measurement covariance. For each observed marker, the probability that the marker is generated by the model probability distribution is computed. The observed marker with the highest probability is associated with the model marker. This approach simultaneously deals with swapped and unlabeled markers.
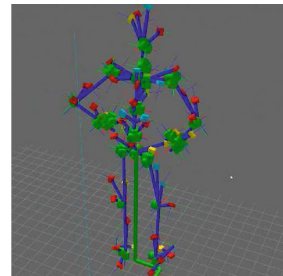
To handle missing markers, the observation vector size varies at each time step, based on the number of markers observed. Missing markers therefore do not contribute to the measurement update; only the observable markers and the state evolution model are used to estimate the joint pose. Information from the observable markers and the estimates of the joint pose and velocity allow the filter to smoothly deal with temporary marker occlusions.
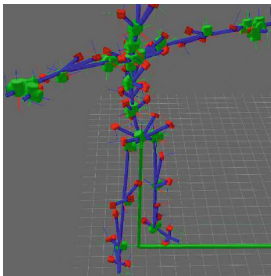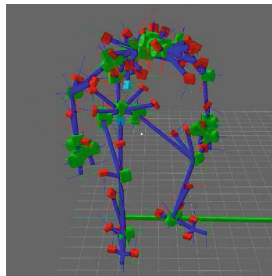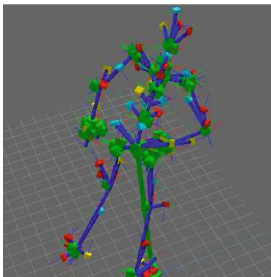


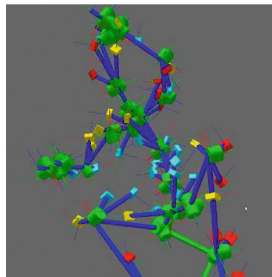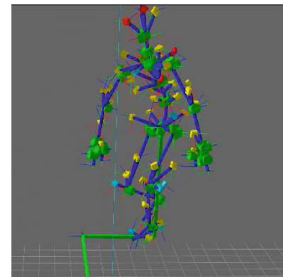(a) Initial T-pose.          (b) Seated pose.          (c) Standing with another actor to the left.

(d) Pushed by other actor into an acted fall.     (e)  After  fall,  the back  and  markers  near spine/torso is occluded.     (f) Standing up after fall and restoring T-pose.

**Fig. 2.** Sequence of frames of sample sequence: two actors interact and one actor then falls to the ground, occluding the posterior markers. The actor remains on a mattress for around twenty seconds before standing up. Visible, attached markers shown as red boxes. Cyan boxes refer to markers that are missing. Yellow boxes refer to identified mislabeled markers.

## 3    Experimental Results

The proposed approach based on marker data is evaluated with a dynamic motion capture dataset with significant occlusions, marker-swapping and unlabeled markers. Figure 2 illustrates an exemplar sequence, consisting of two actors – both outfitted with markers – interacting briefly followed by an acted fall, with significant marker occlusions during ground contact.

We do not assume that any labels are correct due to the possible swaps and frequent occlusions in the raw data, and the procedure described in Sect. 2.3 is applied to all marker measurements at each time step to perform labeling prior to pose estimation. In the initial frame, observed markers are first assigned to the actors – and their associated model markers – of the motion capture scene. An example sequence illustrating marker trajectories and the recovered pose is shown in Fig. 3.
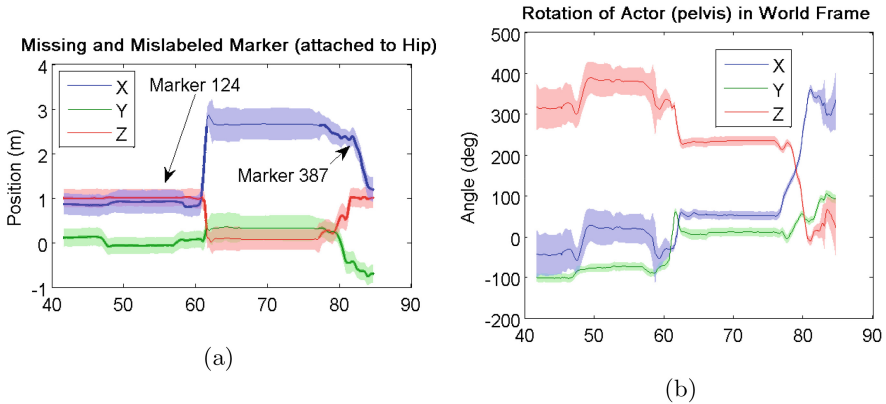


**Fig. 3.** A sample marker trajectory (a) illustrating the position of an occluded and mislabeled marker, together with the corresponding recovered joint angle trajectory (b). One standard deviation (covariance estimate by EKF) is shown as the filled region surrounding the estimate. Estimates made by EKF (solid) and observed marker data (thick dotted, if visible) are both shown.

The proposed approach is compared to off-line pose estimation and a Jacobian inverse based approach [15]. For off-line pose estimation, pre-processing is applied to fill-in missing marker data, remove spurious markers and correct any mislabeling, and the pose is then estimated using global optimization. The Root Mean Squared Error (RMSE) between the measured marker positions and their estimates based on the recovered joint angles and the forward kinematics are compared in Table 1.

As can be seen from Table 1, the proposed method significantly outperforms the Jacobian-based approach, and achieves performance comparable to the off-line method. In particular, performance for the right hand is significantly

**Table 1.** Root mean squared errors (cm) of the off-line, Jacobian inverse and the proposed EKF-based methods. Standard deviation of errors (cm) are in parentheses. The marker that was visible for most of the dataset was chosen on each extremity for these calculations.

| Marker subset | Post-processed | Jacobian | EKF |
|---|---|---|---|
| Left Hand | 3.95 (1.87) | 3.65 (0.38) | **2.34** (**0.25**) |
| Right Hand | 3.47 (1.51) | 14.46 (2.58) | **2.11** (**0.33**) |
| Left Foot | **1.79** (0.18) | 2.88 (0.64) | 2.60 (**0.13**) |
| Right Foot | **2.00** (**0.33**) | 4.28 (1.12) | 2.56 (0.45) |
| Head | 5.43 (2.20) | 3.06 (0.37) | **2.84** (**0.18**) |
| Right Shoulder | 3.51 (1.83) | 9.60 (0.86) | **1.61** (**0.05**) |
| Left Shoulder | **2.52** (1.15) | 3.09 (0.76) | 3.96 (**0.56**) |

improved; the right hand is occluded for over ten seconds followed by mislabeling of the marker which the Jacobian inverse based approach cannot recover from. The proposed method detects the incorrect label and reassigns the marker appropriately.

Also noteworthy is the standard deviation of the error, shown in Table 1 within parentheses. The proposed approach generally minimizes this deviation,
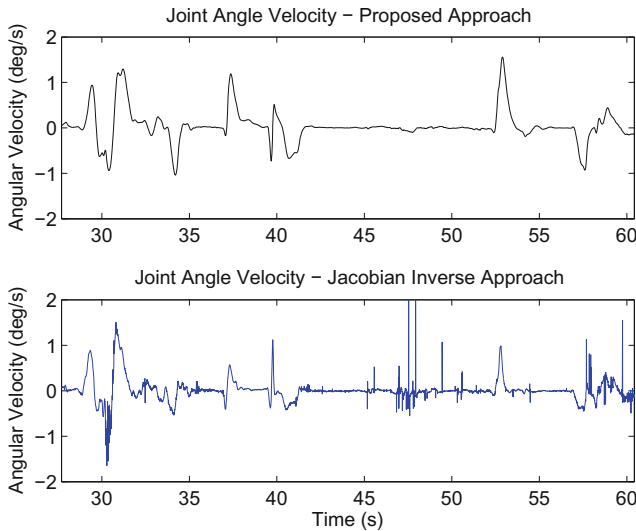


**Fig. 4.** Marker jumps in the raw data impact the continuity of joint angle velocities produced by the Jacobian inverse approach (bottom). The proposed approach (top) preserves continuity by throwing away data that is unlikely to match the prediction made by EKF.
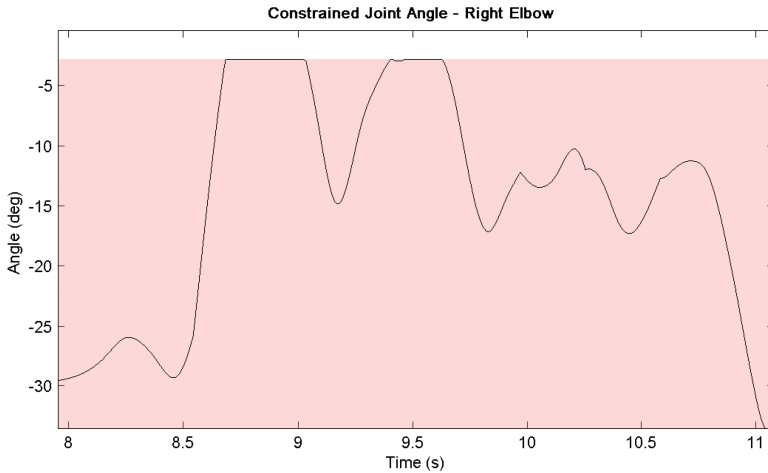
**Fig. 5.** Joint angles produced for the right elbow joint. Range of motion permitted is filled in with red.

indicating smoother and more consistent generated motion as opposed to that of the Jacobian inverse approach. One possible reason for this can be the marker jumps found in the raw data. Marker jumps are where markers erroneously jump a few meters yet remain visible. Since the Jacobian inverse approach cannot effectively determine which data to reject, it generates visible discontinuities in the resulting joint angle velocities, shown in Fig. 4. In addition to rejecting these jumping markers the proposed approach may temporarily assign a nearby, unassigned marker as per Sect. 2.3.

Figure 5 illustrates the constrained estimation, taking into account joint limits for the elbow joint, which is limited by a 150° range. The over-extension of the elbow joint is prevented during the occlusion time frames.

The proposed IMU only approach was compared to the marker based EKF. Participants were asked to march in place while wearing IMU sensors at the waist, thighs, and ankles. Three markers were placed on each IMU to estimate their orientation with respect to the limbs. Markers were also placed on the ankles, knees, and hips; these were used to estimate the joint centers and link lengths to generate the kinematic models as well as for ground truth inverse kinematics using the proposed marker EKF approach. Figure 6 shows the right knee joint angle estimation of the IMU EKF and Periodic-EKF [13] approaches compared to the marker based estimation, the proposed methods achieve an accuracy of 2.79 and 2.07 degrees RMSE respectively. IMU only estimation achieves accuracy comparable to that of camera based motion capture and can be utilized in environments for which camera based systems are not suitable.
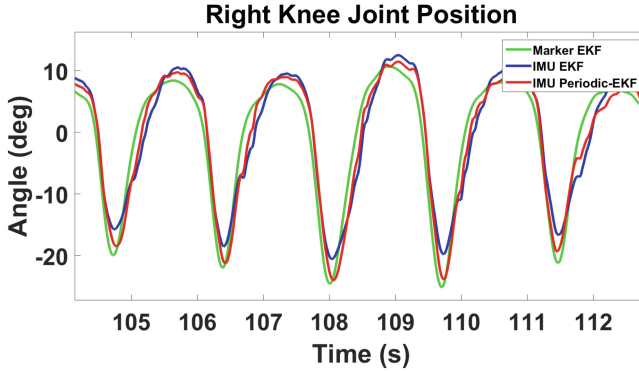
**Fig. 6.** Right knee joint angle estimation with IMU EKF and Periodic-EKF compared to the marker based EKF for marching motion. The IMU based approaches are comparable in accuracy to the motion capture. Due to periodic nature of marching, the IMU Periodic-EKF converges to an accurate motion model and improves estimation.

## 4 Conclusions and Future Work

In this paper, a comprehensive framework for online pose estimation was developed based on the Extended Kalman filter. The proposed approach models the human body as an articulated skeleton, and estimates the body position and orientation in space, together with the joint positions and velocities via fusion of a motion model and noisy measurements. Imperfect motion capture measurements, including missing and mislabeled markers, as well as IMU measurements, can be incorporated into the same framework. The proposed method is evaluated on a variety of datasets and demonstrates improved performance over state-of-the-art systems.

## References

1. Kulić, D., Venture, G., Yamane, K., Demircan, E., Mizuuchi, I., Mombaur, K.: Anthropomorphic movement analysis and synthesis: a survey of methods and applications. IEEE Trans. Robot. **32**(4), 776–795 (2016)
2. Motion analysis. http://www.motionanalysis.com
3. Vicon. https://www.vicon.com/
4. Aristidou, A., Lasenby, J.: Real-time marker prediction and CoR estimation in optical motion capture. Vis. Comput. **29**(1), 7–26 (2013)
5. Meyer, J., Kuderer, M., Müller, J., Burgard, W.: Online marker labeling for fully automatic skeleton tracking in optical motion capture. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 5652–5657. IEEE (2014)
6. Maycock, J., Rohlig, T., Schroder, M., Botsch, M., Ritter, H.: Fully automatic optical motion tracking using an inverse kinematics approach. In: 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), pp. 461–466 (2015)

7.  Steinbring, J., Mandery, C., Pfaff, F., Faion, F., Asfour, T., Hanebeck, U.D.: Real-time whole-body human motion tracking based on unlabeled markers. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (2016)
8.  Roetenberg, D., Luinge, H., Slycke, P.: Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. Xsens Motion Technologies BV, Technical report (2009)
9.  El-Gohary, M., McNames, J.: Shoulder and elbow joint angle tracking with inertial sensors. IEEE Trans. Biomed. Eng. **59**(9), 2635–2641 (2012)
10. Lin, J.F., Kulić, D.: Human pose recovery using wireless inertial measurement units. Physiol. Measur. **33**(12), 2099–2115 (2012)
11. Gupta, N., Hauser, R.: Kalman filtering with equality and inequality state constraints. arXiv preprint (2007). arXiv:0709.2791
12. Bonnet, V., Daune, G., Joukov, V., Dumas, R., Fraisse, P., Kulić, D., Seilles, A., Andary, S., Venture, G.: A constrained extended kalman filter for dynamically consistent inverse kinematics and inertial parameters identification. In: IEEE-RAS International Conference on Biomedical Robotics and Biomechatronics, pp. 952–957 (2016)
13. Joukov, V., Bonnet, V., Karg, M., Venture, G., Kulić, D.: Rhythmic EKF for pose estimation during gait. In: IEEE-RAS International Conference on Humanoid Robots, pp. 1167–1172 (2015)
14. Welch, G., Bishop, G.: An introduction to the kalman filter. Technical report TR 95–041, Department of Computer Science, University of North Carolina at Chapel Hill (2006)
15. Yamane, K., Nakamura, Y.: Natural motion animation through constraining and deconstraining at will. IEEE Trans. Visual. Comput. Graph. **9**(3), 352–360 (2003)