TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

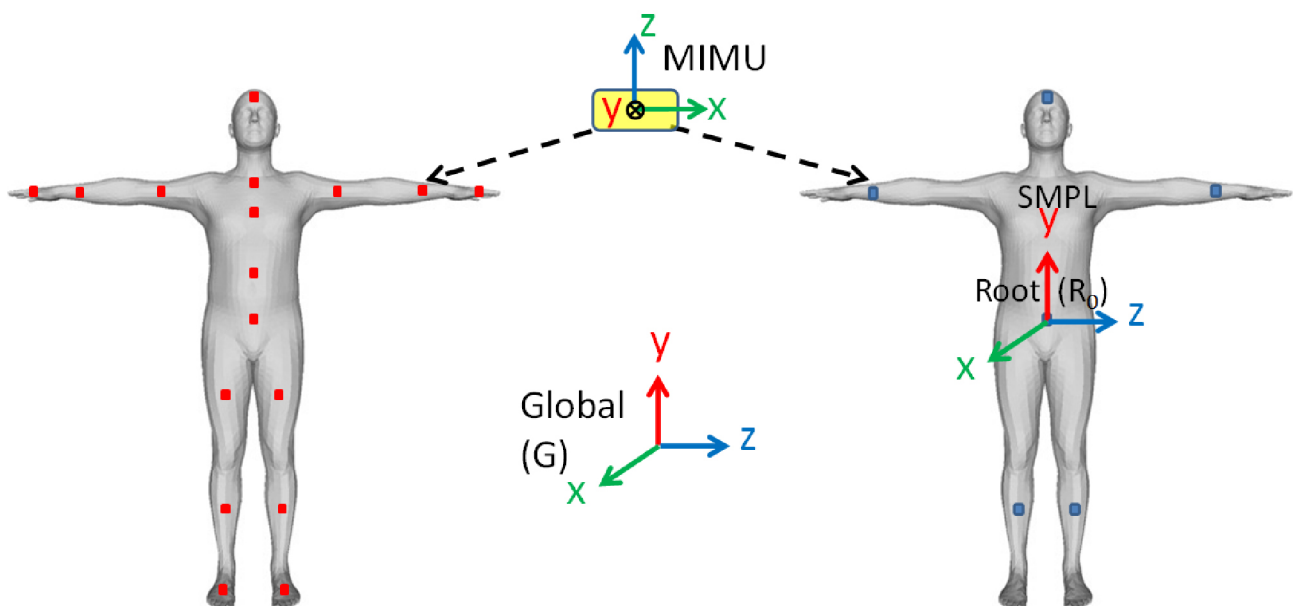# IMPROVED SENSOR FUSION AND DEEP LEARNING OF 3D HUMAN POSE FROM SPARSE MAGNETIC INERTIAL MEASUREMENT UNIT

Hammad Tanveer Butt

Improved Sensor Fusion and Deep Learning of 3D Human Pose From Sparse Magnetic Inertial Measurement Units

Dissertation approved by the
Department of Computer Science
Technische Universität Kaiserslautern
for the award of the Doctoral degree
Doctor of Engineering (Dr.-Ing.)
to

Hammad Tanveer Butt

Date of Defence: 04. November 2022

Dean: Prof. Dr. Jens Schmitt
Reviewer: Prof. Dr. Didier Stricker
Reviewer: Prof. Dr. Imran Shafi

DE-386

Abstract

3D joint angles based human pose is needed for applications like activity recognition, musculoskeletal health, sports biomechanics and ergonomics. The microelectromechanical systems (MEMS) based magnetic-inertial measurement units (MIMUs) can estimate 3D orientation. Due to small size, MIMUs can be attached to the body as wearable sensors for obtaining full 3D human pose and this system is termed as inertial motion capture (i-Mocap). But the MIMUs suffer from sensor errors and disturbances, due to which orientation estimated from individual MIMUs can be erroneous. Accurate sensor calibration is essential and subsequently alignment of these sensors to body segments must also be precisely known, which is called sensor-to-segment calibration. Sensor fusion is employed to address the disturbances and noise in MIMUs. Many state-of-art inertial motion capture approaches ignore the magnetometer and only use IMUs to reduce the error arising from inhomogeneous magnetic field. These algorithms rely on kinematic constraints and assumptions regarding joints and are based on IMUs located on the adjacent body segments. The full body coverage requires 13-17 such units and can be quite obtrusive. The setting up and calibration of so many wearable sensors also take time.

This thesis focuses on 3D human pose estimation from a reduced number of MIMUs and deals with this problem systematically. First we propose an accurate simultaneous calibration of multiple MIMUs, which also learns the uncertainty of individual sensors. We then describe a novel sensor fusion algorithm for robust orientation estimation from an MIMU and for updating sensors calibration online. The residual errors in both sensor calibration and fusion can result in drift error in the joint angles. Therefore, we present anatomical (sensor-to-segment) calibration in which an orientation offset correction term is updated and used for online correction of residual drift in individual joint angles. Subsequently we demonstrate that 3D human joint angle constraints can be learned using a data-driven approach in a high dimensional latent space. Owing to temporal and joint angle constraints, it is possible to use only a reduced set of sensors (as opposed to one sensor per segment) and still obtain 3D human pose. But the spatial and temporal prior learning from data is often limited due to finite set of movement patterns in most datasets. This introduces uncertainty while estimating 3D human pose from sparse MIMU sensors. We propose a magnetometer robust orientation parameterization and a data-driven deep learning framework to predict 3D human pose with associated uncertainty from sparse MIMUs. The model is evaluated on real MIMU data and we show that the uncertainty predicted by the trained model is well-correlated with actual error and ambiguity.

Acknowledgment

Contents

1. Introduction

Estimation of 3D human pose is an important goal in computer vision, human motion capture, augmented/virtual reality (AR/VR), human-computer interfaces (HCI) and biomechanical analysis both for clinical and non-clinical use. The applications are as diverse as posture and gait analysis, rehabilitation exercises, sports analytics, task ergonomics, man-machine interaction, character animation and control in 3D games and simulations. Either extrinsic sensors like cameras or body fixed ego-centric sensors are used for 3D human pose estimation. In later category, inertial and magnetic sensors have become quite common with the advent of Microelectromechanical systems (MEMS) in recent years. With miniature size, low cost and low power requirements, body worn magnetic-inertial measurement unit (MIMU) sensors are an excellent choice for non-invasive human motion capture, biomechanical analysis and body tracking in everyday ambulatory settings. This technology is called inertial motion capture (i-Mocap) or wearable motion capture. It is based on consumer grade magnetic-inertial measurement units (MIMUs). Each consumer grade MIMU comprises of low cost and small size MEMS based 3-axis magnetometer, 3-axis rate gyro and 3-axis accelerometers. Compared with camera based optical 3D pose estimation (Figure 1.1), body worn inertial motion capture is robust to occlusion and also suitable for pose estimation in the wild.

(a)   (b)
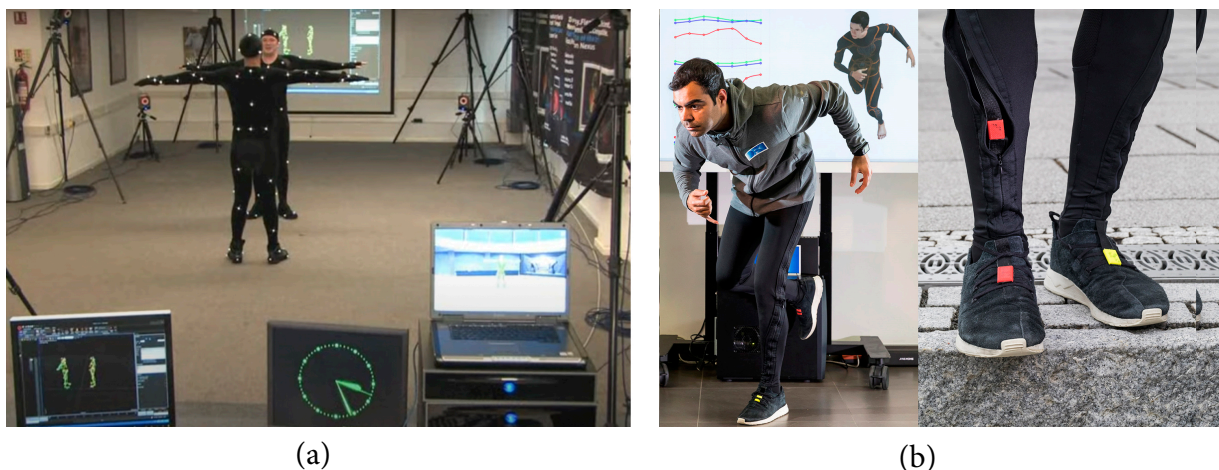
Figure 1.1.: Optical versus Inertial Motion Capture.

The human body is assumed as comprising of rigid body segments articulated at joints and one sensor per segment is sufficient to compute 3D joint angle if adjacent segment orientations are known. The main limitation and challenge to i-Mocap systems however is long term drift of sensors. Being low cost, these MEMS based sensors are not precise and robust. Despite

calibration, residual errors and parameter drifts are common. The MEMS based sensors show degraded performance when run for long time, especially indoors. This is due to the fact that the accelerometer/ magnetometer are used to complement rate gyro for orientation estimation, but the disturbances registered by these sensors lead to sensor fusion errors. The error in heading angle or yaw is particularly significant due to persistent nature of magnetic inhomogeneity in the environment. This ultimately results in inaccurate and drifting joint angle estimates between body segments which would need correction. Thus accurate calibration and robust sensor fusion [15, 9] is essential prerequisite to inertial human motion sensing.

Figure 1.2.: Xsens Link (wired nodes) and Awinda (wireless nodes) i-Mocap Systems [1].

The human body has constrained degrees of freedom (DoF) and temporal coherence and smoothness is an important feature of human motion. Many existing kinematic based i-Mocap frameworks therefore use predefined constraints to reduce measurement errors or drifts [143, 134]. The number of sensors or special clothing makes i-Mocap rather obtrusive. Commercially available i-Mocap systems like Xsens (Figure 1.2) use up to 13-17 sensors [1]; one per each body segment for full kinematic capture. The setting up and calibration of so many wearable sensors take time. Reducing the number of sensors on body makes the 3D pose estimation less obtrusive and thus this modality can be used for daily activity monitoring, ergonomics and wearable health applications easily. The data driven approach to 3D human pose estimation using reduced sensors has been demonstrated using deep learning [52] on a large synthetic dataset. However the 3D pose estimation with reduced sensors and a learned model usually depend greatly on the correlation in the activity. Thus the predictions are often inaccurate when inputs are outside the domain of training data. In such a case, an estimate of uncertainty of

predicted 3D pose becomes important.

1.1. Motivation

The main aim of this thesis is to investigate the problem of inertial motion capture (i-Mocap) using sparse or reduced number of MIMU sensors in an end-to-end manner and systematically address the challenges mentioned in the introduction. This is accomplished in the framework of data-driven deep learning and by improved adaptive filtering and calibration. The objective is to arrive at a robust framework in which the intrinsic drift and extrinsic disturbances of MIMU sensors due to inhomogeneous magnetic field do not lead to an unbounded error in 3D human pose. Our research is driven by the motivation that sparse number of MIMU sensors, with their miniature size and overall reduced power requirement makes such a configuration a desirable choice for wide range of ambulatory applications. A few motivating examples can be the use of sparse wearable sensors for monitoring physical and rehabilitation exercises or their use for measuring the ergonomic performance or in biomechanical analysis of sports activity. As the number of wearable sensors is reduced, it becomes less obtrusive for the users and even challenging activities can be monitored. But the reduced number of sensors means higher uncertainty and we address this by a data driven learning of temporal priors. Our approach provides a well calibrated uncertainty of predicted 3D pose for this scenario. The data-driven estimation of kinematic uncertainty of 3D human pose may be used for uncertainty driven information fusion with other modalities or with the output of other uncertainty based algorithms.

1.2. Problem Definition and Approach

This thesis deals with the problem of 3D human pose estimation from a sparse or reduced set of 9-axis IMUs (also called MIMUs). Each MIMU unit or sensor node comprises of low cost consumer grade MEMS based 3-axis magnetometer, 3-axis rate gyro and 3-axis accelerometers. The sparse configuration which is investigated, comprises of six MIMU sensor nodes, one each worn by the subject on left wrist, right wrist, left lower leg, right lower leg, lower spine and head (Figure 1.3). The sensor at lower spine is treated as reference or root sensor. The five sensors located at lower/upper limbs and head provide orientation measurements w.r.t. root sensor which is used to predict full 3D human pose. The 3D orientation of each MIMU node is estimated using a robust sensor fusion algorithm from magnetic and inertial sensors. The sensors are assumed to be rigidly attached to body segments and sensor-to-segment calibration is carried out using an initial I-pose. We have approached the entire problem in a systematic way as follows:

1. We employ a wearable MIMU based body tracking system developed as part of BIONIC (Body Information On Integrated Chip) project at WearHealth Group of German Research Center for Artificial Intelligence (DFKI), Kaiserslautern. The system is shown in Figure 1.4. It is designed using low cost state-of-art hardware and MEMS sensors to reduce errors and improve computational latency. The system allows for high rate data acquisition

and sensor fusion at low power budget. The package size of sensing nodes is small, and we have used textile wires which make it very flexible. Thus entire system can be easily integrated with body worn suit/pants, making it ideal for data acquisition in ambient settings. The aim of using an in-built i-Mocap sensors (compared to commercial i-Mocap system) is to implement and test the embedded algorithms developed during the research for improved sensor calibration and sensor fusion **(Chapters 3 and 4)**.
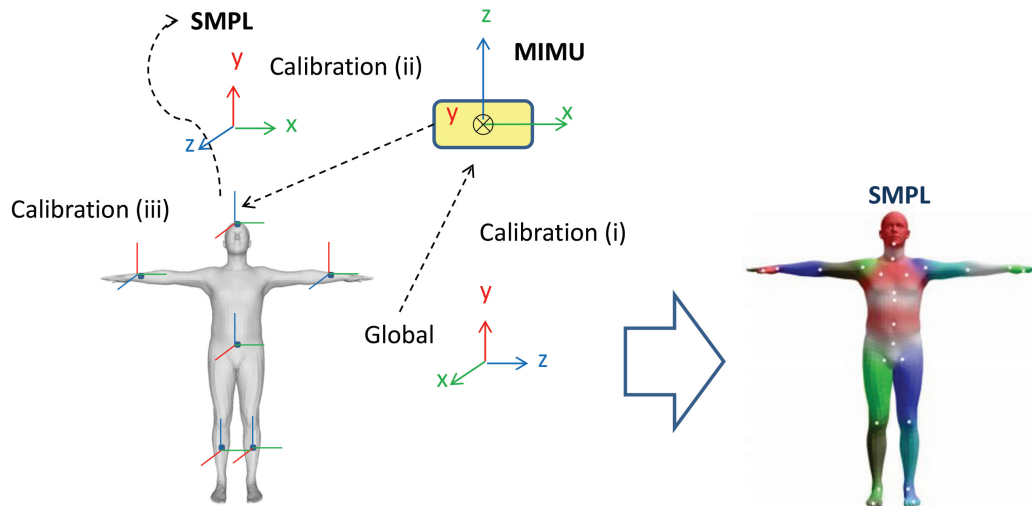
Figure 1.3.: Problem Definition: 3D human pose from sparse MIMUs.
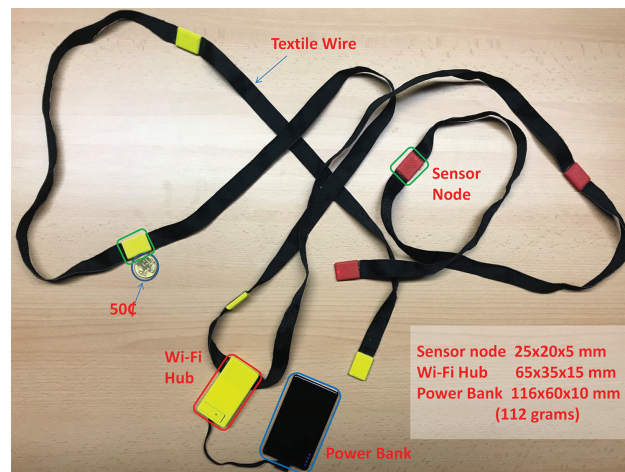
Figure 1.4.: Complete Flexible Harness based MIMU Tracking System

2. The MEMS MIMUs require calibration of systematic errors of bias, sensitivity, non-orthogonality and misalignment, which vary with temperature and duration of use. We propose and test an *offline* method for easy and fast calibration of these sensors together

in an optimal way, while mounted on a single platform using kinematic constraints. The uncertainty of sensor output is also learned as part of our model. Compared with the existing state-of-art methods, our algorithm gives more consistent readings of all MIMUs and our framework also predicts the associated uncertainty of the sensor output (**Chapter 3**). The uncertainty prediction of individual sensors is particularly helpful in the sensor fusion.

3. Wearable systems for human motion capture based on low cost MEMS MIMUs show degraded performance when these systems are run for long time, especially indoors. The limitation arises due to the inhomogeneous magnetic field in the indoor environment, the high dynamic accelerations of body segments and due to bias variation over time and also with temperature. In our approach, we perform sensor fusion and online self-recalibration of accelerometers, magnetometers and rate gyros in a distributed algorithm, in order to provide an accurate orientation estimate. An auto-calibration algorithm runs on each sensing node and provides correction to gyro biases. Together a hybrid sensor fusion and calibration algorithm is run on the hub for every sensor node. Our sensor fusion (**Chapter 4**) uses a novel adaptive covariance based EKF which makes it robust to both dynamic body accelerations as well as inhomogeneous magnetic field. It also intelligently updates the calibration of accelerometer and magnetometer, when disturbances are low.

4. Accurate joint angle estimation using MIMUs requires not only robust sensor fusion, but also a precise anatomical (sensor-to-segment) calibration. Any error ultimately results in inaccurate and drifting joint angle estimates between body segments that would require some means of correction. In order to account for any residual drift of joint angles, we propose a novel correction term in our anatomical formulation that performs online correction of residual drift in individual joint angles and updates it as an orientation offset. This offset correction for joint angle is performed automatically when the limb or extended torso are in neutral pose and this condition is judged using accelerometers. Overall our approach achieves precise orientation estimates in highly dynamic conditions and avoids drift or error accumulation due to inhomogeneous magnetic fields (**Chapter 5**).

5. After we have achieved a robust performance of i-Mocap using body worn sensors on adjacent body segments (traditional i-Mocap) by addressing various sources of error, we then demonstrate that 3D human skeleton has joint limit constraints that can be effectively learned in latent space using a data-drive approach from 3D motion capture (mocap) datasets. We formulate joint constraints with a parameterization (s_1, s_2, τ) of swing-twist space to improve upon the limitations of previous 3D angle representations. Our parameterization is applied on Human3.6M dataset to create (s_1, s_2, τ) map for each joint. These maps are used to generate 'synthetic' datasets in entire joint space. The neural network discriminators are then trained on synthetic dataset to learn valid/invalid 3D joint rotations. The accuracy of these discriminators have been found varying between $[95\% - 99\%]$ for different joints. Our work validates that neural networks can model and learn 3D human joint angles in high dimensional latent space with high accuracy (**Chapter 6**).

6. The ability of neural networks to learn 3D human joint angles from 3D mocap datasets is exploited for 3D human pose estimation from sparse sensors data. We propose a deep learning based framework that learns data-driven temporal priors to perform 3D human pose estimation from six body worn Magnetic Inertial Measurement units (MIMUs). Our work estimates 3D human pose with associated uncertainty from sparse body worn sensors **(Chapter 7)**. We derive and implement a 3D angle representation that eliminates yaw angle (or magnetometer dependence) and show that 3D human pose is predicted even with reduced pitch/roll information from sparse sensors with well calibrated uncertainty which is correlated with ambiguity and actual errors. We have demonstrated our results on two real sensor datasets; DIP-IMU [52] and Total Capture [136] and have come up with the state-of-art accuracy on both. Our work confirms that the main limitation of sparse sensor based 3D human pose prediction is the lack of temporal priors. Therefore fine-tuning on a small synthetic training set of target domain, improves the accuracy.

1.3. Contributions

The main contribution of this research is the systematic study of the 3D human pose estimation from sparse body worn MIMU sensors using a data-driven approach, while also predicting the associated uncertainty. The related technical contributions are as follows:

- An easy and fast method for calibration of multiple MIMU units simultaneously and together in an optimal way using mutual kinematic constraints. This calibration model also learns uncertainty of sensor outputs both in static and dynamic condition. This contribution is based on **Chapter 3** and peer reviewed in a journal paper:

  – Hammad Tanveer Butt, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy, Ankit Kumar and Didier Stricker, **"Simultaneous End User Calibration of Multiple Magnetic Inertial Measurement Units with associated Uncertainty,"** *IEEE Access*, Volume 9, pp. 26468-26483, 2021.

- A robust senor fusion and online calibration algorithm for MIMU units. It uses a novel adaptive covariance based EKF which makes it robust to both dynamic body accelerations as well as inhomogeneous magnetic fields. It intelligently updates the calibration of accelerometer and magnetometer, when disturbances are low. This contribution is based on **Chapter 4** and presented in following peer reviewed conference papers:

  – Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Maria Alejandra Sanchez and Didier Stricker, **"Development of High Rate Wearable MIMU Tracking System Robust to Magnetic Disturbances and Body Acceleration,"** In *Intelligent Systems Conference (IntelliSys), London, UK*, 2019. (**Best scientific paper award**)

  – Hammad Tanveer Butt, Mathias Musahl, Manthan Pancholi, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Intelligent Sensor Fusion with Online Distributed MIMU Calibration for Wearable Motion Capture (Poster),"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

- A novel correction term in anatomical (sensor-to-segment) calibration that performs on-line correction of drift in individual joint angles and updates it as an orientation offset. This offset correction for joint angle is performed automatically when the limb or extended torso are in neutral pose and this condition is judged using accelerometers. This contribution is based on **Chapter 5** and presented in following peer reviewed conference paper:

    – Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Inertial Motion Capture Using Adaptive Sensor Fusion and Joint Angle Drift Correction,"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

- An efficient data-driven learning of human joint constraints with a swing-twist parameterization (s_1, s_2, τ) of 3D joint angles to improve upon the limitations of previous 3D angle representations, and thus demonstrate that neural networks can model and learn 3D human joint angles in high dimensional latent space with high accuracy. This contribution is based on **Chapter 6** and presented in following peer reviewed conference paper:

    – Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker, **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"** In *International Conference on Machine Vision Application (MVA), Tokyo, Japan*, 2019. ([†] First two authors contributed equally in this work.)

    An extended version of same work was published in following peer reviewed journal paper:

    – Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker. **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"**, *IPSJ Transactions on Computer Vision and Applications*, Volume 11, pp. 1-9, 2019. ([†] First two authors contributed equally in this work.)

- Data driven deep learning of 3D human pose from temporal priors and sparse or reduced number of body worn sensors (only six) using a magnetometer robust orientation representation. the model also learns and predicts the associated uncertainty of joint angles of 3D human pose. This contribution is based on **Chapter 7** and peer reviewed in a journal paper:

    – Hammad Tanveer Butt, Bertram Taetz, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy and Didier Stricker, **"Magnetometer Robust Deep Human Pose Regression with Uncertainty Prediction Using Sparse Body Worn Magnetic Inertial Measurement Units,"** *IEEE Access*, Volume 9, pp. 36657-36673, 2021.
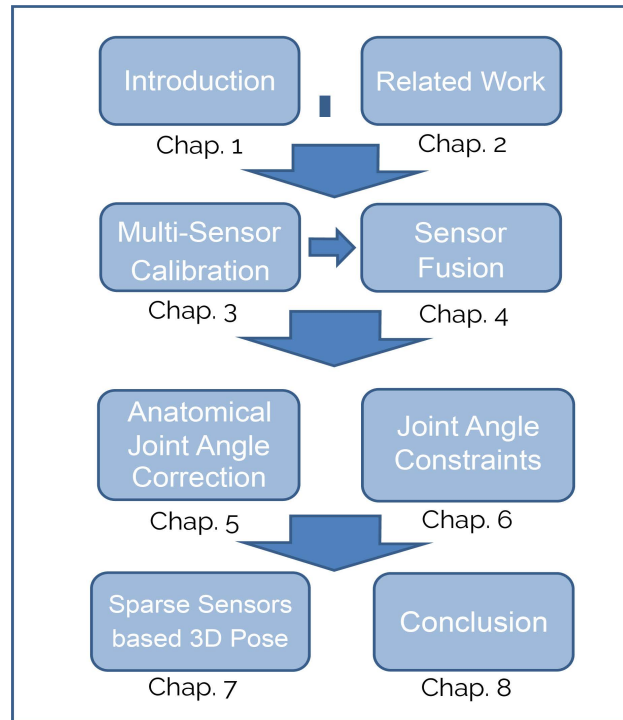
Figure 1.5.: Organization of the Thesis

1.4. Organization of the Thesis

This thesis is organized in eight chapters (Figure 1.5). **Chapter 1** is an introduction to the problem of 3D human pose estimation using magnetic-inertial sensors (i-Mocap) and defines it in the context of sparse or reduced body worn MIMUs. **Chapter 2** presents a broad review of related work in the field regarding sensor calibration, sensor fusion, joint angles and 3D human pose estimation. **Chapter 3** deals with the calibration of all sensors in an MIMU. After basic sensor models are discussed, this chapter presents an easy and fast method for calibration of multiple MIMUs simultaneously. A set of mutual kinematic constraints are used for this purpose and the calibration model also learns to predict the uncertainty of individual sensor outputs. **Chapter 4** then describes an adaptive covariance based Extended Kalman Filter (EKF) for 3D orientation estimation. The proposed EKF algorithm performs the sensor fusion of rate gyro, accelerometer and magnetometer which is robust to both body accelerations and inhomogeneous magnetic field, and also performs online calibration of accelerometer and magnetometer. **Chapter 5** presents anatomical (sensor-to-segment) calibration in which an orientation offset correction term is updated and used for online correction of residual drift in individual joint angles. **Chapter 6** highlight our work on data-driven learning of human joint constraints with a swing-twist parameterization (s_1, s_2, τ) of 3D joint angles. In this chapter, we demonstrate that neural networks can model and learn 3D human joint angles in a high dimensional latent space with high accuracy. **Chapter 7** presents our main work on a magnetometer robust orientation

parameterization and a data-driven deep learning framework to predict 3D human pose with associated uncertainty from sparse MIMUs. The deep learning model is trained based on temporal priors from synthetic MIMU data obtained from Mocap datasets. We evaluate the results and find that the main limitation arises due to lack of temporal priors. It is shown that fine-tuning on a small synthetic MIMU data representative of target domain, greatly improves accuracy. The uncertainty predicted by the trained model is found to be well-correlated with actual error and ambiguity. The associated uncertainty is high when reduced magnetometer robust orientation parameterization is used. **Chapter 8** then provides the conclusion and future direction of our work.

Overall chapters 3, 4 and 5 form **part-I** of this dissertation, dealing with improvements in traditional i-Mocap pipeline. The chapters 6 and 7 focus on data driven and deep learning approach to i-Mocap and form **part-II** of this dissertation.

1.5. Publications

Most of the work presented in this thesis has been accepted and presented in peer-reviewed conferences or journals. In the following, we provide a list of the publications derived from this work:

Journal Publications:

1. Hammad Tanveer Butt, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy, Ankit Kumar and Didier Stricker, **"Simultaneous End User Calibration of Multiple Magnetic Inertial Measurement Units with associated Uncertainty,"** *IEEE Access*, Volume 9, pp. 26468-26483, 2021.

2. Hammad Tanveer Butt, Bertram Taetz, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy and Didier Stricker, **"Magnetometer Robust Deep Human Pose Regression with Uncertainty Prediction Using Sparse Body Worn Magnetic Inertial Measurement Units,"** *IEEE Access*, Volume 9, pp. 36657-36673, 2021.

3. Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker. **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"**, *IPSJ Transactions on Computer Vision and Applications*, Volume 11, pp. 1-9, 2019.
([†] First two authors contributed equally in this work.)

Conference Publications:

1. Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Maria Alejandra Sanchez and Didier Stricker, **"Development of High Rate Wearable MIMU Tracking System Robust to Magnetic Disturbances and Body Acceleration,"** In *Intelligent Systems Conference (IntelliSys), London, UK*, 2019. (**Best scientific paper award**)

2. Hammad Tanveer Butt, Mathias Musahl, Manthan Pancholi, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Intelligent Sensor Fusion with Online Distributed MIMU Calibration for Wearable Motion Capture (Poster),"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

3. Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Inertial Motion Capture Using Adaptive Sensor Fusion and Joint Angle Drift Correction,"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

4. Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker, **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"** In *International Conference on Machine Vision Application (MVA), Tokyo, Japan*, 2019. - the long version of Conference paper is peer reviewed and published as Journal paper too. ([†] First two authors contributed equally in this work.)

2. Related Work

In this chapter, we provide a broad overview of related work to our thesis. Since a range of issues in the domain of magnetic-inertial sensors (MIMUs) and inertial motion capture (i-Mocap) are dealt within the framework of this thesis, therefore an exhaustive listing of published research for each topic is not the aim here. However for completeness, we have included all the papers which outline the main approaches to the problems within the scope of our thesis and also represent a significant contribution to the domain. This review of related work covers the period from 2001 to 2020.

2.1. MIMU Sensors Calibration

The Inertial and magnetic sensors are used both for orientation estimation and navigation [100, 25]. Low cost MEMS based MIMUs comprise of 3-axis rate gyros, 3-axis accelerometer and a 3-axis magnetometer, also termed as 9-axis IMUs. These sensors experience both systematic error and random noise. The systematic error arises from bias, misalignment, sensitivity etc. and is compensated by performing calibration. Two kinds of sensor calibration have been focused in the past literature, (1) the factory calibration and (2) field calibration. Since low end MEMS sensors employed in MIMUs for i-Mocap need frequent in-field calibration, our main focus in the review would be on it. The stochastic error of MEMS based MIMUs has also been modeled in the past in order to obtain noise parameters and bias models to be used in sensor fusion and therefore related work will be mentioned.

2.1.1. Accelerometer Calibration

The most common method to calibrate 3D accelerometer triad in MEMS magnetic-inertial measurement units is based on multi-positions [129]. A linear or affine model is assumed for 3-axis accelerometer; hence the calibration requires finding three (x,y,z) bias parameters and a 3x3 scale matrix (a total of 12 unknowns). The raw accelerometer data is collected when the IMU is static in different 3D orientations. In this condition the accelerometer only registers acceleration due to gravity. When 3D orientations are precisely known from additional equipment, the reference output readings can be used to find model parameters as a linear least square fit [94]. But in the field, only magnitude of gravity vector is a known constant, hence norm constraint on the accelerometer output is utilized to calibrate it [38]. The former leads to a nonlinear optimization problem, the solution to which is found by Gauss Newton method, until convergence criteria is met. The calibration results are always slightly different due to different noise and bias in each dataset.

In practical terms, the accelerometer has inherent noise, even when it is static. Therefore the raw readings in each static pose are averaged to reduce stochastic effect on calibration. As a minimum only four known positions are sufficient to find 12 unknowns, but optimal results are obtained when uniformly spaced 3D orientations [36, 121] are used, which reduce any procedural errors and overfitting. Since uniform space coverage requires data collection in many equally spaced positions on unit sphere, the cost in terms of time is high.

The accelerometer calibration procedure in literature has mainly focused on gravity as reference. The local estimate of gravity is obtained using Earth's charts, which might also introduce errors. This approach to calibrate has its limitations for navigation and 6D pose estimation, where an accurate estimate of linear acceleration is needed to obtain position by double integration.

2.1.2. Rate Gyro Calibration

Rate gyros are often calibrated independent of other sensor triads in MIMU. The bias offset for a MEMS rate gyro is calibrated simply while it is at rest. A precision turn-table [40] set-up is used to produce precise and constant rotation rate. MIMU is aligned with the axis of rotation of turn-table, so that only one axis of 3-axis rate gyro registers a rotation. Alternately, in the field, a fixed rotation on a horizontal surface is proposed around each rate gyro axis [152], as a means to calibrate it. The authors [96] use the accelerometer signal as reference to calibrate rate gyro, while MIMU is mounted on a wheel. The main problem with all these approaches is that the rate gyro's misalignment with others sensor triads is not addressed. The procedure only calibrates the 3D rate gyro to the package body, if a perfectly aligned mounting on test rig is assumed. These procedural errors introduce uncertainty in the gyro calibration.

2.1.3. Magnetometer Calibration

A simple figure-of-eight [78] procedure is followed to collect data over complete 3D sphere for magnetometer calibration. But the key assumption is that the magnetometer encounters a non-varying magnetic field during the procedure. Using this fact, a norm constraint on the magnetometer output is utilized to calibrate it. This is not true in settings where magnetic field is non-homogenous like indoors environment, and often a new calibration is needed. Also the calibration model of magnetometer is more intricate. Not only external magnetic disturbances, but also on-board magnetic materials and varying currents cause non-linearity and hysteresis effects. The onboard effects are summed up either as Hard iron or Soft-iron [13]. The problem with figure-of-eight procedure is that the inter-sensors alignment offset is not obtained [135]. To overcome this ambiguity, simultaneous accelerometer-magnetometer data collection is done at rest in multi-position. The magnetometer is then calibrated in two stages. First intrinsic magnetometer calibration is accomplished as described above. Then magnetometer to accelerometer triad alignment is done using the constraint that the dip angle (angle between magnetic vector and gravity vector) remains fixed [63]. This constraint again assumes a homogeneous magnetic field.

2.1.4. Calibration Using Reference Sensor

Few authors suggest methods where a 3-axis sensor can be calibrated using another sensor or orientation estimation obtained from a previously calibrated reference MIMU [53]. For example Wu et al. [149] suggest a rate gyro calibration using magnetometer and Olsson et al. [97] propose accelerometer calibration using rate gyro. Similarly Zhu et al. [161] performs a gyroscope aided magnetometer calibration. The algorithm is either implemented as non-linear optimization or a Kalman filter. The obvious advantage is that the inter-sensor triad misalignment is inherently addressed and thus the residual error is reduced during sensor fusion.

The problem with above mentioned algorithms is the assumption that the reference sensor is well-calibrated and neither drifts nor much affected by the disturbances during the procedure. The authors [11] propose switching between calibration of sensors of an MIMU for auto-calibration of one sensor using another in online sensor fusion to address this limitation.

2.1.5. Sensor Models

Majority of previous work assumes a linear or affine model for 3-axis sensors calibration. But Drag et al. [28] calibrated accelerometer using a non-linear neural network model and Wang et al. [142] has done same for magnetometer compass which takes care of non-linearity. Others [6] have used neural networks to explicitly include temperature dependence in the model of sensors.

2.1.6. In-Situ Calibration of Sensors

MEMS sensor are not very robust and parameters drift over time, therefore in-situ or before-use calibration becomes necessary. Since the calibration algorithms have iterative formulation like Kalman Filter or Gauss Newton optimization [26], the same may be adapted to apply on the data obtained in-situ. The authors [72, 43] demonstrate it for accelerometers, while [108, 151] use this approach on magnetometer. The work by [66] and [93] performs online calibration of all three tri-axial sensors in MIMU. There is however limitations to online approach, as highlighted by [83]. Also the quality and space coverage of the data dictate the calibration accuracy on unseen data [46, 141]. In general the algorithms that depend on norm-constraint of 3D sensor require more space coverage and high quality data than algorithms which use constraints aided by reference sensor [101]. But the later assume that reference sensor is well calibrated [11]. Depending on sensor noise and disturbances, the convergence of iterative algorithms can be slow and even steady-state error can be high in online or in-situ approach [11].

2.1.7. Joint Calibration of Sensors

Recently few authors have suggested techniques where multiple tri-axial field sensors or arrays of such sensors can be calibrated together. Papafotis et al. [101] suggest joint calibration and alignment of multiple accelerometers and magnetometers. However the authors [101] assume that all sensors are of the same type, attached to a fixed rigid platform and experience

the same 3-axis excitation. This is not a trivial assumption. First, the slight misalignment of sensors in mounting can change the 3D excitation input. Second, due to gradient the magnetic field is position dependent. The later problem has been dealt by [27], which jointly calibrate multiple magnetometers using average output vector obtained from different magnetometers in an iterative algorithm. However alignment of magnetometers is again assumed a priori in this work. The alignment assumption is relaxed in [122] and [95]. Both these work solves for both the constraints and relative pose between multiple sensors in an array or on a rigid platform. These algorithms are sensitive to initialization of values and lead to high variance of calibration parameters for different sets of data-collection [11].

2.1.8. Stochastic Error Models

The MEMS accelerometer, rate gyro and magnetometer are not only affected by systematic errors of bias, sensitivity, non-orthogonality and misalignment, but also undergo stochastic errors. For example bias instability, scale factor instability and random walk appear in the output of these sensors, even in static condition. In dynamic conditions, the sensors may give different output due to internal dynamics. Apart from that, all sensor outputs also experience zero mean white Gaussian noise. Detailed stochastic models have been proposed for inertial sensors. Petkov et al. [106] used power spectral density (PSD) and Allan Variance (AV) to derive Gaussian Markov (GM) models of rate gyros and accelerometers. El-Diasty et al. [30] used Allan Variance to obtain Angle Random Walk (ARW) and Velocity Random Walk (VRW) and employed Autocorrelation Function (ACF) to derive GM model of bias instability in their inertial navigation. While Nassar et al. [90] employed Autoregressive (AR) models for stochastic errors of inertial sensors. The work by [127, 45, 8] adopts a wavelet variance base approach called Generalized Method of Wavelet Moments (GMWM). This method is more suitable than AV since several stochastic processes are superposed in inertial sensors. Quinchia et al. [109] performed a comparison of different stochastic error modeling approaches like ACF, PSD, AV, AR and wavelet based for GPS/INS integrated navigation. The wavelet approach (GMWM) is extended by [110, 153, 111] for multivariate and multi-sensor scenario by using wavelet cross-covariance. The wavelet cross-covariance determines the dependence between sensors and may be used to construct a virtual sensor with improved navigation accuracy. All the past approaches require explicit modelling of the noise process and determine the model parameters after hours long data-collection in static condition. But the dynamic stochastic error of MEMS sensors can be quite different from static error and are dominated by variation in system dynamics of individual sensors. Therefore the uncertainty of sensor readings obtained using stochastic error models derived from static data is not representative of dynamic error.

2.2. Sensor Fusion for Orientation Estimation

The low cost MEMS based sensors are not very precise and robust. Despite calibration, residual errors and parameters drift are common. In MIMU based human motion capture, we use a 9-axis MIMU on each body segment to register its 3D orientation. The sensor fusion algorithms

are used for orientation estimation from 9-axis MIMU. This task is common to augmented reality, navigation and inertial motion capture, where MIMU sensors are used. The main challenge arises from the long term drift and errors if the magnetic heading is not robust. This is often the case in built up area and indoor environment where magnetic field is spatially inhomogeneous or a varying magnetic disturbance is present on the platform itself. Also if the platform is continuously accelerating/decelerating, it causes error accumulation over time in 3D orientation. Extensive literature exists regarding sensor fusion of rate gyro, accelerometer and a magnetometer to obtain 3D orientation and is discussed below.

2.2.1. Complementary Filters (CF)

A straightforward approach to online sensor fusion for MIMU is to implement complementary filter (CF) for processing rate gyro and accelerometer data and combine it with heading from magnetometer. In this framework, the rate gyro signal is low pass filtered to remove slowly changing bias and accelerometer signal is high pass filtered to eliminate noise and disturbances either implicitly (observer based) or explicitly. Most famous filters in this class are Madgwick filter [79] and Mahony filter [81], which are observer based. The former is also termed as gradient descent orientation filter. In their work [4], the authors reported a comparison of Madgwick and Mahony filters for orientation estimation task. The authors [154] proposed a framework of quaternion based complementary filter for human motion tracking in which factored quaternion algorithm (FQA) is used to obtain 3D orientation from accelerometer and magnetometer which is then combined with angular rate integration from rate gyro. A complementary filter (CF) has its advantages in terms of mathematical simplicity and ease of implementation.

2.2.2. Kalman Filter (KF)

Since the orientation estimation problem is non-linear, Extended and Unscented Kalman Filters (EKF/UKF) are most popular for this task. Both direct and indirect KF have been proposed [118, 24, 19]. The aforementioned Kalman filtering algorithms for MIMU sensor fusion employ quaternion representation for 3D orientation. Therefore renormalization to obtain a valid unit norm quaternion is an essential step. Few works have reported Kalman filtering to obtain orientation estimation using Euler angle representation [157, 41] and rotation matrices (direction cosine matrices DCM) [107] as well. But quaternion representation has been preferred in the past as it avoids the singularity [118].

2.2.3. Robust Orientation Estimation

In general accelerometer readings include dynamic body accelerations in addition to gravity vector and magnetometer in indoor/urban environment measures varying local magnetic field and not the true magnetic north. Thus MIMU based sensor fusion show degraded performance when run for long time, especially indoors [114]. The MIMUs in most commercial i-Mocap systems provide orientation estimates from embedded sensor fusion algorithms in addition to

raw sensor data [1]. However, these algorithms are not very robust to magnetic disturbances [34] and motion induced errors arising from body accelerations . The sensor fusion approaches reported in past literature which are claimed to be robust to magnetic disturbances and fast motion, do not give very satisfactory results except under restricted assumptions [156, 21, 50, 36]. Since magnetometer information is mostly unreliable, few authors [124, 86] have used only accelerometer-gyro pair (IMU) together with implied kinematic constraints on joints to avoid angle drift during human motion capture. Therefore the accuracy and reliability of MIMU based motion capture is still a major concern.

2.2.4. Sensors Calibration by Online Filtering

In orientation estimation using MIMUs, it is vital that rate gyro provides precise estimates, when either accelerometer or magnetometer readings are unreliable. This occurs during the period of high dynamic accelerations and/or inhomogeneous magnetic fields [34]. Being low cost, these sensors are often not very robust. Even if a calibration is performed beforehand, the sensitivity and biases do not remain constant over time. Moreover in many applications, it is not easy for a user to perform high quality sensor calibration. Therefore rate gyro bias parameters are often estimated together with 3D orientation [119]. This however assumes that all the other sensors readings are precise and rate gyro biases are accurately observable using those sensors. The approach by [93], extends it to scale factor estimation of rate gyro in addition to its bias. But if there are residual errors in calibration of other sensors, the rate gyro biases would not converge or bad estimates are obtained [82]. Some authors [97, 63] have performed the online calibration of accelerometer or magnetometer by using rate gyro itself or vice versa [149]. The work by [151] proposes real-time magnetometer calibration using Kalman filtering in which both rate gyro bias and magnetometer parameters are estimated along with 3D orientation. However to the best of our knowledge, no work in the past has addressed the real-time parameter estimation of all the three sensors in MIMUs in one algorithm.

2.2.5. Optimization based Filtering

The work by [61, 60, 62] uses window based optimization to obtain 3D orientation from a MIMU data sequence. A moving horizon window is used and a cost function optimizes the error terms associated with a prior, dynamics model based rate gyro error and accelerometer/magnetometer measurement error. This approach is extended by [37] for visual-inertial sensor fusion.

2.3. Anatomical Calibration for Joint angles

The MIMUs attached to body segments provide 3D orientation of sensor frame w.r.t a global North-East-Up frame. To obtain a 3D joint angle between two body segments, each MIMU on adjacent body segments must be registered and aligned with local bone segment frame. This registration step is called sensor-to-segment calibration. Either a known static pose or

specified functional movements is performed to obtain this alignment between sensor frame [98]. If an absolute heading reference like global North is available, an initial static calibration is straightforward. Normally an I-pose or T-pose is used for this initial calibration [20]. In former the subject stands with his arms by the side and hands pointing down, while in later the subject extends his arms straight sideways with palms pointing down. When only IMUs (less magnetometer) are used, absolute heading is not available, hence a combination of static pose and functional movement is used to align the 3-axis rate gyro with local bone segment frame [75]. The precision of sensor-to-segment calibration depends on how good a static initial pose or functional movement is executed. The authors [58] have used principle component analysis to obtain anatomical calibration of sensors using functional movements. While Nazarahari et al. [91] propose a semi-automatic calibration from gait. In an independent approach [124, 68, 67] have obtained joint axis calibration from rate gyro and accelerometer data, assuming joint constraints. The authors [130] perform an online sensor-to-body calibration using optimization while also estimating the joint angles. The challenge with initial calibration using static pose is that due to residual calibration errors of MIMU sensors, the global North-East-Up frame is not often registered same by each sensor. Also sensor-to-segment calibration may change over time due to sensor drift or skin artifacts [130].

2.4. Data-driven Learning of Human Joint Angle Limits

The intuitive swing-twist decomposition of joint rotation is devised by [7, 64, 7, 125] to simplify the parameterization and reduce computation when deciding a valid/invalid pose or correcting an invalid pose to nearest valid one. However these earlier works have not addressed the interdependence of swing and twist (intra-joint dependence). In a data driven approach, parameterization in terms of an implicit surface on joint trajectory data in quaternion space has also been proposed [48]. The authors further extended their work in [47], to deal with inter-joint dependencies. But the proposed method is computationally expensive and is not intuitive like swing-twist decomposition. Further, the limits are encapsulated directly on captured data, which might lead to unrealistic limits especially in the presence of sparse mocap data. Since only vector part of quaternion is used in [48] , it also ignores information encoded in scalar part, which may lead to ambiguity in joint angle. Moreover, quaternions have dual representation (+q and -q) for same rotation. Thus, a preprocessing step is required over quaternion trajectory to eliminate this duality

More recently, Akhter et al. [3] obtained pose-conditioned joint angle limits from a more controlled mocap dataset performed by trained gymnasts and athletes. They encoded joint limits for swing in a discretized binary occupancy map based on local spherical coordinates of the joint and defined a table-look-up function for valid/invalid poses. Their work also conditioned pose of a child joint based on its immediate parent in a kinematic chain, but it ignores intra-joint dependency of swing and twist. Moreover, a limited set of data is not enough to learn inter-joint dependency i.e. not sufficient mocap data instances of child limb swing are present in the data, conditioned upon a prior pose of parent limb. Therefore, the learning on limited mocap dataset can be detrimental to quality of estimated parent-child inter-joint dependency.

2.5. Inertial Motion Capture (i-Mocap)

2.5.1. Inertial Human Motion Sensing

MEMS MIMUs are used as wearable sensors in the past [143, 84, 131, 86, 73] for full body human motion capture. The human body is assumed as comprising of rigid segments articulated at joints and one sensor per segment is sufficient to compute 3D joint angle. The main limitation and challenge to i-Mocap frameworks however is the long term drift of sensors. Thus accurate calibration and robust sensor fusion [15, 9] are essential prerequisite to inertial human motion sensing. The kinematic and inverse kinematic based i-Mocap frameworks employ optimization [61] or stochastic filtering [87] on 13-17 body worn sensors, one per each body segment to obtain full kinematic pose. Thus the kinematic approaches are obtrusive and computationally expensive.

2.5.2. Sparse Sensing of Human Pose

Past work has shown that owing to kinematic and temporal constraints of 3D human pose, it is possible to use only a reduced set of sensors (as opposed to one sensor per segment) and still obtain 3D human pose; though in general the ambiguity of this ill-posed problem can lead to high uncertainty. The authors [76, 16] demonstrated this with statistical human body model fitted to reduced marker set. A small set of inertial sensors [88, 33] is shown to estimate 3D pose accurately. The work by [5] uses an inverse dynamic solver for joint torques and internal/contact forces which satisfies motion priors and sparse sensor measurements, and thus generates physically plausible human motion. The data-driven approaches using reduced sensors (≤ 6 instead of 13-17) [123, 138] are well suited for online implementation than optimization or constrained stochastic filtering [31, 51]. But good scalability of data driven approaches typically require a large dataset. The authors [52] used deep learning on a large SMPL synthetic dataset for learning 3D human pose from sparse (six) *synthetic* sensors and demonstrated state-of-art results after fine-tuning on real MIMU data. Their work is an extension of earlier optimization based approach [140]. Shallow temporal convolution (TC) and multilayer perceptron (MLP) are shown by [146] to yield similar results but 'jerkiness' error is less using deep learning [52]. Later, the authors employed a Kalman Filter [144] to address this jerkiness from shallow network. Recently Eckhoff et al. [29] have demonstrated the condition for observability in a kinematic chain comprising of double hinge joints using sparse magnetometer free inertial tracking. Their work highlights that a sliding window (moving horizon) filter can estimate kinematics of two connected single hinge joints, using measurements from magnetometer free IMUs only present at the end links.

2.5.3. Learning of Human Motion Manifold

The data-driven learning of human motion has come to fore with large human motion capture datasets and deep learning. The human motion can be represented as a manifold in a latent space. Both convolutional and recurrent neural networks are used to learn human motion

manifolds [49, 57, 103]. This section only identifies the representative work to emphasize the relatedness to our research. In [71] the latent representation of human motion learnt by bidirectional recurrent autoencoder is shown to be robust to both input noise and artifacts. A related issue in human motion manifold learning is minimal representation of 3D joint angles (SO3) so that data-driven learning is not complicated by singularity, duality and discontinuity problems occurring in the input or output data. Earlier work [89] has compared Euler angles, quaternions and more intuitive swing-twist representation. Also authors [159] show that 5-parameter and 6-parameter representations (instead of full 9-parameter rotation matrices) are always continuous and best for deep learning. A minimal versus full 9-parameter representation of 3D joint angles affects the number of parameters in the model and hence the computational cost, both of training and inference.

2.5.4. Data-Driven Uncertainty in Deep Learning

Estimating uncertainty of deep regression is a relatively new research direction. Data-driven learning of uncertainty in deep learning framework has been performed by [69] without assuming any explicit model. Both the uncertainty in the data (aleatoric) and model uncertainty (epistemic) affect the final error in the output of deep model. Few authors have dealt with 3D angle (SO3) uncertainty in the deep learning. The aleatoric SO3 uncertainty is dealt with by negative log likelihood (NLL) cost function and quaternions in the context of 3D rotation by [105] and demonstrated on KITTI visual odometery dataset. Russel et al. [116] extend uncertainty estimation in deep learning to multivariate outputs and used a Kalman filter for training and evaluation. Salinas et al. [120] and Zhu et al. [160] deal with uncertainty estimation in time series forecasting using LSTM. Most frameworks assume independent Gaussian distributions for the outputs and estimate both mean and standard deviation. Wen et al. [151] and Kivaranovic et al. [59] have proposed a distribution free approach and predicted quantiles or prediction intervals of the outputs. The main challenge in learning uncertainty from data in deep learning lies in the calibration of regression uncertainty [65]. The authors [70] propose a framework for calibration of test data uncertainty by scaling of standard deviation (σ) with a scalar value. The robust prediction of uncertainty from deep regression allows detection of unreliable outputs predicted by the deep network.

3. MEMS-Magnetic Inertial Sensors Modeling and Calibration

In this chapter, we provide a fast method to simultaneously calibrate multiple MEMS magnetic inertial measurement units (MIMUs) accurately in the field. In order to obtain 3D human pose using MIMUs, accurate 3D orientation is needed. Thus the sensors must provide precise measurements. To meet this requirement, MIMUs require calibration of systematic errors of bias, sensitivity, non-orthogonality and misalignment, which vary with temperature and use. Even after calibration, the sensors undergo stochastic errors in static and dynamic conditions and thus uncertainty of output must also be modeled. We propose a method for easy and fast calibration of these sensors together in an optimal way, while mounted on a single platform. The precise alignment of sensors is not assumed. Our method calibrates both fixed array of MEMS MIMUs or many independent MIMUs together using kinematic constraints. The uncertainty of sensor output is also learned as part of our model. Compared with existing state-of-art methods, our algorithm gives more consistent readings of all MIMUs and the framework also predicts the associated uncertainty of the sensor output. The uncertainty prediction of individual sensors is particularly helpful in the sensor fusion for accurate 3D orientation estimation. The work presented in this chapter is based on our peer reviewed published research [10].

3.1. Background

MEMS based inertial sensors such as rate gyro, accelerometer and other sensors like magnetometers have become almost ubiquitous in many devices and different application areas. These are often packaged as one unit and called magnetic-inertial measurement unit (MIMU). In many applications like wearable body sensor networks for motion capture or pedestrian navigation, often more than one MIMU unit is used in a distributed fashion or as an array [112, 115]. The precise calibration of these multi-MIMU systems, remains a major challenge, in order to use the data from a set of sensors [83]. The computation of a state or latent features from sensor inputs often implies integration over time, like integration of rate gyro readings for orientation estimation or of accelerometer readings for translation. The sensor errors introduce bias and drift [44] and can be very detrimental for the results of this integration. The inaccurate sensor calibration is a major cause of these systematic errors. First, the MEMS sensors are not very robust over time and the parameters of a sensor model needs to be updated often. Second, a challenge arises when we need to perform this updated calibration in the field without sophisticated equipment or accurate reference. Third, when multiple MIMU units are employed in a distributed fashion or as an array, the calibration must ensure inter-unit consistency of output

for the same input. Fourth, a fast, convenient and real-time calibration scheme is needed in the field, i.e. the calibration of MIMUs must be performed through a simple protocol and with no special arrangements. And lastly, stochastic errors require modeling of uncertainty. Though stochastic models have been proposed [106, 30, 90, 127, 45, 8, 109, 110, 153, 111], but these models require explicit modeling of the noise process. On the other hand, a data-driven model of the uncertainty of sensor output can be learned without an explicit model [116]. The focus of our simultaneous MIMUs calibration approach is to achieve all the above mentioned goals employing mutual auto-calibration of sensors in an MIMU unit, while also reducing the variance of their outputs with sensors of other MIMU units in the distributed network or array by proposing a new algorithm.

3.2. Contribution

Our approach addresses the problem of joint and simultaneous calibration of multiple MIMUs, but uses the constraints different from ones reported in previous work. It is motivated by the fact that when multiple MIMUs are mounted on a single rigid platform, mutual kinematic constraints exists that can be exploited for improving the calibration. For example rate gyros mounted on same rigid platform must register the same absolute magnitude of rate of 3D rotation and a set of accelerometers moved together in one direction, must register same magnitude of acceleration at each instant, irrespective of relative pose. Also the change in measured 3D magnetic vector between two instants must correlate with calibrated rate gyro readings.

Since our algorithm does not assume constant norm of magnetic field, nor need extensive space coverage, it is much simpler. We use a new sensor model based on ensembles of weak models learned with small amount of data and then fine-tuning of the complete sensor model, which leads to improved results. Our model also learns and output the uncertainty of calibrated readings using a data-driven approach. To the best of our knowledge, it is the first time that later has been done using mutual kinematic constraints of multiple MIMUs. We show that this uncertainty correlates well with the error both in static and dynamic conditions. We believe that predicting the uncertainty of calibrated sensor readings can be of value in sensor fusion algorithms, which in the past have assumed a fixed uncertainty based on static data. Our work shows that uncertainty of inertial sensor output is different in static and dynamic condition. The latter arises due to variations in MEMS dynamics and internal filtering of individual sensors.

The main contributions covered in this chapter are,

1. A two-tier sensor calibration model is proposed to reduce the procedural variance in estimation of calibration parameters and account for nonlinearity. The model is based on ensemble approach and fine-tuned on acquired data.

2. A robust calibration procedure based on kinematic constraints is designed so that intra-MIMU and inter MIMU output variance is reduced, when same inputs are applied.

3. No sophisticated external equipment or calibrated input signal is used. An easy-to-fabricate 20- position stable icosahedron platform is designed for in-field data collection

for better evaluation. But the procedure can be implemented even without this platform, by simply mounting the multiple MIMUs or MIMU array on any rigid body which can take up to six stable positions.

4. Data collection protocol is quite simple and of short duration. We perform level rotation, a free fall and arbitrary rotation between six stable orientations. This does not require full space coverage like multi-position [129] or figure of eight [78].

5. To the best of our knowledge, the work presented in this chapter is the first to learn the data-driven uncertainty of calibrated readings from multiple MIMUs using novel mutual kinematic constraints. We show that this uncertainty correlates well with the error, both in static and dynamic conditions.

3.3. Organization of Chapter

This chapter is organized in nine main sections. In the first two sections we have already covered the background and our main contribution to the problem of simultaneous calibration of multiple MIMUs and modelling of their uncertainty. In section 3.4, we describe our wearable MIMU based system built using flexible textile harness (EasyIMP) as well as wireless option (BIONIC) and its plugin software architecture. The system offers high rate data acquisition and plugin implementation of sensor models and algorithms developed in this thesis. In section 3.5, we state the sensor models used for MIMUs and in section 3.6 we present our methodology to calibrate multiple MIMUs using these models. The later section explains both the data collection and our detailed algorithm. In section 3.7, we provide the details of architecture, cost function, training and inference with uncertainty as well as the protocol of experimental evaluation. Towards the end of this chapter we provide the results of our experiments, both using synthetic data and real MIMU data. The former obtains the baseline performance of our new algorithm. In both cases, we compare our algorithm with contemporary methods and show that it achieves better performance. Our model also predicts the uncertainty of calibrated readings, which is found to correlate well with the errors in static and dynamic conditions as well as due to magnetic inhomogeneity.

3.4. Body Worn MIMU System

Both harness based and wireless MIMU nodes are used for human body tracking, however the wireless MIMUs suffer from reliability and synchronization issues, need more power and have bigger form factor for individual nodes. On the contrary, flexible harness based MIMUs remain a viable option for less invasive (due to low form factor), reliable and high rate body worn systems with very low power consumption. The pace of evolution and improvement in MEMS sensors is very fast. Better magnetometers, rate gyros and accelerometer are becoming available, driven by needs of personal mobile devices and ubiquitous computing.

3.4.1. EasyIMP Hardware

The design of our harness based MIMU motion tracking system has partly developed and matured as a 'meta-product' under collaborative EasyIMP project supported by European Union Grant No. 609078. The aim has been *to develop a lightweight and low form factor MIMU based motion tracking system for ambulatory and non-clinical settings, with low power consumption, decent embedded processing power as well as sufficient on-board storage.* Intel Edison® platform is selected as the hub for being low-power, low-cost solution with embedded processing ability and on-board storage.

The sensing nodes are based on Invensense® MPU-9150, integrated 9-axis Motion Tracking device that combines a 3-axis MEMS gyroscope, a 3-axis MEMS accelerometer, a 3-axis MEMS magnetometer and a $Digital Motion Processor^{TM}$. This MPU has external triggering option, which makes it extremely useful when combining multiple sensor units. NXP PCA9501 EPROM module has been added to store calibration data of each node, thus calibration information stays on plug-in sensor node. The data transmission happens through a differential I^2C bus connected to the microcontroller of the Edison development board. Differential I^2C bus can run up to 400 KHz, while data acquisition from the 7x IMU sensors take place at 100 Hz sampling rate in this baseline system. The flexible textile wires are used for both power and communication between hub and sensor nodes. Figure 3.1 shows the Intel Edison® hub connected with textile differential I^2C wire to 7x sensor nodes as well as with textile micro USB wire to the 5V power bank.

The package size of miniaturized sensing nodes is $25 \times 20 \times 5 \ mm^3$, making the system almost non-invasive (for comparison, Xsens MTw wireless series has MIMU size of $47 \times 30 \times 13 \ mm^3$). The flexible textile harness allows easy routing and differential I^2C bus/external trigger of MPU-9150 provides very good synchronization. There is sufficient free memory (up to 1.0 GB) on Intel Edison module to record an outdoor session. The low current consumption feature of the system is tested with 7x IMU nodes at 100 Hz sampling rate and its current consumption was found out to be 300mA, running raw data calibration and sensor fusion algorithm for all 7x nodes onboard and TCP data transfer on Wi-Fi. This system has been reliably tested for lower body tracking with 7x MIMUs and is extendable up to 10 sensors with existing hardware design. Two such sub-systems with synchronization are used for simultaneous upper and lower body tracking.

3.4.2. Plugin Software Architecture

The system has been built with plug-in software architecture in consideration, as shown in Figure 3.2. Intel Edison® platform runs Yocto Linux. The application can be extended with user compiled code at runtime via a dynamically linked shared library plugin interface. The definition of the plugin interface and all the application specific API calls are stored on the device itself. The plugins can be activated via generic protocol on the TCP socket. Various types of data are transmitted to the application layer through Wi-Fi (or USB Ethernet) using TCP protocol. The firmware/software framework gathers the sensor data and runs data calibration/fusion algorithm for orientation estimation. It is written in C++ and can be compiled for different plat-
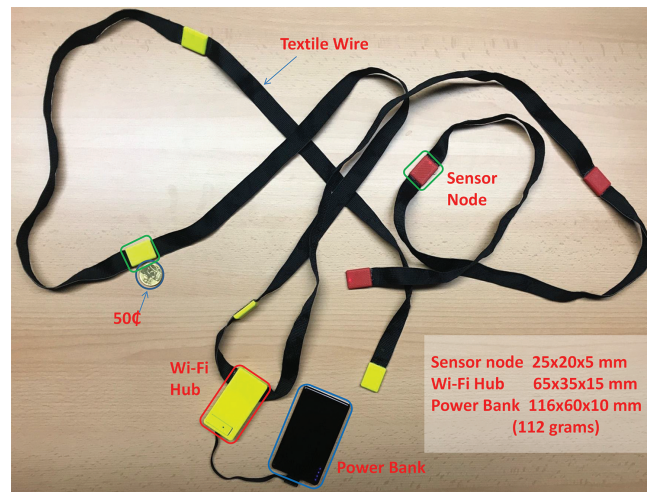
Figure 3.1.: Complete Flexible Harness based EasyIMP MIMU Tracking System

forms with the GNU Compiler. The firmware is set to run on Edison® module as soon as it is accessed by Web API for host application. Both the raw sensor data and orientation estimation from all MIMU nodes are available over Wi-Fi. Our host application can be accessed via TCP protocol, HTTP server and it even has a MQTT (Message Queuing Telemetry Transport) client to ingest data to cloud.
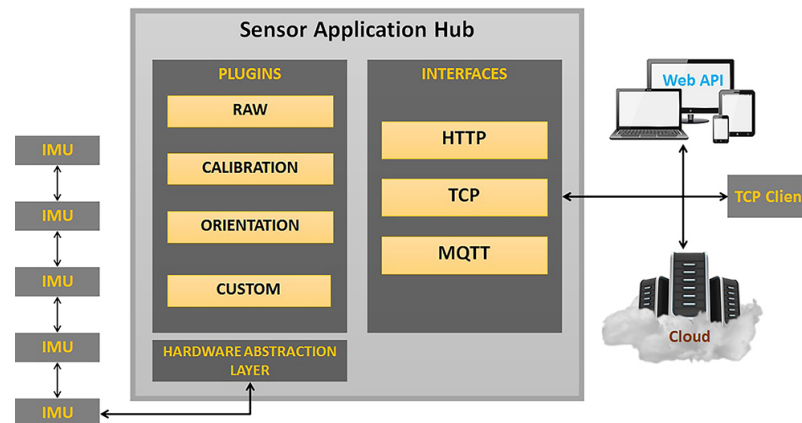
Figure 3.2.: Plugin Software Architecture

3.4.3. Body Information On Integrated Chip (BIONIC) Hardware

In sensor fusion and body tracking it is always required to improve the quality of the data delivered by the sensors, in order to implement more accurate tracking. Therefore we evolved our design to a new hardware platform using newer MEMS sensors with better performance in terms of noise levels and bias stability. It is accomplished as part of Body Information On Integrated Chip (BIONIC) project supported by European Union under Horizon 2020 program. This new design also offers the flexibility of choosing between wireless and wired communication.

The new prototype for the sensor node design is based on the nRF52832 System on Chip (SoC) that combines an ARM Cortex-M4F microprocessor with a 2.4GHz multi-protocol radio supporting Bluetooth Low Energy (BLE) and ANT multicast protocol. The presence of a low power microcontroller in every node provides the possibility to perform local data preprocessing, thus decreasing the computational load on the hub as well as save on intra-system communication bandwidth.

On every node, there is an ICM20602 6-axis Motion Tracking device that combines a 3-axis gyroscope and 3-axis accelerometer in a small $3 \times 3 \times 0.75 \ mm^3$ package. The node also contains the MMC3416PJ ultra small $1.6 \times 1.6 \times 0.6 \ mm^3$, 3-axis magnetic sensor. The nRF52832 SoC communicates through SPI and I2C with the ICM20602 and MMC3416PJ, respectively. The sampling rate of the system is 1 KHz for the accelerometer and gyroscope, and 500 Hz for the magnetic sensor.

The system offers a trade-off between high sample rate and low power consumption, depending on the application and accuracy required, as depicted in Table 3.1 (the hub performs streaming over Wi-Fi for this budget as well).

The improvement in the performance by the new sensors used in this prototype can be observed on Figure 3.3, where the Allan Variances of the gyroscope and accelerometer are plotted for each axis of the MPU9150 (old inertial sensors) and the ICM20602 (new inertial sensors). It can be observed from the plot that the bias stability of the new sensors (both gyro and accelerometer) is considerably enhanced and intrinsic noise baseline is reduced for new accelerometers.
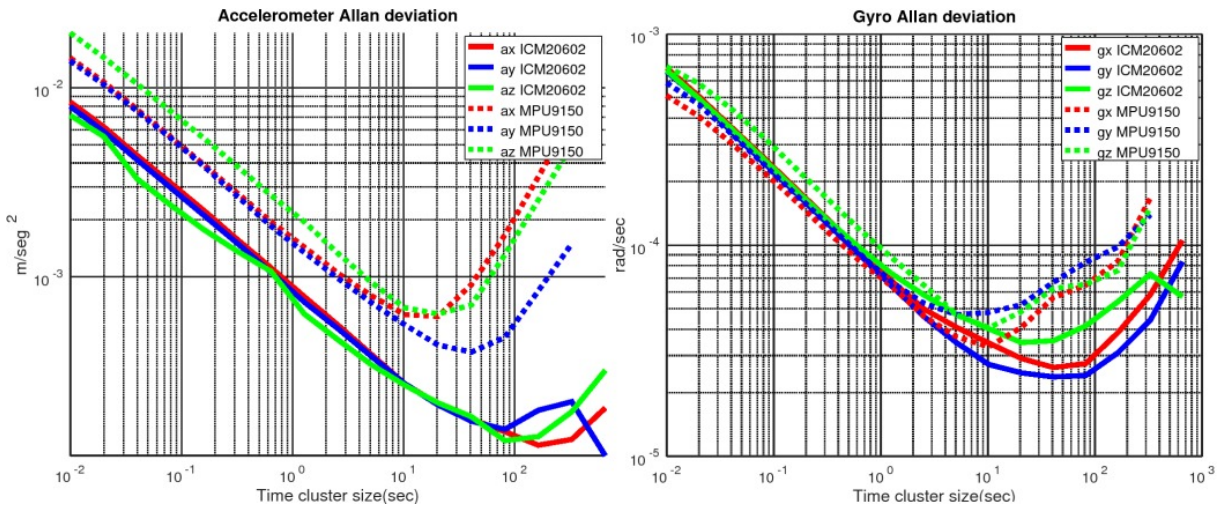

Figure 3.3.: Experimental Comparison of Allen Variance of Rate Gyro and Accelerometer used in our Baseline (MPU9150) and Improved System (ICM20602)

The PCB size (Figure 3.4) of the sensor node for the new prototype is $25.5 \times 18 \times 1.5 \ mm^3$, including a coaxial antenna connector for possible BLE implementations.
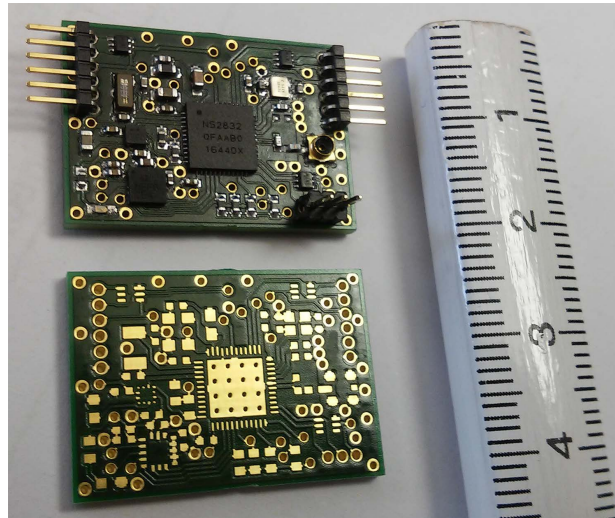
Figure 3.4.: Improved Sensor Node based on nRF52832 System on Chip (SoC) and ICM20602 Motion Tracking

The communication between the new nodes is implemented through textile wires using a 4Mbps SPI Daisy chain protocol. The synchronization is handled by sending a sync signal to every node. Up to 7x nodes can be connected without compromising the maximum output frequency (1 KHz), with the possibility to add more nodes using additional bridge stations between nodes. For comparison, Xsens harness based MVN Link® offer a maximum output rate of 240 Hz (internal update rate is 1000 Hz).

The advantage of the new hardware is that it can be adapted to a wide range of possibilities. The applications which demand the highest output rate can be implemented with the Daisy chain SPI solution using textile harness. On the other hand, the applications that tolerate lower output frequencies can be implemented with the BLE capability that the nRF52832 offers, eliminating the need of Wi-Fi hub, hence decreasing the overall system current consumption (see column 3 in Table 3.1).

Sampling Frequency	Node Current Consumption	Hub Current Consumption
$Idle^*$	$3.8\ mA$	$60\ mA$
$60\ Hz$	$4\ mA$	$100\ mA$
$100\ Hz$	$4.2\ mA$	$110\ mA$
$500\ Hz$	$4.65\ mA$	$125\ mA$
$1000\ Hz$	$6.7\ mA$	$140\ mA$

Table 3.1.: New prototype- Current Consumption (in milliampere, mA) @ 3.3V. *The idle state occurs when Hub is waiting for the start of a TCP transmission on Wi-Fi.

3.5. Sensor Models

Past work [94, 38, 40, 152, 78] has assumed linear or affine sensor models for the 3-axis accelerometer, rate gyro and magnetometer in MIMUs. We follow same to define initial parameters of 'weak learners' in our ensemble based approach for sensor calibration. For the MEMS rate gyro and accelerometer, following linear models relate raw measurements with real physical quantities. We follow the convention in this paper that italic or bold small letters represent vectors and bold capital letters are matrices

$$\boldsymbol{\omega}_{raw} = \mathbf{S}^g \boldsymbol{\omega} + \mathbf{b}_g + \mathbf{w}_g \tag{3.1}$$

$$\mathbf{a}_{raw} = \mathbf{S}^a \cdot \mathbf{R}_b^n (\mathbf{a} + \mathbf{g})_b^n + \mathbf{b}_a + \mathbf{w}_a \tag{3.2}$$

In equations above, $\mathbf{S}$ are 3×3 sensitivity and misalignment matrices and $\mathbf{b}$ are bias terms, while $\mathbf{w}$ represents the white Gaussian noise intrinsic to each sensor. The vector $\boldsymbol{\omega}$ is the angular rate. $\mathbf{R}_b^n$ is a 3×3 rotation matrix that rotates the physical vectors $\mathbf{a}$ and $\mathbf{g}$ represented in navigation frame of reference to body fixed frame of reference. The body fixed frame has its origin at the center of 3-axis sensors triad and its x, y, z axis are orthogonal and aligned with that of MIMU package. The navigation frame of reference has its origin at the initial position of MIMU and its x-axis is aligned with local magnetic north while y-axis is opposite to gravity (and z-axis is orthogonal). The body acceleration $\mathbf{a}$ (in navigation frame) is added to reference gravity vector $\mathbf{g}$. Similarly linear sensor model for magnetometer is.

$$\mathbf{m}_{raw} = \mathbf{S}^m \cdot \mathbf{R}_b^n (\mathbf{m}_n + \mathbf{d}) + \mathbf{b}_m + \mathbf{w}_m \tag{3.3}$$

In magnetometer case, $\mathbf{S}^m$ is a 3×3 matrix that in addition to sensitivity, misalignment and non-orthogonality, also account for soft iron effects. The soft-iron effects are distortions that appear in the magnetometer measurement due to ferrous materials fixed w.r.t body frame. Similarly, $\mathbf{b}_m$ include both the sensor bias as well as hard iron effects due to permanent magnetic sources fixed to body. It is interesting to note that magnetic reference vector for indoor settings, is often very different from the earth's magnetic field. It is also continuously changing due to a position dependent (and time varying) magnetic disturbance $\mathbf{d}$ in the environment [13].

3.6. Proposed Methodology

Every switch-on of MIMU unit causes a change in bias of MEMS sensors. The parameters also vary with warm-up, until the temperature becomes stable. The bias stability of sensors is given by the Allan variance [150]. Our work assumes that the field calibration of MIMUs is done at each switch-on and after a warm-up period of few minutes has elapsed. Then we perform data-collection using three protocols with all the MIMUs mounted on a rigid surface of 20-position icosahedron platform. This platform was developed in-house using cardboard (shown in Figure 3.5). However as the data collection procedure would demonstrate in section

3.6.1, the use of this platform is not mandatory. Though, it has provided us with flexibility during evaluation phase to collect more extensive data for other state-of-art calibration methods for comparison with ours. The collected data is segmented into 'static' and 'dynamic' sets using an automated algorithm [132]. In our work we assume a two-tier sensor model. First (ensemble of) weak linear models are obtained from dataset as explained in section 3.6.2. These models are then used to initialize weights/biases of a hidden layer in a neural network (section 3.7.2). The fusion layer of neural network is a Long Short Term Memory (LSTM) connected to a linear layer for the output and softmax for associated uncertainty. In a fine-tuning step, the whole network is then trained to reduce the error on a pseudo-measurement, obtained from kinematic constraints (section 3.6.2). The model is fine-tuned using negative loss likelihood cost function and the uncertainty is learned as a part of it [105] (section 3.7.4). The overview of our complete framework is depicted in Figure 3.5. The step-by-step procedure is further detailed in a flowchart in Figure 3.6
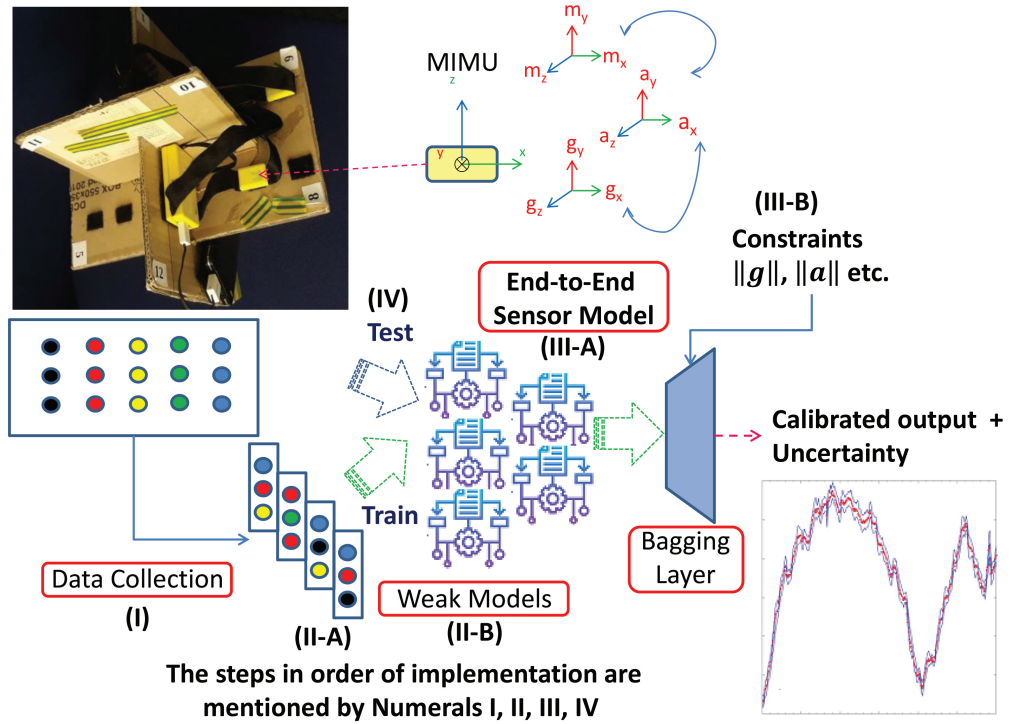
Figure 3.5.: Overview of Our Framework with key Components.

In our ensemble model, **S** define the initial weights and **b** define the initial biases of three hidden neurons in each weak learner (for details, see Figure 3.15. Five separate weak learners are obtained for each sensor, using datasets and algorithm explained in sections 3.6.1 and 3.6.2 (the steps are depicted in Figure 3.8). The activation function of each hidden node is '$tanh$'. The model parameters **S** and **b** are normalized w.r.t maximum scale factor (from datasheet), so that the raw sensor readings ω, **a**, **m**, give an output of ± 1 at full-scale (section 3.7.1). All weakly learned models are combined by an LSTM layer. The parameters of LSTM, output layer along with initial parameters of hidden layer are fine-tuned, using kinematic constraints and a
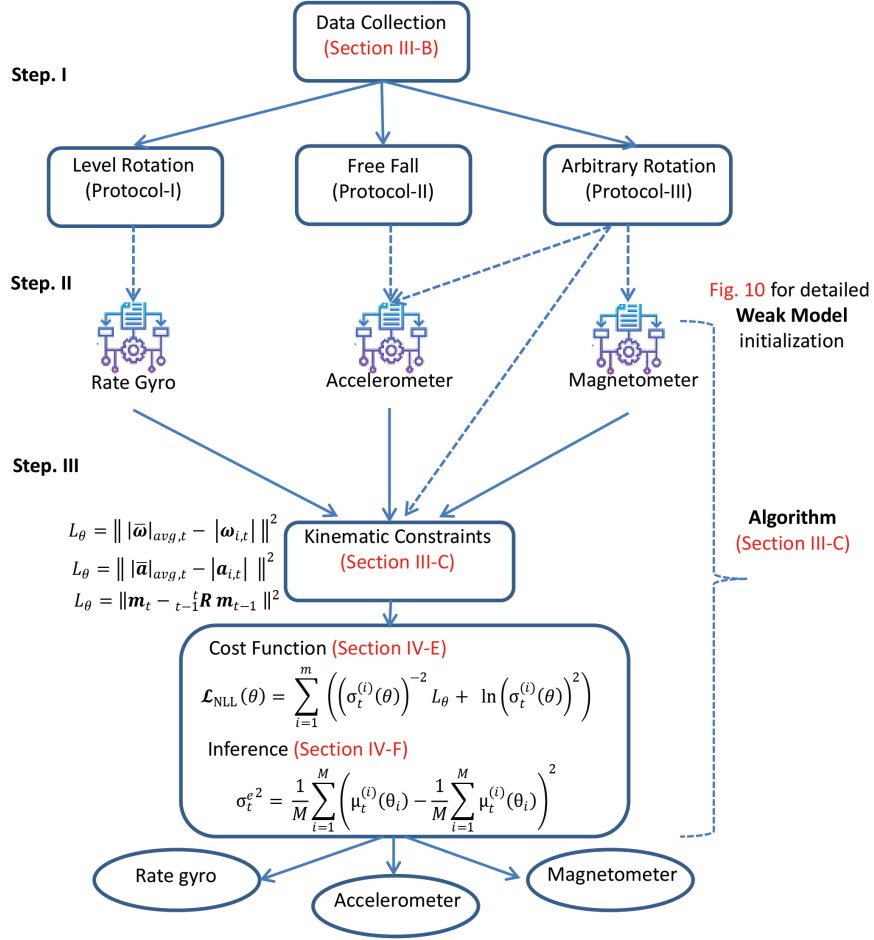
Figure 3.6.: Step-by-Step Procedure in a Flow Chart

'pseudo-measurement', which will be explained in section 3.6.2. The output layer also has additional softmax to output the uncertainty, which is learnt as part of negative log likelihood cost function [116]. The section 3.7.4 would cover uncertainty aware cost function.

3.6.1. Data Collection

The data collection protocol from multiple MIMUs in our calibration method is very simple and can be easily performed in the field within 2-3 minutes. We performed data collection using a 20-position icosahedron platform as shown in Figure 3.7. But this special platform is only used for better evaluation and comparison with multi-position methods [129] in experiments. Any simple rigid platform like a cube-box may be used instead for data collection. For our algorithm, we used only 6 distinct positions of the platform.

The following kinds of data are collected in different stable positions of rigid platform.

1. In Protocol #1, MIMUs mounted on rigid platform are rotated in-plane on a level surface, such as table. The precise attitude of this surface is not important. This is done with rigid
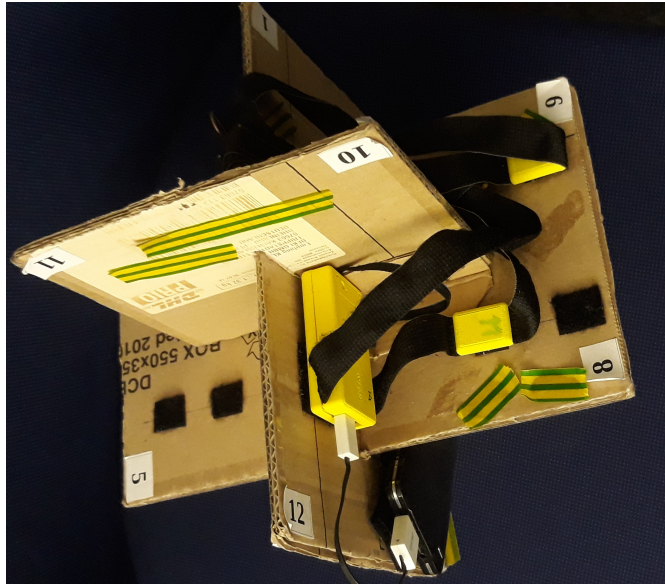
Figure 3.7.: 20-positions Icosahedron Platform for Multiple MIMUs

platform placed in different orientations, to get multiple small datasets. Each rotation should preferably complete more than one complete circle.

2. In Protocol #2, all MIMUs mounted on a rigid platform are allowed to free fall, while the platform is held by an elastic cord to de-accelerate the free fall softly at the end. A free fall period of 0.2-0.3 sec is possible for typical human initiating the fall at shoulder height.

3. In Protocol #3, dynamic and static data are collected. The rigid platform is rotated and placed in different orientations. The dynamic data is recorded while platform is in transition between stable positions. The static data is recorded in stable position. Protocol #3 is also repeated for all 20-positions of our platform to get 'sufficient' data to perform comparison with multi- position methods [129] for accelerometer/ magnetometer calibration.

Protocol #1, 2 and 3 are used to obtain up to five independent datasets in our method. The data from Protocol #1 is used for initial gyro calibration. Protocol #2 is used for computing estimate of accelerometer bias. Protocol #3 is used for final rate gyro calibration and then accelerometer and magnetometer calibration using pseudo- measurements.

3.6.2. Algorithm

Figure 3.8 depicts the data and sequence of steps we use for calibration of rate gyro, accelerometer and magnetometer.

Figure 3.9 depicts the magnetometer data that we obtain in one experiment of Protocol #1, after the raw magnetometer data is processed so that initial magnetometer readings at the start
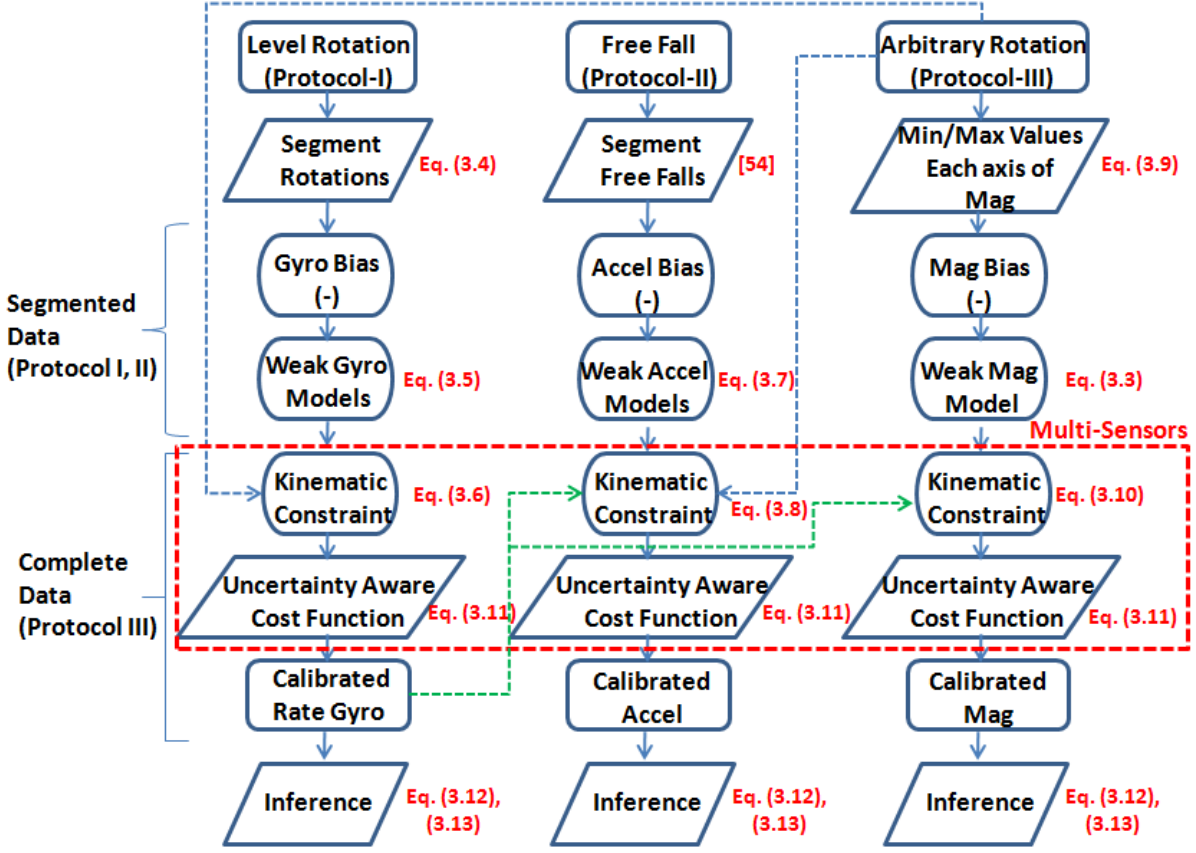
Figure 3.8.: Flow Chart of Complete Algorithm for Calibration

of sequence are subtracted from all subsequent data.

$$\mathbf{m}_{proc} = \{\mathbf{m}_i - \mathbf{m}_{i0}\}_{i=1}^{M} \tag{3.4}$$

In (3.4), M is the number of magnetometers and $\mathbf{m}_{i0}$ is the initial raw reading of i-th magnetometer. Even though different magnetometers are yet uncalibrated, the data sequence is repeated and reaches a minimum (near-zero) in each rotation, i.e. initial magnetometer readings are repeated after one complete rotation. The sensor data segmented between two such near minima (of magnitude or norm of magnetometers) is used to represent one complete rotation of 2π radians.

The segments of data representing one rotation can now be used to calibrate weak model (3.1) of the rate gyros. Although every rate gyro might be tilted differently w.r.t rotation axis of platform, each performs a rotation of 2π radians. First the bias is removed from raw gyro data. The bias $\mathbf{b}_g$ is simply computed as average of rate gyro output during 'static' interval at start and end of rotation data. Then all gyro readings are normalized to get the unit axis $\hat{\omega} = \frac{\omega}{|\omega|}$ around which planar rotation takes place. The unit axis across the rotation data is averaged to take care of noise. We then write a simple matrix equation to relate the complete rotation with

Figure 3.9.: Minima of Magnetometer detect 2π rotation in Protocol #1

time-integrated raw rate gyro readings $\boldsymbol{\omega}_i$ (obtained after subtracting bias $\mathbf{b}_g$).

$$2\pi[\hat{\boldsymbol{\omega}}_i] = \mathbf{S}^g[\hat{\boldsymbol{\omega}}_i] \int_0^t |\boldsymbol{\omega}_i| \Delta t \tag{3.5a}$$

$$\mathbf{Y}_g = \mathbf{S}^g \mathbf{A}_g \tag{3.5b}$$

$$\mathbf{S}^{g^{-1}} = \mathbf{A}_g \mathbf{Y}_g^{-1} \tag{3.5c}$$
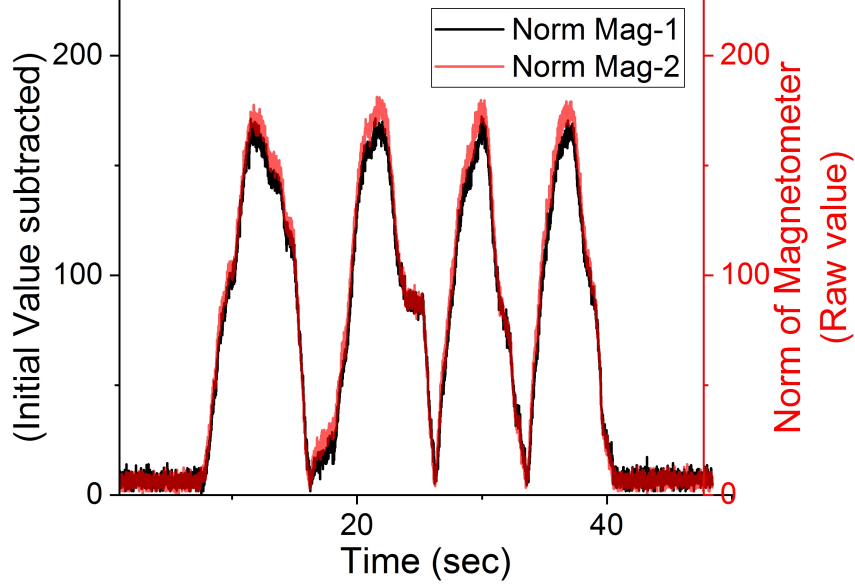
In equations above, $\mathbf{Y}_g = 2\pi[\hat{\boldsymbol{\omega}}_i]$ and $\mathbf{A}_g = [\hat{\boldsymbol{\omega}}_i] \int_0^t |\boldsymbol{\omega}_i| \Delta t$. The scaling and misalignment matrix of rate gyro $\mathbf{S}^g$ has 9 unknowns, therefore we require at least three such complete rotations, while platform is aligned with unique axis $\hat{\boldsymbol{\omega}}$ in different orientations. $\mathbf{Y}_g$ and $\mathbf{A}_g$ are therefore 3×3 matrices with each column representing data of one rotation. The number of rotations performed at one orientation of platform are segmented [132] using minima of (3.4) and we find $\mathbf{S}^{g^{-1}}$ by 'bootstrapping' data from rotations performed at three different orientation and forming triplets for 3×3 matrices $\mathbf{Y}_g$ and $\mathbf{A}_g$. Each $\mathbf{S}^{g^{-1}}$ and $\mathbf{b}_g$ pair obtained from such bootstrapped triplets, using (3.5) is then used to initialize the weak models of rate gyro.

The 'dynamic' data from Protocol #3, provides us with independent kinematic constraint on rate gyro. Since all MIMUs are mounted on same rigid platform, the rate gyros in each MIMU must register the same magnitude $\bar{\omega}$ of rate of rotation (although the axis of rotation of each gyro may be different). Figure 3.10 depicts how the magnitude output of different rate gyros looks like when data is calibrated using only weak models.

Clearly there is dispersion amongst rate gyros due to non-linearity/model uncertainty. Since
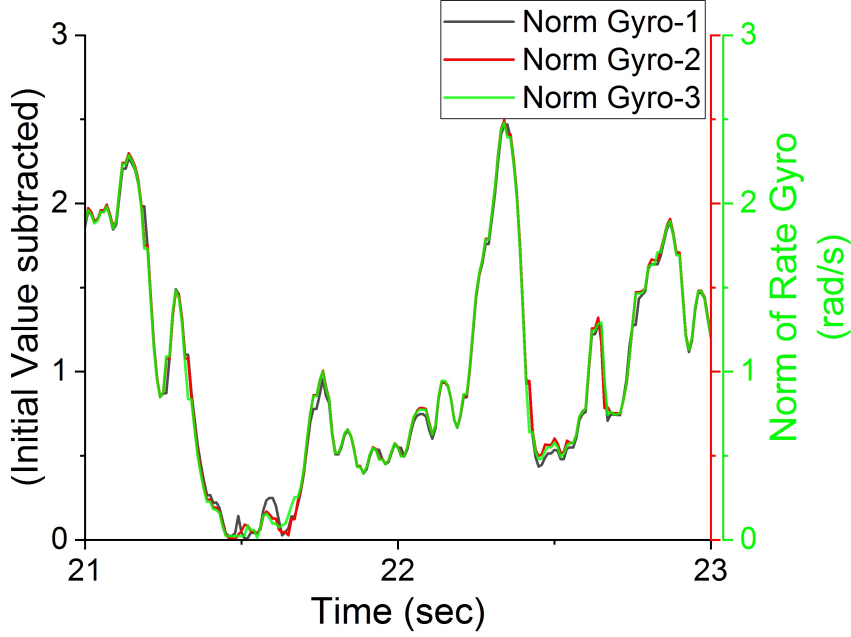
Figure 3.10.: Magnitude Output of Rate Gyros after Weak Calibration

our aim is to reduce this dispersion amongst multiple MIMU outputs, we aim to optimize complete gyro model $F_\omega(\theta)$, using following cost function.

$$L_\theta = ||(|\bar{\boldsymbol{\omega}}|_{avg,t} - |\boldsymbol{\omega}_{i,t}|)||^2 \tag{3.6}$$

where $|\bar{\boldsymbol{\omega}}|_{avg,t}$ is the average magnitude norm of rate gyros of all the MIMUs at each time step and $|\boldsymbol{\omega}_{i,t}|$ is the output of a i-th gyro model. The end-to-end model $F_\omega(\theta)$ is depicted in Figure 3.14. The gradient descent algorithm is used and model is fine-tuned over 'dynamic' data obtained in Protocol #3.

Next we use the data of Protocol #2 and segment the free fall using a detection window adapted from [132]. Figure 3.11 shows the data collected in free fall. We average the accelerometer readings in free fall window. In free fall no gravity is registered, and accelerometer true input is zero. Thus, we obtain the bias of all the accelerometers mounted on the platform this way. Using this bias and diagonal sensitivity matrix obtained using datasheet; we obtain weak models of accelerometers. We perform the free fall protocol five times and obtain five weak models (3.2) for each accelerometer.

Assuming that our rate gyro is fully calibrated in step 1 and 2, we use both the gyro and accelerometer data of Protocol #3, to refine the weak linear models (3.2) of accelerometer. Between each two stable positions, we optimize weak model of accelerometer using a cost function

$$L_{\phi_a} = ||\mathbf{a}_{n+1} - \mathbf{R}_n^{n+1}\mathbf{a}_n||^2 \tag{3.7}$$

Figure 3.11.: Free Fall Window for Bias Estimation of Accelerometer

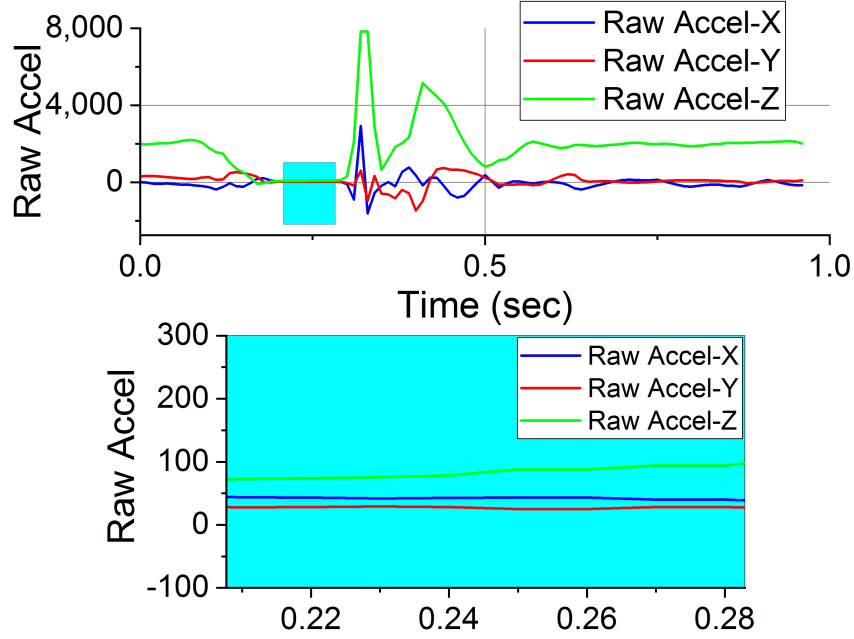where $\mathbf{a}_n$ and $\mathbf{a}_{n+1}$ are average accelerometer readings in two stable positions respectively, obtained using (3.2) from raw accelerometer data and $\phi_a = \{\mathbf{S}^a, \mathbf{b}_a\}$ comprises of model (3.2) parameters to be optimized. The rotation $\mathbf{R}_n^{n+1}$ is obtained by integrating calibrated gyro readings of recorded 'dynamic' data between two positions. This method is preferred over multiposition [129] for calibration of accelerometer, since it needs less data collection and only 4-6 stable positions (the use of 22-position platform is optional). It also intrinsically aligns the 3-axis accelerometer with corresponding rate gyro triad. On the other hand, multiposition calibration need additional step to find out alignment matrix $\mathbf{R}_m^g$ between gyro and accelerometer of each MIMU.

The final accelerometer calibration is performed for complete model $F_a(\theta)$ as depicted in Figure 3.14, using kinematic constraint on accelerometer data collected in Protocol #3. Figure 3.12 depicts the accelerometer data after applying (3.2) that we obtain in one experiment of Protocol #3. Since the accelerometers may be aligned differently w.r.t each other on platform, we only use magnitude value of accelerometer.

All MIMUs are mounted on same platform; therefore during motion, each accelerometer must register same magnitude of acceleration due to kinematic constraint. But as we notice in Figure 3.12, there are variations in acceleration that arise from approximate calibration, non-linearity and dynamic response/filtering of each accelerometer, causing dispersion amongst the outputs in dynamic condition.

In order to reduce the dispersion amongst multiple accelerometers due to non-linearity/model uncertainty, we optimize complete accelerometer model $F_a(\theta)$ using following cost function.
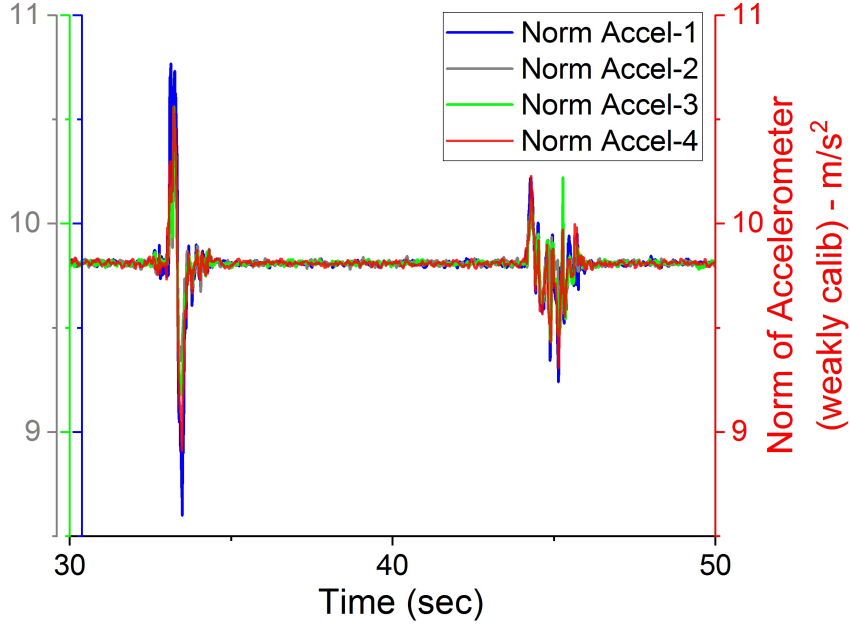
Figure 3.12.: Magnitude of Accelerometers after Weak Calibration.

$$L_\theta = ||(|\hat{\mathbf{a}}|_{avg,t} - |\mathbf{a}_{i,t}|)||^2 \tag{3.8}$$

where $\hat{\mathbf{a}}_{avg,t}$ is the average magnitude norm of accelerometer of all the MIMUs at each time step and $\mathbf{a}_{i,t}$ is the output of a i-th accelerometer model. The end-to-end model $F_a(\theta)$ is depicted in Figure 3.14. The gradient descent algorithm is used and model is fine-tuned over data obtained in Protocol#3.

For the magnetometer calibration, we first obtain an initial estimate of bias/hard iron effect. This is done using the dynamic data of Protocol#3. As shown in Figure 3.13, we get maximum and minimum values along each of 3-axis of every magnetometer as we perform arbitrary rotation. The bias/hard iron for magnetometer for each of 3-axis is then obtained using

$$\mathbf{b}_m = (\max_{i=x,y,z}(\mathbf{m}_{raw}) + \min_{i=x,y,z}(\mathbf{m}_{raw}))/2 \tag{3.9}$$

where max and min is performed on each magnetometer axis readings $i = x, y, z$ to obtain minima and maxima, after outlier rejection. Using this bias/hard iron and diagonal sensitivity matrix obtained using datasheet; we obtain weak models of magnetometer.

The complete magnetometer calibration is done using fully calibrated rate gyro of respective MIMU. Our procedure is robust to magnetic disturbances. The assumption of constant magnetic field is not required, as we rotate the platform in Protocol #3. Neither all magnetometers on a platform experience the same magnetic field due to its spatial gradient. However when the platform is rotated, the value of magnetic disturbance can be assumed not changing between any two time instants (when high sampling rate $>=100$ Hz is used). From (3.3), the variation in
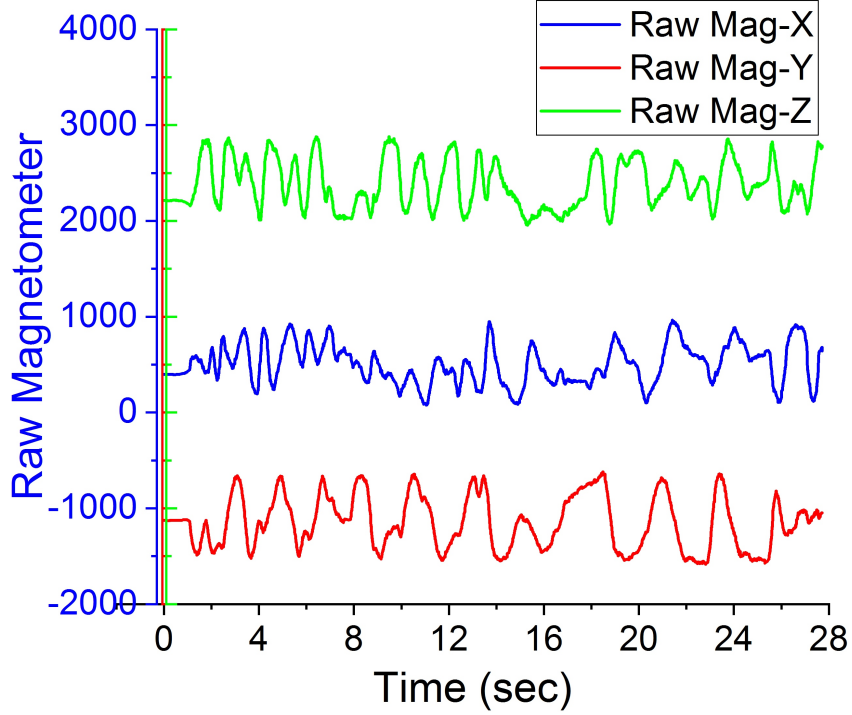
Figure 3.13.: Magnetometer Values from Protocol #3 (dynamic data).

output can be adequately accounted for based on rotation $\mathbf{R}_t^{t+1}$ and parameters $\phi_m = \{\mathbf{S}^m, \mathbf{b}_m\}$. To account for nonlinearity, we instead use the model defined in Figure 3.14 and train it on dynamic data obtained in Protocol#3 end to end, using a cost function

$$L_\theta = ||\mathbf{m}_t - \mathbf{R}_{t-1}^t \mathbf{m}_{t-1}||^2 \tag{3.10}$$

In (3.10), we obtain rotation $\mathbf{R}_{t-1}^t$ between two time instants using rate gyros and only dynamic data of Protocol# 3 is used. The term $\mathbf{R}_{t-1}^t \mathbf{m}_{t-1}$ is taken as pseudo-ground truth. The $\mathbf{m}_t$ is output value at a given time instant from model $F_m(\theta)$. Again compared to multi-position [129] or figure-of-eight [78] methods, our calibration algorithm for magnetometers require much less space coverage. Also the magnetometers are intrinsically aligned with rate gyro in respective MIMU, using (3.10).

3.6.3. Uncertainty Estimation

The uncertainty estimation in our models consists of two parts. The intrinsic covariate uncertainty in the data is called aleatoric uncertainty. It arises due to noise, non-linearity, hysteresis and dynamic response etc. of sensors. The output layer (section 3.7.1) of our model is an LSTM with linear output activation function. This layer has additional neurons to output the aleatoric uncertainty, which is learnt as part of negative log likelihood cost function [116] (described later in section 3.7.4). The model uncertainty comprise of uncertainty in the learned model and ac-

counts for unseen data. It is termed as epistemic uncertainty and is implemented as MC Dropout [39] in our models at time of inference. Both aleatoric and epistemic uncertainty are computed in our algorithm and the model provides overall uncertainty along with calibrated data (section 3.7.5). Our approach is different from other calibration models, where uncertainty is simply based on sensor static noise characteristics[106, 30, 90, 127, 45, 8, 109, 110, 153, 111]. We perform data-driven learning of uncertainty as part of our model.

3.7. Implementation

The data acquisition in our method is implemented in C++ library using propriety hardware (section 3.4.2). The data pre-processing, preparation, training of ensemble models and results evaluation is performed using Matlab 2019a on desktop PC.

3.7.1. Model Architecture

The end-to-end model architecture is shown in Figure 3.14. Our model comprises of weak learners initialized with parameters $\mathbf{S}$ and $\mathbf{b}$, obtained using methodology explained in detail in section 3.6.2. Five weak models are obtained for each sensor using a different subset of bootstrapped data. The weak (linear) model parameters $\mathbf{S}$ and $\mathbf{b}$ of equations (3.1)-(3.3) are normalized w.r.t maximum scale factor (from datasheet), so that raw sensor readings $\boldsymbol{\omega}$, $\mathbf{a}$, $\mathbf{m}$ at full scale provide ± 1 output to the hidden nodes as shown in Figure 3.15.

There are three nodes for each weak learner, which comprise of *tanh* activation function. The bagging layer is an LSTM of size 16 with linear output for both the calibrated sensor readings and associated uncertainty. A drop-out of 0.2 is used. The complete model of each sensor is fine-tuned using cost functions in (3.6), (3.8) and (3.10) for rate gyro, accelerometer and magnetometer respectively. Each cost function is augmented by (3.11) for aleatoric uncertainty.

3.7.2. Weight Initialization

The weight initialization is the key to our models convergence since only a small dataset is acquired in data collection. Instead of random weight initialization, which would then require more data to train, more space coverage and hence longer data-acquisition, we first obtain weak models based on (3.1), (3.2) and (3.3) from reduced data after 'bootstrapping' by using (3.5), (3.7) and (3.9) for rate gyros, accelerometers and magnetometers respectively. We then use parameters $\mathbf{S}$ and $\mathbf{b}$ obtained for each weak model to initialize weights as depicted in Figure 3.15

The weights/biases of output LSTM bagging layer are randomly initialized before training. The entire models are then trained end-to-end using (3.6), (3.8) and (3.10) respectively.

To demonstrate that our weight initialization of weak models performs better than randomly initialized weights of this layer, we also perform a comparison of our models with the same model architecture (single hidden layer of 15 nodes) in which the weights are randomly initialized and trained on same limited data (given in Table 3.2).
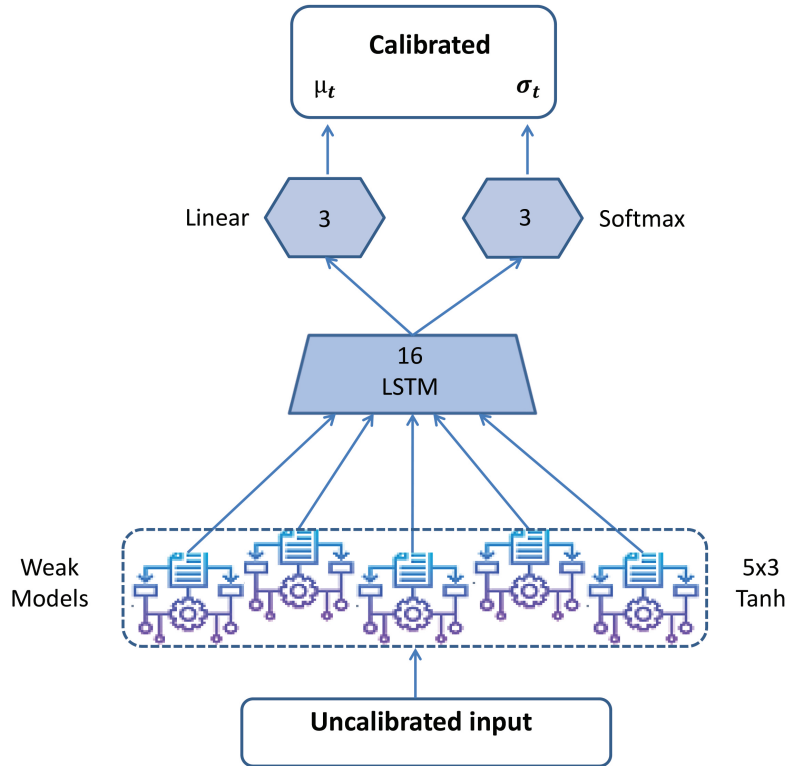
Figure 3.14.: End to End Calibration Model (3-axis Gyro, Accel, Mag).

3.7.3. Pre and Post Processing

The raw sensor input is used directly at the input, and normalization by maximum scale factor is accounted for by initial weights of weak learners. The linear output from LSTM layer of the model (which performs fusion of weak ensembles) is then scaled back by sensitivity factor at full scale obtained from datasheet.

3.7.4. Cost Function For Training

The aleatoric uncertainty deals with covariate uncertainty found in the data for a given model structure. We learn this uncertainty, using negative log likelihood (NLL) cost function [116] that we use for training our models as follows,

$$L_{NLL}(\theta) = \sum_{i=1}^{m}((\sigma_t^{(i)}(\theta))^{-2}L_\theta + \log(\sigma_t^{(i))}(\theta))^2) \qquad (3.11)$$

Here the m is the number of samples in a batch (a batch size of 32 is used). L_θ is the cost obtained using either (3.6), (3.8) or (3.10) for corresponding sensor models. We assume diagonal covariance $\sigma_t I$ and the aleatoric uncertainty is a scalar value obtained at the output that shows degree of confidence in the calibrated output values of sensors.
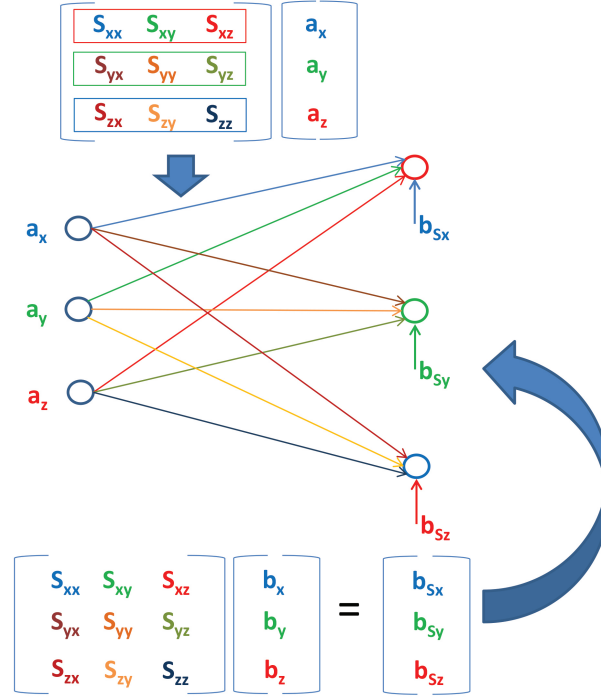
Figure 3.15.: Weight Initialization of Weak Models in Ensemble.

3.7.5. Inference With Uncertainty

Our learned model gives us a calibrated output μ_t and its covariance $\sigma_t^{a^2}$, given a raw sensor input x_t. The later represent aleatoric uncertainty and hence we assign a superscript a in its symbol $\sigma_t^{a^2}$. We train our model using a weight dropout of 0.2, and use the same for Monte Carlo dropout (MC Dropout) [39] at the time of inference to obtain epistemic (model) uncertainty. If $\mu_t^{(i)}(\theta_i)$ is the output for i-th Monte Carlo iteration of model dropout, then we write for epistemic uncertainty,

$$\sigma_t^{e^2} = \frac{1}{M}\sum_{i=1}^{M}(\mu_t^{(i)}(\theta_i)) - \frac{1}{M}\sum_{i=1}^{M}(\mu_t^{(i)}(\theta_i))^2 \tag{3.12}$$

The epistemic uncertainty is then combined with aleatoric part to get overall uncertainty as follows,

$$\sigma_t^2 = \sigma_t^{e^2} + \frac{1}{M}\sum_{i=1}^{M}\sigma_t^{a^2} \tag{3.13}$$

3.7.6. Training

To train end-to-end model shown in Figure 3.14, we have used truncated backpropagation through time (BPTT), and divided our data into subsequences of length 100, discarding those with length ≤ 90, to avoid too much zero padding. We have randomly divided our synthetic data into training, validation and test set in 70/20/10 ratio. Our model was trained using Adam algorithm with an initial learning rate of 0.001, exponentially decayed at rate of 0.9 with decay step 1000. Batch size for training was set at 32. Gradient clipping with a norm 1 was applied to LSTM. The loss on validation set was used as early stopping criteria while training was set to max epoch size of 2000. The model with best error on validation set was saved during the training run.

3.7.7. Experimental Evaluation

We evaluate our algorithm using two kinds of data. First we perform evaluation on synthetic data and then on real MIMU data obtained using two types of MIMU hardware described in section 3.4. The advantage of this two folds evaluation is that in case of synthetic data, we have precise knowledge of true sensed quantities based on which we compare our new model/algorithm with more traditional level rotation [152], multi-position [129] and figure of eight [135] methods used for field calibration of individual sensors. The synthetic ground truth establishes the baseline for the performance of our method. The synthetic data is prepared using simulated rotations and stable positions for all three protocols described for our method, as well as provides the dataset required in other calibration methods. The validation and evaluation is done on an independent test dataset. Zero mean white Gaussian noise is added to synthetic sensor readings. Also different calibration parameters are applied to perform many experiments and we estimate these applied parameters using the methods mentioned above. Our method gives less dispersion from true value in repeated set of experiments as noted in section 3.8.1 , Table 3.2.

We then carry out evaluation of our model on real MIMU data in section 3.8.3. The comparison of performance is made with state-of-art methods for simultaneous multi-sensor calibration [27] and [95] to demonstrate the better performance of our model. Since the true values are not known in this case, we choose a performance metric, where we compare in terms of mean absolute deviation (MAD) of the norm of the outputs of many sensors mounted on same rigid platform and experiencing same acceleration or rotation. We perform this comparison for the sensors, before and after the calibration.

In section 3.8.4, we show that uncertainty of the calibrated output of our model obtained using (3.11), (3.12) and (3.13). It is found well correlated with the high dynamic error and confirms that dynamic error in sensors output is different from error in static condition. Therefore, instead of using a constant covariance based on sensor noise characteristics, we get a more precise data-driven measure of uncertainty from our sensor models.

3.8. Results and Discussion

3.8.1. Performance on Synthetic Data

As mentioned in section 3.7.7, the synthetic data for baseline evaluation of our algorithm, is prepared using simulated rotations and stable positions and zero mean white Gaussian noise is added to it. We compare the performance of our model with level rotation [152] method for rate gyro calibration. For accelerometer we compare our algorithm with multi-position [129] method using equally spaced 20-positions on a unit sphere and for magnetometer we compare with figure-of-eight [135] method on unit sphere.

We report two variants of our model. The weights of our Model-I are initialized as explained in section 3.7.2, while in our Model-II, the weights are randomly initialized for all the layers for same architecture. Ten experiments are performed for each method, in which different calibration parameters are applied to synthetic true data (with noise). Then the calibration parameters are estimated using the above mentioned methods in different experiments with noisy data.

The evaluation is performed on an independent test dataset after each estimated calibration is applied to reconstruct the true sensor signal. Mean absolute error (MAE) is reported w.r.t. 3-axis sensor output as well as its norm. We also report dispersion (standard deviation: std. dev) of the error amongst different experiments. The performance comparison is given in Table 3.2.

It is evident from Table 3.2 that our method outperforms all other state-of-art methods [152], [129] and [135] used for field calibration of the 3-axis triad sensors in an individual MIMU. Although our method uses limited data-collection compared to other methods, it still performs better. The use of kinematic constraints between multiple sensors reduces the estimation error in calibration of individual sensors that arise from noisy readings and other procedural errors. Also the weight initialization using weak models and then fine-tuning (Model-I), outperforms the random weight initialization and then end-to-end training (Model-II) using our method.

The comparison of our calibration method with level rotation [152] for one axis of rate gyro and norm error of 3-axis is shown in Figure 3.16. Although the data-collection protocol is similar in both the methods, the algorithm is quite different. We see that a clear advantage is achieved in the norm error of 3- axis rate gyro using our calibration.

In Figure 3.17, we depict the comparison of our calibration with multi-position method [129] for one axis of accelerometer. The norm error of 3-axis is also shown. We have used equally spaced 20-positions on a unit sphere to generate synthetic accelerometer data for method [129]. The data protocol for our method is explained in section 3.6.1. To synthesize free fall data, we have simply averaged the synthetic accelerometer data after subtracting the gravity vector in body frame for a given applied miscalibration. The test data in Figure 3.17 is shown without noise and no acceleration except that due to gravity is assumed for clarity in the figure. It depicts a change of accelerometer readings as the orientation is changed. Again the norm error using the parameters estimated by our calibration is much less than the ones obtained using multiposition method [129].

Sensor	Method	3-axis Output		Norm Output	
		MAE	**Std. Dev**	**MAE**	**Std. Dev**
Gyro (rad/sec)	Level Rotation [152]	$1.4\ (10^{-2})$	$2.1\ (10^{-1})$	$2.2\ (10^{-2})$	$2.3\ (10^{-1})$
	Ours-I	$\mathbf{6.1}\ (10^{-3})$	$\mathbf{1.2}\ (10^{-2})$	$\mathbf{8.2}\ (10^{-3})$	$\mathbf{3.7}\ (10^{-2})$
	Ours-II	$2.1\ (10^{-2})$	$1.1\ (10^{-1})$	$3.7\ (10^{-1})$	$1.7\ (10^{-1})$
Accel (m/sec^2)	Multi-position [129]	$0.2\ (10^{-2})$	$1.7\ (10^{-1})$	$0.3\ (10^{-2})$	$0.4\ (10^{-1})$
	Ours-I	$\mathbf{3.2}\ (10^{-3})$	$\mathbf{2.5}\ (10^{-2})$	$\mathbf{4.4}\ (10^{-3})$	$\mathbf{1.6}\ (10^{-2})$
	Ours-II	$4.1\ (10^{-2})$	$3.7\ (10^{-1})$	$5.1\ (10^{-1})$	$3.3\ (10^{-1})$
Mag (unity-norm)	Figure-of-Eight [135]	$4.2\ (10^{-2})$	$9.8\ (10^{-1})$	$5.7\ (10^{-2})$	$8.9\ (10^{-1})$
	Ours-I	$\mathbf{7.3}\ (10^{-3})$	$\mathbf{6.5}\ (10^{-2})$	$\mathbf{8.2}\ (10^{-3})$	$\mathbf{9.5}\ (10^{-2})$
	Ours-II	$8.1\ (10^{-2})$	$7.8\ (10^{-1})$	$9.3\ (10^{-1})$	$8.6\ (10^{-1})$

Table 3.2.: Performance Comparison on Synthetic Data. MAE is Mean Absolute Error. Both MAE and Std Dev are computed w.r.t. synthetic ground truth data of rate gyro (rad/s), accelerometer (m/s^2) and magnetometer (normalized w.r.t max scale). Our Model-I weights are initialized with Weak Models and Model-II has randomly initialized weights. The later is fine-tuned, while former is trained fully on the data. Each Method is repeated 10 times with Synthetic dataset prepared from different calibration parameters and different ground truth values.

Figure 3.16.: Performance Comparison of Our calibration (Model-I) with Level rotation method [152] for Rate Gyro on synthetic Test Set.

Figure 3.17.: Performance Comparison of Our calibration (Model-I) with Multi-Position method [129] for Accelerometer on synthetic Test Set.
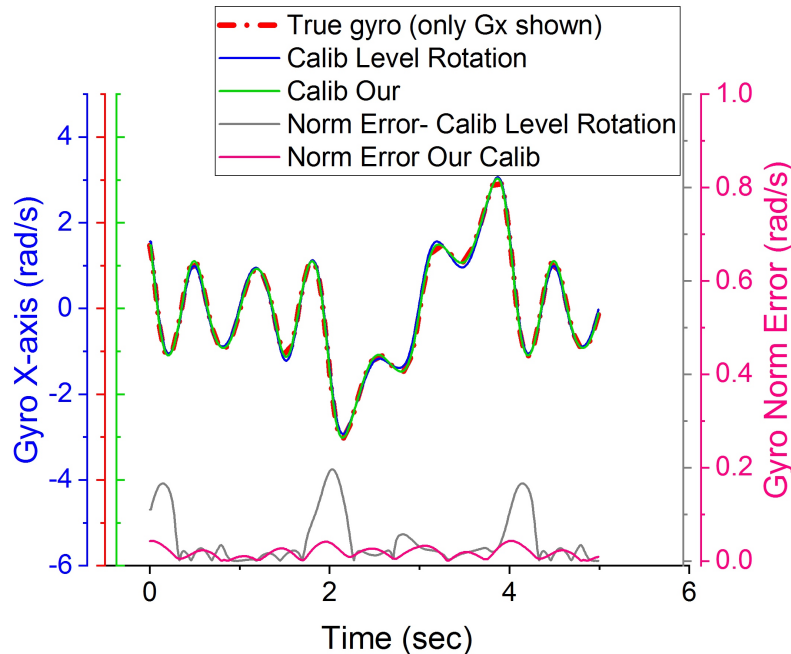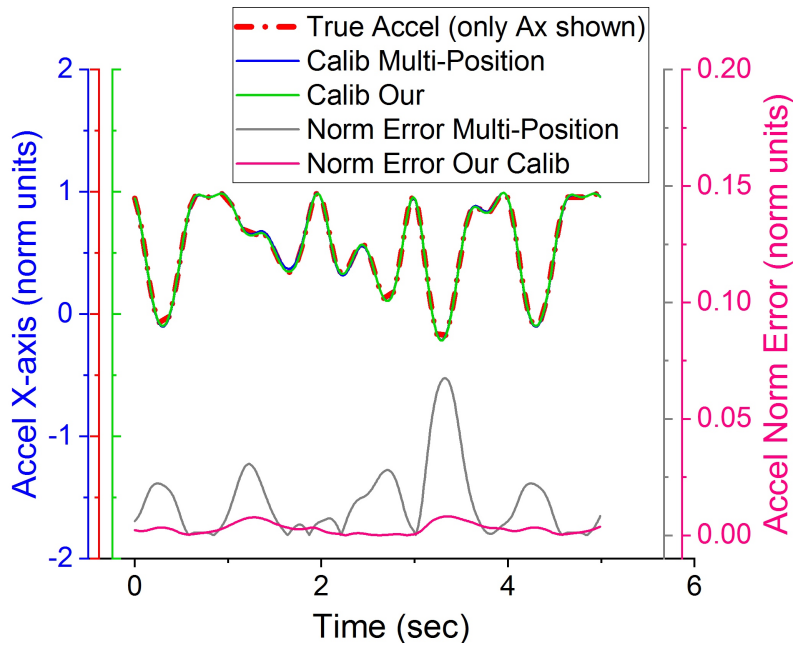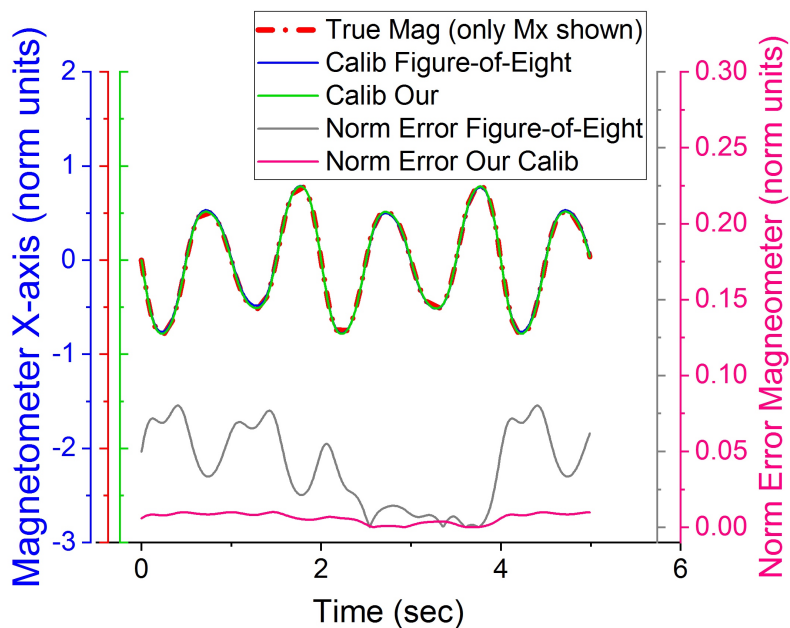

Figure 3.18.: Performance Comparison of Our calibration (Model-I) with Figure-of-Eight method [135] for Magnetometer on synthetic Test Set.

Figure 3.18, shows the comparison of our calibration with method proposed by [135]. It depicts the ground truth and calibrated values for one axis of magnetometer as well as the norm

error of all 3-axis. Again the data-collection for our method is limited to synthetic rotation between six stable positions chosen arbitrarily, same as for accelerometer calibration. The data for [135] is instead collected by rotation between 20-positions equally spaced on a unit sphere.

The miscalibration applied to synthesize magnetometer data used perturbed calibration parameters **S** and **b** in order to simulate the soft-iron and hard iron effects. The results of applying estimated calibration are depicted in Figure 3.18. Again our calibration outperforms the calibration estimated using [135].

The magnetometer test data in Figure 3.17 is shown without noise for clarity and no magnetic disturbance is assumed. But as we would show later in the section 3.8.4 that our procedure is able to calibrate the real magnetometers in inhomogeneous magnetic environment quite well and also outputs the uncertainty that results from such an indoor calibration at different orientations.

3.8.2. Effect On 3D Angle and Acceleration

We further demonstrate the advantage of better estimation of calibration parameters in case of MIMUs on rate gyro integration and linear acceleration in global frame. Figure 3.19 shows the results of rate gyro integration using calibration parameters estimated by our calibration versus level rotation method [152]. Clearly the drift in 3D orientation angles is much less using rate gyro parameters estimated by our calibration model.
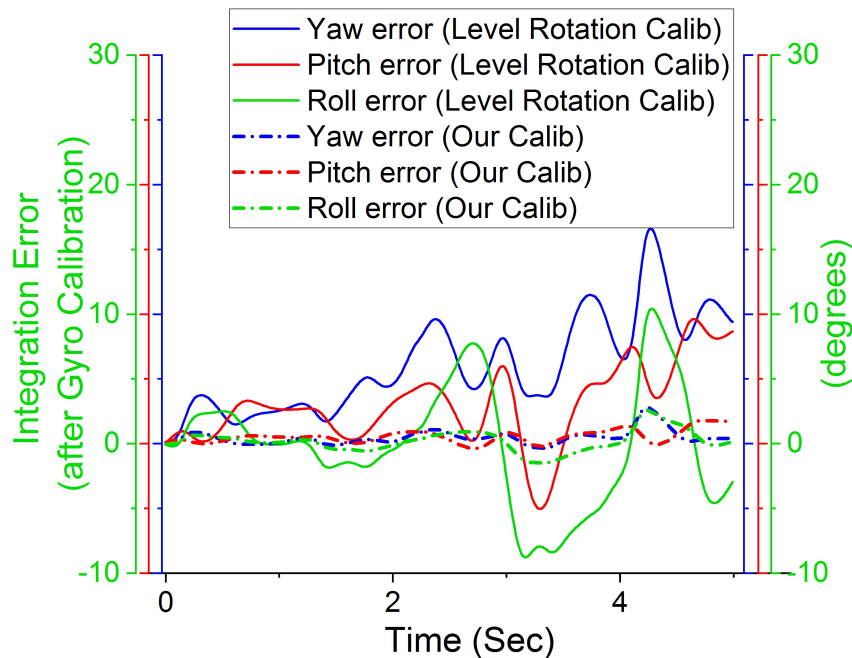

Figure 3.19.: Euler Angle Errors for Rate Gyro Integration using Our calibration (Model-I) and Level rotation method [152] on synthetic data.
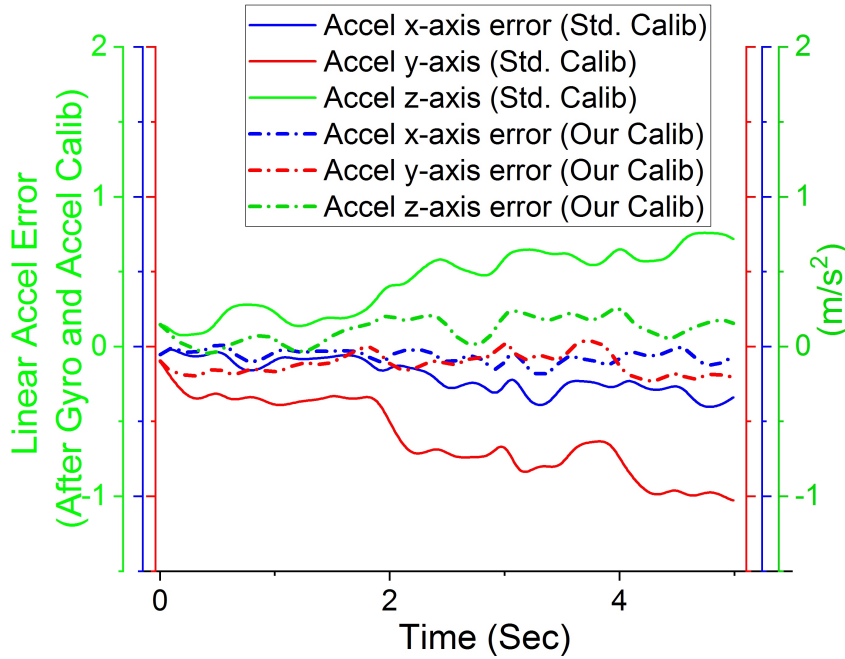
Figure 3.20.: Linear acceleration error (in global frame) using Our calibration (Model-I) and Std. methods [129], [152] on synthetic data.

The improvement in the accuracy of orientation angles leads directly to less error in the linear acceleration as well.

As shown in Figure 3.20, the error drift in linear acceleration computed from orientation and sensor frame acceleration using our calibration is much less than that when methods [129] and [152] are used for sensors calibration.

3.8.3. Performance on Real MIMU Data

The real multi-MIMU dataset used in our evaluation is obtained using two different MEMS sensors of our 'proprietary' MIMU motion tracking systems as noted in section 3.4 . Since the true sensed values are not known in this case, we choose a performance metric, where we compare in terms of mean absolute deviation (MAD) of the norm of the output of seven different sensors mounted on same rigid platform and experiencing same acceleration or rotation. We perform this comparison for the sensors before and after the calibration, in both static and dynamic conditions for the test data obtained using Protocol#3. The comparison of performance is made with the state-of-art methods for simultaneous multi-sensor calibration [27] and [95] in order to demonstrate the better performance of our model. The results are depicted in Table 3.3.

As seen in Table 3.3, our calibration performs better than other simultaneous multi-sensor calibration methods, in reducing the dispersion observed among norm output of different sensors. Since all rate gyros and accelerometers are mounted on same rigid platform, the kinematic constraint applies that the norm value of these sensed quantities, should be same (actual 3-axis

Sensor	Method	Before Calibration		After Calibration	
		Static (MAD)	**Dynamic (MAD)**	**Static (MAD)**	**Dynamic (MAD)**
Gyro	[95]	$2.4\ (10^{-2})$	$3.3\ (10^{-1})$	$1.4\ (10^{-2})$	$2.1\ (10^{-1})$
	Ours-I	same as above	same as above	**0.6** (10^{-3})	**0.3** (10^{-2})
Accel	[95]	$1.6\ (10^{-2})$	$5.7\ (10^{-1})$	$0.3\ (10^{-2})$	$4.3\ (10^{-1})$
	Ours-I	same as above	same as above	**1.5** (10^{-3})	**1.2** (10^{-2})
Mag	[27]	$4.2\ (10^{-2})$	$5.5\ (10^{-1})$	$1.5\ (10^{-2})$	$2.1\ (10^{-1})$
	Ours-I	same as above	same as above	**1.1** (10^{-3})	**1.3** (10^{-2})

Table 3.3.: Performance Comparison on real MIMU Data. MAD is Mean Absolute Deviation. MAD between norm output of seven different sensors mounted on same rigid platform is shown in both static and dynamic condition. Our Model-I in which weights are initialized with Weak Models is compared with [95] and [27]. The units of Gyro, Accelerometer and Magnetometer are rad/s, m/s^2 and mG respectively.

readings depends upon relative orientation). We observe that the deviation among inertial sensors is always more in dynamic condition than the static one. It implies that the error in dynamic condition for inertial sensors is also high compared with static condition. This may arise due to variations in filtering and system dynamics of individual MEMS inertial sensors. In next section, we would demonstrate how the uncertainty predicted for calibrated values by our model correlates well with error observed for inertial sensors in static and dynamic condition.

The deviation of norm output among magnetometers improves after our calibration, but does not show significant difference between static and dynamic case. We conclude that it is only affected by residual error of calibration performed in inhomogeneous magnetic field.

3.8.4. Uncertainty Estimation

The main feature of our model is prediction of uncertainty along with calibrated sensor readings, while regressing from a sequence of raw sensor values. We estimate both aleatoric and epistemic uncertainty and then combine these in overall estimate using (13). Figure 3.21 depicts seven rate gyro readings after calibration and the uncertainty for two gyro outputs.

The norm value of rate gyro outputs is shown and the uncertainty is depicted in terms of standard deviation (δ). The scale chosen for the uncertainty is different. We note that the uncertainty changes between static and dynamic condition

In Figure 3.22, we further show the uncertainty (δ) for two accelerometers as well as calibrated norm readings of seven accelerometers both in static and dynamic condition. Again the uncertainty of accelerometer during static condition is much less than that during dynamic condition. This correlates well with more deviation in the norm values of accelerometers during dynamic condition which may be attributed to slight variations in filtering and system dynamics of individual MEMS accelerometers.

Figure 3.23, shows the norm values of seven magnetometers after calibration and uncertainty (δ) for two magnetometers.
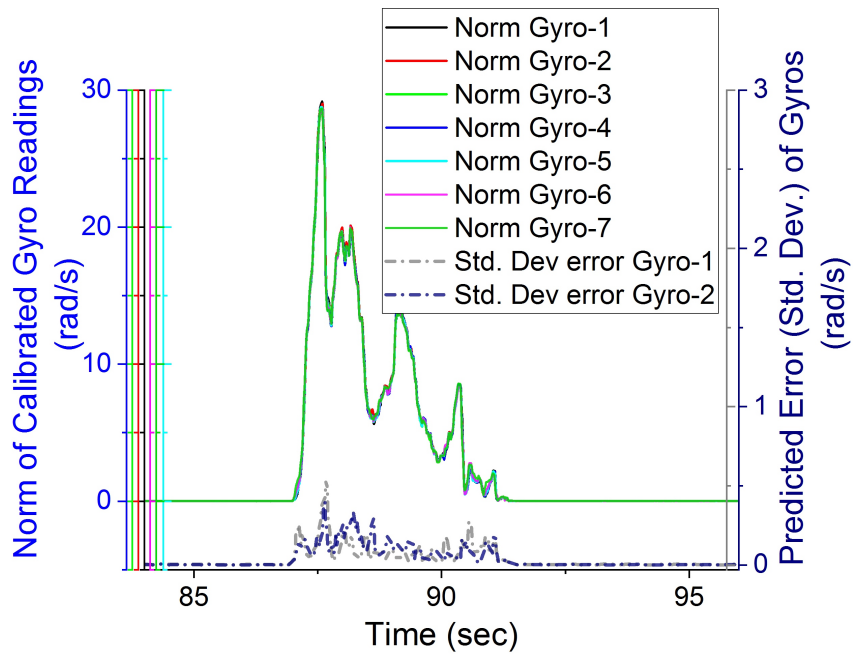
Figure 3.21.: Norm output of seven gyros after Our calibration (Model-I). Uncertainty is shown for two gyros only (for clarity)
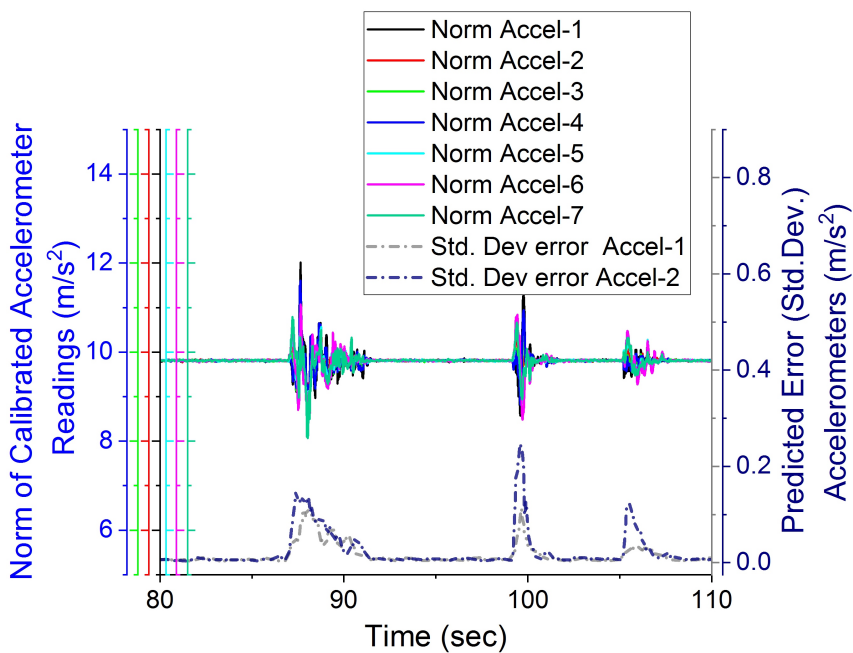

Figure 3.22.: Norm output of seven accelerometers after Our calibration (Model-I). Uncertainty is shown for two accelerometers
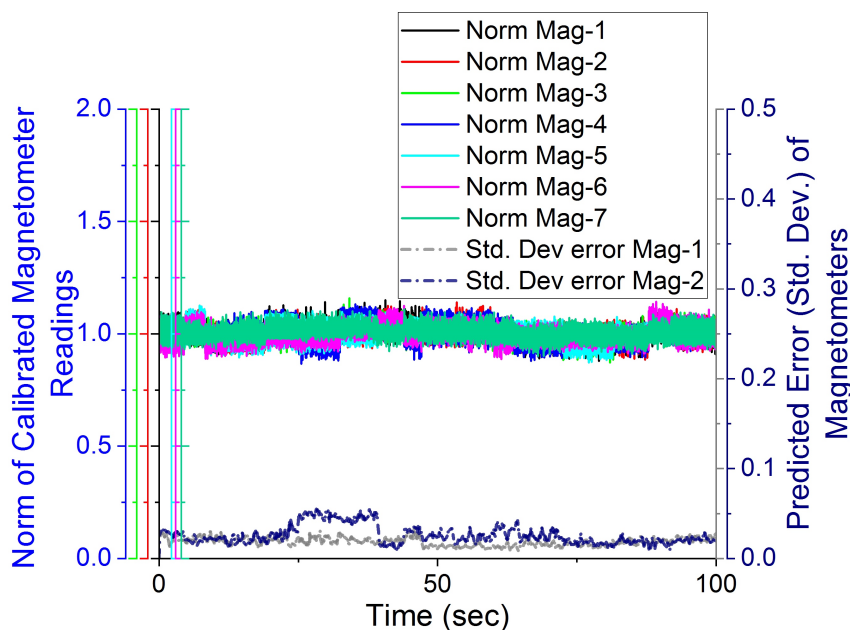
Figure 3.23.: Norm output of seven magnetometers after Our calibration (Model-I). Uncertainty is shown for two magnetometers

In Figure 3.23, we note that the uncertainty in magnetometer readings does not change between static and dynamic condition, instead it correlates with the residual error after calibration. Since the magnetometers calibration is performed in inhomogeneous magnetic field, it is not very precise, but the uncertainty is still predicted accurately to represent this residual error.

3.9. Conclusion

Our work presented in this chapter, proposes a fast and convenient calibration method for simultaneous calibration of multiple MIMU units in the field. Each MIMU comprises of 3-axis rate gyro, accelerometer, and magnetometer and all the units are mounted on a single rigid platform, which can be placed in up to 6 stable positions and also rotated on a level plane in each position. We use a two-tier sensor calibration model based on ensemble approach and apply kinematic constraints w.r.t multiple sensors on a single rigid platform to calibrate it. Our model is initialized using weakly learned linear models from subsets of data. Then the entire model with LSTM fusion layer is fine-tuned on acquired data. The uncertainty (sensor error) is learned as a part of our framework. We perform data-driven learning of uncertainty of sensor values and show that the predicted uncertainty for inertial sensors correlates well with the change of error between static and dynamic conditions. Similarly the uncertainty predicted for magnetometer arises due to residual calibration error in inhomogeneous magnetic field and is also learned as part of model. Compared with existing state-of-art methods, our algorithm gives more consistent readings of all MIMUs and also predicts the associated uncertainty of the sen-

sor output. The uncertainty prediction of individual sensors is particularly helpful in the sensor fusion. In the next chapter, the uncertainty estimated by our calibration model would be used in sensor fusion algorithm in order to demonstrate its advantage in terms of accuracy and error convergence.

4. Sensor Fusion for Orientation Estimation

In this chapter, we present a robust sensor fusion and online sensor calibration algorithm for MIMUs (magnetic-inertial measurement unit) in a body worn inertial human motion capture. Our system hardware is designed for high rate data acquisition and embedded sensor fusion at low power budget and is described before in chapter 3. The proposed sensor fusion algorithm is designed for estimating accurate 3D orientation from 3-axis rate gyro, accelerometer and magnetometer triad. We introduce an adaptive sensor covariance based EKF for sensor fusion, which makes it robust to both dynamic body accelerations as well as inhomogeneous magnetic field. The adaptive covariance method penalizes the bad accelerometer and magnetometer measurements. Compared with existing state-of-art methods, it is well suited for indoor human motion analysis and monitoring highly dynamic motion. The algorithm also performs auto-recalibration of MIMU sensors while in use. The rate gyro bias is intelligently calibrated online during quasi-static time steps using only undisturbed readings of accelerometer/magnetometer, while both accelerometer and magnetometer biases are recalibrated using the residuals obtained in EKF during sensor fusion. The work presented in this chapter is based on our peer reviewed published research [13, 11].

4.1. Background

The 3-axis rate gyro gives only relative 3D rotation w.r.t a reference frame. The absolute 3D orientation needs measurement of at least two non-collinear reference vectors which define a reference frame. In case of MIMU, 3-axis accelerometer and magnetometer measures acceleration due to gravity and local magnetic field respectively. The individual MEMS based 3-axis sensors like rate gyro, accelerometer and magnetometer are not very precise and robust. Despite calibration, residual errors and parameters drift are common. Therefore in MIMU based human motion capture, we employ 9-axis MIMU on each body segment to obtain its 3D orientation. Each 9-axis MIMU comprise of a triad of 3-axis rate gyro, accelerometer and magnetometer and sensor fusion algorithms are implemented for orientation estimation from 9-axis MIMU.

The triad of 3-axis rate gyro, accelerometer and magnetometer in an MIMU can be used to obtain 3D orientation. The accelerometer and magnetometer measurements provide 3D orientation w.r.t magnetic North-East-Up frame of reference (Figure 4.1). This assumes that inter-alignment between 3-axis accelerometer and magnetometer is known a priori. Traditionally TRIAD [117] and QUEST [17] algorithms have been used to obtain 3D orientation from

measurement of two reference vectors. These deterministic algorithms based on two vectors observation are adapted using either Kalman Filter or Least squares [74] for noisy measurements.

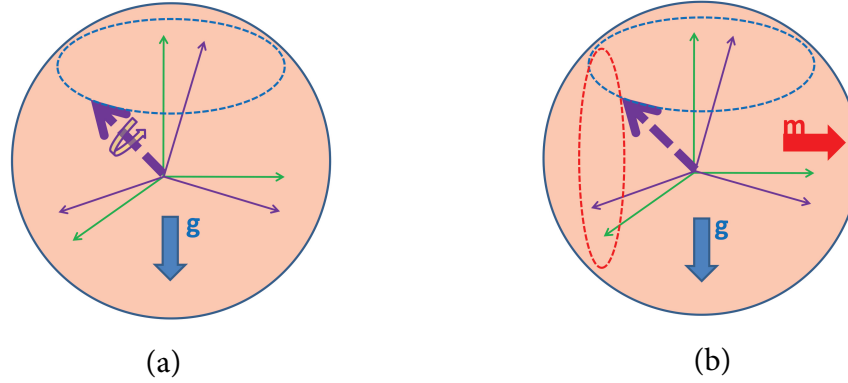(a)   (b)

Figure 4.1.: 3D rotation of a MIMU frame (purple) w.r.t a North-East-Up reference frame (green). The rotation is represented by a dotted (purple) unit axis vector (swing) and a 1D rotation around this axis (twist) *(a)* When only gravity vector (blue) is measured, there is ambiguity in 3D orientation of MIMU frame. All 3D orientations in which locus of axis vector trace the dotted (blue) circle are possible *(b)* When both magnetic and gravity vectors are measured. Each has its individual ambiguity shown by dotted circles (blue and red). But only one intersection of these circles on unit sphere meets the right hand rule, which uniquely determines the 3D orientation

In addition to white Gaussian noise, the measurements of accelerometer and magnetometer are however also affected by external accelerations and magnetic variations respectively (c.f. equations (3.2) and (3.3) in chapter 3). Particularly, the indoor environments show a great degree of magnetic inhomogeneity [23]. Also the MIMU mounted on human body undergo dynamic acceleration whose intensity varies with the movement itself. A rate gyro can measure the change in 3D rotation without being affected by external disturbances. However if rate gyro is used alone to estimate 3D rotation over a time interval, its residual intrinsic bias causes the 3D angle to drift. A sensor fusion algorithm that makes optimal use of all the three sensors, 3-axis rate gyro, accelerometer and magnetometer in an MIMU, provides a more robust 3D orientation.

The problem with existing sensor fusion algorithms is twofold. Complementary filter algorithms [79, 81] assume a heuristic or gradient based weight for the fusion of three sensors while Kalman filter [119] weights are adapted assuming a tuned covariance. In both cases, it is assumed that uncertainty of accelerometer and magnetometer readings remain constant. Secondly, rate gyro bias is estimated together with 3D orientation in some algorithms [118]. This however assumes that magnetometer and accelerometer readings are precise and rate gyro biases are accurately observable using those sensors. But in general accelerometer readings include dynamic body accelerations in addition to gravity vector and magnetometer in indoor/urban environment measures varying local magnetic field and not the true magnetic north. Therefore MIMU based

sensor fusion and gyro bias calibration show degraded performance when run for long time, especially indoors.

For robust sensor fusion, it is vital that rate gyro can provide precise estimates, when either accelerometer or magnetometer readings are unreliable. This occurs during the period of high dynamic accelerations and/or inhomogeneous magnetic fields. On the contrary, the rate gyro calibration can be updated using accelerometer and magnetometer, when these sensors are well calibrated and measurement distortions and disturbances are low. Our goal is to implement an algorithm that simultaneously perform sensor fusion to obtain 3D orientation and also the online recalibration of the triad of sensors in an MIMU. In the past work [92, 35] the sensor covariance have been switched between a baseline value and a high value based on hard constraints to obtain filter weights, but we propose more soft adaptation of sensor covariance and demonstrate that it works better.

4.2. Contribution

Our present work adopts the approach of using a distributed sensor fusion algorithm in which adaptive covariance is implemented for accelerometer and magnetometer in an Extended Kalman Filter (EKF). The self-recalibration feature of our algorithm achieves precise orientation estimates in highly dynamic conditions and avoid drift or error accumulation in inhomogeneous magnetic fields. The main contributions covered in this chapter are,

1. An online calibration of rate gyro bias is accomplished iteratively at sensing MIMU node during quasi-static intervals . It is done using residuals from incremental change in orientation suggested by other sensors. The updated gyro calibration is crucial and achieves good orientation estimates when accelerometer and/or magnetometer measurements are bad.

2. We implement an adaptive measurement covariance based correction step in the sensor fusion algorithm that makes our algorithm robust to estimation errors arising due to highly dynamic conditions and error accumulation in inhomogeneous magnetic field. The adaptive covariance penalizes the bad accelerometer and magnetometer measurements.

3. The accelerometer and magnetometer are calibrated to reduce their residuals with respect to rate gyro prediction in sensor fusion step accomplished on the hub node. Again the use of adaptive covariance ensures that calibration step is least affected by disturbed measurements.

4.3. Organization of Chapter

This chapter is organized in eight main sections. In the first two sections we have covered the background and our main contribution to the problem of sensor fusion and real-time intrinsic calibration of body worn MIMUs. The aim is robust estimation of 3D orientation from each

MIMU under body accelerations and inhomogeneous magnetic environment. In section 4.4, we present our methodology to perform dual (offline/online) calibration and sensor fusion in a distributed algorithm. In section 4.5, we provide the details of implementation like hardware, offline calibration and distributed algorithm for sensor fusion and online calibration. In section 4.6, we describe the protocol of our experiments to evaluate the proposed algorithm. In section 4.7, we present results of high rate data acquisition and comparison of 3D orientation obtained using our sensor fusion algorithm and from gold standard optical tracking. We also demonstrate the robustness of 3D orientation estimation obtained from our algorithm under accelerations and magnetic disturbances compared to results from a baseline algorithm. In both cases, we show that our algorithm achieves better performance than contemporary baseline method. Our algorithm also performs online calibration of sensors which is found to give better results compared to when only offline calibration is performed.

4.4. Methodology

4.4.1. Dual Calibration Concept

The systematic errors intrinsic to the magnetic and inertial sensors, are determined in a calibration procedure. The quality of sensor fusion performed using different sensor modalities is directly dependent on the accuracy achieved in intrinsic calibration of sensors. However the calibration of low cost sensors is not robust and varies slowly. It is not always practical to carry out intricate precise calibration at intervals. In order to solve this issue, we adopt a two-step calibration. First an offline calibration is done to establish the constant parameters like physical misalignment between the sensor triads and to provide an initial estimate of other calibration parameters. Then a more precise online calibration of all the sensors is performed. A distributed algorithm is implemented for online calibration. It partly runs on the sensing node itself and a part of it is implemented as hybrid algorithm with sensor fusion on hub. Our scheme of algorithm is depicted in Figure 4.2.

4.4.2. Calibration models

The systematic errors of 3-axis magnetic and inertial sensors, like scaling, misalignment and non-orthogonality (as well as soft-iron effects) are represented by 3x3 $\mathbf{S}$ matrix and 3x1 bias (and hard iron effects) terms $\mathbf{b}$. These calibration parameters are stored on the sensor node, for later use in sensor fusion. At run-time the calibrated readings (in physical units) are obtained for all sensors as follows.

$$\boldsymbol{\omega}_{cal} = \mathbf{S}^{g^{-1}}(\boldsymbol{\omega}_{raw} - \mathbf{b}_g) = \boldsymbol{\omega} + \tilde{\mathbf{w}}_g \tag{4.1}$$

$$\mathbf{a}_{cal} = \mathbf{S}^{a^{-1}}(\mathbf{a}_{raw} - \mathbf{b}_a) = \mathbf{R}_n^b(\mathbf{a} + \mathbf{g}) + \tilde{\mathbf{w}}_a \tag{4.2}$$
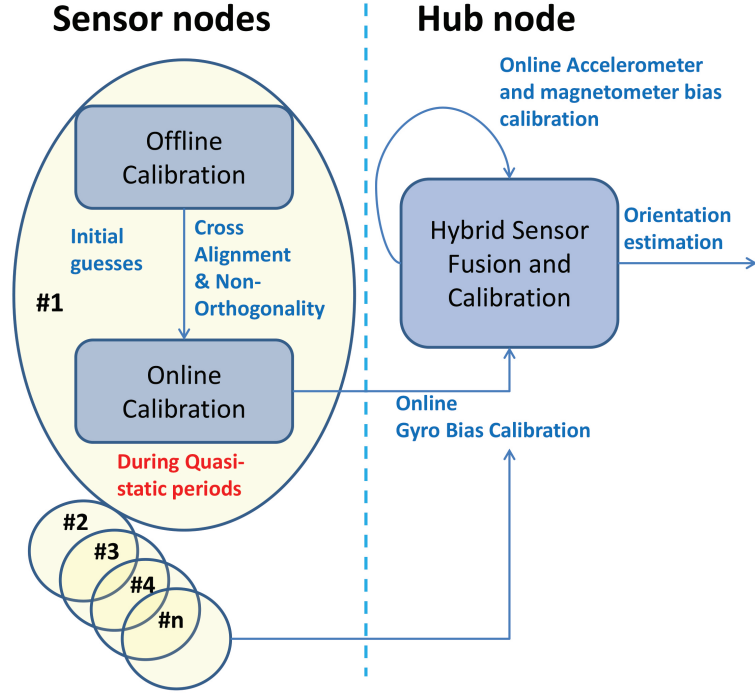
Figure 4.2.: Distributed Sensor Fusion and Calibration

$$\mathbf{m}_{cal} = \mathbf{S}^{m^{-1}}(\mathbf{m}_{raw} - \mathbf{b}_m) = \mathbf{R}_n^b(\mathbf{m}_n + \mathbf{d}) + \tilde{\mathbf{w}}_m \tag{4.3}$$

In equations above, $\mathbf{S}$ are 3×3 sensitivity and misalignment matrices and $\mathbf{b}$ are bias terms, while $\mathbf{w}$ represents the white Gaussian noise intrinsic to each sensor. The vector $\boldsymbol{\omega}$ is the angular rate. $\mathbf{R}_n^b$ is a 3×3 rotation matrix that rotates the physical vectors $\mathbf{a}$ and $\mathbf{g}$ represented in navigation frame of reference to body fixed frame of reference. The body fixed frame has its origin at the center of 3-axis sensors triad and its x, y, z axis are orthogonal and aligned with that of MIMU package. The navigation frame of reference has its origin at the initial position of MIMU and its x-axis is aligned with local magnetic north while y-axis is opposite to gravity (and z-axis is orthogonal). The body acceleration $\mathbf{a}$ (in navigation frame) is added to reference gravity vector $\mathbf{g}$. In magnetometer case, $\mathbf{S}^m$ is a 3×3 matrix that in addition to sensitivity, misalignment and non-orthogonality, also account for soft iron effects. The soft-iron effects are distortions that appear in the magnetometer measurement due to ferrous materials fixed w.r.t body frame. Similarly, $\mathbf{b}_m$ includes both the sensor bias as well as hard iron effects due to permanent magnetic sources fixed to body. It is interesting to note that magnetic reference vector for indoor settings, is often very different from the earth's magnetic field. It is also continuously changing due to a position dependent (and time varying) magnetic disturbance $\mathbf{d}$ in the environment [13].

4.4.3. Sensor Fusion

Body segment orientation is the primary quantity of interest in human motion capture and biomechanical analysis. The MIMU sensor nodes are attached to rigid body segments and orientation is obtained through sensor fusion. Two known non-collinear global reference vectors (like Earth's gravity and magnetic field) provide a global reference frame to define a rigid body orientation. Since the accelerometer measures reference gravity vector, it can give absolute tilt (pitch/roll) information with respect to gravity vector. Similarly horizontal component of a reference magnetic vector $\mathbf{m}_n$ can define magnetic north and gives heading (yaw) information. However as evident from (4.2) and (4.3) (right hand side), both these reference vectors in measurements $\mathbf{a}_{cal}$ and $\mathbf{m}_{cal}$ are affected by disturbance vectors $\mathbf{a}$ and $\mathbf{d}$ respectively. Thus (4.2) and (4.3) even if solved together, do not provide accurate orientation $\mathbf{R}_n^b$. On the other hand, (4.1) shows that angular rate $\boldsymbol{\omega}_{cal}$ does not suffer from any disturbance (if all bias is eliminated). Hence the following rate equation can be integrated over time from a known initial orientation to predict current orientation.

$$\dot{\mathbf{R}}_n^b = \frac{1}{2}\boldsymbol{\Omega} \cdot \mathbf{R}_n^b \tag{4.4}$$

Here $\boldsymbol{\Omega}$ is a 3x3 skew-symmetric matrix obtained from $\boldsymbol{\omega}$. The discretization of integration based on (4.4) is however only approximate and hence small error appears at each time step. Even if gyro bias and calibration is accurately known, integration of Gaussian white noise in gyro readings causes a random walk to appear, even under stationary conditions. Bearing these anomalies in mind, the sensor fusion is the preferred approach for orientation estimation from MIMU sensors.

A robust sensor fusion algorithm ensures that the error converges and the system is also robust to outliers i.e. disturbed accelerometer and magnetometer readings are rejected. The results of past algorithms on account of later goal can still be improved, because when the robustness is checked practically in unrestricted ambulatory settings, drift or random walk error in yaw is very common. This is due to the reason that magnetic disturbances unlike body accelerations are of more permanent nature, as humans move indoor.

Extended Kalman Filter (EKF) has been found computationally cost effective for orientation estimation from noisy MIMU or MARG (magnetic, acceleration and rate gyro) sensors, based on sensor fusion. The unit quaternion representation of orientation is widely popular in these EKF implementations [156], as it avoids Gimbal Lock problem and is less memory intensive than rotation matrix or DCM (Direction Cosine Matrix) representation. It is also very convenient to impose unit norm constraint on a unit quaternion after each step. Our sensor fusion algorithm is also based on a quaternion based EKF formulation. We implement two EKF filters for comparison. **EKF#1** is a baseline sensor fusion algorithm without any measurement covariance adaptation. The process and measurement covariance are assumed constant in this implementation. Whereas **EKF#2** demonstrates our novel covariance adaptation scheme as well as a new magnetometer measurement correction. Thus the difference between two implementations is only in the process and measurement covariance, and will be duly highlighted in description below.

4.5. Implementation

4.5.1. Hardware

Our algorithm is implemented on our wearable MIMU based system built using flexible textile harness (EasyIMP). The complete system is described in detail in preceding chapter in section 3.4. The sensing nodes are based on Invensense MPU-9150 integrated 9-axis Motion Tracking device, connected through a differential I2C bus to the Intel Edison hub, running a proprietary plugin software on Yocto Linux. Our sensor fusion algorithm is implemented as part of plugin software stack.

4.5.2. Offline Calibration

The aim of our offline calibration is to register physical misalignments between sensors and axes non-orthogonality. It also provides a good initial estimate of biases and sensitivities for faster convergence of online calibration. Our work assumes that the offline calibration of MIMUs is performed at each switch-on and after a warm-up period of few minutes has elapsed. Our algorithm [10] for offline calibration of multiple MIMUs simultaneously is described in detail in preceding chapter (chapter 3). We assume that body frame coincides with accelerometer triad, hence $\mathbf{S}^a$ is assumed a lower triangular matrix with decomposition defined as $\mathbf{S}^a = \mathbf{T}^a \cdot \mathbf{S}^a$. The diagonal matrix $\mathbf{S}^a$ represents the sensitivities and unit lower triangular matrix $\mathbf{T}^a$ represents non-orthogonality. We obtain an initial guess of accelerometer bias $\mathbf{b}_a$ from offline calibration while matrix $\mathbf{S}^a$ is treated as constant in sensor fusion. For rate gyros, we define $\mathbf{S}^g = (\mathbf{R}_a^g \cdot \mathbf{T}^g) \cdot \mathbf{S}^g$, where $\mathbf{R}_a^g$ defines the physical misalignment between gyro and accelerometer triads. From calibration, we obtain a combined $(\mathbf{R}_a^g \cdot \mathbf{T}^g)$ alignment matrix and perform QR decomposition to obtain $\mathbf{R}_a^g$ and $\mathbf{T}^g$. The gyro biases $\mathbf{b}_g$ obtained from offline calibration are used as initial guess in online calibration, while $\mathbf{R}_a^g$, $\mathbf{T}^g$ and $\mathbf{S}^g$ are treated constant.

We found out that a calibration of magnetometers and gyro biases prior to use is more reliable. First the residual gyro bias errors are obtained accurately prior to start-up in a rest position as $\mathbf{b}_{g,new}$. Thereafter, an arbitrary motion is performed for the magnetometer calibration to cover enough 3D points. After intrinsic magnetometer calibration, $\mathbf{S}^m$ is still ambiguous in terms of rotation w.r.t inertial sensors triad. Therefore we employ the cost function. $L_{\Psi_m} = ||\mathbf{m}_{n+1} - \mathbf{R}_m^g \cdot \mathbf{R}_n^{n+1} \cdot \mathbf{m}_n||^2$ where Ψ_m comprise of three Euler angles defining the alignment $\mathbf{R}_m^g$. Incremental rotation $\mathbf{R}_n^{n+1}$ is obtained using calibrated gyro readings. Since rate gyro is already aligned with accelerometer, $\mathbf{R}_m^g = \mathbf{R}_m^a$. We obtain $\mathbf{S}_{new}^m = \mathbf{R}_m^g \cdot \mathbf{S}^m$ for full magnetometer offline calibration and alignment. The raw sensor readings are preprocessed using offline calibration parameters stored on sensor node to provide calibrated values. The preprocessed data is then used by sensor fusion algorithm running on the hub node.

4.5.3. Online Calibration at Sensor Node

After the in-situ offline sensor calibration, the MIMUs based motion capture system is worn by the user. As soon as sensor fusion algorithm invokes sensor preprocessing/calibration algorithm

on the sensor node, the later monitors the calibrated sensor outputs for detection of quasi-static states using a detection window [132]. During normal human motion, these states occur often. The gyro bias is estimated by running a simple EKF during quasi-static conditions. The state vector comprises of x,y,z angular rates and its prediction is given by.

$$\widetilde{\mathbf{x}}_k = \widetilde{\boldsymbol{\omega}}_k = \acute{\boldsymbol{\omega}} - \mathbf{b}_g^{k-1} \tag{4.5}$$

Here $\acute{\boldsymbol{\omega}}$ is an average of $\boldsymbol{\omega}_{cal}$ readings of last $k-1$ and current time step k and $\mathbf{b}_q^{k-1}$ is the last estimated rate gyro bias vector. The angular rate during quasi-static conditions is not assumed zero, as body worn sensors may undergo slight motion. The predicted measurements are then given by.

$$\mathbf{g}_k^p = \mathbf{q}_{\triangle k} \oplus \mathbf{g}_{k-1} \oplus \mathbf{q}_{\triangle k}^* \tag{4.6}$$

$$\mathbf{m}_k^p = \mathbf{q}_{\triangle k} \oplus \mathbf{m}_{k-1} \oplus \mathbf{q}_{\triangle k}^* \tag{4.7}$$

where $\mathbf{g}_k^p$ and $\mathbf{m}_k^p$ are predicted unit gravity and magnetic vectors. $\oplus$ represents quaternion multiplication and unit quaternion $\mathbf{q}_{\triangle k}$ represent rotation from $k-1$ to current time step k. It is obtained using estimated angular velocity $\widetilde{\boldsymbol{\omega}}_k$ from (4.5), as $\mathbf{q}_{\triangle k} = [\cos(\frac{|\widetilde{\boldsymbol{\omega}}_k|}{2})\Delta k, \sin(\frac{|\widetilde{\boldsymbol{\omega}}_k|}{2})\Delta k \cdot (\frac{\widetilde{\boldsymbol{\omega}}_k}{|\widetilde{\boldsymbol{\omega}}_k|})]$, and $\mathbf{q}_{\triangle k}^*$ represents conjugate quaternion of $\mathbf{q}_{\triangle k}$. The detailed equations of EKF are covered in next section and not repeated here for brevity. The residuals obtained on (4.6) and (4.7) w.r.t actual measurements during quasi-static period are used to calculate Kalman gain and to correct the state prediction from (4.5). Thereafter, a residual of corrected $\boldsymbol{\omega}_k$ w.r.t $\widetilde{\boldsymbol{\omega}}_k$ is low pass filtered to obtain an updated estimate of $\mathbf{b}_g$. The rate gyro bias is not updated during motion. Instead the residuals of orientation estimation EKF running on hub, are used to update calibration of accelerometer and magnetometer.

4.5.4. Intelligent Sensor Fusion & Calibration (at Hub)

Our sensor fusion algorithm at hub is also based on EKF. We implement two EKF filters for comparison. **EKF#1** is a baseline sensor fusion algorithm based on reported literature [50, 21, 35]. The process and measurement covariance are assumed constant in this implementation. Whereas **EKF#2** demonstrates our novel covariance adaptation as well as a new innovation based update in EKF using magnetometer.

Our state for both EKF at time step k is simply $\mathbf{x}_k = \mathbf{q}_k$, where $\mathbf{q}_k$ is a unit quaternion representing orientation at time step k. We initialize with an arbitrary orientation state $\mathbf{q}_k = [1, 0, 0, 0]$ i.e. body is assumed aligned with reference North-East-Up global frame. The filter converges very fast to true orientation in few times steps. The state covariance $\mathbf{P}_0$ is initialized as diagonal matrix of low values. During this initialization period, average local reference vector $\mathbf{m}_n$ is computed as $\mathbf{m}_{avg} = \bar{\mathbf{m}}_{cal}$ for few time steps.

We perform the state prediction step using a bias corrected average of $\boldsymbol{\omega}_{cal}$ readings of last $k-1$ and current time step k, i.e. $\bar{\boldsymbol{\omega}} = \frac{1}{2}(\boldsymbol{\omega}_k + \boldsymbol{\omega}_{k-1}) - \mathbf{b}_g^k$. This assumes constant angular

acceleration (first order integration) model. Taylor 's expansion of the integration of quaternion equivalent of (4.4) then reduces to (4.8)

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{q}}_k = \mathbf{q}_{k-1} \oplus \mathbf{q}_{\triangle k} \tag{4.8}$$

Here $\oplus$ represents quaternion multiplication, $\mathbf{q}_{\triangle k} = [\cos \frac{|\boldsymbol{\omega}|}{2}\Delta k, \sin \frac{|\boldsymbol{\omega}|}{2}\Delta k \cdot (\frac{\bar{\boldsymbol{\omega}}}{|\bar{\boldsymbol{\omega}}|})]$ represents the change in orientation for this time step and $\tilde{\mathbf{q}}_k$ is the current prediction.

The state covariance is updated by (4.9)

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k \tag{4.9}$$

The $\mathbf{P}_{k|k-1}$ represent the uncertainty of new prediction. $\mathbf{F}_k$ is the Jacobian evaluated for (4.8) at $\mathbf{q}_{k-1}$ and $\mathbf{Q}_k$ is the process covariance arising due to uncertainty of prediction step. In our **EKF#2**, we have adapted $\mathbf{Q}_k$ in relation to $|\bar{\boldsymbol{\omega}}|$ as follows

$$Q_{ii}^k = variane + scale * (max(|\bar{\boldsymbol{\omega}}| - threshold, 0)) \tag{4.10}$$

Now, $\mathbf{Q}_k = diag([Q_{ii}^k])$. The parameters $variance$, $scale$ and $threshold$ can be set separately for (4.10) and later for (4.12-4.13), (4.20-4.21). Typical values of these parameters are not given here (for propriety reasons), but algorithm is found robust for a range of these parameters values. In (4.10), the first term defines a baseline variance due to integration of gyro noise (random walk), while second term defines the uncertainty arising due to calibration and integration errors that scale with $|\bar{\omega}|$ for a known time step. In contrast, **EKF#1** assumes a constant value for process covariance, $\mathbf{Q}_k$.

The correction step with measurements is carried out one by one in both **EKF#1** and **EKF#2**. First the gravity vector in body reference frame is estimated using predicted quaternion for step k

$$\mathbf{g}_k^p = \tilde{\mathbf{q}}_k \oplus \mathbf{g} \oplus \tilde{\mathbf{q}}_k^* \tag{4.11}$$

In 4.11, $\tilde{\mathbf{q}}^*$ represents conjugate quaternion of $\tilde{\mathbf{q}}$. Gravity vector is assumed Up i.e. $\mathbf{g} = [0, 0, 1]$. Now with measured normalized acceleration , $\hat{\mathbf{a}} = (\frac{\mathbf{a}_{cal}}{|\mathbf{a}_{cal}|})$, the residual is obtained $\mathbf{y}_a = (\hat{\mathbf{a}} - \mathbf{g}_k^p)$. Our **EKF#1** implementation always uses a constant value for measurement covariance $\mathbf{R}_k^a$, while measurement covariance in our **EKF#2** algorithm is adapted in two different ways

$$R_{ii}^{a,k} = variance + scale * (max(|(|\mathbf{a}_{cal}| - g_{avg})| - threshold, 0) \tag{4.12}$$

$$R_{ii}^{a,k} = variance + scale * (max((|\boldsymbol{y_a}|) - threshold, 0)) \tag{4.13}$$

So as, $\mathbf{R}_k^a = diag([R_{ii}^{a,k}])$. **EKF#2** approach uses (4.12) during initialization /convergence period, thereafter (4.13) is used. The equation (4.13) is a novel covariance adaptation proposed in this paper and the motivation is explained as follows. When the body segments experience downward and sideward acceleration together, a situation may arise when norm of acceleration vector is still equal to normal gravity i.e. $9.81 \ m/s^2$, however acceleration vector direction is perturbed by significant sideward component. Therefore $\mathbf{y}_a$ which defines Euclidean norm

between actual and predicted measurements, gives a better estimate of how good is the actual measurement. Since in general predicted measurement is never same as actual, we allow a threshold, in which covariance (uncertainty) of actual measurement is kept constant. Our process of adapting measurement covariance is depicted in Figure 4.3.
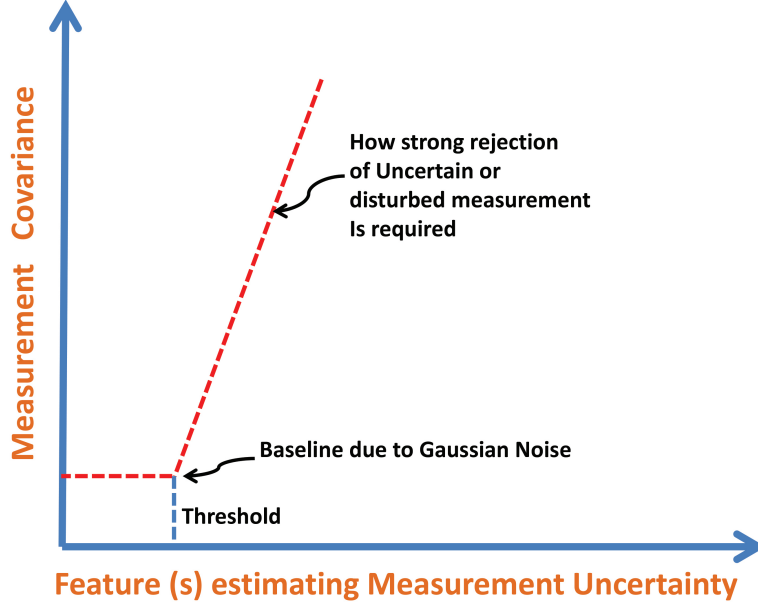

Figure 4.3.: Adaptive Covariance $\mathbf{R}_k^a$ and $\mathbf{R}_k^m$ for Robust EKF

The innovation $\mathbf{S}_k^a$ is now obtained from (4.14)

$$\mathbf{S}_k^a = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k^a \tag{4.14}$$

Again $\mathbf{H}_k$ is the Jacobian of measurement equation (4.11) evaluated at $\tilde{\mathbf{q}}_k$. We get Kalman gain for accelerometer measurement residual, apply error correction to state and update state covariance as follows

$$\mathbf{K}_k^a = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{S}_k^a)^{-1} \tag{4.15}$$

$$\mathbf{X}_k^a = \tilde{\mathbf{X}}_k + \mathbf{K}_k^a \cdot \mathbf{y}_a \tag{4.16}$$

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathbf{K}_k^a \mathbf{H}_k^T) \mathbf{P}_{k|k-1} \tag{4.17}$$

Since quaternion addition does not respect unit norm constraint, a renormalization after (4.16) is essential.

In order to calibrate accelerometer bias using rate gyro, we define loss function $L_{b_a} = ||\mathbf{y}_a(\mathbf{b}_a)||_{\mathbf{S}_k^{-1}}^2 + log \cdot det(\mathbf{S}_k^a)$, where $\mathbf{y}_a$ (defined earlier) is the residual of acceleration vector.

$\mathbf{S}_k^a$ is the innovation covariance from (4.14) and $|| \cdot ||_{\mathbf{S}_k^{-1}}^2$ represents 2-norm weighed by the inverse of covariance matrix. We perform iterative parameter update according to Gauss-Newton nonlinear optimization as follows

$$\boldsymbol{\theta}_{acc}^{k+1} = \boldsymbol{\theta}_{acc}^k - \lambda(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T L_{\boldsymbol{\theta}_{acc}}^k \tag{4.18}$$

where $\boldsymbol{\theta}_{acc} = [b_x^a, b_y^a, b_z^a]$, $J = \frac{\partial L_{\boldsymbol{\theta}_{acc}}^k}{\partial \boldsymbol{\theta}_{acc}^{k,i}}$ is the Jacobian matrix at time k of $L_{\boldsymbol{\theta}_{acc}}^k$ w.r.t each $\boldsymbol{\theta}_{acc}^{k,i}$ and λ_i is the learning rate.

While human motion causes only transient acceleration errors to appear in sensing of gravity vector, the sensing of local magnetic field suffers from lot of local variations and sometimes extreme disturbances [114]. Many works have reported success dealing with magnetic disturbances using threshold on applying magnetometer correction [50, 21, 35]. However, the results reported are often for short duration and obtained by using only rate gyro during periods of magnetometer disturbance. Slight inaccuracies in rate gyro calibration/bias can thus lead to error accumulating over long term, especially indoors. Another approach [18] focuses on modelling, predicting and incorporating magnetic disturbance vector itself. But it is very hard to accurately track magnetic disturbances in the environment. Therefore we have again applied a novel way to deal with it in our **EKF#2** implementation.

In order to remain consistent with other orientation determination algorithms (which assume local magnetic field as North or x-axis), we initialize with magnetic reference vector $\mathbf{m}_n = [\sqrt{m_x^2 + m_y^2}, 0, m_z]$, where m_x, m_y, m_z are obtained from $\mathbf{m}_{avg,0}$ initially. We normalize $\mathbf{m}_n$ i.e. $\hat{\mathbf{m}} = (\frac{\mathbf{m}_n}{|\mathbf{m}_n|})$, then perform measurement update steps in line with equivalent of equations (4.11-4.12) and (4.14-4.17) as done for accelerometer measurement, till convergence to initial orientation is achieved in few time steps. Our **EKF#1** implementation then continues with these steps, but assuming a constant R_k^m instead of adapting covariance by (4.12) . Except after initialization period, we obtain $\mathbf{m}_n = [\sqrt{m_x^2 + m_y^2}, 0, m_z]$ at each next step by using m_x, m_y, m_z obtained from $\mathbf{R}_b^n\mathbf{m}_{k-1}$, where $\mathbf{R}_b^n = (\mathbf{R}_n^b)^T$. This step is equivalent to a reset of magnetic reference vector to its best guess at time step k.

In our **EKF#2** approach, we eliminate the need to explicitly track or reset the reference magnetic vector using a novel approach. Once initialized, we no longer follow measurement equation analogous to (4.11). Instead we use the fact that change in orientation $q_{\Delta k}$ can rotate the magnetic vector m_{k-1} to m_k . Thus our new measurement equation becomes

$$\mathbf{m}_k^p = \mathbf{q}_{\Delta k} \oplus \mathbf{m}_{k-1} \oplus \mathbf{q}_{\Delta k}^* \tag{4.19}$$

Now measurement residual is computed using (4.19) as $\boldsymbol{y_m} = (\hat{\mathbf{m}} - \mathbf{R}_b^n\mathbf{m}_k^p)$. Since $\mathbf{m}_{k-1}$ and $\mathbf{m}_k$ are disturbed by almost same magnetic disturbance, it is implicitly taken care of. The measurement covariance $\mathbf{R}_k^m = diag([R_{ii}^{m,k}])$ is now adapted according to following novel criteria, if $|\bar{\boldsymbol{\omega}}| \leqslant \mu$,

$$R_{ii}^{m,k} = variance + scale * (max((|\boldsymbol{y_m}|) - threshold, 0)) \tag{4.20}$$

else,

$$R_{ii}^{m,k} = variance + scale * (max(|(\theta_k^{incl} - \theta_{k-1}^{incl}| - threshold, 0)) \tag{4.21}$$

where estimate of inclination angle θ^{incl} of local magnetic reference w.r.t gravity is obtained as $\theta_k^{incl} = \cos^{-1}(\frac{\mathbf{g}_k^p \cdot \bar{\mathbf{m}}_{cal}}{||\mathbf{g}_k^p|| \cdot ||\bar{\mathbf{m}}_{cal}||})$. The equation (4.20) employs same Euclidean norm between actual and predicted measurements as described earlier. However, in case of magnetometer, covariance adaptation using this equation is only used if $|\bar{\boldsymbol{\omega}}| \leqslant \mu$. At higher $|\bar{\boldsymbol{\omega}}|$, bad magnetometer readings cannot be discriminated from bad predicted measurements using Euclidean norm of residual. Due to gyro calibration errors, higher $|\bar{\boldsymbol{\omega}}|$ lead to more prediction error. Hence, we employ a more robust criteria of checking for good magnetometer readings using inclination angle of $\mathbf{m}_{k-1}$ and $\mathbf{m}_k$.

The equations analogous to (4.14-4.17) are used for magnetometer based update to obtain innovation covariance, Kalman gain and apply error correction to state. The calibration of magnetometer is iteratively improved by minimizing a loss function defined using magnetometer measurement residual $\mathbf{y}_m$, and its innovation covariance $\mathbf{S}_k^m$, as analogous to accelerometer case using (4.18).

After we obtain measurement corrected states $\mathbf{X}_k^a$ and $\mathbf{X}_k^m$ and corresponding covariance matrices $\mathbf{P}_{k|k}^a$ and $\mathbf{P}_{k|k}^m$, we simply perform the final fusion step as

$$\mathbf{P}_{k|k} = ((\mathbf{P}_{k|k}^a)^{-1} + (\mathbf{P}_{k|k}^m)^{-1})^{-1} \tag{4.22}$$

$$\mathbf{X}_{k|k} = \mathbf{P}_{k|k}[(\mathbf{P}_{k|k}^a)^{-1}\mathbf{X}_k^a + (\mathbf{P}_{k|k}^m)^{-1}\mathbf{X}_k^m] \tag{4.23}$$

The state $\mathbf{X}_{k|k}$ obtained from (4.23) is renormalized to get unit quaternion $\mathbf{q}_k$

4.6. Experimental Protocol

In order to test the performance of our MIMU based motion capture hardware (EasyIMP/BIONIC) and embedded sensor fusion and online calibration algorithm, we performed a series of experiments. In our first experiment, we performed evaluation of 1 KHz high sampling rate of our MIMU system (BIONIC) versus 100Hz sampling to signify the differences and underline the importance of higher rate sampling for highly dynamic body tracking and health monitoring. The data of a single sensor was segmented from 7x sensors data stream which were mounted on a stable platform and moved by right hand of the subject.

In second experiment, The real-time results of our **EKF#2** implemented on Edison module of MIMU tracking system (EasyIMP) were compared with ground truth obtained from sub-millimeter accurate OptiTrack system. The hand-eye calibration for the two systems has been performed using procedure in [99]. The rigid body configuration with placement of optical markers is shown in Figure 4.4. A single sensor with Edison module was used for this experiment.

We initially started from a rest position, carried out first set of sudden movement for computing temporal alignment between two systems. The first set of movements was also employed for hand-eye calibration (spatial frame alignment). Then various set of movements were performed for short-to-moderate duration with intervening rest positions to segment out these movements. The data obtained from two systems (after temporal and spatial frame alignment) were then

Figure 4.4.: Rigid Body setup of MIMU with Infrared Optical Markers for Combined OptiTrack and IMU Tracking

compared. Both fast and slow movements with different degree of amplitude were performed in this experiment. The movements with $max(abs(|\mathbf{a}_{cal}| - g_{avg})) < 0.5g$ were assumed slow, else these were assumed fast.

In order to evaluate the robustness of our sensor fusion algorithm for orientation estimation against body accelerations and magnetic disturbances, we further performed experiment 3 and 4 respectively. In particular performance of our **EKF#1** (baseline) implementation against **EKF#2** were compared. For experiment 3, we again obtained the data of a single sensor segmented from 7x sensors stream, which was mounted on a stable platform and moved by right hand of the subject. Various arbitrary gestures with varying speed were performed, similar to experiment 1. These movement periods were interjected with static rest positions. The magnitude profile of acceleration during these movements is shown in Figure 4.10 (section 4.7). Apart from body vibrations (tremors), significant fast accelerations can be observed to occur during change from one rest position to another in the norm of observed acceleration. We chose intervening rest positions after movement periods, in order that filter settling response can be easily observed in the plots.

For experiment 4, we chose the same set-up, except now we introduced hard and soft magnetic disturbance. The magnitude profile of these magnetic disturbances is easily observable in Figure 4.11 (section 4.7). We introduce these disturbances during rest periods to easily observe the resulting filter behavior. Soft iron disturbances are introduced by bringing in a cubic ferrous block of $5cm \times 5cm \times 5cm$ close to MIMU and are discernible by slight change of magnetic norm in Figure 4.11. For hard iron disturbances we used stack of small Neodymium magnets with varying strength (by adding to or removing from stack) in the range $100 - 500\mu T$ (our observed ambient indoor field strength was in the range $25 - 88\mu T$ and $25\mu T$ is mapped as unit norm in plot). Hard iron disturbance can be observed as sharp spike in Figure 4.11.

In our last experiment, we performed evaluation of the online calibration of our MIMU system to signify the variation of parameters with time and underline the importance of performing

it online. The data of a single sensor was obtained mounted on a stable platform. The stable positions allowed us to obtain ground truth (GT) value of rate gyro biases (Figure 4.11) for comparison.

4.7. Results and Discussion

4.7.1. Advantage of High Rate Inertial Sensors Data for Sensor Fusion

The results of our first experiment are depicted in Figure 4.5 and 4.6 for accelerometer and rate gyro respectively for 100 Hz and 1000 Hz sample rate. It is important to observe in Figure 4.5 that the high frequency features present at 1000 Hz are quite under-sampled at 100 Hz in acceleration data. Therefore high frequency features which are clinically significant for diagnosis may not be observed at 100 Hz.
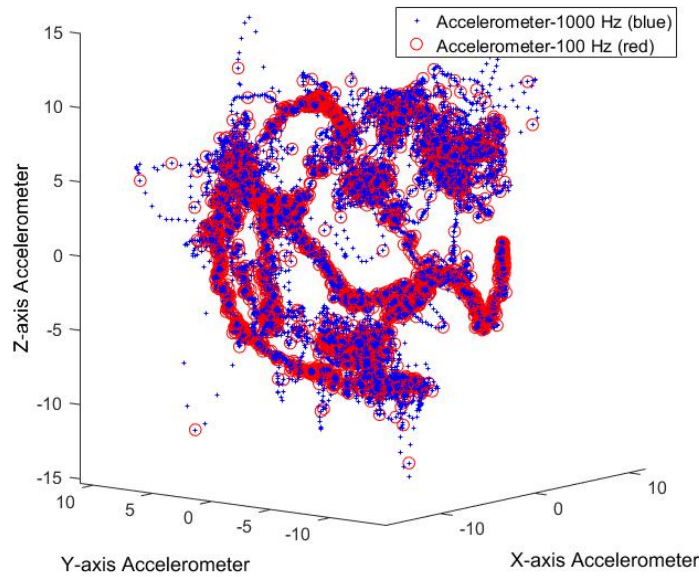
Figure 4.5.: Accelerometer sampling at 100 Hz leads to sparse features in region of fast motion/body tremors etc.

Apparently in Figure 4.6, rate gyro readings at 100 Hz appear good approximation of trajectory at 1000 Hz. However since in sensor fusion (section 4.5.4), the Equation (4.8) represents only a first order integration model of (4.4), we observe that the assumption regarding constant angular acceleration is not valid between two samples of rate gyro measurements at 100Hz. Hence during periods of fast motion, orientation integration error increases.

Since the resulting integration error is cumulative, (4.8) or any other approximate discrete model for integration of (4.4) provides better accuracy with higher data rate. Even with sensor

fusion, during periods (or time steps) of bad magnetometer readings, our heading (yaw) estimates depend largely on rate gyro integration. Thus depending on accuracy and application, we must choose an appropriate rate of data acquisition from accelerometer and rate gyro.
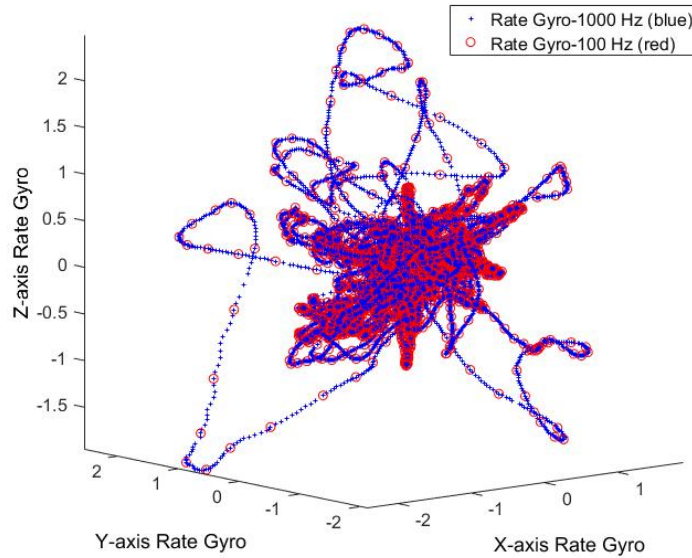

Figure 4.6.: Sampling of rate gyro at 100 Hz, while trajectory at 1000 Hz shows angular acceleration is not constant

4.7.2. Comparison of MIMU tracking with OptiTrack System

In second experiment, we ran our **EKF#2** implemented on Edison® module of MIMU tracking system and the ground truth was simultaneously obtained from submillimeter accurate OptiTrack® system. The error in Euler angles w.r.t ground truth over multiple test sequences data is summarized in Table 4.1.

Euler Angle (ZXY)	At Rest	Slow Movements	Fast Movements
Yaw Error	$\pm 0.01°$	$\pm 0.5°$	$\pm 5°$
Pitch Error	$\pm 0.0025°$	$\pm 0.25°$	$\pm 4.25°$
Roll Error	$\pm 0.15°$	$\pm 1.25°$	$\pm 5°$
Note: 95% of errors lie within reported limits.			

Table 4.1.: Error between OptiTrack and MIMU tracking @100Hz Sampling rate

It shows that the error increases significantly for fast movements, but 95% of time the error remains in the range of $\pm 5°$. This finding is in agreement with [113] , which reports that accuracy of IMU tracking depends upon both the amplitude and frequency of movement. Our

error performance is also better than that reported for Xsens MTw sensors, which only report static and dynamic RMS accuracy instead of 95% range. Xsens MTw series static accuracy for pitch/roll is reported 0.5° RMS and for yaw is 1.0° RMS, dynamic accuracy for pitch/roll is reported 0.75° RMS and for yaw is 1.5° RMS (RMS is root mean square). Compared with Xsens MTw series, we also report our accuracy for high rate dynamic movements of large amplitude (column-3 in Table 4.1), where filtering performance is crucial.
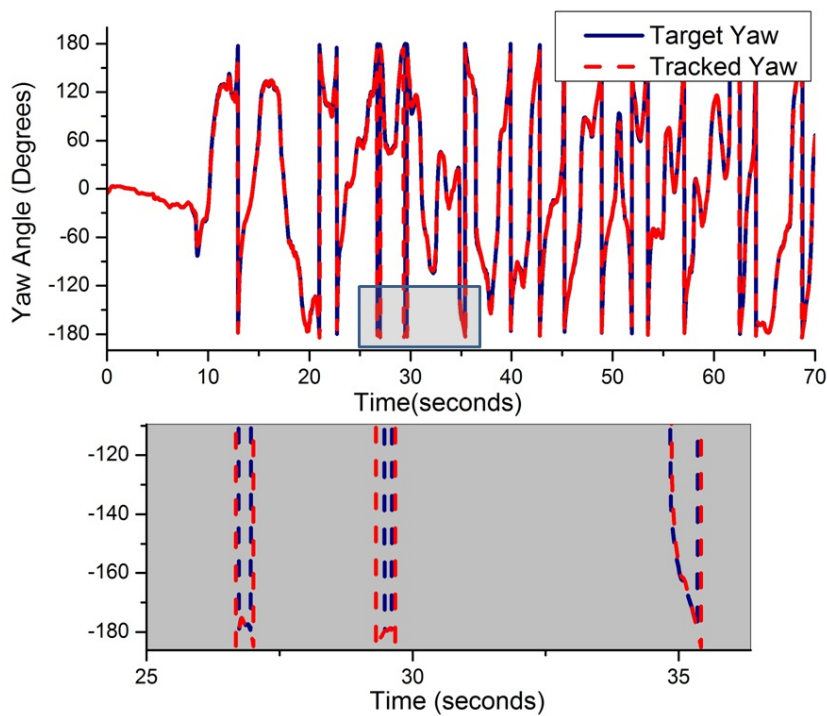
Figure 4.7.: Comparison of OptiTrack® and MIMU Tracking (Yaw Angle- ZXY order)

Figure 4.7-4.9 display our results for yaw, pitch and roll angles (ZXY order) tracked via both OptiTrack® and MIMU based systems for test duration of 10 min. Only a part of sequence comprising of fast movements (comprising of 70 seconds) is shown for better visualization of amplitude changes. The inset plot in each figure tries to depict the error between two systems more clearly.
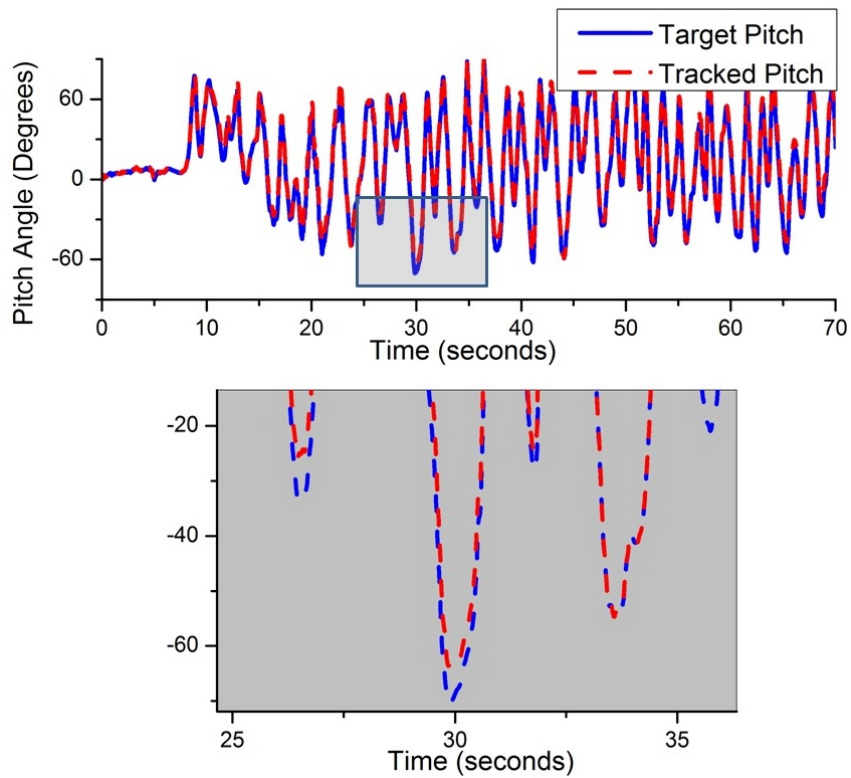
Figure 4.8.: Comparison of OptiTrack® and MIMU Tracking (Pitch Angle- ZXY order)
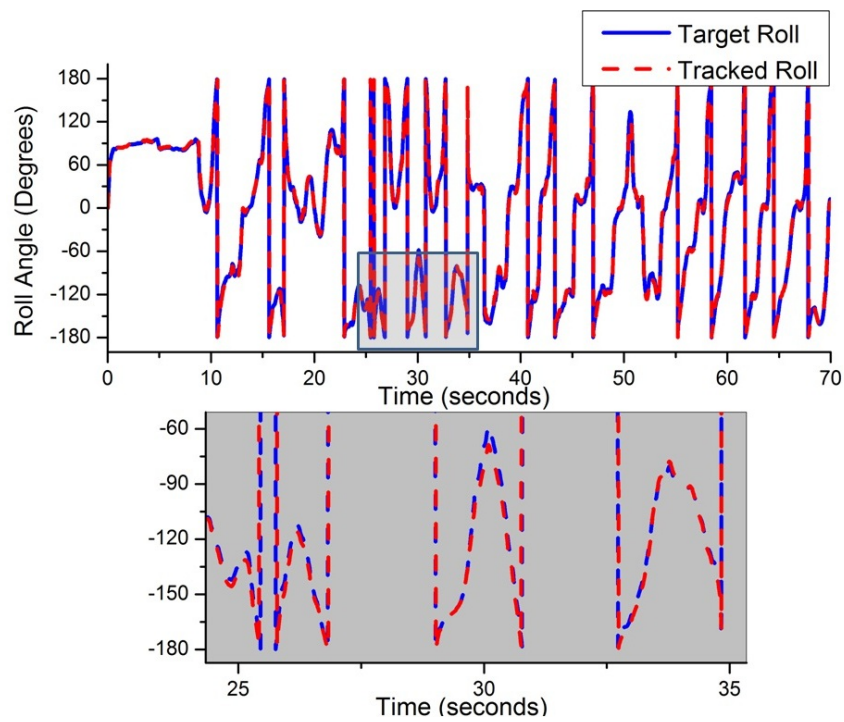

Figure 4.9.: Comparison of OptiTrack® and MIMU Tracking (Roll Angle- ZXY order)

4.7.3. Robustness against Body Acceleration

The accurate real-time performance of **EKF#2** obtained in experiment 2, lies in strong rejection of acceleration induced errors and magnetic disturbances. It is compared with **EKF#1** algorithm in experiment 3 and 4 to elaborate this feature further. Figure 4.10 describes the performance comparison of **EKF#2** with **EKF#1** in presence of strong accelerations. Only pitch angle output is shown, as both pitch/roll are affected by the acceleration in a similar manner. Also for ease of visualization, **EKF#1** and **EKF#2** outputs are shown with an offset (these overlap in real). It can be seen that pitch angle output of **EKF#1** is affected by even small accelerations (even due to tremors/vibrations) while that of **EKF#2** is smooth. Also fast convergence of filter algorithm can still be seen despite strong rejection of acceleration, during pitch angle change to new static pose.
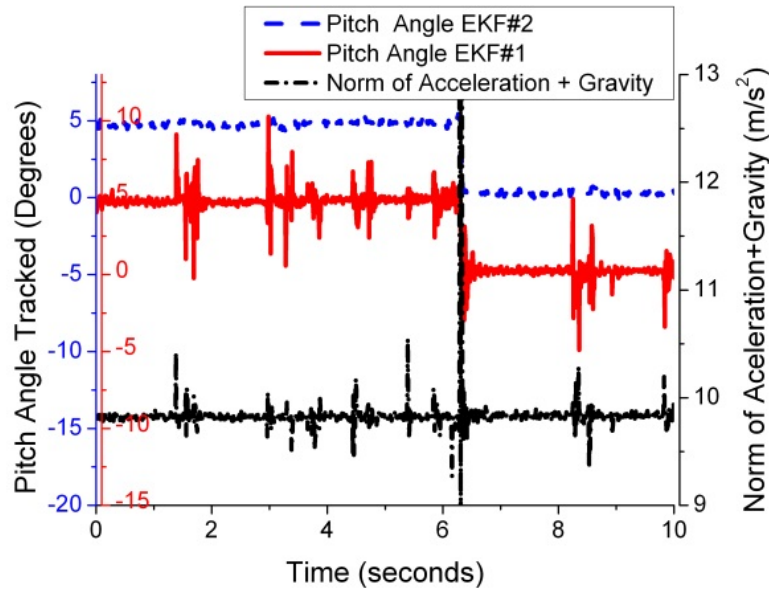

Figure 4.10.: Performance Comparison of **EKF#1** and **#2** (outputs shown with offset), Rejection of Acceleration Errors

4.7.4. Performance in presence of Magnetic disturbances

A strong magnetic disturbance was introduced in experiment 4, as shown by changes in the norm of local magnetic field in Figure 4.11 (soft iron disturbances appear as smaller spikes in norm value, while hard iron disturbance induce a sharp spike in norm of magnetic field). Since the magnetic disturbance predominantly affect the heading (yaw) angle, therefore only a plot of yaw angle is shown. Again for ease of visualization, **EKF#1** and **EKF#2** outputs are shown with an offset (these overlap in real). The yaw angle estimation by **EKF#1** is not robust to these magnetic variations, while **EKF#2** has been shown to be less perturbed, despite magnetic disturbance.
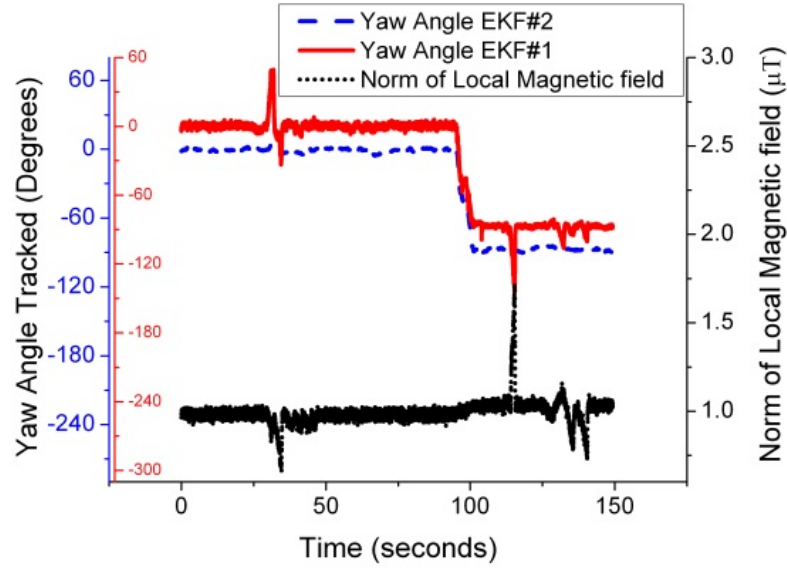
Figure 4.11.: Performance Comparison of **EKF#1** and **#2** (outputs shown with offset), Rejection of Magnetic Disturbances

4.7.5. Performance of Online Calibration

The results of our first experiment are depicted in Figure 4.12, 4.13 and 4.14 for rate gyro biases, accelerometer and magnetometer calibration respectively. Figure 4.12 clearly shows that ground truth (GT) for rate gyro biases (obtained offline by segmenting static periods manually) vary over time. Our online estimate of biases is close to actual values. We also notice that quasi-static instants occur often during human motion for this scheme to work. The experiment has been run for 30 minutes (1800 sec). The maximum deviation error of estimated gyro bias observed during this time is given in Table 4.2.

Maximum Deviation	With Online Calibration	Without Online Calibration
Gyro Bias Error	$\pm 10^{-4}$ rad/s	$\sim 10^{-2}$ rad/s
Accel Norm Deviation	$\pm 0.08\%$	$\pm 2.3\%$
Mag Norm Deviation	$\pm 0.1\%$	$\pm 3.4\%$

Table 4.2.: Error With and Without Online Calibration

Similarly in Figure 4.13, we obtain accelerometer bias iteratively using (4.18) and evaluate calibration for 11 values of quasi-static accelerometer samples. Our estimated bias improves the norm of accelerometer readings, which is a criterion for good calibration. The individual axes of accelerometer show only slight change before and after calibration. However, it still represents significant improvement in terms of pitch/roll errors (since these angles are obtained as a tangent function of the ratio of different accelerometer axes). All the values of acceleration

(and its norm) are shown in arbitrary units (a.u.) referenced to local magnitude of acceleration due to gravity.
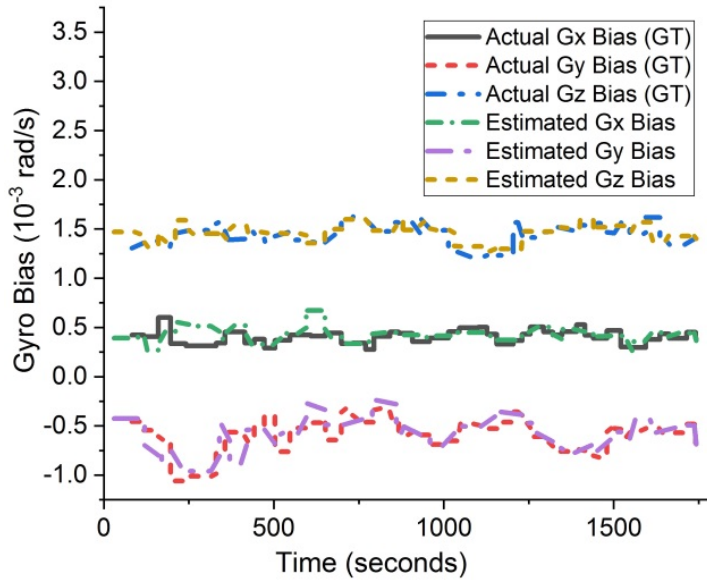

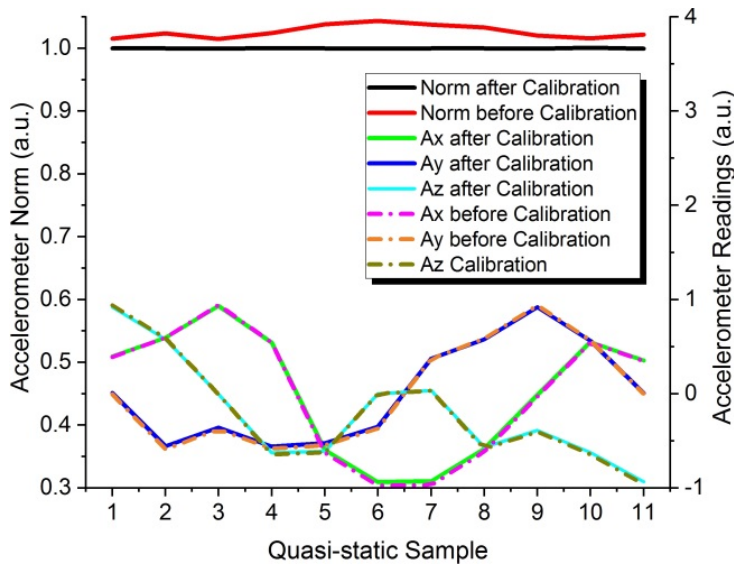Figure 4.12.: Estimated and Actual Rate Gyro Biases (shown in $\sim 10^{-3}$ rad/s)

Figure 4.13.: Online Calibration of Accelerometer (shown in normalized arbitrary units a.u.)

Figure 4.14 displays the results of online calibration of magnetometer. Again 11 values of distorted magnetometer readings are shown for comparison. Using (4.18) for magnetometer bias, we are able to improve the norm as well as correct bias/hard iron distortions. As it can be

seen in Figure 4.14 that the magnetometer readings of all 3-axis were distorted before calibration which improved after calibration. All the values of magnetometer (and its norm) are shown in arbitrary units (a.u.) referenced to local magnitude of magnetic field.
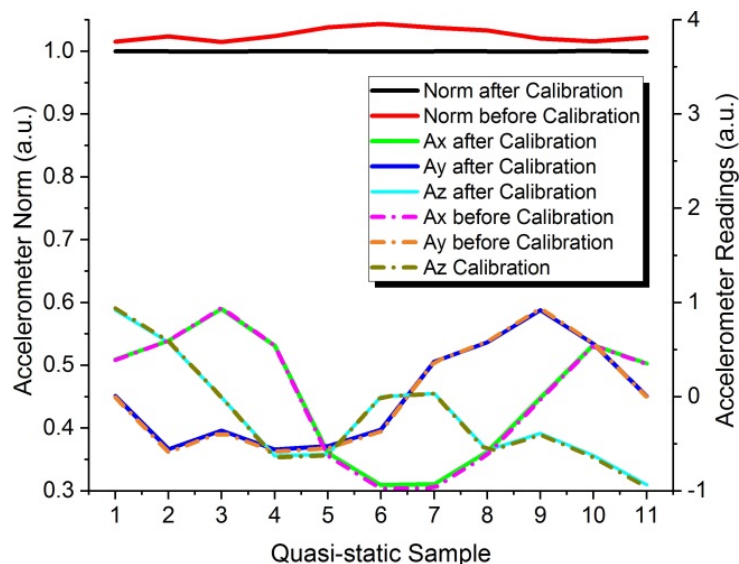

Figure 4.14.: Online Calibration of Magnetometer (shown in normalized arbitrary units a.u.)

Table 4.2 outlines the maximum deviation from local norm for accelerometer and magnetometer, both with and without online calibration during the 30 minutes experiment.

4.8. Conclusion

In this chapter, we presented an online self-calibration algorithm based on intelligent sensor fusion for MIMU based human motion capture. We implemented this distributed algorithm using improved hardware on a low power high rate wearable MIMU based body tracking system, both on the hub and sensor nodes. The high rate of inertial data provides intrinsic accuracy to sensor fusion as well as capture high frequency features for clinical diagnostics and biomechanical analysis. The implementation is based on real-time embedded EKF based algorithm with low computational latency. It uses quasi-static conditions for precise calibration of rate gyro biases at sensor node. Thereafter, an intelligent hybrid sensor fusion and calibration algorithm is run at hub. Using the innovation residuals in EKF, it performs online calibration of accelerometer and magnetometer. The sensor fusion also performs orientation estimation using novel adaptive sensors covariance, which makes it robust to both body accelerations and magnetic disturbances. To the best of our knowledge, our work is the first to address online calibration of all 3-sensors in MIMU in one algorithm.

5. Inertial Motion Capture with Joint Angle Drift Correction

In this chapter we describe the integration of our flexible textile wire based MIMU tracking system (EasyIMP) with a body worn suit and a method to calibrate the extrinsic sensor-to-body segment orientations for inertial motion capture. The precise sensor-to-segment calibration is essential to obtain accurate joint angles for 3D human pose. This is also called anatomical calibration. Our method comprises of an initial static I-pose and a functional calibration step. The former also performs the alignment of the global reference frame seen by each MIMU. Due to residual errors, the global frame w.r.t which each MIMU gives 3D orientation, might be slightly misaligned from sensor-to-sensor, even after intrinsic calibration and sensor fusion are performed (covered in chapters 3 and 4). We demonstrate that a novel correction term in our anatomical formulation takes care of this initial misalignment. Also uncertainties accumulate during motion capture arising from either sensor degradation or attachment artifacts, the same correction term is updated and performs online correction of drift in individual joint angles as an orientation offset. This offset correction for joint angle is performed automatically when the limb or extended torso are in neutral quasi-static pose and this condition is judged by using accelerometers. In comparison to state-of-art anatomical calibration methods, our approach is more robust and avoids drift or error accumulation during inertial motion capture. The work presented in this chapter is based on our peer reviewed published research [12].

5.1. Background

The ambulatory motion capture and gait analysis using wearable MEMS based magnetic-inertial measurement units (MIMUs) is challenging. The multi-sensor fusion and intrinsic calibration is performed to obtain precise 3D orientation of MIMUs. To obtain complete 3D human pose, MIMUs are placed on body segments. Therefore, precise registration of each MIMU w.r.t anatomical body frame is also required. An effective and robust anatomical (sensor-to-segment) calibration procedure is therefore a pre-requisite in inertial motion capture (i-Mocap) pipeline.

From an anatomical standpoint, the human body is considered multi-body system comprising of rigid body segments articulated at joints as shown in Figure 5.1. The kinematic of human body is fully described by the position of a root (normally a midpoint on hip bone below spine is assumed reference point or root), its 3D rotation in global reference frame and the degrees of freedom of each joint in the articulated body. The later are called joint angles. In an inertial motion capture these joint angles are obtained as relative orientation between two anatomical frames which are defined on adjacent rigid body segments, as shown in Figure 5.2.
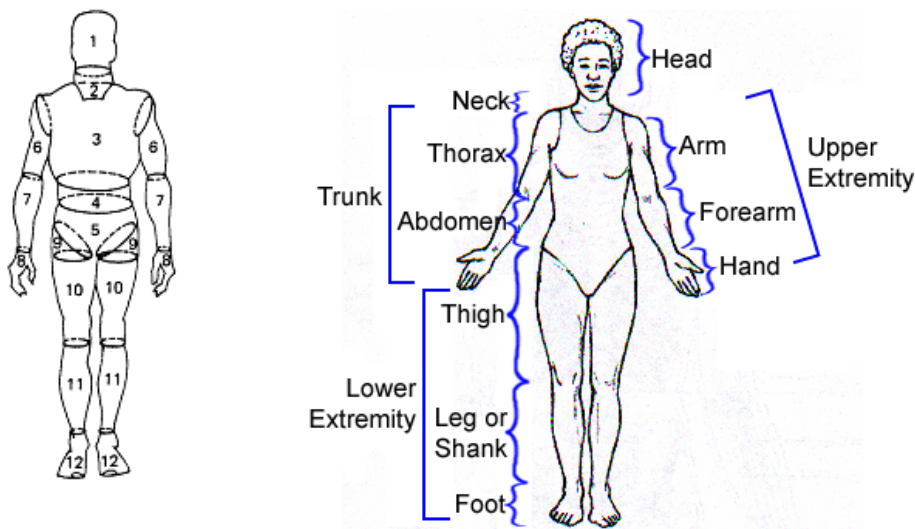
Figure 5.1.: Human body as articulation of rigid segments at its joints

In general the anatomical reference frame of a body segment is not always aligned with corresponding MIMU body frame. The 3D orientation offset between the two is assumed fixed if MIMU is rigidly attached to the body segment, and is called sensor-to-segment (S2S) offset. This orientation offset must be accurately registered in an anatomical calibration so that the sensor measured orientation can be related to the corresponding anatomical frame. The joint angles is then determined between pair of adjacent anatomical frames of body segments. This pipeline is depicted in Figure 5.2 and would be further explained analytically in Section 5.5.

The challenge of MEMS based inertial motion capture is two-folds. First the sensors show degraded performance when run for long time, especially indoors. This is due to the fact that assumption of no acceleration except gravity and homogenous magnetic field no longer holds, when the motion is being performed. The rate gyro is used to complement the accelerometer/ magnetometer for orientation estimation. Although intrinsic calibration and sensor fusion (covered in chapters 3 and 4) is performed but the integration of rate gyro residual bias as well as noise eventually causes the sensor fusion estimates to drift gradually. The errors in heading angle or yaw is particular significant due to persistent nature of magnetic inhomogeneity in the environment. The second challenge arises from the extrinsic sensor-to-segment calibration. Practically, there is always a residual error in the anatomical calibration performed and it also slightly changes over time due to slippage and other motion artifacts. This ultimately results in inaccurate and drifting joint angle estimates between body segments that would require a method of determination and correction.

Anatomical approaches in the past [124, 68, 67, 130] have dealt to overcome the above mentioned limitations. However in order to account for any residual drift of joint angles, we propose a novel correction term in our anatomical formulation that performs online correction of drift in individual joint angles and updates it as an orientation offset. This offset correction for joint
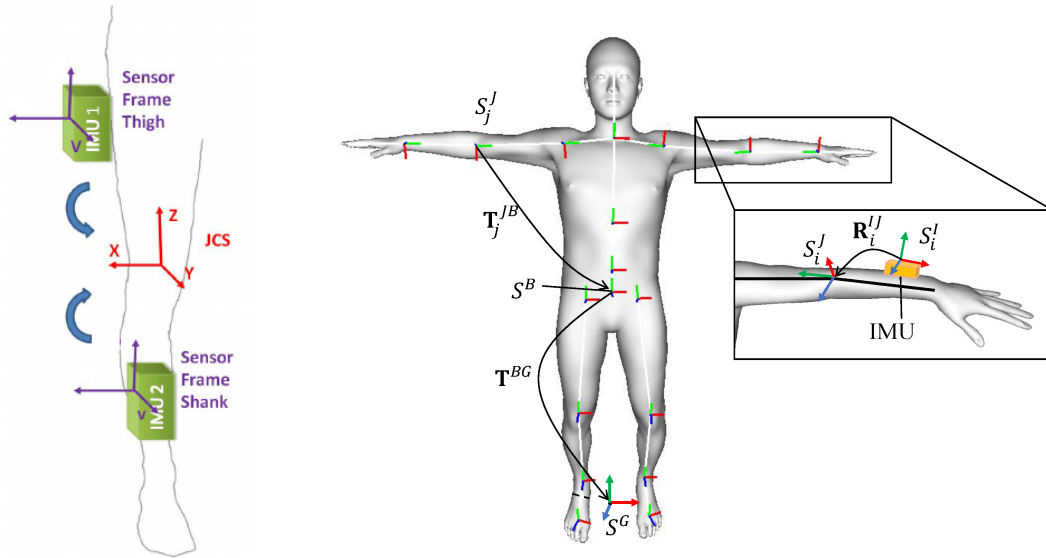
Figure 5.2.: Joint angle computation in Joint Coordinate System (JCS) and Definition of different Reference Frames for Anatomical Calibration [55]. The sensor-to-segment orientation $\mathbf{R}^{IJ}$ is the offset registration between inertial sensor frame $\mathbf{S}^I$ and joint frame $\mathbf{S}^J$

angle is performed automatically when the limb or extended torso are in neutral quasi-static pose and this condition is judged using accelerometers. Overall our approach achieves precise orientation estimates in highly dynamic conditions and inhomogeneous magnetic fields (chapters 3 and 4) as well as avoids drift or error accumulation due to residual and dynamic artifacts during inertial motion capture (chapter 5).

5.2. Contribution

Our approach propose the static and dynamic error correction in anatomical calibration performed during inertial motion capture (i-Mocap) for joint angle estimation from adjacent MIMUs. It is motivated by the observation that different body segments go to neutral or zero pose during the normal movement or gait. The main contributions covered in this chapter are,

1. A novel anatomical model is assumed in which an offset term comprising of all residual errors is introduced. It is determined when joint angle is zero or at neutral I-pose.

2. We perform anatomical calibration for i-Mocap using both initial static I-pose and functional movement. The former also performs the alignment of the global reference frame seen by each MIMU during initial step. Due to residual errors, the reference frame may be misaligned from sensor-to-sensor.

3. During i-Mocap, the offset term is updated to reset the corresponding joint angle to zero when neutral condition is encountered again. This is judged using accelerometers in a quasi-static pose. This online correction prevents any error accumulation over time in our approach.

5.3. Organization of Chapter

This chapter is organized in eight main sections. In the first two sections we have covered the background and our main contribution to the area of anatomical calibration for i-Mocap while addressing residual error and drift correction in joint angle estimation. The aim is robust estimation of 3D joint angles from MIMUs placed on adjacent body segments. These MIMUs are affected by residual errors of intrinsic calibration as well as gradual drift errors of sensor fusion arising from body accelerations and inhomogeneous magnetic environment. In section 5.4, we describe the integration of our flexible textile wire EasyIMP inertial body tracking system with a body worn close fit suit and pants. In section 5.5, we present our methodology of inertial motion capture (i-Mocap), comprising of anatomical model, its calibration and online joint angle drift correction based on novel offset term in our model. In section 5.6, we explain the protocol of our experiments to validate the performance of proposed anatomical calibration and online joint angle drift correction. In section 5.7, we discuss the results of our experiments and demonstrate that despite magnetic inhomogeneity and body accelerations, the full body tracking using our anatomical algorithm shows no performance degradation or drift in joint angles over time.

5.4. Body Integration

Our EasyIMP hardware (section 3.4) is based on harness based design, in with MIMU nodes are connected with an Intel Edison® hub and a battery pack using textile wires (Figure 3.1). The flexibility of textile harness allows easy routing. The advantage of using textile wires for both data and power sharing is that the system can be conveniently integrated with body worn suit (upper body) and/or pants (lower body) as shown in Figure 5.3. The body-fit suit and pant is made of stretchable fabric. Enough length of flexible wire between sensor nodes allows it to stretch and twist while staying integrated with body-fit suit/pant, while nodes are located on different body parts. Thus any maneuver or extreme activity can be performed without hindrance. The weight of sensor nodes and hub is almost negligible and a small weight contribution comes from battery pack itself.

The package size of miniaturized sensing nodes is $25 \times 20 \times 5\ mm^3$, making the system almost non-invasive (for comparison, Xsens MTw wireless series has MIMU size of $47 \times 30 \times 13\ mm^3$). There is sufficient free memory (up to 1.0 GB) on Intel Edison module to record an outdoor session. The calibration information stays on plug-in sensor node and an EPROM module stores this calibration data. The data transmission happens through a differential I^2C bus that

can run up to 400 KHz. The data acquisition from the individual MIMU sensors take place at 100 Hz sampling rate in this baseline system.

Figure 5.3.: Body Integration of MIMU Motion Capture System

5.5. Proposed Methodology

5.5.1. Anatomical Calibration

In the following algorithm, we assume that the inertial sensor body fixed frame $\mathbf{S}^I$ has its origin at the center of 3-axis sensors triad and its x, y, z axis are orthogonal and aligned with that of MIMU package. Therefore we may also denote $\mathbf{S}^I$ as $\mathbf{S}^S$. The global navigation frame of reference $\mathbf{S}^G$ has its origin at the initial position (0,0,0) and its x-axis is aligned with local magnetic north while y-axis is opposite to gravity (and z-axis is orthogonal). The joint coordinate frame (JCS), $\mathbf{S}^J$ is fixed at the anatomical joint and oriented according to bony landmarks definition of each joint according to International Society of Biomechanics (ISB) [147, 148]. We define $\mathbf{S}^L$ to be a $\mathbf{S}^J$ frame translated to the midpoint of distal body segment (limb) from the respective joint position, so that each joint angle is then a 3D rotation between two such adjacent frames $\mathbf{S}^{L1}$ and $\mathbf{S}^{L2}$.

At any time step k, the orientation $\mathbf{R}_L^G$ of a given limb (segment) of body w.r.t. global earth fixed frame can be obtained from orientation measured by MIMU sensor $\mathbf{R}_S^G$ rotated by a fixed rotation between sensor to limb (segment) $\mathbf{R}_L^S$.

$$\mathbf{R}_L^G = \mathbf{R}_S^G . \mathbf{R}_L^S \tag{5.1}$$

The rotation $\mathbf{R}_L^S$ is due to sensor placement w.r.t. body segment or limb frame of reference. If the sensor is rigidly attached to body segment, this rotation remains nearly fixed during motion/ articulation and is also termed as sensor-to-segment (S2S) calibration. Note that the global earth fixed frame $\mathbf{S}^G$ assumed by each sensor may not be same due to residual errors in accelerometer and magnetometer calibration. The joint angle $\mathbf{R}_{L1}^{L2}$ between any two limbs or body segments is therefore calculated as

$$\mathbf{R}_{L1}^{L2} = \mathbf{R}_{L1}^{G1}.\mathbf{R}_{G1}^{G2}.\mathbf{R}_{G2}^{L2} \tag{5.2}$$

where the novel term $\mathbf{R}_{G1}^{G2}$ is introduced in our full body calibration framework to signify an offset $\mathbf{R}_{offset}$ arising from residual errors and uncertainties and is updated over time, as will be shown in next section. The complete sensor-to-segment (S2S) calibration in our framework is obtained in two steps: (1) a static zero reference pose and (2) functional or dynamic calibration.
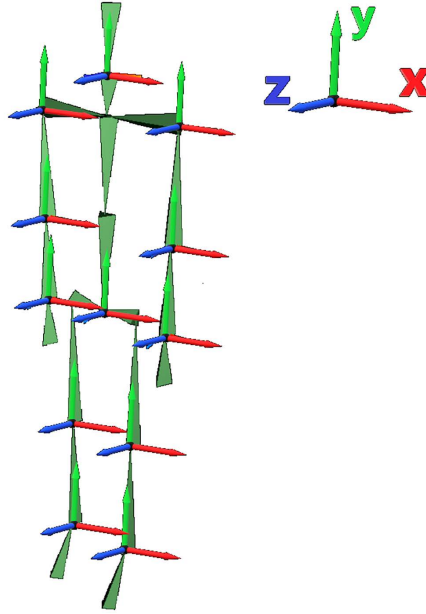

Figure 5.4.: Skeleton Coordinate System in I-Pose

At step-1 the subject performs a static zero reference pose for 1-2 seconds. In reference pose, he stands upright with hands by the sides and pointing down. By definition (as shown in Figure 5.4), the limb frame $\mathbf{S}^L$ is aligned with global earth fixed frame $\mathbf{S}^G$ while in zero or reference pose. The average normalized accelerometer reading of each MIMU sensor $\widehat{\boldsymbol{a}}^S$ (on each limb) is calculated. It defines the unit vector $\widehat{\boldsymbol{y}}$ or y-axis of sensor in respective limb frame $\mathbf{S}^L$, since both frames $\mathbf{S}^G$ and $\mathbf{S}^L$ are aligned.

$$\widehat{\boldsymbol{y}}^L = \widehat{\boldsymbol{y}}^G = (\widehat{\boldsymbol{a}}^S)_{avg} \tag{5.3}$$

In step-2, the subject is asked to perform the functional or dynamic motion of the limb. For

each of his leg, he performs a rotation of straight leg at hip joint from a flexion to extension. Similarly for each arm, a rotation of straight arm is performed at shoulder joint from back to front. For head and extended torso, the subject simply bends forward while head is in-line with spine. These rotations must take place in a plane around a fixed axis of rotation. We obtain unit vector $\widehat{\boldsymbol{n}}$ that defines the axis of rotation from average normalized rate gyro readings of each MIMU sensor.

$$\widehat{\boldsymbol{n}}^L = \widehat{\boldsymbol{n}}^G = (\widehat{\boldsymbol{\omega}}^S)_{avg} \tag{5.4}$$

Due to error of performance, this axis of rotation $\widehat{\boldsymbol{n}}$ may not be orthogonal to unit vector y obtained from (5.3) so we also perform orthogonalization using cross-product as follows

$$\widehat{\boldsymbol{z}}^L = \widehat{\boldsymbol{n}}^L \times \widehat{\boldsymbol{y}}^L \tag{5.5}$$

$$\widehat{\boldsymbol{x}}^L = \widehat{\boldsymbol{y}}^L \times \widehat{\boldsymbol{z}}^L \tag{5.6}$$

In the end we obtain a set of unit vectors $\widehat{\boldsymbol{x}}$, $\widehat{\boldsymbol{y}}$, and $\widehat{\boldsymbol{z}}$ that defines the complete orientation of IMU sensor w.r.t the reference frame of limb or body segment it is attached to, i.e.

$$\mathbf{R}_L^S = [\widehat{\boldsymbol{x}}^L \widehat{\boldsymbol{y}}^L \widehat{\boldsymbol{z}}^L] \tag{5.7}$$

5.5.2. Joint Angle Drift Correction

As explained in (5.2) due to residual calibration error, there remains an offset term $\mathbf{R}_{G1}^{G2}$ between the global frames assumed by each pair of MIMU sensor. We decompose this for each joint angle in two parts as $\mathbf{R}_{G1}^{G2} = \mathbf{R}_G^{G2}.R_{G1}^G$. where $\mathbf{S}^G$ is assumed as reference frame of MIMU mounted at pelvis. We obtain the corrected $\mathbf{R}_{Li}^G = \mathbf{R}_{Gi}^G.\mathbf{R}_{Li}^{Gi}$ for each limb (segment), by calculating the offset $\mathbf{R}_{offset} = \mathbf{R}_{Gi}^G$. This is done after static and dynamic (functional) calibration steps. We simply obtain offset $\mathbf{R}_{offset}$ as follows at zero pose

$$\mathbf{R}_{offset} = \mathbf{R}_{Gi}^G = \mathbf{R}_{pelvis}^G.\mathbf{R}_{Gi}^{Li} \tag{5.8}$$

Since at zero or reference pose, both the pelvis reference frame $\mathbf{S}^{pelvis}$ and all limb (segment) reference frames $\mathbf{S}^{Li}$ are aligned by definition, (5.8) simply provides us with required offset. The same offset is then regularly update using (5.8), as we encounter quasi-static conditions during articulation/ motion, i.e., when all the body segments return to same alignment as in zero or reference frame. This quasi-static zero pose condition is detected using accelerometer. We have initial $(\widehat{\boldsymbol{a}}^s)_{avg,0}$ from (5.3) for IMU on each body segment during static zero reference pose. During motion capture, we calculate a weighted moving average of accelerometer unit vectors at each time step k as follows

$$(\widehat{\boldsymbol{a}}^s)_{avg,k} = \frac{\frac{([\sum_{k-n-1}^{k-1} \widehat{\boldsymbol{a}}^s])}{n}.(n-1) + (\widehat{\boldsymbol{a}}^s)_k}{n} \tag{5.9}$$

Where we maintain a moving window of size n and $(\widehat{a}^s_k)$ is the new normalized accelerometer unit vector at time step k. For quasi-static pose we evaluate the condition on all body segments such that

$$||(\widehat{a}^s)_{avg,k} - (\widehat{a}^s)_{avg,0}|| < \epsilon_0 \tag{5.10}$$

In (5.10), ϵ_0 is a threshold defining Euclidean distance. We then perform an offset correction according to (5.8), if in addition to (5.10) being true, at least one of joints shows that its yaw angle (Euler angle Y in XYZ order by our definition) differs from zero by a threshold of $\pm\epsilon_1$ degrees. This ensures that a small posture deviations from zero pose at rest do not lead to complete reset to initial zero pose.

5.6. Experimental Evaluation

In chapter 4 we have already evaluated the single MIMU orientation obtained using our sensor fusion algorithm for accuracy and robust orientation. This is done in a series of experiments. First the real-time orientation estimation of our sensor fusion algorithm implemented on MIMU was compared with ground truth obtained from submillimeter accurate OptiTrack system, after temporal and spatial frame alignment. Then we also checked the performance of sensor fusion under body acceleration and magnetic interference.

The error and long term drift is minimal using our sensor fusion and calibration algorithms. However residual error remains due to uncertainty and may cause full body tracking error to grow with time. Our full body calibration and joint angle estimation algorithm presented in this chapter addresses it.

Therefore we implemented an experiment to evaluate the end-to-end performance of our human motion capture. In particular the robustness of our approach in presence of the body accelerations and magnetic disturbances was checked during gait analysis. We performed gait capture of normal human subject indoors in a free walk. The walk was performed in a corridor where magnetic disturbances were paramount and vary strongly with position. The subject walked at varying speed to cause different degree of body accelerations and amplitudes. The magnitude profiles of dynamic body accelerations and magnetic disturbances during this experiment are easily observable in Figures 5.9-5.11 respectively (section 5.7).

5.7. Results and Discussion

5.7.1. Results of Motion Capture

The gait capture was performed indoors in a long corridor in a more unrestricted and realistic setting to evaluate joint angles of lower body. The comparison with OptiTrack® ground truth was excluded for gait analysis, since the corridor which provides an ideal walkway to evaluate the effect of magnetic disturbances, did not allow easy setup of OptiTrack® system with sufficient accuracy. Also the comparison of single MIMU orientation between our sensor fusion

and OptiTrack® system has already been made in terms of accuracy in an earlier experiment (chapter 4). Therefore, the focus of current experiment was to observe for residual drift in joint angle estimation and demonstrate the robustness of our end-to-end algorithm under indoor conditions. The experiment was conducted for 30 minutes, and data for different subjects at different walking speeds was obtained in gait sequences. For more clarity, only a limited set of data from one sequence is shown in Figures 5.5-5.11 . Further discussion will be based on these plots.
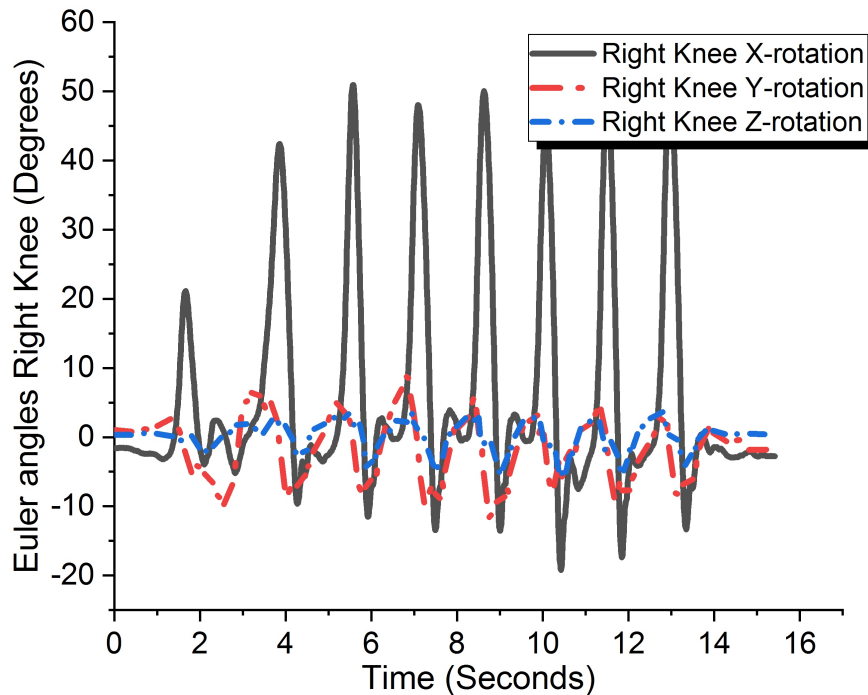
Figure 5.5.: Right Knee Joint Euler Angles (XYZ)

Figure 5.5 shows the right knee joint angle obtained during our experiment. It is depicted as Euler angle in XYZ order. The X rotation represents the knee flexion/extension, Y-rotation is the internal rotation and Z-rotation is abduction/adduction. One can instantly observe the advantage of our inertial motion capture over kinematic constraint based approaches [124, 68, 67]. Often a kinematic constraint of one degree of freedom (1-DOF) is assumed for knee joint i.e. internal rotation and abduction angles are assumed zero. But one can observe appreciable internal rotation and abduction angles in Figure 5.5. More comprehensive biomechanical models do allow more DOF for knee joint [31, 51, 87], but at computational cost and/or reduced accuracy/drift compared to our algorithm.

In a magnetically disturbed environment, the drift error is expected to appear in Y-rotation (related to yaw or heading angle). If the rate gyro biases of MIMUs situated at upper and lower legs are not compensated well and instead the drift of two gyros reinforce each other around Y-axis, a significant drift component would appear in Y-rotation of knee joint. This drift would also appear in Y-rotations of respective MIMUs in Figure 5.6.
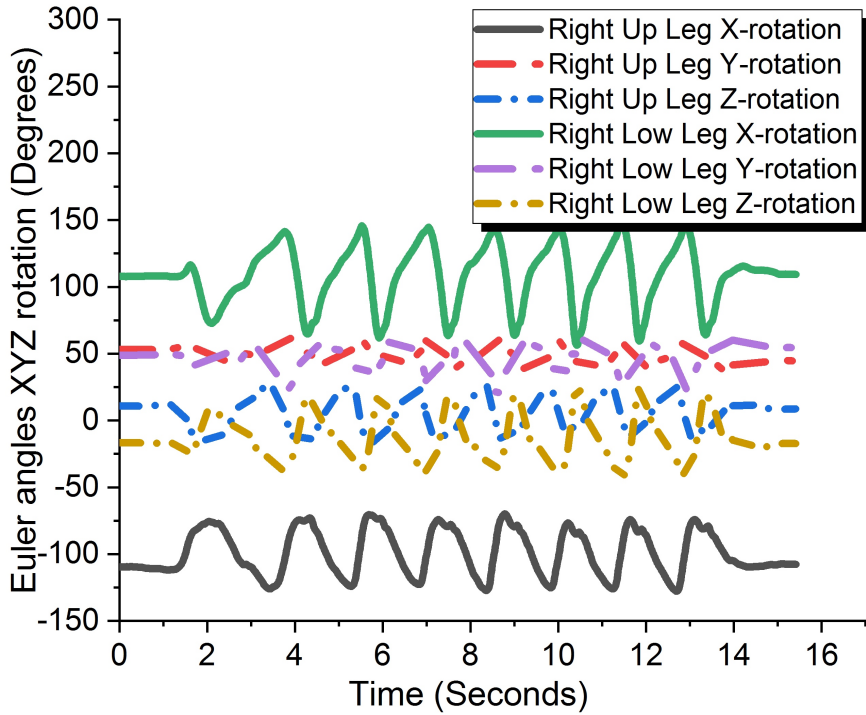
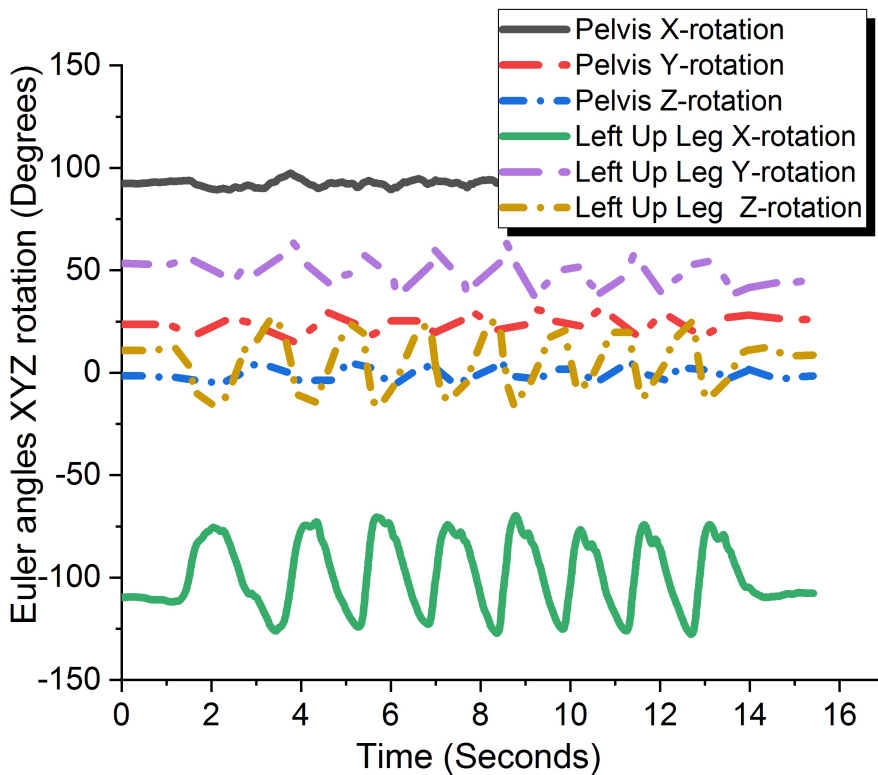Figure 5.6.: Euler Angles (XYZ) Output of MIMUs attached to Right Up and Low Leg

Figure 5.7.: Euler Angles (XYZ) Output of MIMUs attached to Pelvis and Left Up Leg

But we observe no significant drift in individual MIMUs in Figure 5.6, nor for the knee joint in Figure 5.5. This shows that our sensor fusion algorithm correctly estimate the rate gyro biases and thus our orientation estimation remains valid even under magnetic disturbances and body accelerations. More importantly, the negligible drift of individual MIMUs do not reinforce in case of right knee joint and hence online correction of drift is not seen in Figure 5.5.

The lack of long term drift can also be noticed in Y-rotation (internal rotation) of left hip joint as shown in Figure 5.8. But in this case, we observe that offset correction of joint angle takes place according to (5.8) when the threshold of $\pm\epsilon_0$ degrees is exceeded. Again the respective MIMUs at Pelvis and Upper Leg register only negligible drift, as visible in Figure 5.7. But overall the drifts of individual MIMUs reinforce and causes hip joint angle to drift. This is addressed by our proposed algorithm through offset correction of joint angle. To elaborate upon these results, next section would only focus on one MIMU located at Pelvis.
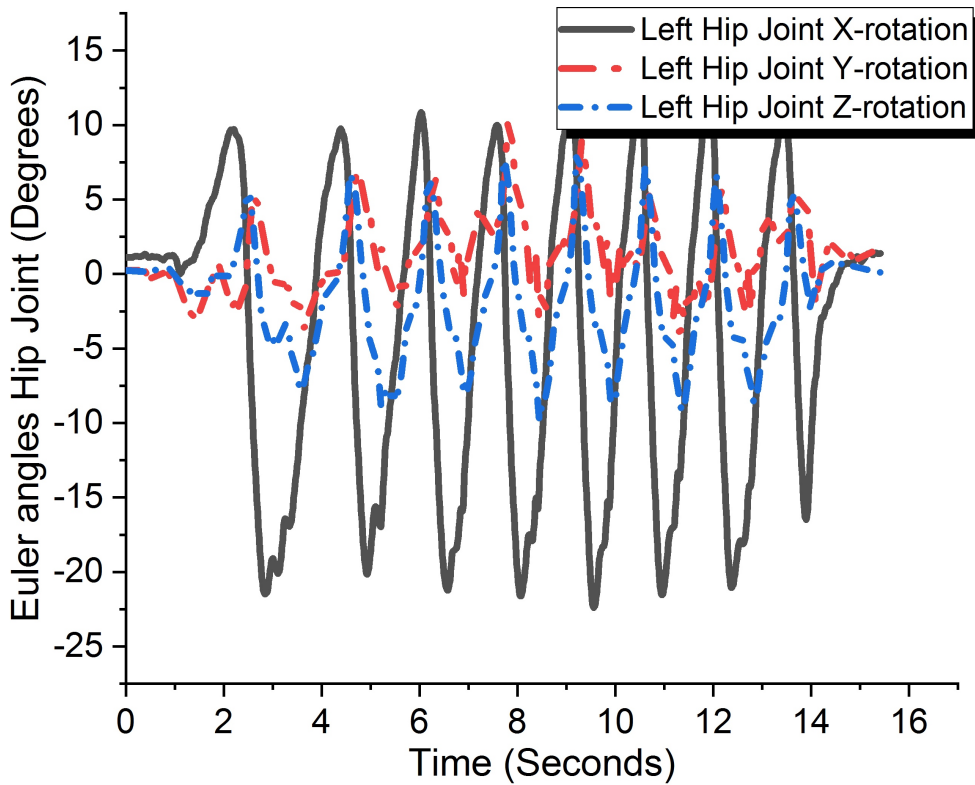
5.7.2. Performance Against Acceleration and Magnetic Disturbances

In order to appreciate the robustness of our sensor fusion algorithm against body acceleration and magnetic disturbance in the real scenario of gait experiment, we focus on the MIMU located at Pelvis. In the absence of OptiTrack® ground truth, the bias compensated rate gyro integration provides us with a pseudo ground truth for a short duration i.e. in a sequence of few seconds, the bias of rate gyro is precisely obtained and assumed constant during integration to obtain 3D

angle. Figures 5.9 and 5.10 plots the pitch (X) and roll angle (Z) respectively obtained in ZXY order of Euler angle.
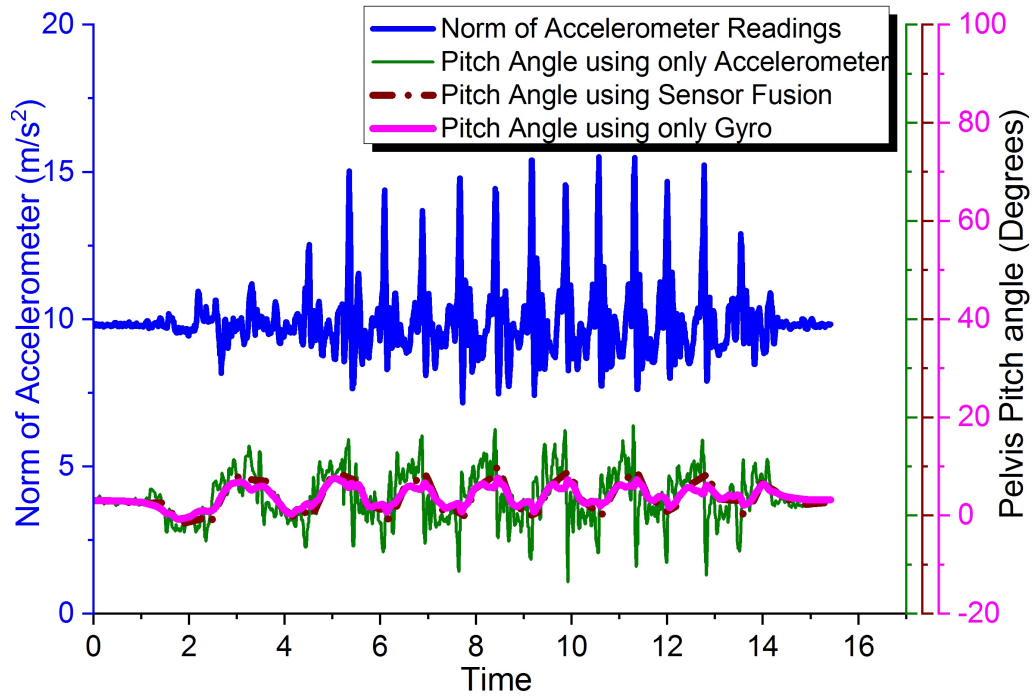

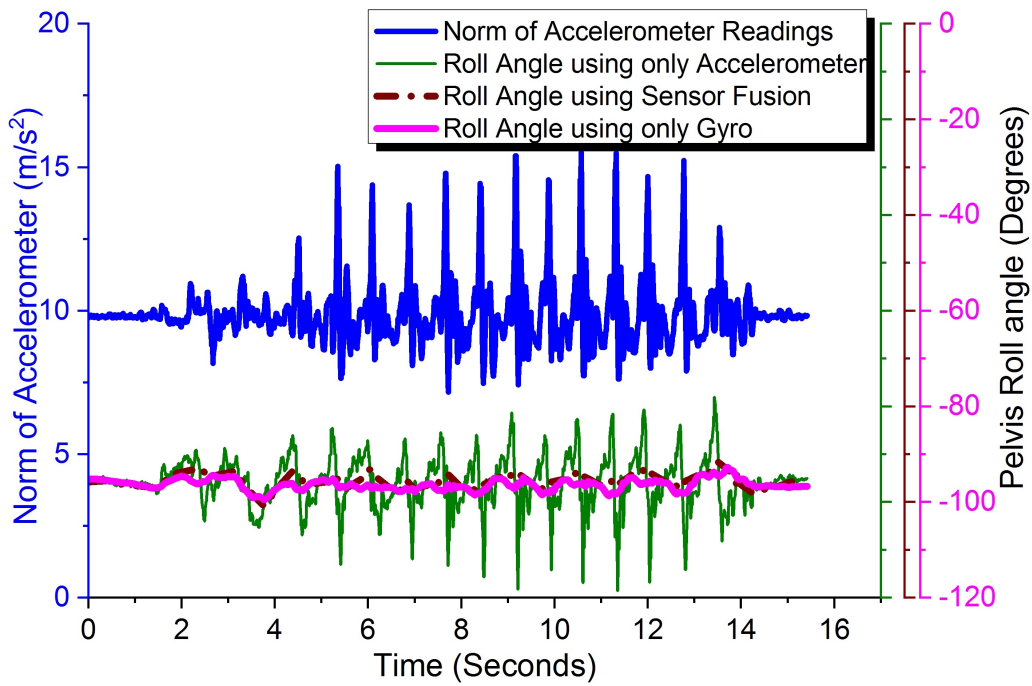Figure 5.9.: Pitch Angle (X) Output of Pelvis MIMU

Figure 5.10.: Roll Angle (Z) Output of Pelvis MIMU

We plot pitch/roll angle calculated using (1) only the accelerometer, (2) with our sensor fusion algorithm and (3) with bias compensated rate gyro integration respectively. It is apparent that pitch/roll from accelerometer alone is too noisy, despite that the MIMU was located at Pelvis, which do not experience large body accelerations during walking (norm of accelerometer readings is shown). On the contrary bias compensated rate gyro integration gives smooth pitch/roll angles (which represents pseudo ground truth but only for short duration). Our sensor fusion algorithm results are close to bias compensated rate gyro and thus results in only minor error without long term drift.

In Figure 5.11, we report the yaw or heading (Y) angle of Pelvis MIMU as obtained from (1) attitude corrected magnetometer, (2) using our sensor fusion algorithm and (3) bias compensated rate gyro integration respectively. It is in yaw angle that we observe the major advantage of our sensor fusion/gyro bias compensation and joint angle drift correction. As depicted by the magnetic norm, the magnetic field is continuously varying as we traverse along the corridor. This results in corresponding (though not exactly correlated) strong variations in yaw/heading angle, if only attitude compensated magnetometer is used. In contrast, both bias compensated rate gyro integration and sensor fusion (which almost totally depend upon rate gyro under these conditions), keep good heading angle during the walk. No significant drift is visible in this figure (showing 16 seconds of gait), however in practice the slight drift that appears, is quickly addressed using dynamic update of joint angle offset given by (5.8) as soon as the quasi-static pose recurs.
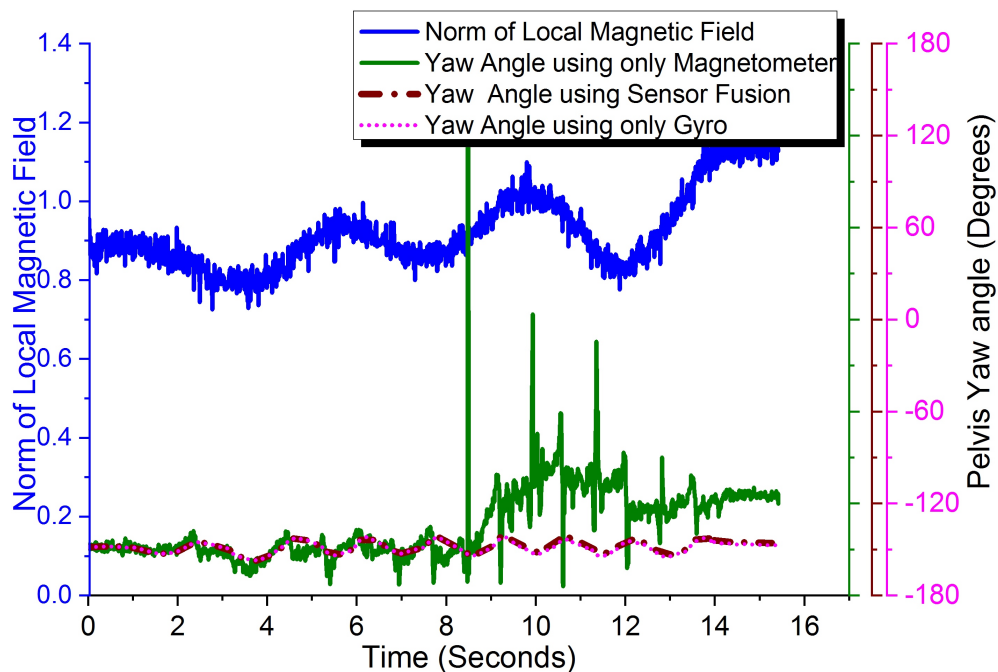

Figure 5.11.: Yaw Angle (Y) Output of Pelvis MIMU

5.8. Conclusion

In this chapter, we have proposed a complete anatomical formulation for inertial motion capture (i-Mocap). Our joint angle estimation from adjacent MIMUs, addresses the problem of static and dynamic error in i-Mocap. The full body calibration framework eliminates these errors by incorporating a novel joint angle offset correction. It not only takes care of initial misalignment of MIMUs (static error) but also the dynamic error that accumulate during motion capture arising from either sensor degradation or attachment artifacts. The correction term in the anatomical model is updated and performs online correction of drift in individual joint angles. This action is performed automatically when the limb or extended torso are in neutral quasi-static pose. The former condition is judged by using accelerometers. Overall our approach achieves precise orientation estimates in highly dynamic conditions and inhomogeneous magnetic fields (chapters 3 and 4) as well as avoids drift or error due to residual and dynamic artifacts during inertial motion capture (chapter 5). In comparison to state-of-art anatomical calibration methods, our approach is more robust and prevents long term error accumulation during inertial motion capture.

6. Learning 3D Joint Angle Constraints from Motion Capture Datasets

In chapters 3, 4 and 5, we focused on the robust performance of inertial motion capture (i-Mocap) pipeline using sensors worn on adjacent body segments (traditional i-Mocap) by addressing its various sources of error. We improved the sensor calibration and fusion algorithms and also proposed an anatomical calibration framework to deal with joint angle drift correction online. Overall these three chapters form part-I of this dissertation, dealing with improvements in traditional i-Mocap pipeline. The current and next chapter would focus on data driven and deep learning approach to i-Mocap and form part-II of this dissertation. In this chapter, we demonstrate that 3D human skeleton has joint limit constraints that can be effectively learned in a latent space using a data-drive approach from 3D motion capture (mocap) datasets. We formulate the joint constraints with a parameterization (s_1, s_2, τ) of swing-twist space to improve upon the limitations of previous 3D angle representations. Our parameterization is applied on Human3.6M dataset to create (s_1, s_2, τ) map for each joint. These maps are used to generate synthetic datasets in the entire joint space. The neural network discriminators are then trained on synthetic dataset to learn valid/invalid 3D joint rotations. The accuracy of these discriminators have been found varying between $[95\% - 99\%]$ for different joints. Our work validates that neural networks can model and learn 3D human joint angles in a high dimensional latent space with high accuracy. The work presented in this chapter is based on our peer reviewed published research [89].

6.1. Background

Human body is an articulated rigid body, comprising of fixed length bone segments connected by anatomical joints. The set of 3D positions together with orientations of these joints define the complete '3D pose' of the human body. Using forward kinematics (assuming fixed bone lengths), 3D positions of all joints are obtained simply from a set of 3D joint rotations. For practical human motion capture, the anatomical joints can be approximated by simpler mechanical ones. For example both shoulder and hip joint complex are represented by a 3-DOF spherical (ball and socket) joint, while wrist and ankle are assumed 2-DOF universal joints and elbow and knee joints are either assumed 1-DOF revolute or hinge joint or 2-DOF universal joint. The elbow and wrist share an additional DOF that can be referred to either joint. The three types of mechanical joint used in modeling the articulated human body are depicted in Figure 6.1.
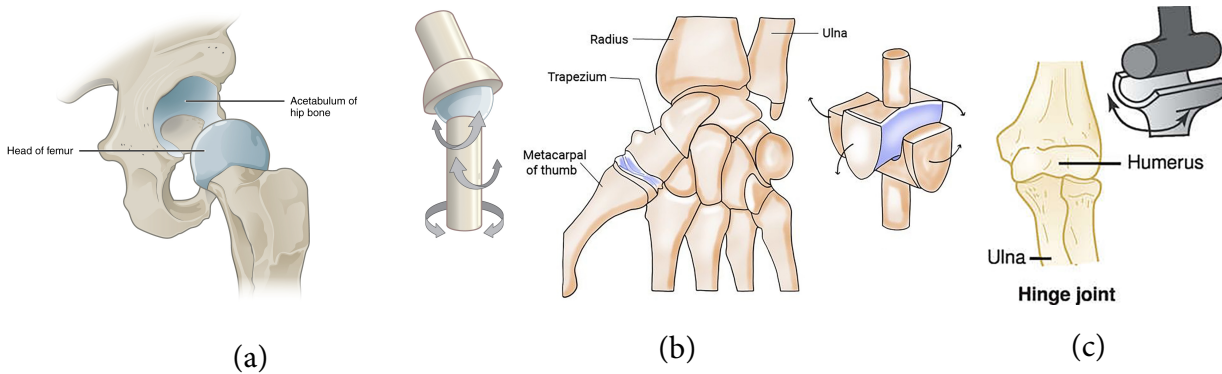
Figure 6.1.: Approximation of Human joints as Mechanical joints (a) hip joint as 3-DoF ball and socket joint (b) wrist joint as 2-DoF universal joint and (c) knee joint and 1-DoF hinge joint

In general, all the body joints have limited range of motion due to anatomical restrictions which are expressed as 'joint constraint'. This provides an additional measure of enforcing 'valid' human pose in cases of incomplete or noisy set of pose parameters or if the problem is ill-defined.

Since the MIMU based sensor fusion and gyro bias calibration show degraded performance when run for long time, especially indoors and the magnetometer information is mostly unreliable (chapters 4 and 5), few authors have used only accelerometer-gyro pair (IMU) together with implied kinematic constraints on joints to avoid angle drift [124, 68, 67]. The challenge however lies in the fact that due to reduced degrees of freedom or constraints in assumed anatomical model, only a best fit of joint angles is obtained which is not always true in biomechanical sense. The accuracy and reliability of MIMU based motion capture is therefore still a major concern.

The forward kinematic of articulated 3D human pose from a set of joint rotations has an advantage compared to obtaining it directly as a set of 3D joints positions. If only 3D joints positions are known, there is a many-to-one ambiguity especially in terms of twist around a body segment. The anomaly is not apparent in a stick based visualization of 3D human pose, but would nevertheless lead to shape distortions around joints, if a surface mesh is applied. The twist motion (along kinematic chain) is thus essential to predict the correct 3D orientation of body parts at extremity such as hands, head or feet. The twist error is not reflected in the 3D joint position error and often goes unreported in the current literature on 3D human pose estimation, which uses the later error metric.

The literature in computer graphics (CG), deals with 3D human pose in more perceptive manner. The character animation and realistic motion synthesis requires that human body meets the anatomical constraints on its articulation. These constraints are often represented in swing-twist formulation of human pose. For example the work by [3] uses the swing part and incorporate joint limits as pose priors in estimation of human pose.

3D human pose estimation using deep learning based kinematic models has produced state

of art *robust* results [56, 85, 158] compared to unconstrained regression [102, 133]. The deep learning methods unite the strength of the large human motion datasets and hierarchical feature learning. In former case, a kinematic layer or module is used to incorporate structured prediction and constraints, but the learning is end-to-end and the intermediate kinematic vector remains latent. The error itself is minimized on the output 3D joint rotations and not the latent space. The inclusion of a kinematic layer/module helps to learn a realistic 3D human pose meeting anatomical constraints.

An enormous corpus of human motion capture (mocap) datasets have emerged to train deep neural networks. Recently, [137] has even employed a synthetic human dataset for better training of deep neural networks. But synthesis of *realistic* human motion for long run is an open problem itself [104]. In the past, a *'penalty term'* for exceeding joint limits was used, but human joint constraints have been worked out from more limited datasets [2, 3]. However to the best of our knowledge, no recent attempt has been reported to obtain more realistic and statistically significant joint limits from current datasets or to integrate data-driven joint limits rigorously in deep learning.

The artifacts and outliers in the mocap data (ground truth) can be pruned using joint constraints inferred directly from respective datasets and imposing accurate joint constraints as *priors* is valuable for 3D human pose estimation, realistic human motion animation and generation of *synthetic* human motion datasets. But a data-driven learning of joint constraints from mocap datasets is still very challenging. Presence of false outliers, voids and sparse data makes it imperative to choose a representation of joint space, which is robust to such artifacts. Moreover instances of invalid class are missing in mocap data and need to be synthesized for balanced learning.

The parameterization of joint constraints is dependent on the choice of representation for joint rotations. The simplest representation of joint rotation is in terms of Euler angles. A straightforward approach to implement joint constraints is to assume minimum and maximum limits on Euler angles (box-model). However, not only this approach is too crude to capture realistic anatomical joint limits, but also it is not unique and do not provides contiguous cluster of valid poses in joint space [32]. The Euler angles also involve two singularities and often causes the problem of Gimbal lock [7]. Due to these restriction, the box-model based joint limit though simple is not an optimal choice.

Aforementioned limitations of previous works form the motivation of our research presented in this chapter. Instead of quaternion or Euler angles, we use exponential map parameterization to represent joint rotations in present work. The exponential map representation avoid singularities as well as non-contiguous clusters [32] of mocap data. It also avoids the duality problem in representation as described earlier. We derive our joint constraints using more comprehensive Human 3.6 Million (H36M) dataset [54] and compared these to Akhter et al [3]. We also encoded lower and upper bounds of twist angle τ on occupancy map (s_1, s_2) of swing, as an additional parameter to be learned. This addresses the intra-joint dependency to overcome limitation of previous work.

6.2. Contribution

Realistic estimation and synthesis of articulated human motion must satisfy anatomical constraints on joint angles. In this chapter, a data driven approach is used to learn human joint limits from 3D motion capture(mocap) datasets. The main contribution (see Figure 6.2) of our paper are:

1. A new (s_1, s_2, τ) representation based on exponential map for swing-twist decomposition of a joint rotation is used for data-driven learning. Compared to other 3D joint angle representations (Euler angle and quaternion), the new representation gives a contagious cluster of valid 3D rotation in joint space.

2. The discretized *look-up maps* were obtained for each joint angle (in new representation) from mocap data, to deal with voids and sparsity.

3. Using look-up maps, labels were *synthesized* for both valid and invalid instances in joint space. A balanced valid/invalid labeled dataset is thus obtained.

4. Finally, a *balanced learning* of neural network discriminators is performed on the synthesized datasets of each joint.

Our three-parameter representation is unambiguous and free of anomalies, which makes it easier to learn a joint rotation manifold. Thus, neural network discriminators trained on our representation achieve good accuracy and qualitative precision with very few parameters.

6.3. Organization of Chapter

This chapter is organized in seven main sections. In the first two sections we have covered the background and our main contribution to the data-driven learning of 3D joint angle constraints of human motion. In section 6.4, we cover the salients of our methodology and compare it to [3]. The use of Human3.6M(H36M) dataset [54] is justified for data-driven learning of constraints of each joint, for it covers extensive range of daily human activities. The swing-twist decomposition is also described in detail. In section 6.5, we provide the details of implementation like look-up maps discretization of 3D joint space and generation of a balanced synthetic dataset with valid/invalid pose labels. In section 6.6, we present the results of discriminator learning and also show the examples of 3D poses classified as valid or invalid based on joint constraints.

6.4. Proposed Methodology

We formulate joint constraints with a new parameterization (s_1, s_2, τ) of swing-twist space using exponential maps to overcome limitations of previous representations. The swing or joint range of motion is defined by a contiguous (s_1, s_2) map. The twist of body segment (τ) is

represented as an additional parameter varying continuously on (s_1, s_2) map with a lower and upper bound. Our parameterization is applied on Human3.6M dataset to create (s_1, s_2, τ) map for each joint. These maps are used to generate 'synthetic' datasets in entire joint space. The neural network discriminator(s) are then trained on synthetic dataset to learn valid/invalid joint rotations. The accuracy of these discriminators have been found varying between $[95\% - 99\%]$ for different joints. To further validate precision-accuracy trade-off, we qualitatively evaluate poses classified as valid/invalid, using an interactive tool. The end-to-end approach is depicted in Figure 6.2.
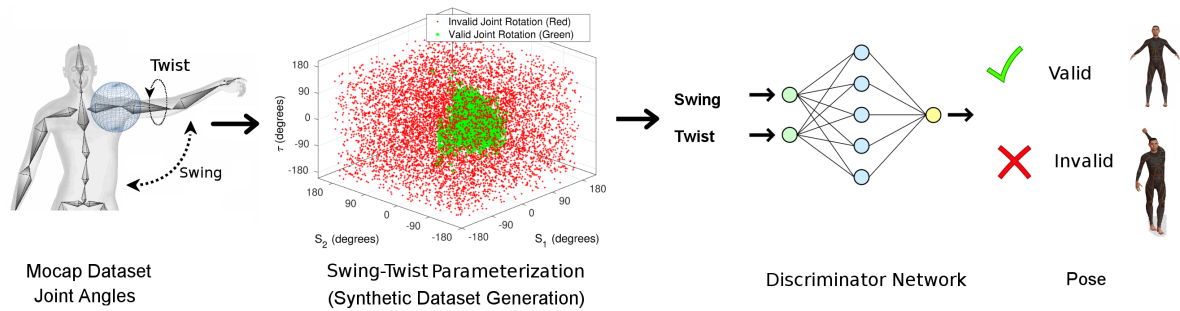
Figure 6.2.: Overview of Framework: The joint angles from motion capture datasets are transformed into (s_1, s_2, τ) swing-twist representation. Synthetic dataset is generated to train neural network discriminators (using look-up maps not shown in figure). Discriminator network learns joint constraints to classify valid/invalid joint rotations.

3D joint constraints for routine activities are learned by using Human3.6M (H36M) dataset [54]. Our joint constraints parameterization is based on swing-twist formulation, but we perform the swing-twist decomposition on quaternion rotation. Later we re-parameterize the swing (expressed in quaternion form) to a two parameter representation (s_1, s_2) and express twist as a simple angle 'τ', using exponential map. Thus, our parameterization of joint space is (s_1, s_2, τ). The inclusion of twist in our parameterization enforces complete limits on joint rotation.

In comparison, Akhter et al. [3] have captured a dataset based on extreme range of joint movements. Moreover, the authors ignore twist around axis as they parameterize joint rotation from local 3D Cartesian coordinates to the spherical coordinates(ϕ, θ) representing swing of body segment. Thus, they do not provide for any intra-joint dependency of swing-twist or twist limits in their framework. In current work, we compare our joint limits (swing part) derived from a range of normal activities performed by human subjects (in H36M) with extreme range limits provided by [3].

Often spherical coordinate representation is used for swing, but in our work we have simply used exponential map for swing as explained in section 6.4.1. This has saved us the computation of 3D Cartesian coordinates of child limb from rotation and later a conversion to spherical coordinates on unit sphere. We have discovered that instead of Euler angles or quaternion or angle/axis, the swing-twist decomposition of joint rotation keeps the entire motion trajectory 'contiguous' [32] in the joint manifold.

6.4.1. Swing-Twist Decomposition

Our (s_1, s_2, τ) parameterization is done by decomposing the quaternion rotation of joint into orthogonal components i.e. swing and twist, where the twist is w.r.t an axes of intrinsic (rotated) frame. For example if y-axis of rotated frame is reference aligned with child limb, the swing-twist decomposition of a rotation in terms of quaternions is expressed in the following way:

Let $\boldsymbol{\eta} := \theta \widehat{\boldsymbol{w}}$, with $\widehat{\boldsymbol{w}} := (w_x, w_y, w_z), \|\widehat{\boldsymbol{w}}\| = 1$ represents the angle-axis representation of a 3D-rotation with $q_s := \cos(\theta/2)$, $q_x := w_x \sin(\theta/2)$, $q_y := w_y \sin(\theta/2)$, $q_z := w_z \sin(\theta/2)$, then corresponding quaternion $\mathbf{q} := (q_s, q_x, q_y, q_z)$ can be decomposed as $\mathbf{q}_{twist} * \mathbf{q}_{swing}$, where $\mathbf{q}_{twist} := (c_y, 0, s_y, 0)$, $\mathbf{q}_{swing} := (c_{zx}, s_x, 0, s_z)$. Then, $\boldsymbol{\eta}_{twist} := (0, \tau, 0)$ and $\boldsymbol{\eta}_{swing} := (s_1, 0, s_2)$ are respective exponential map counter-parts, and using the following quaternion properties:
(i) $\|\mathbf{q}_{twist}\| = 1$, (ii) $\|\mathbf{q}_{swing}\| = 1$ and (iii) $\mathbf{q} = \mathbf{q}_{twist} * \mathbf{q}_{swing}$
we get,

$$\tau = 2 \cdot \arctan\left(\frac{s_y}{c_y}\right),$$

$$s_1 = 2 \cdot \arctan\left(\frac{\sqrt{s_x^2 + s_z^2}}{c_{zx}}\right) \cdot \frac{s_x}{\sqrt{s_x^2 + s_z^2}},$$

$$s_2 = 2 \cdot \arctan\left(\frac{\sqrt{s_x^2 + s_z^2}}{c_{zx}}\right) \cdot \frac{s_z}{\sqrt{s_x^2 + s_z^2}}$$

where,

$$c_y = \left(\frac{q_s}{\sqrt{q_s^2 + q_y^2}}\right), s_y = \left(\frac{q_y}{\sqrt{q_s^2 + q_y^2}}\right), c_{zx} = \left(\frac{(q_s^2 + q_y^2)}{\sqrt{q_s^2 + q_y^2}}\right)$$

$$s_x = \left(\frac{(q_s \cdot q_x - q_y \cdot q_z)}{\sqrt{q_s^2 + q_y^2}}\right), s_y = \left(\frac{(q_s \cdot q_z + q_x \cdot q_y)}{\sqrt{q_s^2 + q_y^2}}\right)$$

Using the above re-parameterization, we get exponential map based swing components (s_1, s_2) and the twist component τ of a joint rotation. This representation is convenient as it is (i) un-ambiguous (no duality), (ii) avoids non-contiguous regions/jumps in joint trajectory and (iii) requires no additional constraints like unit norm etc. Consequently, it provides optimal joint representation for learning valid/invalid poses.

6.5. Implementation

The human pose in Euler angle representation provided by H36M dataset are transformed into (s_1, s_2, τ) representation, as described in section 6.4. The resulting joint motion space is continuous, contiguous and well-behaved for discriminative learning (For comparison of Euler angle, quaternion and swing-twist joint space see Figures 6.3-6.5).
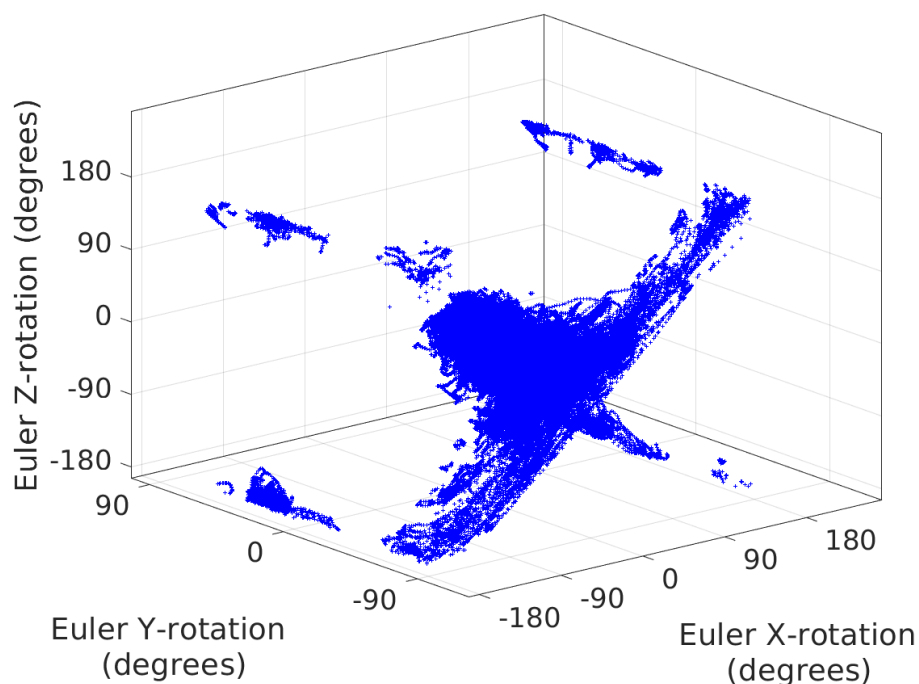
Figure 6.3.: The artifacts in Euler Angle space of Right Hip (Right-Up-Leg) show non-contiguous clusters of valid joint rotation. Also boundary jumps at 180 degrees (for Z and X) and 90 degrees for Y rotation are visible
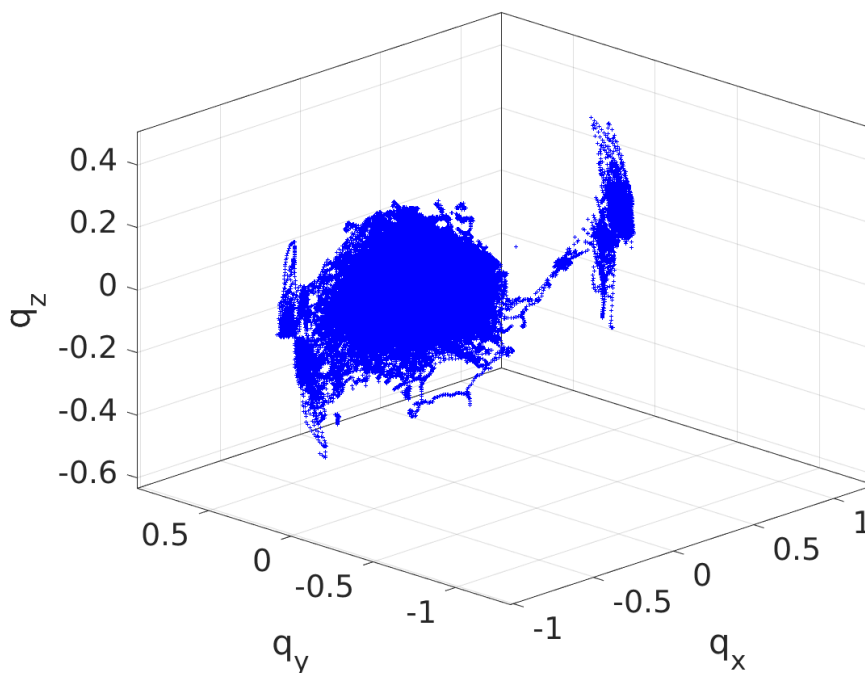

Figure 6.4.: Quaternion (vector part) joint rotation space of Right Hip (Right-Up-Leg) used by Herda et al. [47] shows duality of valid joint rotation. Anti-podal points +q & -q in the cluster represent same rotation
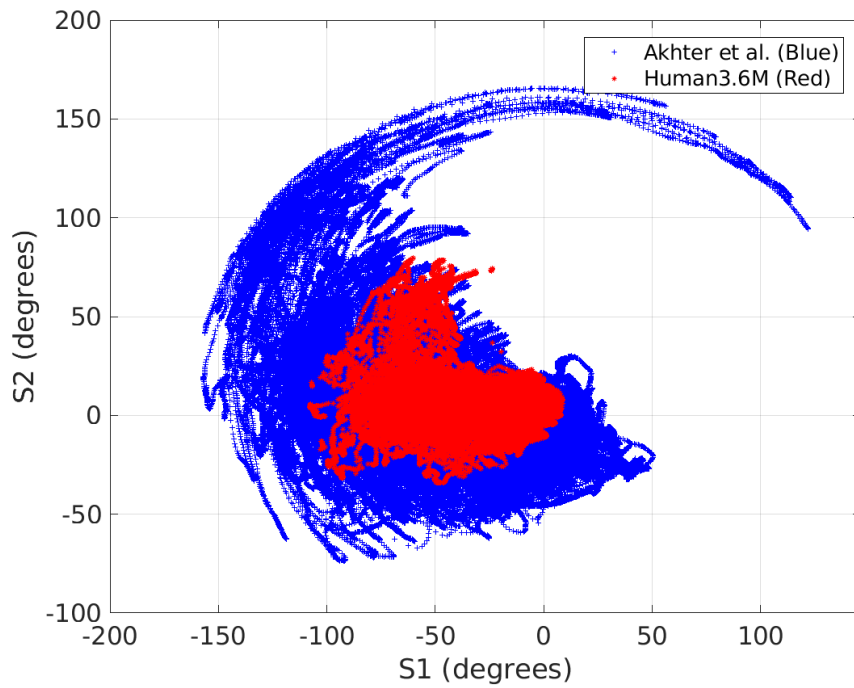
Figure 6.5.: The comparison of right hip joint's range of swing present in H36M [54] and Akhter et al[3] dataset
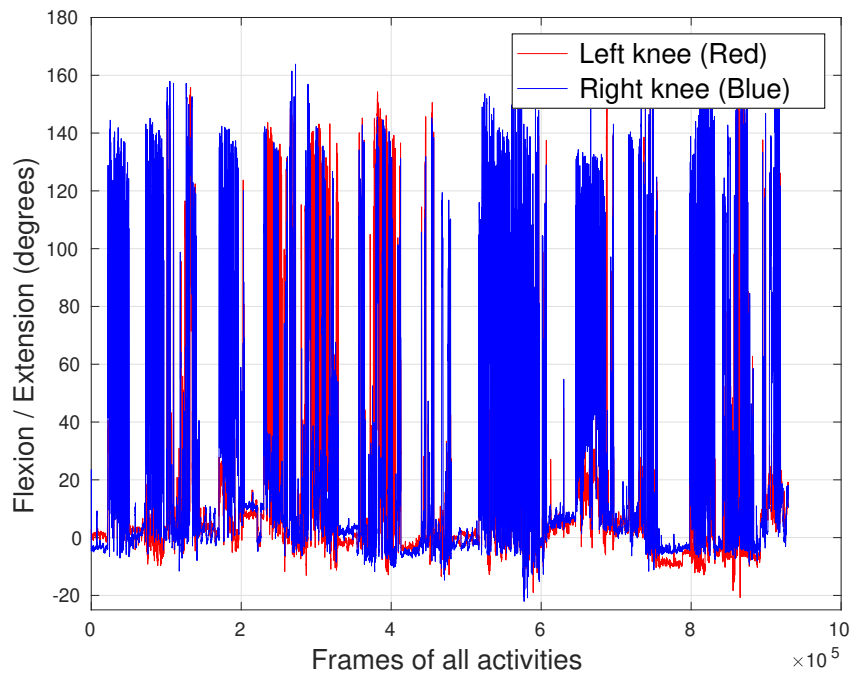
Figure 6.6.: The knee joint rotation anomalies in Akhter et al. dataset [3]. Flexion is positive and Extension is negative knee joint angle.

It is found that considerable outliers are present in mocap dataset provided by Akhter et al [3]. For instance, left and right knee flexion of all subjects performing various activities is as shown in Figure 6.6. It clearly shows an impractical knee extension (upward) of up to -20 degrees for certain motion sequences in dataset. Moreover, we believe that limited set of mocap data of [3] is not enough to learn inter-joint dependency, i.e. insufficient instances of swing of the child limb exists conditioned upon a prior swing of parent limb.

The comparison of swing limits for right hip joint (Right Up leg) from Akhter et al. dataset [3] and H36M dataset [54] in swing space (s_1, s_2) is illustrated in Figure 6.5. The swing limits obtained using H36M although more restrictive but are true representative of human motion involved in routine activities. Also, these limits lie well within the extreme range of motion of [3]. Therefore, we restrict our learning only to H36M dataset.

6.5.1. Look-Up Maps for Synthetic Dataset

Let $\mathbb{M}$ be the set of all samples of H36M dataset for joint J, represented in (s_1, s_2, τ) space . We define a 2D grid (s^i, s^j) on swing subspace, where $s^i, s^j \in \{-180, \ldots, 180\}$ and a discrete mapping as follows

$$
\mathrm{D}_s(s_1, s_2) := \begin{cases} (s^i, s^j) & \text{iff } s_1 \in [s^i, s^{i+1}] \\ & \text{and } s_2 \in [s^j, s^{j+1}] \\ undefined & \text{otherwise} \end{cases} \tag{6.3}
$$

Now for each grid square (s^i, s^j), if at least one mocap sample $\in \mathbb{M}$ lies in it, we term it as valid swing. The discretized swing occupancy map is thus defined by

$$
\mathrm{O}_s(s^i, s^j) := \begin{cases} 1 & \text{iff } \exists\, (s_1, s_2, \tau) \in \mathbb{M} \\ & \text{such that } D_s(s_1, s_2) \mapsto (s^i, s^j) \\ 0 & \text{otherwise} \end{cases} \tag{6.4}
$$

With each valid occupancy bin, we further associate minimum and maximum bounds of τ, based on all the instances of mocap data occurring in that bin

$$
\mathrm{L}_t(s^i, s^j) := \begin{cases} [\tau_{min}, \tau_{max}] & \forall\, (s_1, s_2, \tau) \in \mathbb{M} \\ & s.t.\ O_s(s^i, s^j) = 1, \\ & \tau_{min} = \lceil \tau \text{ and } \tau_{min} = \lfloor \tau \\ undefined & \text{otherwise} \end{cases} \tag{6.5}
$$

In order to obtain smooth twist angle bounds, we apply 3x3 Gaussian kernel(s) on (s_1, s_2, τ_{min}) and (s_1, s_2, τ_{max}) surfaces. Now given any (s_1, s_2, τ), we define a validity function in the fol-

lowing way

$$\text{isValid}(s_1, s_2, \tau) = \begin{cases} 1 & \text{iff } O_s \circ D_s(s_1, s_2) = 1 \\ & \text{and } \tau \in [\tau_{min}, \tau_{max}], \\ & [\tau_{min}, \tau_{max}] = L_t \circ D_s(s_1, s_2) \\ 0 & \text{otherwise} \end{cases} \tag{6.6}$$

We generate synthetic ground truth of valid/invalid samples in (s_1, s_2, τ) space of a Joint J, by uniform random sampling on $s_1, s_2, \tau \in [-180°, 180°]$ and using and using isValid() function defined in (6.6).

6.5.2. Training Discriminators

A large portion is occupied by invalid instances in complete joint rotation space (for any given joint). Therefore, the sets representing valid and invalid classes are imbalanced. We perform random sampling of valid and invalid instances to obtain 25% and 75% sets respectively. The sampled sets are equally spaced labeled training instances in entire joint space. An example of such sampled joint space for right hip joint (right up leg) is as shown in Figure 6.7. A balanced ($50\%/50\%$) subset is selected from these sampled valid/invalid instances using random sampling. The subset is again divided into training, validation and test samples in a ratio ($70\%; 15\%; 15\%$). For comparison, we also tested over-sampling of valid class and undersampling of invalid class to obtain balanced subset independently. However, the sampling proposed above leads to least variation of performance, when evaluation is performed on multiple test sets. The neural network discriminator are trained for each joint, using an equal number of valid/invalid instances in each batch. We train an ensemble of 20 discriminators for each joint to evaluate performance. We also qualitatively evaluate the ability of neural network to discriminate valid/invalid joint rotations by means of an interactive validation tool (Figure 6.10). The tool also helps us to ascertain any anomalous examples found in motion capture datasets contradictory to the learned joint limits.

6.6. Results and Discussion

The learning of a smooth implicit function by neural network represents the valid range of motion for a specific joint i.e valid swing and twist component of joint rotation. We achieve over 95% accuracy (on all joints discriminators) using only one hidden layer of 10 neurons in Fully Connected Neural Network (FCN) (results shown in Table 6.1). The table reports the result obtained for best among 20 classifier trained for each joint. The number of optimal hidden nodes obtaining best result is also reported for each joint in Table 6.1. A choice of 4, 8, 16 and 32 hidden nodes is tested. The optimal number of hidden nodes is decided when performance plateau is reached. The network with 8 hidden nodes is found optimal for most of the joints.
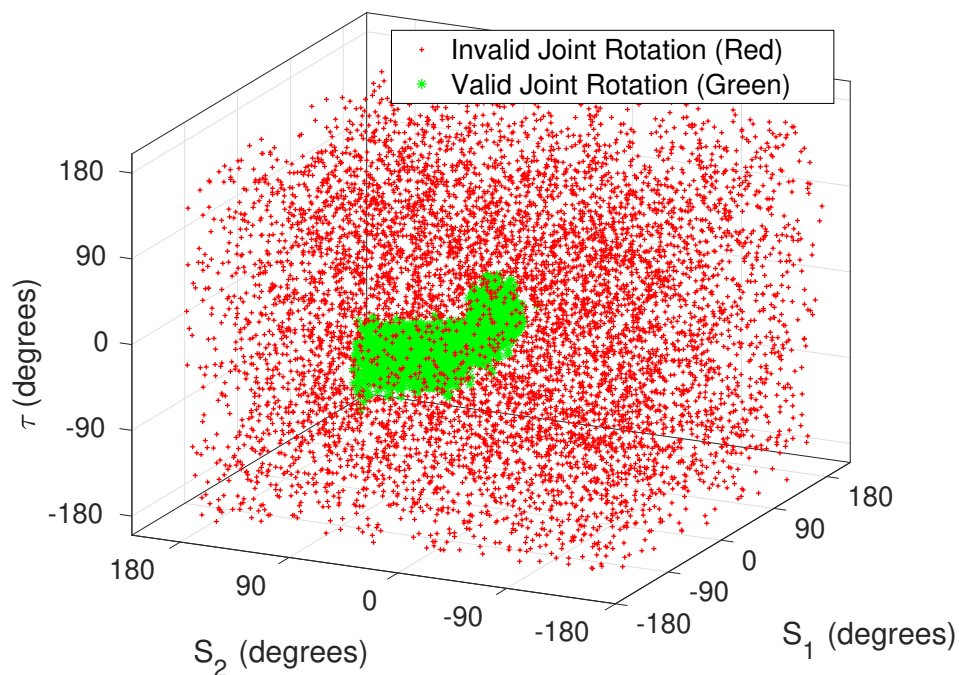
Figure 6.7.: Generated synthetic dataset for right hip joint with valid (green) and invalid(red) samples

Joint Name	Accuracy %	False Valid %	False Invalid %	Optimal hidden layer
RightupLeg	98.7	0.3	1.0	8
RightFoot	97.7	1	1.3	8
LeftUpLeg	98.9	0.3	0.8	8
LeftFoot	98.2	0.6	1.2	8
Spine	99.4	0.1	0.5	16
Spine1	99.4	0	0.6	16
Neck	94.4	2.3	3.3	8
Head	97.7	0.9	1.4	8
LeftShoulder	97.9	0.7	1.5	16
LeftArm	96.8	0.5	2.7	8
LeftHand	94.7	2.1	3.3	8
RightShoulder	95.1	1.5	3.3	16
RightArm	95.5	1.63	2.9	8
RightHand	95.0	2.0	3.0	8

Table 6.1.: The joint discriminator networks classification accuracy on synthetic datasets of respective joint rotations

In our view, this becomes feasible due to our choice of joint rotation parameterization, which is free of ambiguities or discontinuities in joint manifold. The discretization scheme proposed on swing map (s_1, s_2) for complete range of τ rotation is also more intuitive. It avoids the limitations mentioned by [47] in his work on quaternion based joint limits. The representation of joint manifold in Euler angle space and quaternion space is highlighted in Figures 6.3 and 6.4 respectively. The anomalies of both representations can be noted versus our proposed representation shown in Figure 6.5. The discretization of joint space in quaternion and Euler angles to label invalid region is not straightforward as evident from these figures.

The output of discriminator also shows that learning is smooth and effectively accounts for outliers in the data. The corresponding outliers in *synthetic* dataset which are learned as invalid by discriminator are shown in Figure 6.8. However, the accuracy on 'synthetic' ground truth is not an absolute figure of merit, due to false outliers, voids and sparse data. Hence precision of valid/invalid classification is further evaluated qualitatively using an interactive tool.
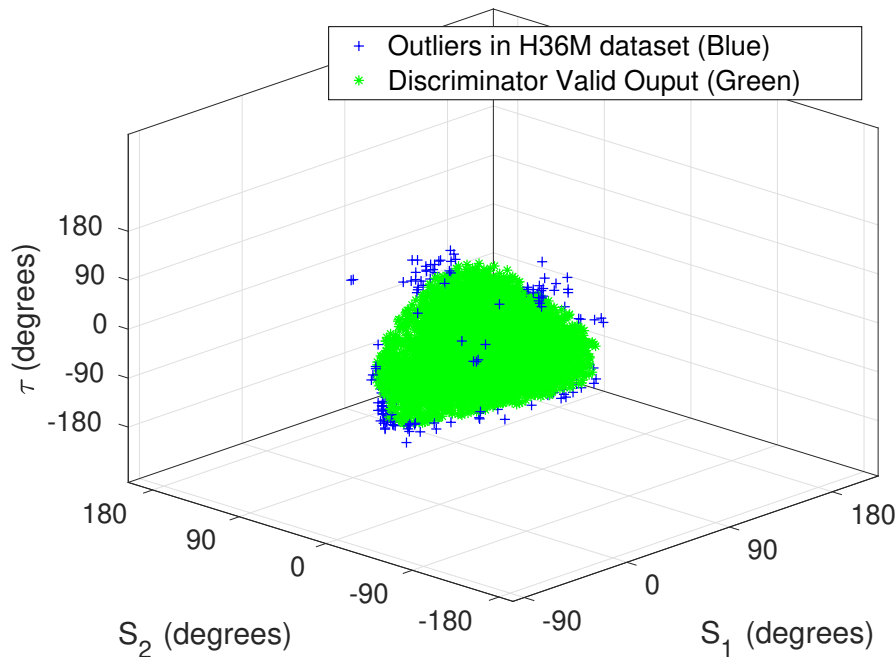
Figure 6.8.: Discriminator network output of valid left shoulder joint rotations in (s_1, s_2, τ) space

The output of our interactive pose validation tool is as depicted in Figure 6.9. The corresponding input frame from motion capture file is played in a window, which can be interrupted at any time and a particular joint can be set interactively to represent a particular (s_1, s_2, τ) value, outside valid range. The displayed pose gets a label as either valid or invalid using trained discriminators. The BioVision Hierarchical data(BVH) file of a motion sequence can also be played continuously in the visualization tool while invalid frames with improbable joint
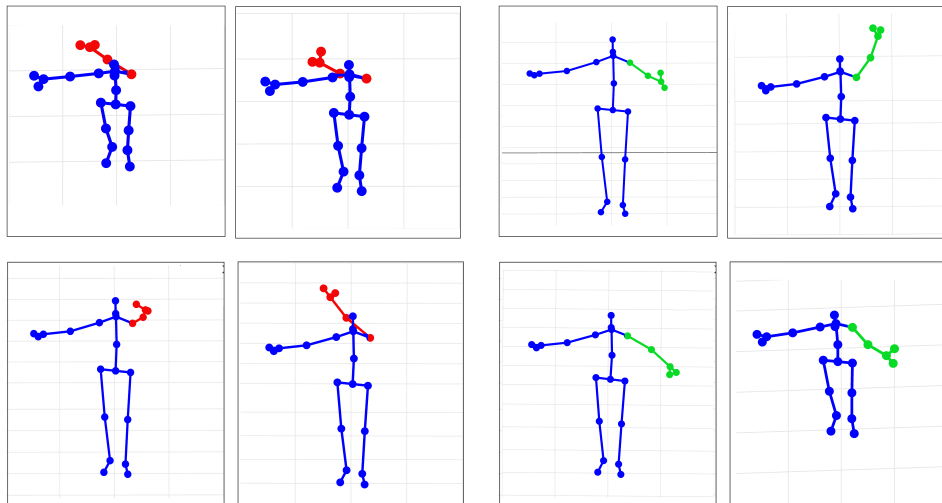
rotations are recorded in an output file.

Figure 6.9.: Interactive validation tool output showing invalid left arm poses (in red) and valid left arm poses (in green)
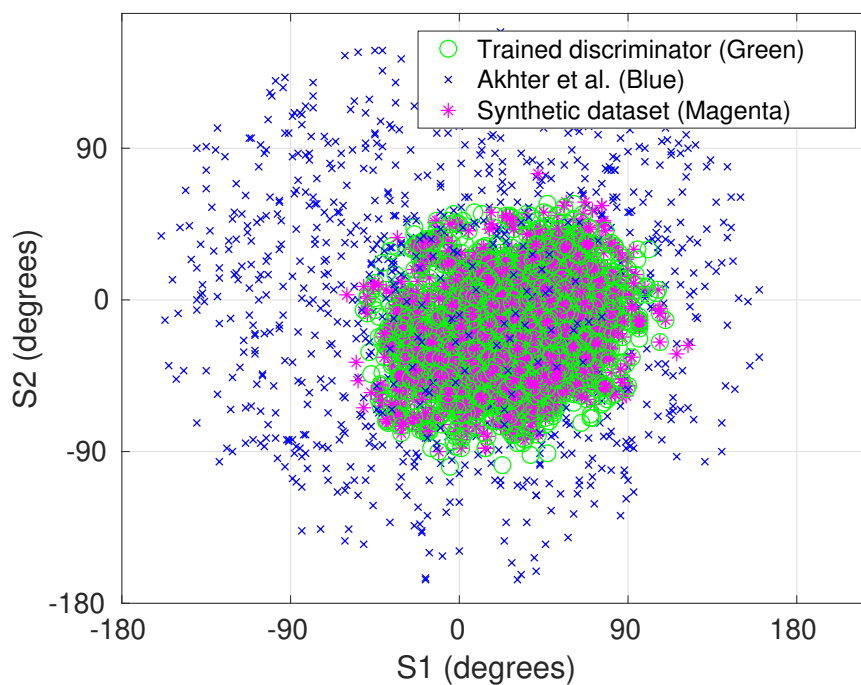

Figure 6.10.: The Valid left shoulder joint rotation of (i) Trained discriminators output (green) (ii) Akhter et al. dataset (blue) [3] and (iii) Synthetic dataset (magenta)

Figure 6.10 shows a comparison for a test data of left shoulder joint, discriminated using Akhter et al [3] isValid() look-up function and our trained network. All poses classified as valid by our network are also found valid by [3] . However, since [3] does not take into account twist, the joint rotations classified as invalid by our network on account of twist are still valid in [3]. Akhter et al. isValid() function also allows unrealistic swing of left shoulder as almost whole (90%) of swing space is classified as valid. On the contrary, valid region in swing space learned by our discriminator is more plausible. These two factors account for high instances (36.65%) of left shoulder poses classified as invalid by our network while still being classified as valid by [3].

The *intra-joint* (parent-child) dependency of range of motion in a hierarchy of joints [3, 48] is also implicitly learned in our approach. The previous works focuses only on the swing component and does not consider the inter dependency of swing and twist. Our approach emphasizes the coupling of parent-child arising from twist itself. For example, a constraint on the twist of upper limb avoids *unrealistic* pose of the lower limb by twist propagation along its kinematic chain. Thus, presented with any (s_1, s_2, τ) parameterized joint rotation, our trained neural network(s) are able to decide for the valid or invalid joints independently.

6.7. Conclusion

The articulated movement of human body is constrained by the limits on 3D joint rotations. In this chapter, we presented a new formalism to learn joint limits implicitly by data driven approach. By using more intuitive swing-twist representation of 3D joint angles, we obtained balanced and more realistic synthetic datasets from real mocap data of human motion. The discriminator(s) trained using this synthetic data showed high accuracy to discriminate valid joint rotations from invalid ones.

Using our analysis of 3D joint rotation limits, we believe that the existing vision based 3D human pose dataset lack full coverage of joint manifolds to learn the complete range of motion. Since the learning is performed using a data-driven approach, the quality of learned constraints depends on the range of motion available in the mocap data. Therefore, we would be extending the current method by capturing more data and using other publicly available motion capture datasets in next chapter.

Our visualization tool can guide collection of motion data close to constraints. A more comprehensive qualitative study to validate synthetic invalid poses with the help of human evaluation would also result in better ground truth generation. Nevertheless, the learned joint constraints from existing datasets are useful to represent normal range of human motion especially vital for social affordance applications like vision based human-robot interaction. It also provides meaningful limits to monitor the rehabilitation of patient's with joint injuries.

7. Deep Learning of 3D Human Pose from Sparse Magnetic Inertial Measurement Units

In this chapter, we propose a deep learning based framework that learns data-driven temporal priors to perform 3D human pose estimation from six body worn Magnetic Inertial Measurement units sensors. Our work estimates 3D human pose with associated uncertainty from sparse body worn sensors. We derive and implement a 3D angle representation that eliminates yaw angle (or magnetometer dependence) and show that 3D human pose is still obtained from this reduced representation, but with increased uncertainty. Our framework is based on Bi-directional recurrent autoencoder. It performs with reasonable accuracy, when a sliding window is used at inference time, instead of full sequence (offline mode). The major contribution of our research is that 3D human pose is predicted even with reduced pitch/roll information from sparse sensors with well calibrated uncertainty which is correlated with ambiguity and actual errors. We have demonstrated our results on two real sensor datasets; DIP-IMU dataset and Total capture and come up with state-of-art accuracy on both. Our work confirms that the main limitation of sparse sensor based 3D human pose prediction is the lack of temporal priors. Therefore fine- tuning on a small synthetic training set of target domain improves the accuracy. The work presented in this chapter is based on our peer reviewed published research [14].

7.1. Background

The number of sensors or special clothing makes inertial motion capture (i-Mocap) quite obtrusive. Commercially available i-Mocap systems like Xsens use upto 13-17 sensors; one per each body segment for full kinematic capture [1]. Figure 1.2 (chapter 1) shows such an i-Mocap system. The full body kinematic is obtained using 3D orientation from MIMU sensors located on all adjacent body segments (one per segment) as explained in chapter 5. The setting up and calibration of so many wearable sensors take time. Each sensor node comprises of magnetic-inertial measurement unit (MIMU) and it employs sensor fusion to get an orientation estimate and linear acceleration in a global frame. The human body has constrained degree of freedom and temporal coherence and smoothness is an important feature of human motion. Many existing kinematic or inverse kinematic based i-Mocap frameworks, therefore uses predefined constraints to reduce measurement errors or drifts [143, 134]. In past research [88, 33], a small set of inertial sensors is shown to estimate 3D pose to a reasonable accuracy. The data-driven

approaches using reduced sensors ($\leq$ 6 instead of 13-17)[123, 138] are more suitable for ambulatory data capture than full kinematic approach [31, 128, 140, 77]. Recently authors [29] demonstrated the condition for observability in a kinematic chain comprising of double hinge joints using sparse magnetometer free inertial tracking. Their work highlights that a sliding window (moving horizon) filter can estimate kinematics of two connected single hinge joints, using measurements from magnetometer free IMUs only at the end links. The scalability of data driven approach to 3D human pose estimation using reduced sensor set has been demonstrated using deep learning [52] and a large SMPL synthetic dataset. In [71] the latent representation of human motion learnt by bidirectional recurrent autoencoder is shown to be robust to both input noise and missing data. Figure 7.1 shows the comparison of full kinematic i-Mocap with sparse body worn sensors based i-Mocap.

(a)   (b)
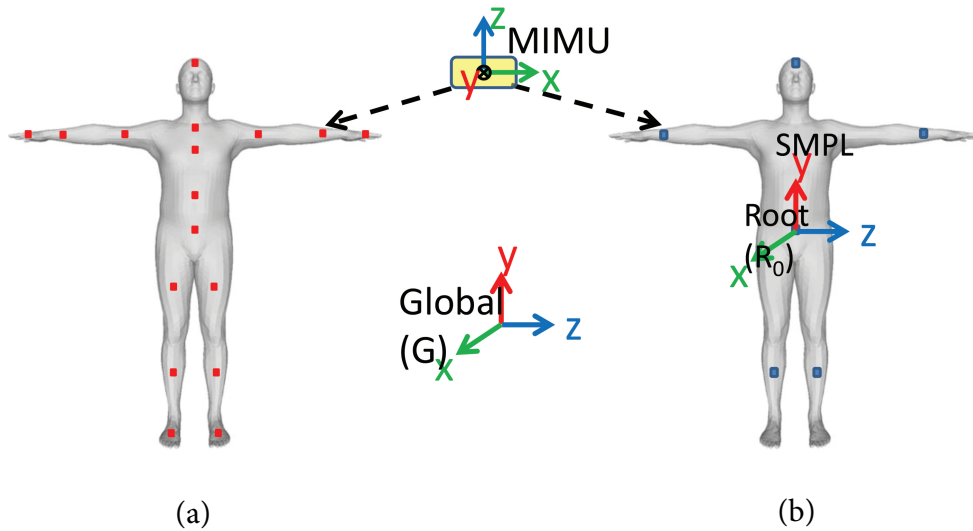
Figure 7.1.: Comparison of (a) Full kinematic i-Mocap (note the placement of 17 sensors on body), with (b) Sparse body worn sensors based i-Mocap (6 sensors at key body segments).

Reducing the number of sensors and flexible placement on body makes the 3D pose estimation less obtrusive and thus this modality can be used for daily activity monitoring, ergonomics and wearable health more easily. Owing to kinematic and temporal constraints of 3D human pose, it is possible to use only a reduced set of sensors (as opposed to one sensor per segment) and still obtain 3D human pose. However the 3D pose estimation with reduced sensors and a learned model usually depend greatly on the correlation in the data. Thus the predictions with inputs outside the training data are often inaccurate and in general ambiguity of this ill-posed problem can lead to high uncertainty. In such a case, an estimate of uncertainty of predicted 3D pose becomes important.

Estimating uncertainty of deep regression is relatively new research direction. Data-driven learning of uncertainty in deep learning framework has been performed by [69] without as-

suming any explicit model. Both the uncertainty in the data (aleatoric) and model uncertainty (epistemic) affect the final error in the output of deep model. The aleatoric uncertainty is dealt with by negative log likelihood (NLL) cost function in the context of 3D rotation by [105] and demonstrated on KITTI visual odometery dataset. Russel et al. [116] extend uncertainty estimation in deep learning to multivariate outputs. Salinas et al. [120] and Zhu et al. [160] deal with uncertainty estimation in time series forecasting using LSTM. Most frameworks assume independent Gaussian distributions for outputs and estimate both mean and standard deviation. The main challenge in learning uncertainty from data in deep learning lies in calibration of regression uncertainty [65]. The robust prediction of uncertainty from deep regression allows detection of unreliable outputs.

A related issue in human motion manifold learning is minimal representation of 3D joint angles (SO3) so that data-driven learning is not complicated by singularity, duality and discontinuity problems occurring in input or output data. Earlier work [89] has compared Euler angles, quaternions and more intuitive swing-twist representation. Also authors [159] show that 5-parameter and 6-parameter representations (instead of full 9-parameter rotation matrices) are always continuous and best for deep learning.

7.2. Contribution

Our work presented in this chapter focuses on deep learning based uncertainty aware framework that learns data-driven temporal and spatial priors in a latent manifold to perform constrained 3D human pose estimation from sparsely worn Magnetic Inertial Measurement units (MIMU) sensors as input. It is motivated by the fact that human motion has spatial and temporal constraints which may be learnt in latent space and such a latent space representation is then robust to missing or noisy information in the input space [71]. Our work differs from [52] and [146] in that it also reports the data-driven uncertainty estimation of 3D human pose from reduced sensors. It also does not use acceleration as input like [52], which is shown to improve generalization to real MIMU datasets. Also our work finds out that the 'jerkiness' which is reported by shallow networks approach [146] results from fixed finite temporal context. We train a bidirectional recurrent autoencoder and at inference time use a temporally skewed time window for real-time prediction with minimum 'jerkiness'.

Our trained model is found to be frame rate agnostic to input. A significant improvement that we make over [52] is use of 6D parameterization for input 3D orientation and exponential map for 3D joint angles at the output. The later allows us to predict the uncertainty in interpretable terms (radians) directly at the output of network without any post processing. Both rotation matrix and quaternion require an orthogonalization step and incorporation of upper limit of uncertainty is not straightforward in either representation, especially when a parameter is near unity.

Apart from uncertainty estimation, most salient aspect of our work is that we develop a robust input calibration and show that even a reduced orientation (comprising of only pitch/roll) from a sparse set of body worn MIMUs is 'sufficient' to estimate 3D human pose. Our model works in the inhomogeneous magnetic field, by ignoring the heading or yaw information but it

shows more uncertainty in this case. The estimation of kinematic uncertainty of 3D human pose obtained from sparse sensors may be used for uncertainty driven information fusion from other sensor modalities [42], readjust the body posture or to capture 'correct' contact interaction with environment [155].

The main contribution of this work are,

1. Data-driven uncertainty estimation of 3D human pose from reduced sensors.

2. A robust deep learning model which leads to a straightforward generalization to real sensor data, by training on *synthetic* data.

3. Though our work uses full orientation obtained using magnetometer as input, we also show that a reduced orientation (comprising only of pitch/roll) from a sparse set of body worn MIMUs is *sufficient* to estimate 3D human pose, but with increase in uncertainty. Thus dependence on magnetometer can be eliminated, which is desired in magnetically disturbed indoor environment or when IMU without magnetometers are used. The former problem is not addressed by [52, 146], but is a well-known limitation of i-Mocap.

4. As compared to [52, 146], our work shows that linear acceleration if used as input reduces the generalization to real sensors, due to different skeletal lengths and variable placement of sensors on real subjects. Our model achieves better performance in terms of generalization and accuracy than state-of-art [52] by not using the linear acceleration from sensors.

5. Also compared to existing state-of-art data-driven approaches [52, 146], we not only predict 3D human pose from a reduced number of sensors, but also provide a well calibrated estimate of uncertainty. To the best of our knowledge, our work is the first to implement and discuss uncertainty of 3D human pose estimation obtained using deep regression from sparse MIMUs.

6. Our work demonstrates a magnetometer robust 3D human pose estimation using reduced or sparse sensors for the first time. Previous work [143, 131, 84, 86, 73] has addressed the problem of magnetometer free i-Mocap in the context of full body worn sensors (typical 13-17 sensors).

7.3. Organization of Chapter

This chapter is organized in eight main sections. In the first two sections we have covered the background and our main contribution regarding inertial motion capture (i-Mocap) from sparse body worn MIMUs with associated uncertainty prediction of 3D human pose. Our data-driven model works with or without magnetometers. In section 7.4, we explain the detailed methodology including synthetic and real MIMU datasets, the sensor placement, input calibration and parameterization of input/output as well as approximate Bayesian model used in our approach. In section 7.5 we describe our implementation comprising of model architecture, pre

and post-processing, cost function for aleatoric uncertainty, training and inference with epistemic uncertainty. In section 7.6, we present the results on synthetic and real MIMU datasets and discuss performance. We also perform ablation studies like analyzing the performance of model with reduced 3-parameter input, effect of time window size and undersampling of frame rate at inference. Finally we describe the uncertainty estimation using both 6-parameter and 3-parameter input trained models and show that even with 3-parameter input (no magnetometer or yaw), our model is able to predict the 3D human pose but with increased uncertainty and error. We further investigate the covariate and domain shift of real MIMU datasets w.r.t. synthetic dataset (on which model is trained after data augmentation), in order to validate generalization performance. We show that while there is no significant covariate shift between real and synthetic MIMU datasets, domain shift does exist. Therefore we conclude that fine-tuning on temporal patterns which are under-represented in previous training, not only improves the error but also calibrate the uncertainty estimation from the model. Since we have found no significant magnetic interference in the real MIMU datasets we have used simulated magnetic interference to demonstrate better performance of reduced 3-parameter input model compared to that of 6-parameter model that employ full 3D orientation from sensors.

7.4. Proposed Methodology

In our work we train a deep bidirectional recurrent autoencoder to learn a rich set of temporal priors for human pose in latent space using SMPL dataset [80]. The model is driven using an input of five body segment orientations (left wrist, right wrist, left leg, right leg, head), normalized with respect orientation of the root of human skeleton. For training, these inputs are *synthesized* using forward kinematics from 3D mocap data. The model outputs a full human pose (joint angles) in SMPL space, less rotation and translation of root. Only major joint angles in SMPL pose are estimated, assuming no 3D rotation at other joints. The major SMPL joints include Left Hip, Right Hip, Spine0, Left Knee, Right Knee, Spine1, Spine2, Neck, Left Clavicle, Right Clavicle, Head, Left Shoulder, Right shoulder, Left Elbow and Right Elbow. The overview of our framework is shown in Figure 7.2. The trained model has then been tested on two real MIMU datasets in section 7.6. The results of our framework are also compared with existing frameworks (SIP [140] and DIP [52]) on same datasets. A robust input parameterization is further suggested to reduce the yaw drift/errors due to magnetometer in section 7.6.3. Ablation studies are performed to understand the limitations of our approach and to come up with a more optimal model in sections 7.6.4 to 7.6.9.

7.4.1. Synthetic Dataset

The deep learning for human motion requires an abundant dataset comprising of many subjects, varying movements and activities, at different temporal speeds and range of motion of human joints. Many motion capture datasets are available, but these do not use standardized 3D skeleton and to the best of our knowledge only few [136], [139] include data from inertial sensors with associated calibration. Thus [52] have used a large synthetic dataset developed based on
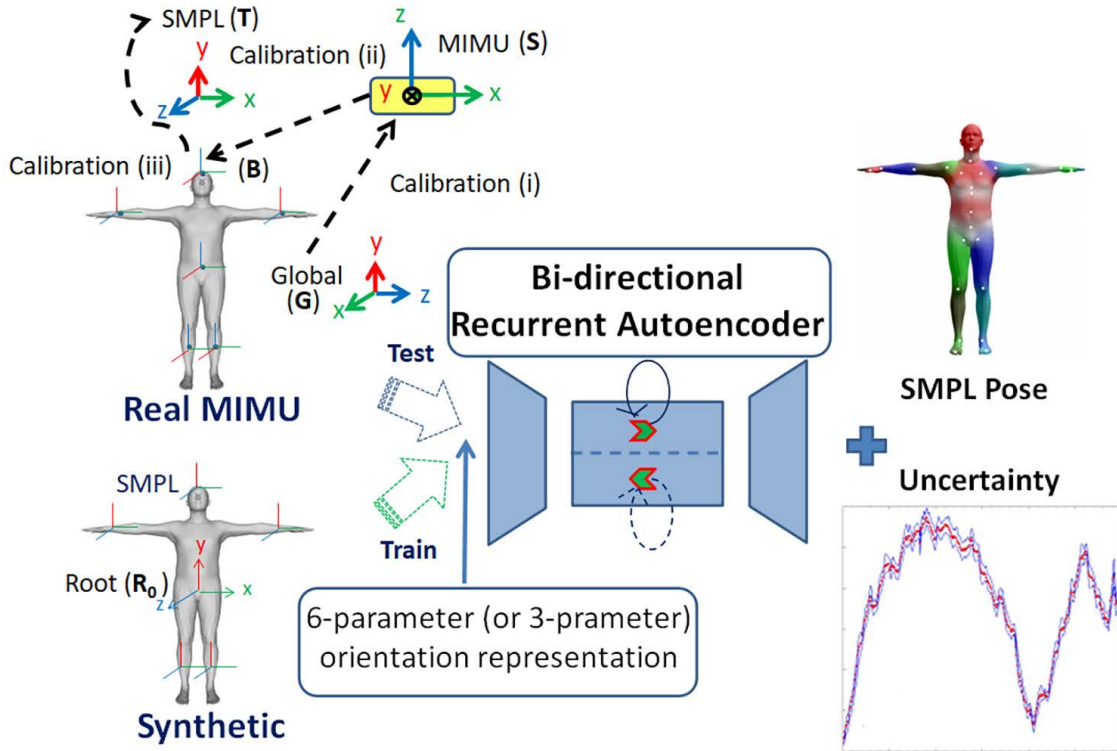
Figure 7.2.: Overview of our framework with key contributions.

SMPL for training their model. We have employed the synthetic dataset which is made publically available by [52]. But we checked this dataset for quality using [89] and also added data augmentation (as explained later) to achieve a more robust training of our models. This SMPL dataset includes Human3.6 [54], CMU human activities [22], Human Eva [126], Joint Limits [3] and eight other datasets all transformed to SMPL skeleton using AMASS framework [80]. The frame rate of standardized SMPL dataset is 60Hz. The body segment orientations are then *synthesized* using forward kinematics from root sensor. The orientation of root sensor is simply obtained from SMPL pose. Assuming that inertial sensors body frames are aligned with human body segment (i.e. sensor-to-segment alignment is identity), the *synthesized* body segment orientations then represent sensor orientations as well. For augmenting this ideal senor data, we have introduced both zero mean Gaussian white noise as well as random impulse noise to ideal sensor orientations. For each sequence in the dataset, we also introduced a drift in yaw angle of 3D orientation (based on random small bias value integrated over time). Our model was initially trained on raw ideal synthetic data and then fine-tuned on noise/ drift augmented data.

7.4.2. Real MIMU Sensor Datasets

Two real MIMU sensor datasets are used for testing our model trained on *synthetic* augmented dataset. DIP-IMU dataset has been provided by [52] as open source. Total capture MIMU

dataset [136] is a publically available dataset with MIMU orientations and calibration. Both these datasets have an advantage that [52] have performed testing on these and our results are thus directly comparable. Also in case of total capture dataset, SMPL ground truth pose are obtained using AMASS framework [80]. In comparison [146] has also used own real MIMU dataset, but their ground truth poses are not in SMPL and instead are based on biomechanical model of Xsens MVN. All these datasets are obtained using Xsens motion tracking MIMUs, hence apart from experimental conditions or calibration accuracy, the test results demonstrated on DIP-IMU and total capture would be applicable to [146] as well.

7.4.3. Sensor Placement

Six MIMU sensors are placed near left wrist, right wrist, left lower leg, right lower leg, lower spine and head. The sensor at lower spine is treated as reference or root sensor. The five sensors located at lower/upper limbs and head provide orientation measurements w.r.t. root sensor which is used to predict full 3D human pose. An alternate sensor configuration with lower legs sensors moved to feet and wrist sensors moved to hands, i.e. the end effectors of kinematic chain, is tested to be worse in performance.

Another interesting configuration is with sensors located on left/ right upper arms and left/right upper legs. Theoretically, this predicts shoulder joint and hip joint angles with perfect accuracy. However it is completely unable to constrain and predict the elbow and knee joints during arbitrary movements.

The acceleration readings from these sensors depend on their exact location w.r.t joint axis and center of rotation. Although [52] has used the sensor accelerations like their earlier work [140], we notice that the accelerations vary based on where the sensors are mounted on skeleton and hence are not a reliable input feature. In case of [140], they got better results with accelerations (SIP versus SOP), because they obtained SMPL model shape with laser scans. The training in [52] uses synthetic data from a standard SMPL skeleton and authors are able to generalize to real subjects with real sensors only after fine-tuning. Even on synthetic data, we show that the error obtained by [52] is more than our work, which only uses sensor orientations.

7.4.4. Calibration of Input Data

Both the *synthetic* and real MIMU 3D orientation data needs proper calibration before it can be used as input to the model. The orientation of the root of the human skeleton varies as subjects perform movement. But the overall 3D human pose is invariant to the rotation and translation of the root. Thus the orientation of five end effectors (left wrist, right wrist, left leg, right leg, head) is also normalized w.r.t. orientation of the root (base of the spine near hip) in SMPL according to (7.1), where $\mathbf{T}$ and $\mathbf{B}_i$ represents the SMPL frame of reference and reference frame attached to bone segment respectively as shown in Figure 7.2 and $\mathbf{R}_0$ is frame of reference fixed to root. $\mathbf{R}_t^{R_0 B_i} : \mathbf{B}_i \rightarrow \mathbf{R}_0$ is the rotation matrix from root frame to respective bone frame at time instant t and same convention holds for other rotation matrices representing orientations. As evident we need 3D orientations from root to SMPL and bone to SMPL to compute our

synthetic orientation input to the model using (7.1). This is trivial for *synthetic* data, since both the orientations can be obtained from SMPL ground truth pose using forward kinematics.

$$\mathbf{R}_t^{R_0 B_i} = (\mathbf{R}_t^{T R_0})^{-1}.\mathbf{R}_t^{T B_i} \tag{7.1}$$

But when real MIMU sensors are employed, the frame of reference of sensor $\mathbf{S}_i$ is not always aligned to reference frame of body segment $\mathbf{B}_i$. We need to compute a sensor-to-segment calibration matrix $\mathbf{R}^{B_i S_i}$: $\mathbf{S}_i \rightarrow \mathbf{B}_i$. This is done at startup time using a *static* calibration pose. We have implemented static I-pose (see chapter 5) for this calibration due to ease of implementation for elderly and functionally impaired subjects. More details are available in [12]. Also due to residual intra-sensor startup and calibration errors, each MIMU sensor after sensor fusion provides an orientation, $\mathbf{R}^{I_i S_i}$, where the *perceived* inertial frame I_i is slightly offset from actual global inertial frame of reference $\mathbf{G}$. Hence we also need to obtain $\mathbf{R}^{G I_i}$: $\mathbf{I}_i \rightarrow \mathbf{G}$ for each sensor at startup. The full calibration of real MIMU data then proceeds in following steps,

$$\mathbf{R}_t^{G S_i} = \mathbf{R}^{G I_i}.\mathbf{R}_t^{I_i S_i} \tag{7.2a}$$

$$\mathbf{R}_t^{G B_i} = \mathbf{R}_t^{G S_i}.(\mathbf{R}^{B_i S_i})^{-1} \tag{7.2b}$$

$$\mathbf{R}_t^{R_0 B_i} = (\mathbf{R}_t^{G R_0})^{-1}.\mathbf{R}_t^{G B_i} \tag{7.2c}$$

$$\bar{\mathbf{R}}_t^{R_0 B_i} = \mathbf{R}^{T G}.\mathbf{R}_t^{R_0 B_i} \tag{7.2d}$$

We obtain $\mathbf{R}^{TG}$ from root sensor as opposed to head sensor (unlike [52]) at initial body model calibration. This is found more robust to inter and intra subject variations.

As opposed to [52, 146], we have not used acceleration for reasons discussed before and later in results section, we demonstrate the advantage of this choice as well. Hence the calibration of acceleration is not discussed.

7.4.5. Input and Output Parameterization

Two different 3D angle parameterization have been used in our work for input body segment orientations and output human pose respectively and the motivation for using both is explained in this section.

Earlier work by [52] employed full 9-parameter rotation matrix for both input and outputs, whereas [145] have used quaternions. Both rotation matrix and quaternion require norm constraints since underlying degree of freedom (DoF) is only three in case of 3D rotation. This is accomplished as post-processing of output in [52] and [146] and thus cause additional jittering error. Moreover the input is over-parameterized in case of either quaternions (4 parameters) or rotation matrix (9 parameters). This redundancy both of input and output also increases the number of model parameters, which increase the training and test time by order of $O(n)$, where n is number of model parameters. On the other hand, if a minimal 3DoF parameterization is

used like Euler angles, it introduces gimbal lock, singularity and discontinuity issues in input space as highlighted in [89].

We have employed for input orientations, a 6-parameter representation (motivated by [159]). A complete 3x3 rotation matrix is over-parameterized. But cross product of any two rows and columns in a right handed order leads to third row/column. The authors in [159] chose first two columns for its 6D representation, since it uniquely determines the remaining column. Our 6-parameter representation (derived in Appendix A) is a simple extension of the fact that yaw, pitch and roll angles (which completely define 3DoF) can be conveniently derived from rotation matrix as follows (ZYX order),

$$\gamma = \tan^{-1}(\frac{r_{21}}{r_{11}}) \tag{7.3a}$$

$$\theta = \tan^{-1}(\frac{-r_{31}}{\sqrt{r_{32}^2 + r_{33}^2}}) \tag{7.3b}$$

$$\phi = \tan^{-1}(\frac{r_{32}}{r_{33}}) \tag{7.3c}$$

Here, γ, θ, ϕ are yaw, pitch and roll angles respectively and $r_{11}, r_{21}, r_{31}, r_{32}, r_{33}$ are components of 3x3 rotation matrix. As evident only 5 components of 3x3 rotation matrix are sufficient to obtain 3DoF. However to avoid an indeterminate case (Appendix A) we also include r_{22} to make a 6-parameter representation. Compared to [159], our 6-parameter representation can also be used to derive a reduced 3-parameter representation derived from (7.3) comprising of r_{31}, r_{32}, r_{33} which only account for pitch and roll angle information. Though reduced 3-parameter representation increases the ambiguity and uncertainty, the motivation of it is *justified* for reasons discussed later in section 7.6.3.

For output human pose (joint angles), we have directly used exponential map (3-parmeters) representation of SMPL. By SMPL definition of human skeleton, the joint angles in exponential map representation are always continuous owing to joint constraints (explained in Appendix B). It is not only a minimal DoF representation for human pose, but also allows us to learn uncertainty directly in quantitative terms in the output space. If either rotation matrices or quaternions are used for output, then post-processing would be needed to obtain uncertainty measure in radians or degrees. The calibration of uncertainty in later case would also be complicated.

7.4.6. Approximate Bayesian Model

Given a training dataset, $D = \{(x_t, y_t)^i\}_{i=1}^N$, where each training example $(x_t, y_t$) is a sequence or time series, we learn an approximate Bayesian model to infer probability distribution $P(y_t|x_t)$ of 3D human pose from a sequence of sensor orientations x_t . If we assume a Normal distribution with diagonal covariance for the full 3D human pose $P(y_t|x_t) = N_y(\mu_t, \sigma_t^2 I)$, we can write our model as in (7.4), where our model **F** performs an approximate Bayesian inference using p perturbations of its parameters θ to obtain mean pose μ_t and its diagonal covariance vector σ_t^2 , given a sequence of inputs x_t over a sliding time window $(t-n : t+m)$. Here, n are input samples from past and m are future samples; h_t represents the latent state of the model

for a given sequence of inputs.

$$N_y(\mu_t, \sigma_t^2 I) = \mathbf{F}_{\mu_t,\sigma_t}(\{x_t\}_{t-n}^{t+m}; h_t, \{\theta\}_{1:p}) \tag{7.4}$$

In deep learning framework, this model is implemented as a shallow (but wide) MLP neural network [146], a 1D CNN or a recurrent temporal network [52]. In case of former two, the input time window remains fixed and cannot be changed at inference. To retain flexibility at inference time and learn a compact model, we have implemented a recurrent model which propagates the latent state h_t forward-backward recurrently and compose $(h_t^F \circ h_t^B)$ at time t to get the output x_t. It is given as follows by equation,

$$N_y(\mu_t, \sigma_t^2 I) = \mathbf{F}_{\mu_t,\sigma_t}(\{x_t\}_{t-n}^{t+m} \circ \{h_t|h_{t+1}\}_{t+m}^{t-n}; \{\theta\}_{1:p}) \tag{7.5}$$

Consistent with previous literature [120, 160], the aleatoric uncertainty σ_t^2 is learnt directly as an output alongside the mean pose μ_t, as an attentive regularization term in negative log likelihood (NLL) cost function (to be described later). The aleatoric uncertainty adapts to variance of data in the domain of training set. The model (epistemic) uncertainty is obtained using approximate Bayesian inference as explained above and for that Monte-Carlo dropout (MC Dropout) [39] is used at inference time. Ensemble approach [69] is also possible for epistemic uncertainty but is computationally more intensive. The composition of two types of uncertainty (aleatoric and epistemic) is also described later in section 7.5.6.

7.5. Implementation

We have implemented our deep learning framework in Tensorflow 1.15.2 and Python 3.6 on a Desktop computer with GPU-Nvidia GTX 1060 and CUDA 10.1. The data pre- processing, preparation and results evaluation is performed using Matlab 2019a.

7.5.1. Model Architecture

For the model given in (7.5), we have implemented Bidirectional LSTM auto-encoder with two stacked hidden Bi-LSTM layers each of size 512 (the size of our latent state). The input to Bi-LSTM is fully connected layer of size 512 with a drop-out of 0.2 and ReLU function. A fully connected output layer after Bi-LSTM stacked layers is chosen with size 150 and linear output to obtain mean pose μ_t. The same layer with independent weights predicts the diagonal covariance vector σ_t. The overall architecture is shown in Figure 7.3. A forward pass on this bi-layer Bi-LSTM comprise of four sub-passes as explained in Figure 7.4.

3D human Pose
(15x3) + (15x3)

μ_t σ_t

Linear   Softmax

150

512

512

512

Sensors orientation (5x6 or 5x3)
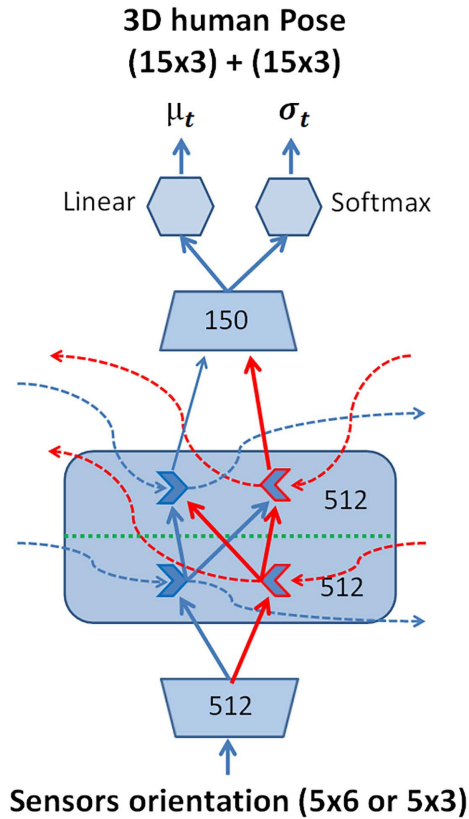
Figure 7.3.: Our Bidirectional Recurrent Autoencoder (BiRAE) Model.

7.5.2. Preprocessing

Huang et al. [52] employed and tested different normalization schemes on the input to their model. We have also tested such normalization of input, but no significant difference is noticed with or without normalization. Since we are only using orientations (and no accelerations), both the 6- parameter input representation and output 3D pose in exponential map (in radians) vary in a small range around zero, and normalization has no advantage.

7.5.3. Postprocessing

Of 24 joints in SMPL model, only 15 major joints are predicted by our model which includes: Left Hip, Right Hip, Spine0, Left Knee, Right Knee, Spine1, Spine2, Neck, Left Clavicle, Right Clavicle, Head, Left Shoulder, Right shoulder, Left Elbow and Right Elbow. Since the rest of joints are located forward of sensors positions at limbs, these are not predicted and substituted by unit rotation as a post- processing step to get the full 3D human skeleton.

We predict the 3D pose directly in exponential map parameterization of SMPL, and obtain the uncertainty directly in radian for each component of exponential map representation at the output. Unlike quaternions [146] or rotation matrices [52], the uncertainty prediction using our proposed representation is smooth and does not need Unscented transform in post-processing.

7.5.4. Cost Function for Aleatoric Uncertainty

The aleatoric uncertainty deals with covariate uncertainty found in the training data for a given model structure. If the test data also lies within the domain of training data, it is sufficient to use aleatoric uncertainty. We learn this uncertainty, using a *regularization* term in negative log likelihood (NLL) cost function that we use for our training as follows,

$$L_{NLL}(\theta) = \sum_{i=1}^{m} \sum_{t=t-n}^{t+m} ((\sigma_t^{(i)}(\theta)^{-2}||y_t^i - \mu_t^{(i)}(\theta)||^2 + \ln(\sigma_t^i(\theta)^2) \quad (7.6)$$

Here the inner sum is over the temporal subsequence $(t - n : t + m)$ and m is the number of subsequences in a batch. Since we assume only diagonal covariance $\sigma_t I$ for the full 3D human pose, our cost function is simplified compared with multivariate case of [116]. Also we assume that Euclidian distance approximate SO(3) error in exponential map space. This assumption is true for small errors upon convergence of training.

Huang et al. [52] includes sensor accelerations as additional input. To force the network learning on this predictor, [52] introduced an auxiliary task during training and their model was also forced to reconstruct the accelerations. Due to inherent problems with acceleration that we noted before, we have not used it as predictor. But we have used this *auxiliary* loss as motivation to reconstruct 3D angular rate readings when we opted for reduced 3-parameter representation, which only account for pitch and roll angle information. Discarding yaw information, we used 3D angular rate readings as additional input and also used auxiliary reconstruction loss on the later input; by this we got slightly better results (Table 7.1) than using 3-paremeter pitch/roll representation alone.

7.5.5. Training

The sequences in our synthetic dataset used for training have variable length. Due to limitation of GPU memory size, we use truncated Back propagation through time (BPTT), and divide our sequences in synthetic data into sub-sequences of length 300, discarding those with length ≤ 200, to avoid too much zero padding. We have randomly divided our synthetic data into training, validation and test set in 70/20/10 ratio. We also later perform testing on validation data, to understand better the poor learning of some pose subsequences.

Our model was trained using Adam algorithm with an initial learning rate of 0.001, exponentially decayed at rate of 0.9 with decay step 5000. Batch size for training was set at 16. Gradient clipping with a norm 1 was applied to Bi-LSTM training. The loss on validation set was used as early stopping criteria while training was set to max epoch size of 2000. The model with best error on validation set was saved during the training run.
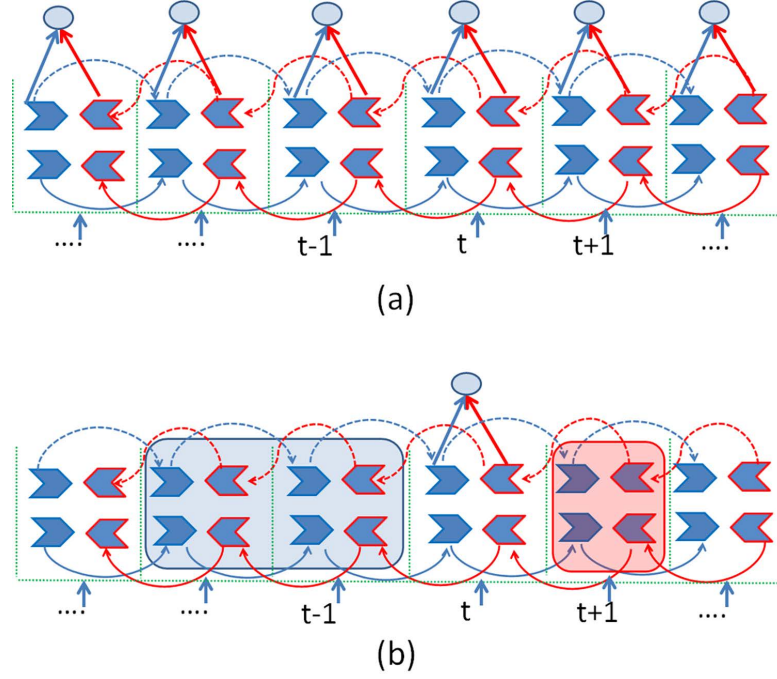
Figure 7.4.: (a) Forward pass in BiRAE (offline/ training). Blue are the forward cells. Red are backward cells. (b) Sliding Time Window Inference in BiRAE. The past (blue window) and future (red window) subsequence is used for real-time prediction of current time step. The size of future window and computation time define the latency.

7.5.6. Inference with Epistemic Uncertainty

We carry out the evaluation using the trained model in two ways. In offline end-to-end inference mode, we use a maximum batch size for which the test dataset is divisible by an integer, for fast evaluation. For inference, four end-to-end sequential sub-passes are performed on the trained Bidirectional LSTM model as depicted in Figure 7.4a (two passes per Bi-LSTM layer).

At each time t, we obtain two hidden vectors after four sub-passes and compose them ($h_t^F \circ h_t^B$) to obtain the output. This gives us mean pose μ_t and its diagonal covariance vector σ_t^2, given a sequence of inputs x_t. The later represent aleatoric uncertainty and hence we assign a superscript a in its symbol σ_t^a. Our model is trained using a dropout of 0.2 for regularization, and we use the same for Monte Carlo dropout (MCDropout) at the time of inference to obtain epistemic (model-based) uncertainty. This is essential to deal with out-of-domain data which was not seen in training. If $\mu_t^{(i)}(\theta_i)$ is the mean pose prediction for ith Monte Carlo iteration of model dropout , the we write for epistemic uncertainty,

$$\sigma_t^{e2} = \frac{1}{M} \sum_{i=1}^{M} [\mu_t^{(i)}(\theta_i) - \frac{1}{M} \sum_{i=1}^{M} \mu_t^{(i)}(\theta_i)]^2 \tag{7.7}$$

The epistemic uncertainty is then combined with aleatoric part to get overall uncertainty as follows,

$$\sigma_t^2 = \sigma_t^{e2} + \sum_{i=1}^{M} \sigma_t^{a2} \tag{7.8}$$

The problem with end-to-end bidirectional estimation as shown in Figure 7.4a is that it can only be carried out offline. For a real-time application, we define a sliding time window with past frames and future frames and only predict the output $(\mu_t, \sigma_t^2 I)$ at time t, after four sequential sub-passes are performed within that time window, as shown in Figure 7.4b. The advantage of our online approach is clearly evident over Shallow temporal convolution (TC) and multi-layer perceptron (MLP). Our BiRAE model can be trained end to end on sequences and then desired time window or unroll can be selected at the run-time (see also Figure 7.7). The former models take only fixed time window and hence their scope and performance is limited.

7.6. Results and Discussion

We analyze the baseline performance of our trained model first on the synthetic dataset to validate the extent of learning on *ideal* data, in order to choose the best performing architecture. Then we compare results of our best performing model on real MIMU data, mainly with state-of-art, DIP [52] who have also tested on the same real MIMU datasets, i.e. DIP-IMU and Total Capture [136]. The results of *offline* mode of model inference are presented in this part unless otherwise specified. Since code and data of [146] has not been publically made available and it does not use SMPL, it is left out of comparison. We also perform ablation studies for self-comparison between different variants of our model on real MIMU data. In particular we discuss and compare the results of 6-parameter and reduced 3-parameter representation and the significance of each. Next, the effect of real-time window length/ configuration is also discussed based on results of *online* mode versus *offline* mode. Lastly, we report the uncertainty estimation obtained using our framework and discuss it. We also discuss the effect of covariate and domains shift, and show that sensor noise in real MIMU data of DIP-IMU and Total capture is not significant to cause covariate shift. But the trained model performs poorly for those data sequences for which similar ones are absent in training data. This is identified as the main limitation of pose tracking based on sparse sensors. We also show how much fine tuning can help address this problem. Since the real MIMU data used in evaluation is not much perturbed, we create simulated magnetic perturbation in yaw part of 3D orientation and demonstrate the better performance obtained using 3- parameter model in this scenario.

The metrics we have used throughout is mean per joint angle error (MPJAE) or per joint angle error of individual joints. Positional error is not used for two reasons; first we are interested in 3D human pose which is agnostic to scale of the skeleton and thus can be used for biomechanical or activity ergonomics across subjects, Secondly we observe that the position of 3D joints alone loses the information of *twist* along a body segment [89] and hence is not useful for the above mentioned target applications focused by us.

7.6.1. Performance on Synthetic MIMU Data

After training our model, most important aspect to investigate was how well it performed on ideal sensors of *synthetic* dataset. This sets a baseline on which we can then evaluate the performance of real MIMU datasets. We performed this evaluation on 10% test set drawn at random from *synthetic* dataset which is not used in training. The comparison was made between a model trained on synthetic data by DIP [52] and variants of our method. The results in Table 7.1 demonstrate the effectiveness of 3D angle parameterization chosen by us for our model both for the input and output, which is different from [52] who chose 9-parameter rotation matrix for both and also included sensors acceleration. We also build a model that uses 6D parameterization of [159] for both input and output. In summary, results of our model with 6-parameter input and exponential map 3D pose are slightly better on synthetic data (our trained model performs much better on real MIMU data, as shown in next section) than DIP [52] and at par with using 6-parameters of [159]. It may be noted that our model parameters are also 20% less than [52]. The performance comparison is depicted in Table 7.1

Model	Distal	Tracking	Other
DIP (with Accel) [52]	12.30	7.17	9.18
Our (No Accel, 6-param) [159]	11.97	6.55	8.88
Our (No Accel, 6-param proposed)	**11.54**	**6.07**	**8.23**
Our (No Accel, 3-param proposed)	14.74	10.36	12.18
Our (No Accel, 3-param +rate)	14.37	10.13	11.91

Table 7.1.: Performance on Synthetic Data showing Mean Per Joint Absolute error (MPJAE) in degrees. DIP model [52] uses 9-parameter sensor orientation input and 9-parameter 3D human pose output, with acceleration. All Our models use exponential map 3D human pose. The input parameterization used is different as discussed in text. Last model also include rate of orientation change of a tracking sensor w.r.t root sensor. The model using 6-parameters of [159], uses this representation for both input/output. Distal Joints are shoulder and hip joints. Tracking joints are proximal to where sensors are located (Knee, Elbow, Head). Other include all others.

Once we evaluated the results on individual sequences in test data, we found that mean per joint angle error (MPJAE) is particularly high for certain sequences as shown in Figure 7.5 and it is uncorrelated with the length of sequences.

We investigated the failure cases further for 3D angle estimation of a joint as shown in Figure 7.6. It was seen that the model converges to correct 3D angle at the start of sequence and also predicts the accurate values for periodic movement of joint, however intermittently it shows large error. We infer that main limitation of 3D pose prediction from sparse sensors comes from the imbalance or absence of certain subsequences in the training dataset. Detailed reasons with experimental evaluation would be discussed later in section 7.6.8 on fine tuning.
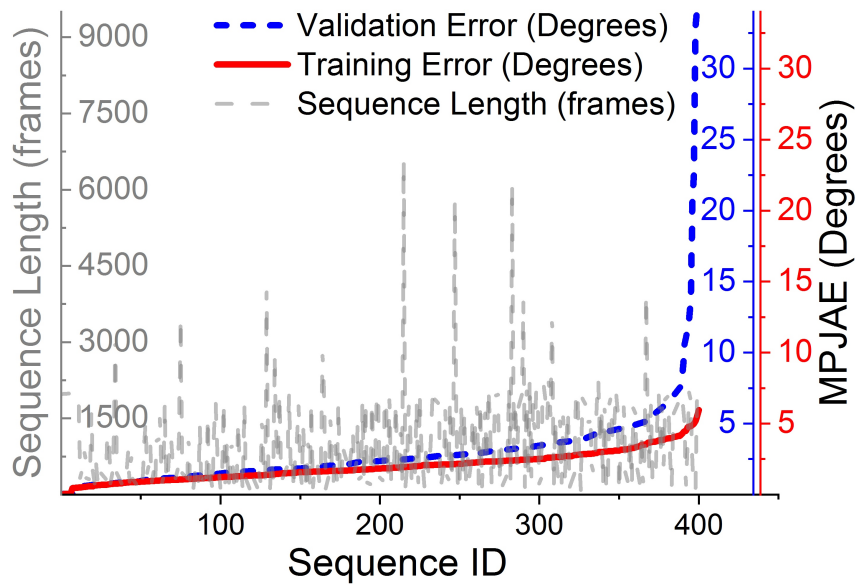
Figure 7.5.: Performance of Trained Model on synthetic Test Data (Blue dotted). For comparison, the error on a random subset of training data is also shown (red). The sequences are shown in order of ascending error. The x-axis is sequence number.
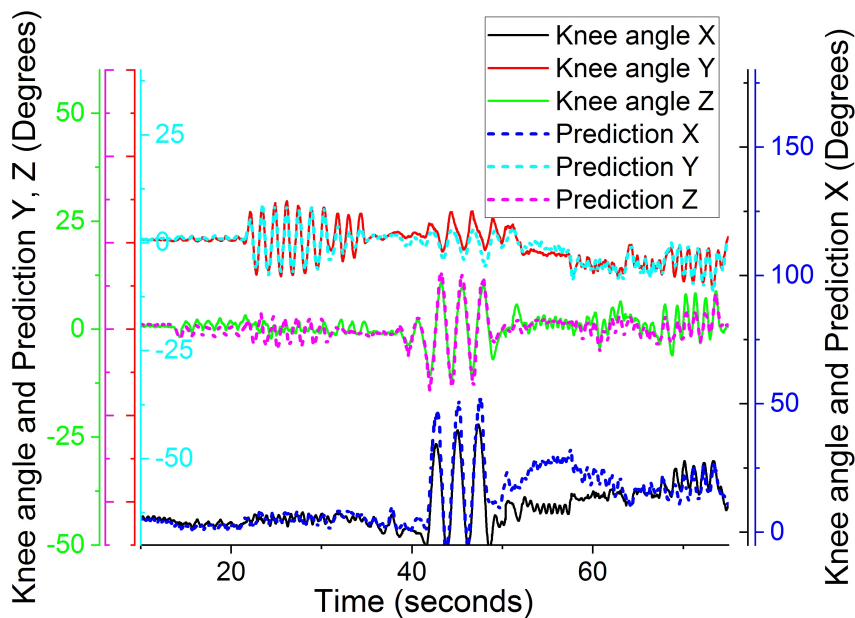
Figure 7.6.: Trained model performance on 3D angles of a Joint in Synthetic Test data. Intermittent failure on *unseen* sub-sequence is observed.

7.6.2. Performance on Real MIMU Data

The real MIMU dataset used in our evaluation are DIP-IMU and Total capture, we have used these datasets because these are publically available and two prior works [52] and [140] have reported results on the same datasets. For comparison of our trained model with [52] and [140] in this paper, we therefore use these datasets. The quality of sensor orientations obtained in these datasets has been evaluated by using forward kinematics on SMPL ground truth in either case. This point will be discussed in section 7.6.7 on covariate and domain shift. We report better accuracy with and without fine-tuning on real datasets compared with DIP [52] and SIP/SOP [140], and results are depicted in Table 7.2

Model	Data	Distal	Tracking	Other
DIP [52]	DIP-IMU(No Fine)	34.27	24.06	25.82
	DIP-IMU(Fine)	19.13	14.77	16.77
Our 6-param Model	DIP-IMU(No Fine)	29.15	4.55	10.92
	DIP-IMU(Fine)	18.47	4.81	9.28
Our 3-param Model	DIP-IMU(No Fine)	34.41	16.62	21.08
	DIP-IMU(Fine)	23.23	14.20	17.26
Our 6-param Model	Total capture(No Fine)	16.06	12.83	14.83
	Total capture(Fine)	15.21	4.16	10.17
Our 3-param Model	Total capture(No Fine)	18.49	15.33	17.04
	IDP-IMU(Fine)	17.23	10.11	14.23
SOP [140]	DIP-IMU		22.78	
	Total capture		22.18	
SIP [140]	Total capture(No Fine)		24.00	
	Total capture(Fine)		16.98	

Table 7.2.: Performance on Real MIMU Data showing Mean Per Joint Absolute error (MPJAE) in degrees. Distal Joints are Left/right shoulder and hip joints. Tracking joints are proximal to where sensors are located (Knee, Elbow, Head). Other include all other major joints predicted with the model.

The model errors in case of real MIMU data are large compared to *synthetic* training data. As would be shown in the sections 7.6.7 and 7.6.8 on fine tuning, it is attributable less to the sensor errors than to the lack of similar subsequences in the training data. We have earlier seen that *synthetic* test data also has such subsequences on which error is large, for these are under-represented in training data. We have trained our model on *synthetic* data augmented with noise and disturbances, which might explain its better performance without fine-tuning on real MIMUs. But disturbances such as magnetic conditions and bias integration are non-stationary in nature. It is shown (chapters 4 and 5, [12]) that sensor fusion might be affected greatly by magnetic disturbances, while the orientation errors due to body acceleration are bounded.

We therefore also tested a variant of our model with a reduced 3-parameter input from sensors which only represent pitch/roll, i.e. magnetometer or heading information was ignored. It will be discussed in next section. We however point out that DIP model [52] performs very poorly on DIP-IMU data without fine tuning, and attribute it to linear acceleration used by [52]. For training data, they used a standard body model and fixed sensor positions, which explain why the results on synthetic data in Table. 7.1 obtained by DIP [52] are comparable to our model. But in real scenario, linear accelerations vary based on where the sensors are mounted on skeleton and hence are not a reliable input feature. In case of [140], the authors got better results with accelerations (SIP versus SOP), because they obtained SMPL model shape with laser scans and measured sensor position accurately for the subjects.

7.6.3. Effect of Reduced 3-Parameter Input

Our model with 3-parameter sensor input takes into account only the pitch/roll information and drops the other components related to yaw angle information from 6-parameter representation. This provides intrinsic robustness against magnetic disturbances. Although magnetometers are still used for sensor fusion and proper calibration to fit in with framework, the yaw information is finally not available to the model for training or at inference time. The comparison of 3-parameter versus 6-parameter input on real MIMU datasets (Table 7.2), shows that the use of 3-parameter representation degrades the performance. The periodic activities like walking etc. are still predicted well using 3-parameter model. The advantage of using 3-parameter input model over 6- parameter is not obvious on real MIMU datasets used in this study, since these datasets are not highly perturbed by magnetic interference or a yaw angle drift (Figure 7.12). More challenging datasets (with SMPL ground truth and magnetically perturbed MIMU data) are not available at the time of this study. Later in this paper, we show a comparison on simulated strong magnetic interference on DIP-IMU to demonstrate the value of 3-parameter model in such situations.

We have further augmented the 3-parameter sensor pitch/roll input with 3D angular rate. It is derived as additional 3-vector from the difference of two consecutive 3D orientations in case of each of five sensors readings using quaternion formalism given in [13]. As explained in section 7.5.4, we use an auxiliary reconstruction loss on this later input; and by this we got better results than using 3- paremeter pitch/roll representation alone. Instead of raw sensor acceleration or rate gyro readings, we have used 3- parameter pitch/roll and 3-parameter angular rate normalized w.r.t root frame of reference, as described above. The advantage of this approach is that we need all input parameters in root sensor frame of reference and our body calibration procedure as implemented in section 7.4.4, works without any change.

7.6.4. Effect of Time Window at Inference

As pointed out earlier by [146], the main difference between shallow approaches using limited temporal context and recurrent neural networks is the *Jerkiness* that appears in the motion. While [52] has demonstrated that Bidirectional LSTM can be used for real-time 3D pose re-
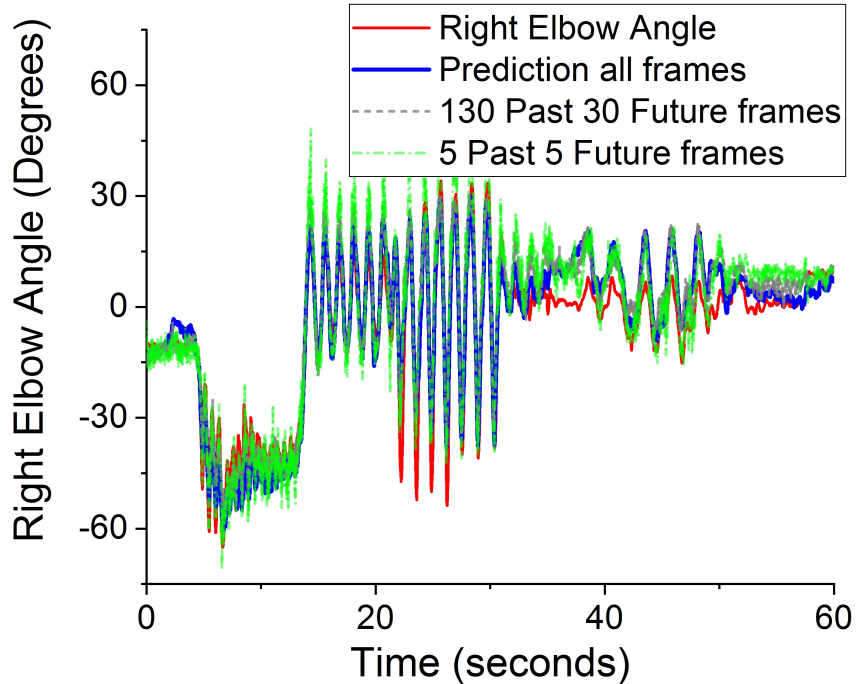
Figure 7.7.: Comparison of *Jerkiness* for Different Time Window Sizes (Online vs. Offline mode).

gression in online mode with limited context of past and future frames, they have not discussed the issue of *Jerkiness*. We also carry out this evaluation for our model and compare the online mode with offline mode for different configurations of time window in Figure 7.7.

We carried out a grid search in variable increments over a range of [0,500] of past/future frames used in pose prediction. A window size of [130,30] for past/future frames gives accuracy comparable to offline mode. The latency in this case is only 0.3 sec (at frame rate of 60Hz) plus computation time. The *Jerkiness* depicted in Figure 7.7 for this window configuration, is also acceptable. The end-to-end offline mode is however recommended for all non-real time applications.

7.6.5. Flexible Frame Rate at Inference

It is interesting to point out that at the time of inference our trained model is flexible to frame rate of input. We trained our 6-parameter model on *synthetic* dataset obtained at 60Hz. But at time of inference, we performed undersampling of recorded data from 60Hz to 30Hz and even 15Hz. We show that the degradation in performance is only gradual as reported in Table 7.3. We infer that the learned recurrent model is only acting by propagating latent state in time and is not affected by its actual rate. The gradual increase in error occurs from loss of high rate component. In comparison, the shallow MLP model proposed by [146] not only depends on exact window configuration but also the frame rate on which it is trained. There is no flexibility

to change either at the inference time.

Frame Down-sampling	Distal	Tracking	Other
Total capture - Original (60Hz)	16.06	12.83	14.83
Total capture - Down Sampled(30Hz)	16.67	13.08	15.11
Total capture - Down Sampled(15Hz)	18.60	14.59	16.72
Total capture - Down Sampled(6Hz)	20.67	17.64	19.62

Table 7.3.: Frame Rate Ablation Study. The error for Distal, Tracking and Other joints is shown in Mean Per Joint Absolute error (MPJAE) in degrees. Distal Joints are Left/right shoulder and hip joints. Tracking joints are proximal to where sensors are located (Knee, Elbow, Head). Other include all other major joints predicted with the model.

7.6.6. Uncertainty Estimation

The main feature of our work is estimation of 3D human pose uncertainty, while regressing from a sequence of sparse MIMU sensors. We estimate both aleatoric and epistemic uncertainty and then combine these in overall estimate using (7.8). Figure 7.8 depicts the output 3D angle with uncertainty ($\pm 1\delta$) bounds for a test sequence from synthetic data. It is clear that uncertainty on *synthetic* data from test set is predicted very well.
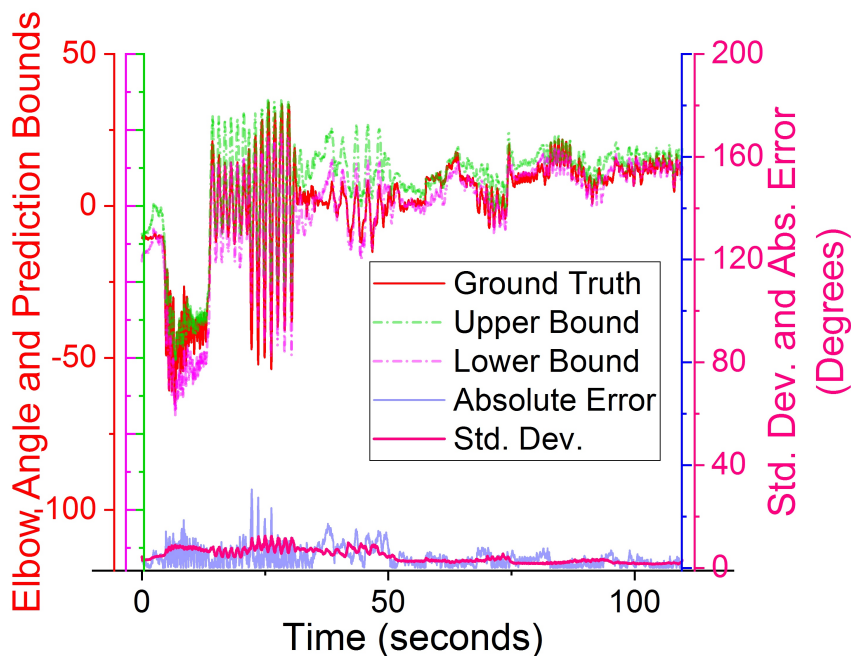
Figure 7.8.: Uncertainty estimation and Actual Error on Synthetic Test data sequence, using 6-parameter trained Model.

In Figure 7.9, we further show uncertainty ($\pm 1\delta$) bounds for a sequence from real Total capture dataset.
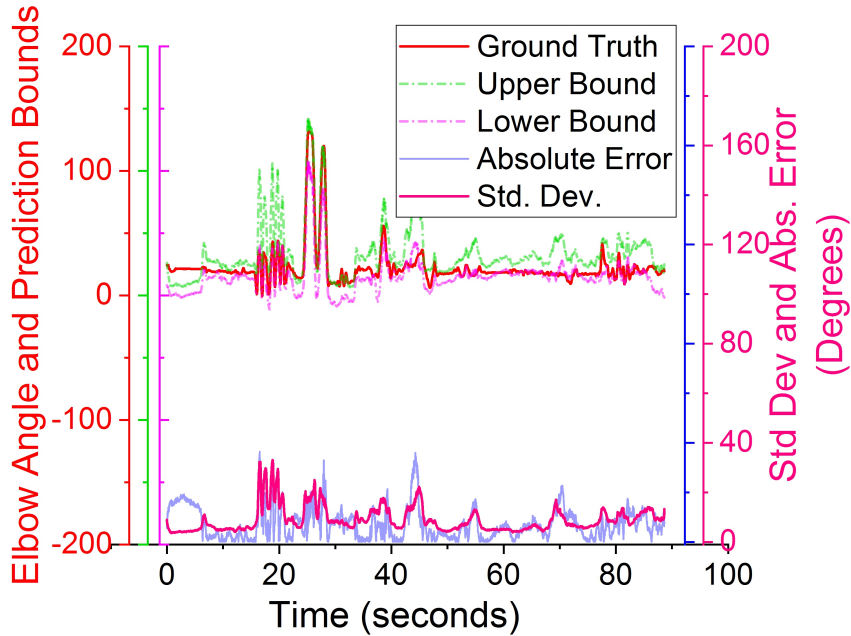
Figure 7.9.: Uncertainty estimation and Actual Error on Total capture data sequence, using 6-parameter trained Model.

Again our uncertainty learned using synthetic data scales well to real MIMU data. The investigation of uncertainty reveals that it correctly scales with the actual error. Since we assume a parametric model of uncertainty and predict Normal distribution with diagonal covariance for the full 3D human pose $N_y(\mu_t, \sigma_t^2 I)$, the actual error must lie within ($\pm 1\delta$) bounds, 68% of the time. We also note that both the synthetic sequence in Figure 7.8 and Total capture sequence in Figure 7.9, do not start at zero initial pose of SMPL (T-pose), yet model converges to low error after first few frames (at inference time, model hidden state is always initialized as zero at start of sequence). The uncertainty prediction during initial frames is not always accurate, but then it gets better. The uncertainty and absolute error is more when the amplitude of movement is more. The absolute error is stochastic (max: ± 20 deg), but the mean absolute error is low ($5°$). The model attempts to replicates the movement patterns.

For a comparison Figure 7.10 and 7.11 show the results obtained with 3-parameter model for the same sequences in *synthetic* and Total capture data respectively. We have noted in Table 7.1 and Table 7.2 that 3-parameter model performance is lower than 6-parameter model. We can also identify from Figure 7.10 and 7.11 that the uncertainty and error of prediction from 3-parmeter model is slightly higher, but it still predicts the movement and changes reasonable well. We demonstrate later in section 7.6.9 that in case the sensor data is highly perturbed by magnetic interference (yaw angle), the performance of 3- parameter model is more robust than 6-parameter model.
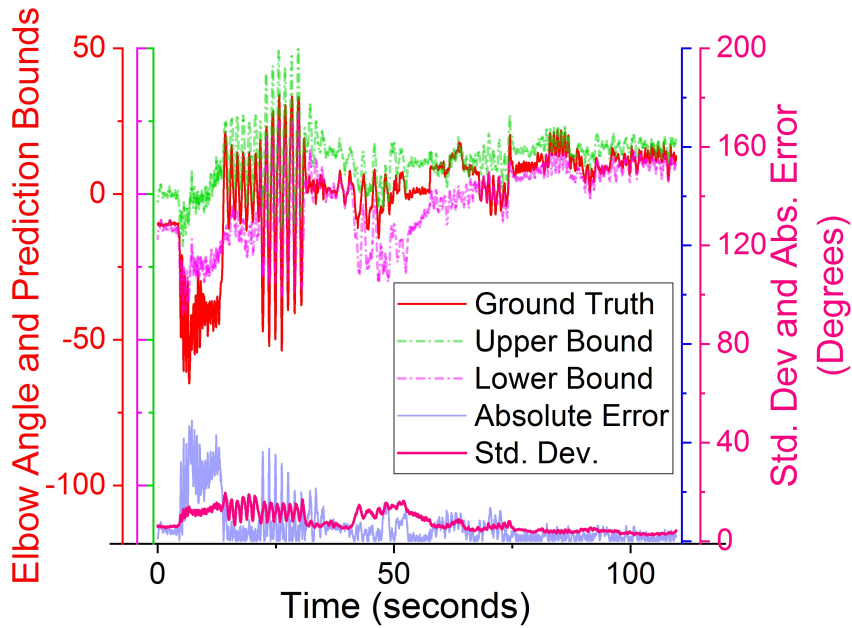
Figure 7.10.: Uncertainty estimation and Actual Error on Synthetic Test data sequence, using 3-parameter trained Model.
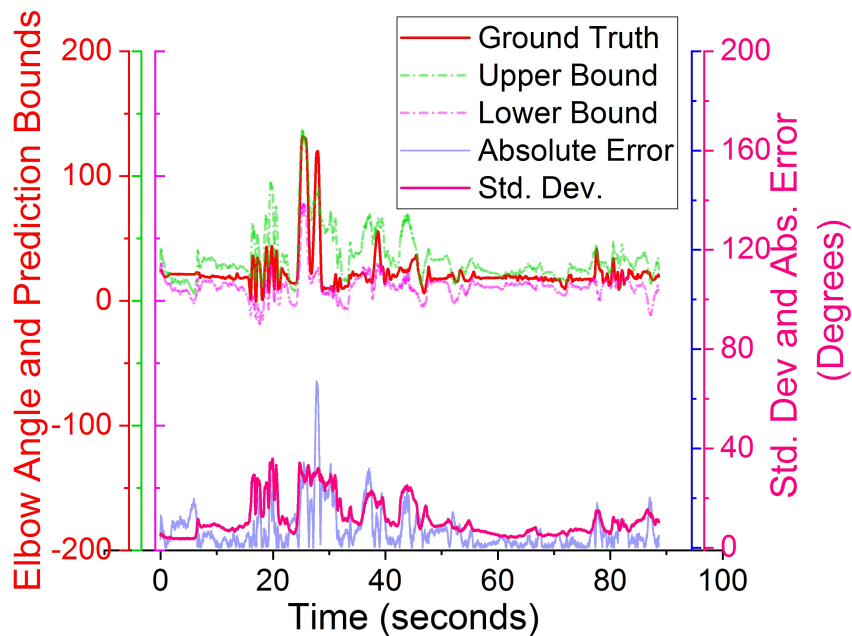
Figure 7.11.: Uncertainty estimation and Actual Error on Total capture data sequence, using 3-parameter trained Model.
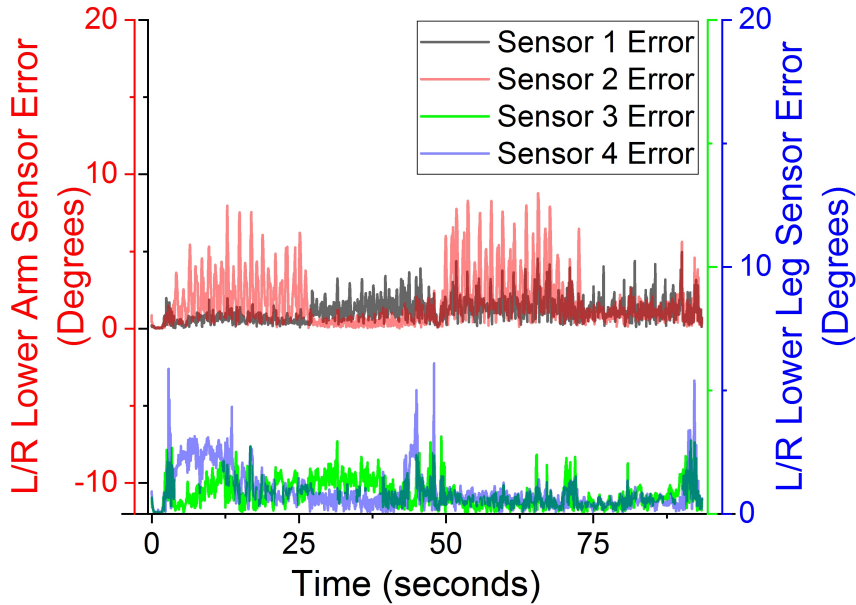
Figure 7.12.: The error between real and ideal Sensor Orientation for a sequence in DIP-IMU dataset.

7.6.7. Covariate and Domain Shift of Real Data

Real MIMU data is never the same as ideal sensor data we created in synthetic dataset. Although we did data augmentation of synthetic data as well by adding noise and disturbances, but in fact the real artifacts are non-stationary and more persistent. So it is obvious to investigate covariate and domain shift that exist in real data. Since we found that DIP-IMU dataset shows highest error when our model is not fine-tuned, we focused on this dataset for analysis.

We synthesized an ideal MIMU orientation, from ground truth pose available for DIP-IMU using forward kinematics. In order to check how the real sensor data differed from *synthetic* data on which we trained our model, we obtain the angular difference between the real and *synthetic* orientations of respective sensors in SO(3) space, using axis- angle metric. In Figure 7.12, we depict the angular difference between ideal sensor data that we should have for a given 3D pose and real sensor orientations for a typical sequence in DIP-IMU dataset. It is important to note that despite real sensors in DIP-IMU are perturbed, the nature and severity of perturbation is only 5-10 degrees and there is no drift. This alone does not explain the high error that we get on DIP- IMU, when we estimate 3D pose from this dataset using a synthetic data trained model. Although not shown here, but sensor errors on Total capture are also of same order, yet we obtain much better 3D pose estimation on Total capture, even without fine tuning.

7.6.8. Effect of Fine Tuning

Since no significant covariate shift was found out by comparing the real and ideal sensor data, we investigated the domain shift of input sequences. Earlier, we observed for synthetic test

data, that certain sequences displayed unusually high error (Figure 7.5). Similarly, we hypothesize that the temporal patterns in the DIP-IMU dataset are also different than the sequences on which our model is trained; therefore the error is more on DIP-IMU. In order to validate this hypothesis, we first tested the performance of a synthetic data trained model (without fine tuning) on real DIP-IMU data and then using the synthesized ideal DIP-IMU data. The results are depicted in Figure 7.13. It is evident that both the real and ideal MIMU data result in almost same error for different sequences in DIP-IMU test data. Therefore we conclude that the noise and data augmentation that we used during training, make the model robust against real sensor errors.
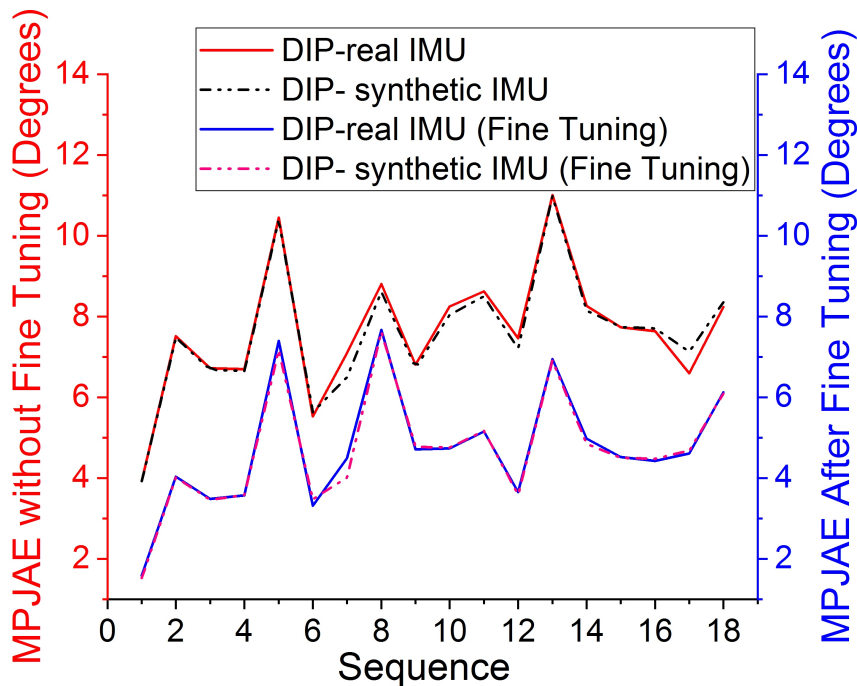
Figure 7.13.: Performance of Trained Model before and after Fine Tuning on DIP-IMU sequences.

Next, we fine-tuned the model on a subset of *synthetic* DIP-IMU data (not the same on which test results are depicted in Figure 7.13). Using this fine-trained model, we again tested for the error on DIP-IMU test sequences, using both ideal sensors and real sensor readings. The results again shown in Figure 7.13, clearly depict a decrease in error. Therefore we conclude that fine-tuning on a class of activities in a dataset, obtains better error performance on test data from the same, especially if the temporal patterns in such a dataset are under-represented in previous training.

It is also interesting to note the effect of fine-tuning on uncertainty estimation in case of DIP-IMU. Figure 7.14 shows the output 3D joint angle with uncertainty ($\pm 1\delta$) bounds for a test sequence from DIP-IMU, estimated using a trained model without fine-tuning. Unlike Total capture dataset, where the estimation error was lower and uncertainty also scaled well with the

estimation error (Figure 7.9), we observe that in case of DIP-IMU not only error is higher but also the uncertainty is underestimated.
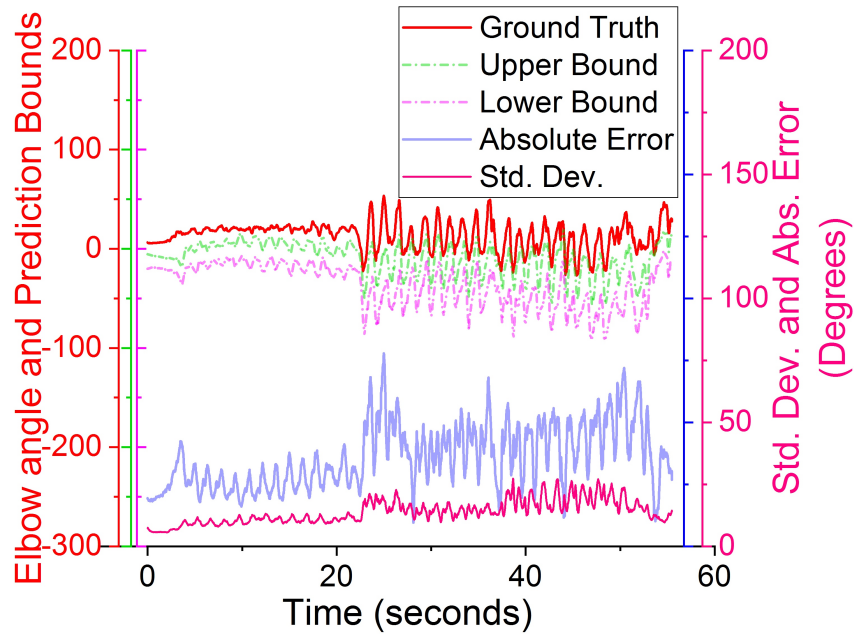
Figure 7.14.: Uncertainty estimation and Actual Error on DIP-IMU data sequence, using trained 6-param Model (without Fine-Tuning).
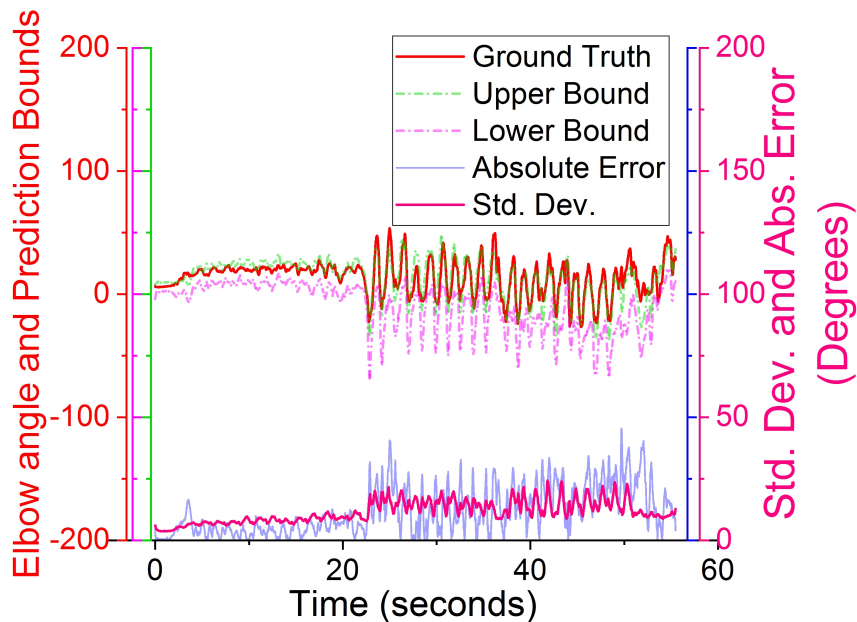
Figure 7.15.: Uncertainty estimation and Actual Error on DIP-IMU data sequence, using trained 6-param Model (after Fine-Tuning).

We then depict the results in Figure 7.15 for same sequence of DIP-IMU, obtained using a model fine-tuned on training data of DIP-IMU. Clearly not only the error has much reduced in this case, but uncertainty is also better calibrated now, after fine-tuning. As demonstrated, only a small training set even with *synthetic* sensor orientations is sufficient to achieve good accuracy and well-calibrated uncertainty on unseen temporal patterns. For instance, the ground truth for a set of training exercises can be obtained in an elaborate optical or inertial capture set-up. The *synthetic* sensor data can then prepared from ground truth poses. Once the pre-trained model is fine- tuned using this small dataset, it can then be deployed in arbitrary setting with real sensors (only six) for inference.

7.6.9. Simulated Magnetic Interference

We see from Figure 7.12 that DIP-IMU data is not strongly perturbed by magnetic interference (error between 5-10 degrees only). The covariate shift for Total capture data has been found out to be of same order. But in general the MIMU orientation is strongly affected by magnetic interference especially indoors and it can be quite significant [13]. Also when only IMU (rate gyro/accelerometer) is used without magnetometer, it leads to constant drift in yaw part of 3D angle. It is with these considerations in mind, that we trained a model on 3-parameter (pitch/roll) representation of sensor input. Since the two real MIMU datasets are not affected much by either yaw drift or strong interference, we tested the 3-parameter versus 6-parameter model on DIP-synthetic IMU data (as obtained in previous section), after it is corrupted by simulated [12] yaw drift and strong magnetic interference. The results of performance comparison are shown in Table 7.4. The models used are first fine-tuned for DIP-synthetic IMU data without magnetic perturbations.

Model	Distal	Tracking	Other
Our 6-param Model (Fine-tuned on DIP-IMU)	18.27	4.04	9.12
Our 3-param Model (Fine-tuned on DIP-IMU)	22.54	13.10	16.24
Our 6-param Model (Tested on DIP-IMU perturbed)	**45.70**	**24.02**	**35.13**
Our 3-param Model (Tested on DIP-IMU perturbed)	23.11	13.90	17.01

Table 7.4.: Performance under Magnetic Interference showing Mean Per Joint Absolute error (MPJAE) in degrees. Distal Joints are Left/right shoulder and hip joints. Tracking joints are proximal to where sensors are located (Knee, Elbow, Head). Other include all other major joints predicted with the model.

As evident from Table 7.4, when DIP-IMU data is perturbed by yaw drift and strong magnetic interference, the performance of 6-parameter model is much worse than 3-parameter model. There is almost no significant degradation in performance of 3-parameter model for perturbed (yaw) and unperturbed DIP-IMU data. However when perturbation is negligible (or none), 6-parameter model indeed performs better than 3-parameter model, as shown earlier.

7.7. Limitations

As discussed in section 7.6.8, the major limitation of our work is the dependence on temporal priors due to sparse sensors information. Although we used a synthetic dataset rich in activities and movements with data augmentation, it is not exhaustive for data-driven learning. We have find out that fine-tuning on a small subset of activities or movements on which prediction is required, addresses this limitation quite well. Another limitation of our work is that we did not have an inertial motion capture dataset with sequences collected in homogenous magnetic field as well as in highly perturbed magnetic environment. The DIP-IMU and Total Capture datasets have been found out to be very *clean* (both employ full 13-sensor MVN algorithm and then isolate 6-sensors data for evaluation). In real there would be only 6 sensors on the body and hence sensor fusion (orientation estimation) performed on individual nodes would not close the gap between synthetic and real MIMU data so well. In order to do more realistic testing, we had to simulate the magnetic interference to test the robustness of our proposed 3-parameter (pitch/roll) sensor input model. Although we have demonstrated its robustness vis a vis full 6-parameter sensor orientation in magnetically perturbed scenario, this still needs to be evaluated more with real perturbations. A dataset collection with SMPL ground truth in magnetically challenging environments is our next goal.

7.8. Conclusion

Our work proposes an uncertainty aware bi-directional deep recurrent model to estimate 3D human pose from only six magnetic-inertial measurement units (MIMUs). To the best of our knowledge, our model is the first to provide uncertainty estimation for this problem in deep learning framework. Our model output the 3D human pose directly in exponential map representation of SMPL. This avoids the renormalization of output as required in case of quaternions or rotation matrices. Also the estimation of uncertainty and its interpretation is straightforward. The definition of SMPL avoids the discontinuities in exponential map, owing to joint constraints. For sensor input, we propose a new 6-parameter representation for 3D orientation, which avoids the singularity and ambiguity in input space. In order to deal with magnetic perturbations, we further introduce a 3-parameter representation for input sensor orientation. This ignores the yaw part in 3D orientation. Our results show that even this reduced 3-parameter (pitch/roll) representation accomplishes 3D human pose estimation but with higher uncertainty. The uncertainty estimated as a part of our model output, is found to be well correlated with error. We test our model on two real MIMU datasets and show that the major limitation in sparse sensor based 3D human pose estimation is the need to train on representative motion sequences, on which prediction is required. Our model can be used both in offline mode for end-to-end bi-directional inference or in online-mode using a moving window over inputs at run-time.

8. Conclusion and Outlook

In this thesis, we have investigated the problem of inertial motion capture (i-Mocap) using sparsely worn MEMS Magnetic inertial measurement unit (MIMU) sensors in an end-to-end manner and systematically addressed the challenges at different levels. This was accomplished in the framework of data-driven deep learning and by improved adaptive filtering and calibration. Each MIMU node comprises of a triad of 3-axis magnetometer, 3-axis accelerometer and 3-axis rate gyro. Using these sensors, we obtain 3D human pose in terms of joint angles of articulated human skeleton. The objective was to arrive at a robust framework in which the intrinsic drift of MIMU sensors and extrinsic disturbances due to inhomogeneous magnetic field do not lead to an unbounded error in 3D human pose. We addressed five main research questions regarding i-Mocap in our thesis,

1. How can we improve the sensor fusion and calibration of individual MIMUs in the presence of inhomogeneous magnetic field, body accelerations and long term parameter drift, so that the uncertainty and error in estimated 3D orientation is reduced?

2. Can the sensor-to-segment offset and drift in 3D joint angles be dealt online by a dynamic anatomical correction in a sensor-to-body calibration framework?

3. Does deep learning provide an effective way to learn anatomical constraints on 3D joint angles from existing motion capture data?

4. Does the data-driven model that performs deep learning on motion capture (Mocap) datasets, implicitly learns the anatomical constraints in a latent manifold and thus allows predicting 3D human pose from a reduced set of MIMUs? Can deep learning also obtain an estimate of uncertainty of 3D human pose?

5. How good is the 3D human pose estimation based on data-driven approach from reduced sensors if we ignore the yaw angle obtained from magnetometers? Does the uncertainty of predicted 3D human pose increase in this case?

Our thesis took a systematic approach to this problem and found that the answer to first three questions had a direct impact on the deep learning performed to answer the last two research questions. Therefore we first addressed the various sources of errors in intrinsic sensor calibration, sensor fusion and extrinsic sensor-to-body segment calibration of MIMUs used in i-Mocap. For each part of the problem we proposed new algorithms in order to improve the pipeline. To conclude, our research answers each of the above research questions as follows,

1. We showed that a robust 3D orientation estimation is obtained from individual MIMUs under magnetic uncertainty and parameter variations using our proposed sensor fusion and calibration algorithms.

2. We also proposed an anatomical calibration framework to deal with joint angle drift correction from a pair of MIMUs online.

3. Subsequently we demonstrated that the joint constraints of human motion can be learned using Mocap data in a latent space. An intuitive swing-twist representation allowed us to synthesize a balanced dataset for learning and gave the best results.

4. In the end, we trained a bidirectional recurrent auto-encoder model for 3D human pose estimation from sparse (only six) body worn MIMUs. Our deep learning model also performed the uncertainty prediction of 3D human pose that it outputs. We showed that the main limitation of 3D human pose estimation from sparse body worn MIMU sensors lies in lack of representative motion in the training dataset. But this limitation can be overcome by fine-tuning the model on a small synthetic dataset of activities on which the prediction is required.

5. We trained and tested our model using both 6-parameter (full 3D orientation) and reduced 3-parameter (magnetometer free, no yaw) input from the sensors. The later representation was found to be more robust in magnetically inhomogeneous environment. Although the output error and predicted uncertainty is more when 3-parameter representation is used, but our model still predicts the 3D human pose robustly without any drift.

In summary, our research presented in this thesis confirms the hypothesis that a data-driven deep learning based model can learn to predict 3D human pose from sparsely worn MIMUs (placed on distal body segments). The reduced number of sensors means higher uncertainty and we address this by a data driven learning of temporal priors and joint constraints. We train our deep learning model on a large synthetic dataset and test it on real MIMU data. Our results demonstrate that the 3D pose estimation from reduced sensors using a learned model, depends greatly on the correlation in the activity. Thus the predictions are often inaccurate when inputs are outside the domain of training data. In such a case, an estimate of uncertainty of predicted 3D pose becomes important. Therefore our model takes the first step using deep learning framework to also predict the uncertainty of 3D pose predictions from sparse sensors. This uncertainty is found to be well correlated with actual prediction error in our experiments.

8.1. Future Directions

8.1.1. Fusion with Other 3D Pose Algorithms

Our approach provides a well calibrated uncertainty of predicted 3D pose from sparsely worn MIMUs. The data-driven estimation of kinematic uncertainty of 3D human pose may be used

for uncertainty driven information fusion with other modalities or with the output of other uncertainty based 3D human pose estimation algorithms. Therefore our work can be extended to take advantage of the uncertainty prediction performed in a deep learning framework.

8.1.2. More Challenging Datasets

The major limitation of data-driven model that estimate 3D human pose using sparse MIMUs is found to be the inadequate training data. We have already employed known Mocap datasets of 3D human pose made public by other authors in order to synthesize a comprehensive dataset of SMPL poses as well as MIMU orientations. Also two real MIMU datasets are used for testing and evaluation. Still we have found that our real MIMU data is not sufficiently perturbed by magnetic variations. This was a major impediment to demonstrate effectiveness of our approach using reduced 3-parameter representation (which excludes the yaw angle from magnetometer). In future a more diverse and expanded dataset of human motion capture with real MIMU readings, collected in a more challenging magnetic environment may further improve the deep learning model performance. Our approach in this thesis however obtains the baseline performance on the state-of-art datasets and also presents the groundwork for uncertainty aware deep learning for i-Mocap domain.

8.1.3. Sensor Fusion and Calibration Algorithms

In addition, the sensor fusion and calibration algorithms proposed in this research for MIMUs show better results compared to the state-of-art. These algorithms can be employed in related research on pedestrian navigation, unmanned autonomous vehicles (UAV) and robot navigation, visual-inertial fusion, human computer interaction (HCI), virtual and augmented reality (VR/AR) and other application areas where the MEMS MIMUs are already in use. This would also enable a more aggressive testing of our proposed algorithms in the associated domains.

8.1.4. Hardware Improvement

Although MIMU hardware design has not been the main subject of this thesis, yet working with the sensor calibration and fusion algorithms has provided some practical insights. Since all the wearable MIMU nodes have been packaged in a plastic body (in both EasyIMP and BIONIC), it is found out that the heat dissipation and temperature control must become an area of focus in future hardware development. For example, a non-magnetic heat sink may be used to efficiently dissipate the heat and thus maintain a constant temperature for MIMUs. This would improve the stability of bias and sensitivity parameters. Also the sensor calibration model and algorithm can be improved by explicitly including the temperature dependence.

8.1.5. Application Areas

The development of specific applications based on i-Mocap from sparse sensors was not the part of this thesis. However the sparse i-Mocap pipeline has been improved in this research and now

specific application areas can be targeted. Our future aim is to use the deep learning framework in the remote monitoring of rehabilitation exercises using sparse body worn MIMUs. Similarly the work can be extended towards human-computer interaction, sports biomechanics and virtual and augmented reality (VR/AR). The use of sparse body worn sensors is very attractive, as it makes this modality less obtrusive and more practical for 3D human pose estimation in the wild.

A. Appendix A

In this appendix, we derive our 6-parameter representation from 3x3 rotation matrix. We also demonstrate that theoretically, it is at par with 6-parameter representation of [159], but gives us an advantage that we can easily obtain a reduced 3-parameter representation from our 6-parameters, by masking three parameters in it.

The 3x3 rotation matrix is an over-complete representation of a 3D angle. There are six constraints on its 9 parameters, which reduce it to 3DoF. These constraints arise from vector cross-product of its three rows and columns. The authors [159] have used this fact to define a 6-parameter representation. They show that the remaining 3-parameters can be uniquely determined from the first two columns of a rotation matrix by a right handed cross-product.

Motivated by [159], we also define a 6-parameter representation using cross-product constraints. We first identify 5 parameters comprising first column and last row of a rotation matrix. Using a cross product of column 1 and 2 and 3 and 1 respectively, we obtain following

$$
\left(c_{21}c_{32} - c_{31}r_{22}\right) = r_{13} \tag{A.1a}
$$

$$
\left(c_{11}c_{32} - c_{31}r_{12}\right) = -r_{23} \tag{A.1b}
$$

$$
\left(c_{11}c_{22} - c_{21}r_{12}\right) = c_{33} \tag{A.1c}
$$

$$
\left(c_{31}c_{23} - c_{21}r_{22}\right) = c_{32} \tag{A.1d}
$$

where $c_{11}, c_{21}, c_{31}, c_{32}, c_{33}$ are the components of our 6-parameters representation which are same as the corresponding components of 3x3 matrix. The unknown components $r_{12}, r_{13}, r_{22}, r_{23}$ of rotation matrix can be unambiguously obtained from (A.1). The only exception is when $c_{11}, c_{21}, c_{31}, c_{32}, c_{33} := [0, 0, 1, 0, 0]$, where above equations converge to a single solution, whereas infinite number of solution exist. Therefore in order to resolve this ambiguity, we also include $c_{22} = r_{22}$ in our representation.

The advantage of our 6-parameter representation vis a vis [159] is that we can uniquely identify, the components which correspond to yaw, pitch and roll, as given by equation (7.3) and hence drop the components c_{11}, c_{21} and c_{22}, in order to obtain a yaw-free 3-parameter representation that we propose to be used for magnetically perturbed environments. The results in Table 7.1 (chapter 7) show that our 6-parameter representation is at par with 6-parameter representation proposed by [159].

B. Appendix B

In this appendix, we justify the use exponential map representation for the output 3D human pose of our model, instead of quaternion [146], rotation matrix [52] or 6-paramater [159]. We show that SMPL skeleton, owing to joint constraints of human body does not present any discontinuity in exponential map representation of 3D joint angles. Since demonstrating this rigorously for human joints is non-trivial, we chose the data-driven approach [89] and check if by performing extreme range of human motion, any of the 3D joint angle reach their limits $[-\pi, \pi]$ radians in exponential map representation of SMPL. We did this analysis for the complete SMPL dataset and found out that none of the joints ever reach the limits of $[-\pi, \pi]$ radians and that discontinuity does not occur. In Figure B.1 for clarity we show only the sequences in which extreme range of motion of a joint is performed. The SMPL data for this was obtained from both Joint Limit [3] and Total capture [136] , in which extreme motions are performed.
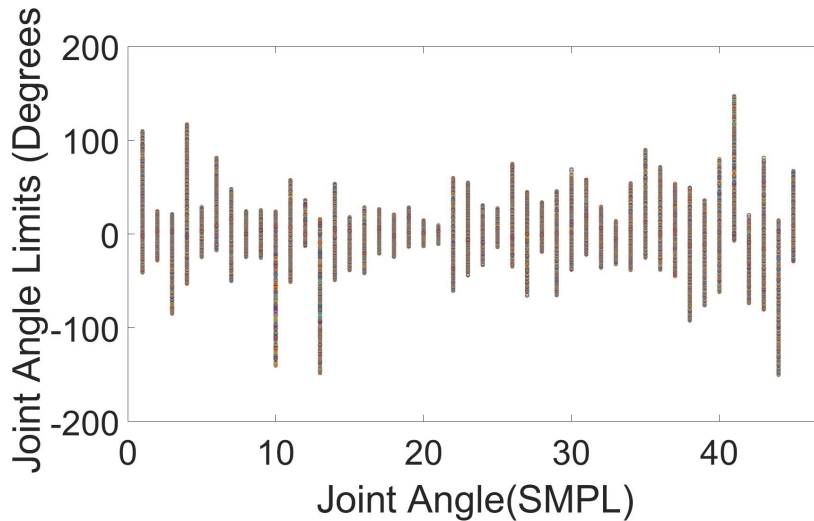
Figure B.1.: Joint Limits Constraint in Exponential Map (SMPL)

Clearly there is no discontinuity seen in SMPL data. The maximum range of motion occurs for x and y-component of Left/Right Knee and Elbow Joints in SMPL exponential map representation respectively but these are still within [-150,150] degrees. The shoulder and hip joints are well within [-100,100] degrees in SMPL exponential map representation.

There is an obvious advantage of regressing 3D human pose directly in exponential map representation of SMPL. No orthogonalization of model output is needed like quaternions or rotation matrix. . The uncertainty which is predicted as a part of output can be also directly
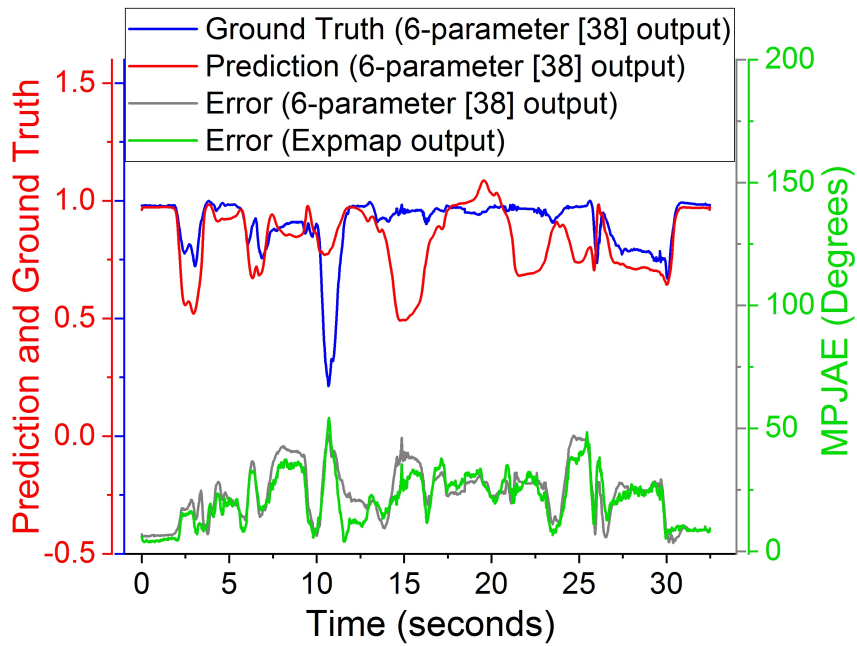
Figure B.2.: Exponential Map Output versus 6-parameter [38] Output

interpreted. If 3D angle output is estimated in quaternions or rotation matrix, it is observed that few components violate the constraint of unit norm (Figure B.2). The orthogonalization is therefore necessary as a post-processing or additional step. This on average leads to slightly more error compared to exponential map as shown in Figure B.2 .

List of Figures

List of Tables

Bibliography

[1] Xsens homepage - motion capture products- mvn awinda, mvn link. *12, 26, 111, 147*

[2] AHMED, E., ONORINA, K., PRAMOD, M., MUHAMMAD, J. N. M., AND STRICKER, D. Fully automatic multi-person human motion capture for vr applications. In *EuroVR (EuroVR-2018)* (2018). *99*

[3] AKHTER, I., AND BLACK, M. J. Pose-conditioned joint angle limits for 3d human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1446–1455. *27, 98, 99, 100, 101, 104, 105, 109, 110, 116, 145, 149*

[4] ALAM, F., ZHAIHE, Z., AND JIAJIA, H. A comparative analysis of orientation estimation filters using mems based imu. In *Proceedings of the International Conference on Research in Science, Engineering and Technology, Dubai, UAE* (2014), pp. 21–22. *25*

[5] ANDREWS, S., HUERTA, I., KOMURA, T., SIGAL, L., AND MITCHELL, K. Real-time physics-based motion capture with sparse sensors. In *Proceedings of the 13th European conference on visual media production (CVMP 2016)* (2016), pp. 1–10. *28*

[6] ARAGHI, G. Temperature compensation model of mems inertial sensors based on neural network. In *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, pp. 301–309. *23*

[7] BAERLOCHER, P., AND BOULIC, R. Parametrization and range of motion of the ball-and-socket joint. In *Deformable avatars*. Springer, 2001, pp. 180–190. *27, 99*

[8] BALAMUTA, J. J., MOLINARI, R., GUERRIER, S., AND YANG, W. A computationally efficient framework for automatic inertial sensor calibration. *IEEE Sensors Journal 18*, 4 (2017), 1636–1646. *24, 32, 48*

[9] BLESER, G., TAETZ, B., MIEZAL, M., CHRISTMANN, C. A., STEFFEN, D., AND REGENSPURGER, K. Development of an inertial motion capture system for clinical application: Potentials and challenges from the technology and application perspectives. *i-com 16*, 2 (2017), 113–129. *12, 28*

[10] BUTT, H. T., MUSAHL, M., SANCHEZ, M. A., MURTHY, P., ANKIT, K., AND STRICKER, D. Simultaneous end user calibration of multiple magnetic inertial measurement units with associated uncertainty. *IEEE Access 9* (2021), 26468–26483. *31, 67*

153

[11] BUTT, H. T., PANCHOLI, M., MUSAHL, M., MURTHY, P., SANCHEZ, M. A., AND STRICKER, D. Intelligent sensor fusion with online distributed mimu calibration for wearable motion capture (poster). In *2019 22th International Conference on Information Fusion (FUSION)*, IEEE, pp. 1–8. *23, 24, 61*

[12] BUTT, H. T., PANCHOLI, M., MUSAHL, M., MURTHY, P., SANCHEZ, M. A., AND STRICKER, D. Inertial motion capture using adaptive sensor fusion and joint angle drift correction. In *2019 22th International Conference on Information Fusion (FUSION)* (2019), IEEE, pp. 1–8. *83, 118, 127, 136*

[13] BUTT, H. T., PANCHOLI, M., MUSAHL, M., SANCHEZ, M. A., AND STRICKER, D. Development of high rate wearable mimu tracking system robust to magnetic disturbances and body acceleration. In *Proceedings of SAI Intelligent Systems Conference* (2019), Springer, pp. 1178–1198. *22, 38, 61, 65, 128, 136*

[14] BUTT, H. T., TAETZ, B., MUSAHL, M., SANCHEZ, M. A., MURTHY, P., AND STRICKER, D. Magnetometer robust deep human pose regression with uncertainty prediction using sparse body worn magnetic inertial measurement units. *IEEE Access 9* (2021), 36657–36673. *111*

[15] CEREATTI, A., TROJANIELLO, D., AND DELLA CROCE, U. Accurately measuring human movement using magneto-inertial sensors: techniques and challenges. In *2015 IEEE International Symposium on Inertial Sensors and Systems (ISISS) Proceedings* (2015), IEEE, pp. 1–4. *12, 28*

[16] CHAI, J., AND HODGINS, J. K. Performance animation from low-dimensional control signals. In *ACM SIGGRAPH 2005 Papers*. 2005, pp. 686–696. *28*

[17] CHENG, Y., AND SHUSTER, M. D. Improvement to the implementation of the quest algorithm. *Journal of Guidance, Control, and Dynamics 37*, 1 (2014), 301–305. *61*

[18] CHESNEAU, C.-I., HILLION, M., HULLO, J.-F., THIBAULT, G., AND PRIEUR, C. Improving magneto-inertial attitude and position estimation by means of a magnetic heading observer. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2017), IEEE, pp. 1–8. *71*

[19] CHIELLA, A. C., TEIXEIRA, B. O., AND PEREIRA, G. A. Quaternion-based robust attitude estimation using an adaptive unscented kalman filter. *Sensors 19*, 10 (2019), 2372. *25*

[20] CHOE, N., ZHAO, H., QIU, S., AND SO, Y. A sensor-to-segment calibration method for motion capture system based on low cost mimu. *Measurement 131* (2019), 490–500. *27*

[21] DAPONTE, P., DE VITO, L., RAPUANO, S., RICCIO, M., AND PICARIELLO, F. Compensating magnetic disturbances on marg units by means of a low complexity data fusion

algorithm. In *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings* (2015), IEEE, pp. 157–162. *26, 68, 71*

[22] DE LA TORRE, F., HODGINS, J., BARGTEIL, A., MARTIN, X., MACEY, J., COLLADO, A., AND BELTRAN, P. Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. *116*

[23] DE VRIES, W., VEEGER, H., BATEN, C., AND VAN DER HELM, F. Magnetic distortion in motion labs, implications for validating inertial magnetic sensors. *Gait & posture 29*, 4 (2009), 535–541. *62*

[24] DEL ROSARIO, M. B., KHAMIS, H., NGO, P., LOVELL, N. H., AND REDMOND, S. J. Computationally efficient adaptive error-state kalman filter for attitude estimation. *IEEE Sensors Journal 18*, 22 (2018), 9332–9342. *25*

[25] DONG, M., YAO, G., LI, J., AND ZHANG, L. Calibration of low cost imu's inertial sensors for improved attitude estimation. *Journal of Intelligent Robotic Systems* (2020), 1–15. *21*

[26] DORVEAUX, E., VISSIÃ¨RE, D., MARTIN, A.-P., AND PETIT, N. Iterative calibration method for inertial and magnetic sensors. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, IEEE, pp. 8296–8303. *23*

[27] DORVEAUX, E., VISSIÃ¨RE, D., AND PETIT, N. On-the-field calibration of an array of sensors. In *Proceedings of the 2010 American Control Conference*, IEEE, pp. 6795–6802. *24, 51, 56, 57, 151*

[28] DRAGANOVÃ¡, K., LAÅ¡Å¡Ã¡K, M., PRASLIÄKA, D., AND KÃ¡N, V. Attitude-independent 3-axis accelerometer calibration based on adaptive neural network. *Procedia Engineering 87* (2014), 1255–1258. *23*

[29] ECKHOFF, K., KOK, M., LUCIA, S., AND SEEL, T. Sparse magnetometer-free inertial motion tracking–a condition for observability in double hinge joint systems. *arXiv preprint arXiv:2002.00902* (2020). *28, 112*

[30] EL-DIASTY, M., AND PAGIATAKIS, S. Calibration and stochastic modelling of inertial navigation sensor errors. *Journal of Global Positioning Systems 7*, 2 (2008), 170–182. *24, 32, 48*

[31] EL-GOHARY, M., HOLMSTROM, L., HUISINGA, J., KING, E., MCNAMES, J., AND HORAK, F. Upper limb joint angle tracking with inertial sensors. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2011), IEEE, pp. 5629–5632. *28, 91, 112*

[32] ENGELL-NØRREGÅRD, M., AND ERLEBEN, K. Estimation of joint types and joint limits from motion capture data. *99, 101*

[33] EOM, H., CHOI, B., AND NOH, J. Data-driven reconstruction of human locomotion using a single smartphone. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 11–19. *28, 111*

[34] FAN, B., LI, Q., AND LIU, T. How magnetic disturbance influences the attitude and heading in magnetic and inertial sensor-based orientation estimation. *Sensors 18*, 1 (2018), 76. *26*

[35] FAN, B., LI, Q., WANG, C., AND LIU, T. An adaptive orientation estimation method for magnetic and inertial sensors in the presence of magnetic disturbances. *Sensors 17*, 5 (2017), 1161. *63, 68, 71*

[36] FANG, B., CHOU, W., AND DING, L. An optimal calibration method for a mems inertial measurement unit. *International Journal of Advanced Robotic Systems 11*, 2 (2014), 14. *22, 26*

[37] FIEDLER, F., BAUMBACH, D., BÖRNER, A., AND LUCIA, S. A probabilistic moving horizon estimation framework applied to the visual-inertial sensor fusion problem. In *2020 European Control Conference (ECC)* (2020), IEEE, pp. 1009–1016. *26*

[38] FROSIO, I., PEDERSINI, F., AND BORGHESE, N. A. Autocalibration of mems accelerometers. *IEEE Transactions on Instrumentation and Measurement 58*, 6 (2008), 2034–2041. *21, 38*

[39] GAL, Y., AND GHAHRAMANI, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (2016), pp. 1050–1059. *48, 50, 120*

[40] GAO, W., FANG, X., AND LIU, F. Analysis on the influence of three-axis turntable nonorthogonal error on gyro calibration of sins. In *2012 IEEE International Conference on Mechatronics and Automation*, IEEE, pp. 2429–2434. *22, 38*

[41] GARCIA, R. V., KUGA, H. K., AND ZANARDI, M. C. F. Unscented kalman filter applied to the spacecraft attitude estimation with euler angles. *Mathematical Problems in Engineering 2012* (2012). *25*

[42] GILBERT, A., TRUMBLE, M., MALLESON, C., HILTON, A., AND COLLOMOSSE, J. Fusing visual and inertial sensors with semantics for 3d human pose estimation. *International Journal of Computer Vision 127*, 4 (2019), 381–397. *114*

[43] GLUECK, M., OSHINUBI, D., SCHOPP, P., AND MANOLI, Y. Real-time autocalibration of mems accelerometers. *IEEE Transactions on Instrumentation and Measurement 63*, 1 (2013), 96–105. *23*

[44] GRAMMENOS, A., MASCOLO, C., AND CROWCROFT, J. You are sensing, but are you biased? a user unaided sensor calibration approach for mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2*, 1 (2018), 1–26. *31*

[45] GUERRIER, S., MOLINARI, R., AND SKALOUD, J. Automatic identification and calibration of stochastic parameters in inertial sensors. *Navigation: Journal of The Institute of Navigation 62*, 4 (2015), 265–272. *24, 32, 48*

[46] HAN, K., HAN, H., WANG, Z., AND XU, F. Extended kalman filter-based gyroscope-aided magnetometer calibration for consumer electronic devices. *IEEE Sensors Journal 17*, 1 (2016), 63–71. *23*

[47] HERDA, L., URTASUN, R., AND FUA, P. Hierarchical implicit surface joint limits for human body tracking. *Computer Vision and Image Understanding 99*, 2 (2005), 189–209. *27, 103, 108, 149*

[48] HERDA, L., URTASUN, R., FUA, P., AND HANSON, A. Automatic determination of shoulder joint limits using quaternion field boundaries. *The International Journal of Robotics Research 22*, 6 (2003), 419–436. *27, 110*

[49] HOLDEN, D., SAITO, J., AND KOMURA, T. A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 1–11. *29*

[50] HU, J.-S., AND SUN, K.-C. A robust orientation estimation algorithm using marg sensors. *IEEE Transactions on Instrumentation and measurement 64*, 3 (2014), 815–822. *26, 68, 71*

[51] HU, X., YAO, C., AND SOH, G. S. Performance evaluation of lower limb ambulatory measurement using reduced inertial measurement units and 3r gait model. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)* (2015), IEEE, pp. 549–554. *28, 91*

[52] HUANG, Y., KAUFMANN, M., AKSAN, E., BLACK, M. J., HILLIGES, O., AND PONS-MOLL, G. Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics (TOG) 37*, 6 (2018), 1–15. *12, 16, 28, 112, 113, 114, 115, 116, 117, 118, 120, 121, 122, 124, 125, 127, 128, 145*

[53] ILEWICZ, W., AND NAWRAT, A. *Direct method of IMU calibration.* Springer, 2013, pp. 155–171. *23*

[54] IONESCU, C., PAPAVA, D., OLARU, V., AND SMINCHISESCU, C. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence 36*, 7 (2014), 1325–1339. *99, 100, 101, 104, 105, 116, 149*

[55] KAICHI, T., MARUYAMA, T., TADA, M., AND SAITO, H. Resolving position ambiguity of imu-based human pose with a single rgb camera. *Sensors 20*, 19 (2020), 5453. 85

[56] KANAZAWA, A., BLACK, M. J., JACOBS, D. W., AND MALIK, J. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 7122–7131. 99

[57] KATIRCIOGLU, I., TEKIN, B., SALZMANN, M., LEPETIT, V., AND FUA, P. Learning latent representations of 3d human pose with deep neural networks. *International Journal of Computer Vision 126*, 12 (2018), 1326–1341. 29

[58] KIM, J., RYU, M.-H., CHOI, H., AND YANG, Y. Anatomy calibration of inertial measurement unit using a principle component analysis. *International Journal of Bio-Science and Bio-Technology 5*, 6 (2013), 181–190. 27

[59] KIVARANOVIC, D., JOHNSON, K. D., AND LEEB, H. Adaptive, distribution-free prediction intervals for deep neural networks. *arXiv preprint arXiv:1905.10634* (2019). 29

[60] KOK, M. *Probabilistic modeling for sensor fusion with inertial measurements*, vol. 1814. Linköping University Electronic Press, 2016. 26

[61] KOK, M., HOL, J., AND SCHÖN, T. An optimization-based approach to human body motion capture using inertial sensors. In *19th World Congress of the International Federation of Automatic Control (IFAC), Cape Town, South Africa, August 24-29, 2014* (2014), International Federation of Automatic Control, pp. 79–85. 26, 28

[62] KOK, M., HOL, J. D., AND SCHÖN, T. B. Using inertial sensors for position and orientation estimation. *Foundations and Trends in Signal Processing 11*, 1-2 (2017), 1–153. 26

[63] KOK, M., HOL, J. D., SCHÃ¶N, T. B., GUSTAFSSON, F., AND LUINGE, H. Calibration of a magnetometer in combination with inertial sensors. In *2012 15th International Conference on Information Fusion*, IEEE, pp. 787–793. 22, 26

[64] KOREIN, J. U. *A geometric investigation of reach*. MIT press, 1986. 27

[65] KULESHOV, V., FENNER, N., AND ERMON, S. Accurate uncertainties for deep learning using calibrated regression. *arXiv preprint arXiv:1807.00263* (2018). 29, 113

[66] KÃ¶PPE, E., AUGUSTIN, D., LIERS, A., AND SCHILLER, J. Self-calibration-method for an inertial navigation system with three 3d sensors. In *2014 International Symposium on Inertial Sensors and Systems (ISISS)*, IEEE, pp. 1–4. 23

[67] LAIDIG, D., MÜLLER, P., AND SEEL, T. Automatic anatomical calibration for imu-based elbow angle measurement in disturbed magnetic fields. *Current directions in biomedical engineering 3*, 2 (2017), 167–170. 27, 84, 91, 98

[68] LAIDIG, D., SCHAUER, T., AND SEEL, T. Exploiting kinematic constraints to compensate magnetic disturbances when calculating joint angles of approximate hinge joints from orientation estimates of inertial sensors. In *2017 International Conference on Rehabilitation Robotics (ICORR)* (2017), IEEE, pp. 971–976. *27, 84, 91, 98*

[69] LAKSHMINARAYANAN, B., PRITZEL, A., AND BLUNDELL, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems* (2017), pp. 6402–6413. *29, 112, 120*

[70] LAVES, M.-H., IHLER, S., FAST, J. F., KAHRS, L. A., AND ORTMAIER, T. Well-calibrated regression uncertainty in medical imaging with deep learning. In *Medical Imaging with Deep Learning* (2020). *29*

[71] LI, S., ZHOU, Y., ZHU, H., XIE, W., ZHAO, Y., AND LIU, X. Bidirectional recurrent autoencoder for 3d skeleton motion data refinement. *Computers & Graphics 81* (2019), 92–103. *29, 112, 113*

[72] LI, Y., NIU, X., ZHANG, Q., ZHANG, H., AND SHI, C. An in situ hand calibration method using a pseudo-observation scheme for low-end inertial measurement units. *Measurement Science and Technology 23*, 10 (2012), 105104. *23*

[73] LIGORIO, G., BERGAMINI, E., TRUPPA, L., GUAITOLINI, M., RAGGI, M., MANNINI, A., SABATINI, A. M., VANNOZZI, G., AND GAROFALO, P. A wearable magnetometer-free motion capture system: innovative solutions for real-world applications. *IEEE Sensors Journal* (2020). *28, 114*

[74] LIGORIO, G., AND SABATINI, A. M. Dealing with magnetic disturbances in human motion capture: A survey of techniques. *Micromachines 7*, 3 (2016), 43. *62*

[75] LIGORIO, G., ZANOTTO, D., SABATINI, A., AND AGRAWAL, S. A novel functional calibration method for real-time elbow joint angles estimation with magnetic-inertial sensors. *Journal of biomechanics 54* (2017), 106–110. *27*

[76] LIU, G., ZHANG, J., WANG, W., AND MCMILLAN, L. Human motion estimation from a reduced marker set. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games* (2006), pp. 35–42. *28*

[77] LIU, H., WEI, X., CHAI, J., HA, I., AND RHEE, T. Realtime human motion control with a small number of inertial sensors. In *Symposium on interactive 3D graphics and games* (2011), pp. 133–140. *112*

[78] LUO, S., PANG, H., ZHANG, Q., CHEN, D., PAN, M., AND LUO, F. Calibration strategy and generality test of three-axis magnetometers. *Measurement 46*, 10 (2013), 3918–3923. *22, 33, 38, 47*

[79] MADGWICK, S. O., HARRISON, A. J., AND VAIDYANATHAN, R. Estimation of imu and marg orientation using a gradient descent algorithm. In *2011 IEEE international conference on rehabilitation robotics* (2011), IEEE, pp. 1–7. *25, 62*

[80] MAHMOOD, N., GHORBANI, N., TROJE, N. F., PONS-MOLL, G., AND BLACK, M. J. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 5442–5451. *115, 116, 117*

[81] MAHONY, R., HAMEL, T., AND PFLIMLIN, J.-M. Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on automatic control 53*, 5 (2008), 1203–1218. *25, 62*

[82] MARTIN, H., GROVES, P., AND NEWMAN, M. The limits of in-run calibration of mems and the effect of new techniques. The Institute of Navigation. *26*

[83] MARTIN, H., GROVES, P., AND NEWMAN, M. The limits of inârun calibration of mems inertial sensors and sensor arrays. *NAVIGATION: Journal of The Institute of Navigation 63*, 2 (2016), 127–143. *23, 31*

[84] MARUYAMA, T., TADA, M., SAWATOME, A., AND ENDO, Y. Constraint-based real-time full-body motion-capture using inertial measurement units. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2018), IEEE, pp. 4298–4303. *28, 114*

[85] MEHTA, D., SRIDHAR, S., SOTNYCHENKO, O., RHODIN, H., SHAFIEI, M., SEIDEL, H.-P., XU, W., CASAS, D., AND THEOBALT, C. Vnect: Real-time 3d human pose estimation with a single rgb camera. *ACM Transactions on Graphics (TOG) 36*, 4 (2017), 44. *99*

[86] MIEZAL, M., BLESER, G., SCHMITZ, N., AND STRICKER, D. A generic approach to inertial tracking of arbitrary kinematic chains. In *Proceedings of the 8th international conference on body area networks* (2013), pp. 189–192. *26, 28, 114*

[87] MIEZAL, M., TAETZ, B., AND BLESER, G. Real-time inertial lower body kinematics and ground contact estimation at anatomical foot points for agile human locomotion. In *2017 IEEE international conference on robotics and automation (ICRA)* (2017), IEEE, pp. 3256–3263. *28, 91*

[88] MOUSAS, C. Full-body locomotion reconstruction of virtual characters using a single inertial measurement unit. *Sensors 17*, 11 (2017), 2589. *28, 111*

[89] MURTHY, P., BUTT, H. T., HIREMATH, S., KHOSHHAL, A., AND STRICKER, D. Learning 3d joint constraints from vision-based motion capture datasets. *IPSJ Transactions on Computer Vision and Applications 11*, 1 (2019), 1–9. *29, 97, 113, 116, 119, 124, 145*

[90] NASSAR, S., AND EL-SHEIMY, N. Accuracy improvement of stochastic modeling of inertial sensor errors. *Zeitschrift tfur Geodäsie Geoinformation und Landmanagement (ZfV) 130* (2005), 146–155. *24, 32, 48*

[91] NAZARAHARI, M., AND ROUHANI, H. Semi-automatic sensor-to-body calibration of inertial sensors on lower limb using gait recording. *IEEE Sensors Journal 19*, 24 (2019), 12465–12474. *27*

[92] NAZARAHARI, M., AND ROUHANI, H. Adaptive gain regulation of sensor fusion algorithms for orientation estimation with magnetic and inertial measurement units. *IEEE Transactions on Instrumentation and Measurement 70* (2020), 1–13. *63*

[93] NEMEC, D., JANOTA, A., HRUBOÅ¡, M., AND Å IMÃ¡K, V. Intelligent real-time mems sensor fusion and calibration. *IEEE Sensors Journal 16*, 19 (2016), 7150–7160. *23, 26*

[94] NEZ, A., FRADET, L., LAGUILLAUMIE, P., MONNET, T., AND LACOUTURE, P. Comparison of calibration methods for accelerometers used in human motion analysis. *Medical engineering physics 38*, 11 (2016), 1289–1299. *21, 38*

[95] NILSSON, J.-O., SKOG, I., AND HÃ¤NDEL, P. Aligning the forcesâeliminating the misalignments in imu arrays. *IEEE Transactions on Instrumentation and Measurement 63*, 10 (2014), 2498–2500. *24, 51, 56, 57, 151*

[96] OLIVARES, A., OLIVARES, G., GORRIZ, J., AND RAMIREZ, J. High-efficiency low-cost accelerometer-aided gyroscope calibration. In *2009 International Conference on Test and Measurement*, vol. 1, IEEE, pp. 354–360. *22*

[97] OLSSON, F., KOK, M., HALVORSEN, K., AND SCHÃ¶N, T. B. Accelerometer calibration using sensor fusion with a gyroscope. In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, IEEE, pp. 1–5. *23, 26*

[98] PACHER, L., CHATELLIER, C., VAUZELLE, R., AND FRADET, L. Sensor-to-segment calibration methodologies for lower-body kinematic analysis with inertial sensors: A systematic review. *Sensors 20*, 11 (2020), 3322. *27*

[99] PANCHOLI, M., DIMITROV, S., SCHMITZ, N., LAMPE, S., AND STRICKER, D. Relative translation and rotation calibration between optical target and inertial measurement unit. In *International Conference on Sensor Systems and Software* (2016), Springer, pp. 175–186. *72*

[100] PAPAFOTIS, K., AND SOTIRIADIS, P. P. Exploring the importance of sensors' calibration in inertial navigation systems. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (2020), IEEE, pp. 1–4. *21*

[101] PAPAFOTIS, K., AND SOTIRIADIS, P. P. Multiple accelerometers and magnetometers joint calibration and alignment. *IEEE Sensors Letters 4*, 3 (2020), 1–4. *23*

[102] PARK, S., HWANG, J., AND KWAK, N. 3d human pose estimation using convolutional neural networks with 2d pose information. In *European Conference on Computer Vision* (2016), Springer, pp. 156–169. *99*

[103] PARK, Y., MOON, S., AND SUH, I. H. Tracking human-like natural motion using deep recurrent neural networks. *arXiv preprint arXiv:1604.04528* (2016). *29*

[104] PAVLLO, D., GRANGIER, D., AND AULI, M. Quaternet: A quaternion-based recurrent model for human motion. *CoRR abs/1805.06485* (2018). *99*

[105] PERETROUKHIN, V., WAGSTAFF, B., AND KELLY, J. Deep probabilistic regression of elements of so (3) using quaternion averaging and uncertainty injection. In *CVPR Workshops* (2019), pp. 83–86. *29, 39, 113*

[106] PETKOV, P., AND SLAVOV, T. Stochastic modeling of mems inertial sensors. *Cybernetics and information technologies 10*, 2 (2010), 31–40. *24, 32, 48*

[107] PHUONG, N. H. Q., KANG, H.-J., SUH, Y.-S., AND RO, Y.-S. A dcm based orientation estimation algorithm with an inertial measurement unit and a magnetic compass. *Journal of Universal Computer Science 15*, 4 (2009), 859–876. *25*

[108] PYLVÃ¤NÃ¤INEN, T. Automatic and adaptive calibration of 3d field sensors. *Applied Mathematical Modelling 32*, 4 (2008), 575–587. *23*

[109] QUINCHIA, A. G., FALCO, G., FALLETTI, E., DOVIS, F., AND FERRER, C. A comparison between different error modeling of mems applied to gps/ins integrated systems. *Sensors 13*, 8 (2013), 9549–9588. *24, 32, 48*

[110] RADI, A., BAKALLI, G., GUERRIER, S., EL-SHEIMY, N., SESAY, A. B., AND MOLINARI, R. A multisignal wavelet variance-based framework for inertial sensor stochastic error modeling. *IEEE Transactions on Instrumentation and Measurement 68*, 12 (2019), 4924–4936. *24, 32, 48*

[111] RADI, A., SHETA, B., NASSAR, S., ARAFA, I., YOUSSEF, A., AND EL-SHEIMY, N. Accurate identification and implementation of complicated stochastic error models for low-cost mems inertial sensors. In *2020 12th International Conference on Electrical Engineering (ICEENG)* (2020), IEEE, pp. 471–475. *24, 32, 48*

[112] RAZAVI, H., SALARIEH, H., AND ALASTY, A. Towards real-time partially self-calibrating pedestrian navigation with an inertial sensor array. *IEEE Sensors Journal 20*, 12 (2020), 6634–6641. *31*

[113] RICCI, L., TAFFONI, F., AND FORMICA, D. On the orientation error of imu: investigating static and dynamic accuracy targeting human motion. *PLoS One 11*, 9 (2016), e0161940. *75*

[114] ROBERT-LACHAINE, X., MECHERI, H., LARUE, C., AND PLAMONDON, A. Effect of local magnetic field disturbances on inertial measurement units accuracy. *Applied ergonomics 63* (2017), 123–132. *25, 71*

[115] ROBERT-LACHAINE, X., MECHERI, H., MULLER, A., LARUE, C., AND PLAMONDON, A. Validation of a low-cost inertial motion capture system for whole-body motion analysis. *Journal of Biomechanics 99* (2020), 109520. *31*

[116] RUSSELL, R. L., AND REALE, C. Multivariate uncertainty in deep learning. *arXiv preprint arXiv:1910.14215* (2019). *29, 32, 40, 47, 49, 113, 122*

[117] RYZHKOV, L., AND STEPURENKO, D. Least squares criterion using to improve the triad algorithm. In *2012 2nd International Conference" Methods and Systems of Navigation and Motion Control"(MSNMC)* (2012), IEEE, pp. 106–108. *61*

[118] SABATINI, A. M. Estimating three-dimensional orientation of human body parts by inertial/magnetic sensing. *Sensors 11*, 2 (2011), 1489–1525. *25, 62*

[119] SABATINI, A. M. Kalman-filter-based orientation determination using inertial/magnetic sensors: Observability analysis and performance evaluation. *Sensors 11*, 10 (2011), 9182–9206. *26, 62*

[120] SALINAS, D., FLUNKERT, V., GASTHAUS, J., AND JANUSCHOWSKI, T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting 36*, 3 (2020), 1181–1191. *29, 113, 120*

[121] SARCEVIC, P., PLETL, S., AND KINCSES, Z. Examining the number of required stationary orientations for efficient accelerometer calibration. In *4th International Conference and Workshop Mechatronics in Practice and Education*, pp. 38–41. *22*

[122] SCHOPP, P., GRAF, H., BURGARD, W., AND MANOLI, Y. Self-calibration of accelerometer arrays. *IEEE Transactions on Instrumentation and Measurement 65*, 8 (2016), 1913–1925. *24*

[123] SCHWARZ, L. A., MATEUS, D., AND NAVAB, N. Discriminative human full-body pose estimation from wearable inertial sensor data. In *3D physiological human workshop* (2009), Springer, pp. 159–172. *28, 112*

[124] SEEL, T., SCHAUER, T., AND RAISCH, J. Joint axis and position estimation from inertial measurement data by exploiting kinematic constraints. In *2012 IEEE International Conference on Control Applications* (2012), IEEE, pp. 45–49. *26, 27, 84, 91, 98*

[125] SHAO, W., AND NG-THOW-HING, V. A general joint component framework for realistic articulation in human characters. In *Proceedings of the 2003 symposium on Interactive 3D graphics* (2003), ACM, pp. 11–18. *27*

[126] SIGAL, L., BALAN, A. O., AND BLACK, M. J. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision 87*, 1-2 (2010), 4. *116*

[127] STEBLER, Y., GUERRIER, S., SKALOUD, J., AND VICTORIA-FESER, M.-P. A framework for inertial sensor calibration using complex stochastic error models. In *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium* (2012), Ieee, pp. 849–861. *24, 32, 48*

[128] SY, L. W., RAITOR, M., DEL ROSARIO, M., KHAMIS, H., KARK, L., LOVELL, N. H., AND REDMOND, S. Estimating lower limb kinematics using a reduced wearable sensor count. *IEEE Transactions on Biomedical Engineering* (2020). *112*

[129] SÃ¤RKKÃ¤, O., NIEMINEN, T., SUURINIEMI, S., AND KETTUNEN, L. A multi-position calibration method for consumer-grade accelerometers, gyroscopes, and magnetometers to field conditions. *IEEE Sensors Journal 17*, 11 (2017), 3470–3481. *21, 33, 40, 41, 45, 47, 51, 52, 53, 54, 56, 147, 151*

[130] TAETZ, B., BLESER, G., AND MIEZAL, M. Towards self-calibrating inertial body motion capture. In *2016 19th International Conference on Information Fusion (FUSION)* (2016), IEEE, pp. 1751–1759. *27, 84*

[131] TAO, G., HUANG, Z., SUN, Y., YAO, S., AND WU, J. Biomechanical model-based multi-sensor motion estimation. In *2013 IEEE Sensors Applications Symposium Proceedings* (2013), IEEE, pp. 156–161. *28, 114*

[132] TEDALDI, D., PRETTO, A., AND MENEGATTI, E. A robust and easy to implement method for imu calibration without external equipments. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3042–3049. *39, 43, 44, 68*

[133] TEKIN, B., MARQUEZ NEILA, P., SALZMANN, M., AND FUA, P. Learning to fuse 2d and 3d image cues for monocular body pose estimation. In *International Conference on Computer Vision (ICCV)* (2017), no. EPFL-CONF-230311. *99*

[134] TEUFL, W., MIEZAL, M., TAETZ, B., FRÖHLICH, M., AND BLESER, G. Validity, test-retest reliability and long-term stability of magnetometer free inertial sensor based 3d joint kinematics. *Sensors 18*, 7 (2018), 1980. *12, 111*

[135] TOMCZYÅSKI, J., MAÅKOWSKI, T., AND KACZMAREK, P. Cross-sensor calibration procedure for magnetometer and inertial units. In *International Conference Automation*, Springer, pp. 450–459. *22, 51, 52, 53, 54, 55, 147, 151*

[136] TRUMBLE, M., GILBERT, A., MALLESON, C., HILTON, A., AND COLLOMOSSE, J. P. Total capture: 3d human pose estimation fusing video and inertial sensors. In *BMVC* (2017), vol. 2, pp. 1–13. *16, 115, 117, 124, 145*

[137] VAROL, G., ROMERO, J., MARTIN, X., MAHMOOD, N., BLACK, M. J., LAPTEV, I., AND SCHMID, C. Learning from synthetic humans. *CoRR abs/1701.01370* (2017). *99*

[138] VLUTTERS, S. Long short-term memory networks for body movement estimation. Master's thesis, University of Twente, 2016. *28, 112*

[139] VON MARCARD, T., PONS-MOLL, G., AND ROSENHAHN, B. Human pose estimation from video and imus. *IEEE transactions on pattern analysis and machine intelligence 38*, 8 (2016), 1533–1547. *115*

[140] VON MARCARD, T., ROSENHAHN, B., BLACK, M. J., AND PONS-MOLL, G. Sparse inertial poser: Automatic 3d human pose estimation from sparse imus. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 349–360. *28, 112, 115, 117, 127, 128*

[141] WAHDAN, A., GEORGY, J., ABDELFATAH, W. F., AND NOURELDIN, A. A technique for fast magnetometer calibration with little space coverage. In *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*, pp. 1098–1104. *23*

[142] WANG, J.-H., AND GAO, Y. A new magnetic compass calibration algorithm using neural networks. *Measurement Science and Technology 17*, 1 (2005), 153. *23*

[143] WEYGERS, I., KOK, M., DE VROEY, H., VERBEERST, T., VERSTEYHE, M., HALLEZ, H., AND CLAEYS, K. Drift-free inertial sensor-based joint kinematics for long-term arbitrary movements. *IEEE Sensors Journal 20*, 14 (2020), 7969–7979. *12, 28, 111, 114*

[144] WOUDA, F. J., GIUBERTI, M., BELLUSCI, G., VAN BEIJNUM, B.-J. F., AND VELTINK, P. H. Improving full-body pose estimation from a small sensor set using artificial neural networks and a kalman filter. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), vol. 33, pp. 10063–10064. *28*

[145] WOUDA, F. J., GIUBERTI, M., BELLUSCI, G., AND VELTINK, P. H. Estimation of full-body poses using only five inertial sensors: an eager or lazy learning approach? *Sensors 16*, 12 (2016), 2138. *118*

[146] WOUDA, F. J., GIUBERTI, M., RUDIGKEIT, N., VAN BEIJNUM, B.-J. F., POEL, M., AND VELTINK, P. H. Time coherent full-body poses estimated using only five inertial sensors: Deep versus shallow learning. *Sensors 19*, 17 (2019), 3716. *28, 113, 114, 117, 118, 120, 121, 124, 128, 129, 145*

[147] WU, G., SIEGLER, S., ALLARD, P., KIRTLEY, C., LEARDINI, A., ROSENBAUM, D., WHITTLE, M., D DâLIMA, D., CRISTOFOLINI, L., WITTE, H., ET AL. Isb recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motionâpart i: ankle, hip, and spine. *Journal of biomechanics 35*, 4 (2002), 543–548. *87*

[148] WU, G., VAN DER HELM, F. C., VEEGER, H. D., MAKHSOUS, M., VAN ROY, P., ANGLIN, C., NAGELS, J., KARDUNA, A. R., MCQUADE, K., WANG, X., ET AL. Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motionâpart ii: shoulder, elbow, wrist and hand. *Journal of biomechanics 38*, 5 (2005), 981–992. *87*

[149] WU, Y., AND PEI, L. Gyroscope calibration via magnetometer. *IEEE Sensors Journal 17*, 16 (2017), 5269–5275. *23, 26*

[150] WU, Y., WANG, J., WANG, X., AND ALQURASHI, M. Stochastic modelling and estimation of inertial sensors. In *China Satellite Navigation Conference (CSNC) 2014 Proceedings: Volume III* (2014), Springer, pp. 499–510. *38*

[151] WU, Y., ZOU, D., LIU, P., AND YU, W. Dynamic magnetometer calibration and alignment to inertial sensors by kalman filtering. *IEEE Transactions on Control Systems Technology 26*, 2 (2017), 716–723. *23, 26, 29*

[152] WU, Z., AND WANG, W. Magnetometer and gyroscope calibration method with level rotation. *Sensors 18*, 3 (2018), 748. *22, 38, 51, 52, 53, 55, 56, 147, 151*

[153] XU, H., GUERRIER, S., MOLINARI, R. C., AND KAREMERA, M. Multivariate signal modeling with applications to inertial sensor calibration. *IEEE Transactions on Signal Processing 67*, 19 (2019), 5143–5152. *24, 32, 48*

[154] YI, C., MA, J., GUO, H., HAN, J., GAO, H., JIANG, F., AND YANG, C. Estimating three-dimensional body orientation based on an improved complementary filter for human motion tracking. *Sensors 18*, 11 (2018), 3765. *25*

[155] YUAN, Q., AND CHEN, I.-M. Kinematic uncertainties in human motion tracking and interaction. In *Advances on Theory and Practice of Robots and Manipulators*. Springer, 2014, pp. 491–499. *114*

[156] YUN, X., AND BACHMANN, E. R. Design, implementation, and experimental results of a quaternion-based kalman filter for human body motion tracking. *IEEE transactions on Robotics 22*, 6 (2006), 1216–1227. *26, 66*

[157] ZHOU, Q., ZHENG, L., YU, G., LI, H., AND ZHANG, N. A novel adaptive kalman filter for euler angle based mems imu/magnetometer attitude estimation. *Measurement Science and Technology* (2020). *25*

[158] ZHOU, X., SUN, X., ZHANG, W., LIANG, S., AND WEI, Y. Deep kinematic pose regression. In *European Conference on Computer Vision* (2016), Springer, pp. 186–201. *99*

[159] ZHOU, Y., BARNES, C., LU, J., YANG, J., AND LI, H. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), CVPR, pp. 5745–5753. *29, 113, 119, 125, 143, 145*

[160] ZHU, L., AND LAPTEV, N. Deep and confident prediction for time series at uber. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (2017), IEEE, pp. 103–110. *29, 113, 120*

[161] ZHU, M., WU, Y., AND YU, W. An efficient method for gyroscope-aided full magnetometer calibration. *IEEE Sensors Journal 19*, 15 (2019), 6355–6361. *23*

Curriculum Vitae

Hammad Tanveer Butt

Education

2010 - 2013 — **National University of Sciences and Technology (NUST)**
Islamabad, Pakistan
MSc Electrical Engineering

1995 - 1999 — **National University of Sciences and Technology (NUST)**
Islamabad, Pakistan
Bachelor of Engineering, Avionics Engineering

Internship

2016 - 2021 — **Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)**
67663 Kaiserslautern, Germany
Augmented Vision Group, Body Sensor Network

Professional experience

2021 — **National University of Sciences and Technology (NUST)**
Islamabad, Pakistan
Assistant Professor

2015 - 2016 — **National University of Sciences and Technology (NUST)**
Islamabad, Pakistan
Deputy Director R&D

2003 - 2005 — **Aviation Engineering School**
Rawalpindi, Pakistan
Faculty member

List of Publications

Journal articles

1. Hammad Tanveer Butt, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy, Ankit Kumar and Didier Stricker, **"Simultaneous End User Calibration of Multiple Magnetic Inertial Measurement Units with associated Uncertainty,"** *IEEE Access*, Volume 9, pp. 26468-26483, 2021.

2. Hammad Tanveer Butt, Bertram Taetz, Mathias Musahl, Maria Alejandra Sanchez, Pramod Murthy and Didier Stricker, **"Magnetometer Robust Deep Human Pose Regression with Uncertainty Prediction Using Sparse Body Worn Magnetic Inertial Measurement Units,"** *IEEE Access*, Volume 9, pp. 36657-36673, 2021.

3. Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker. **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"** *IPSJ Transactions on Computer Vision and Applications*, Volume 11, pp. 1-9, 2019.
   ([†] First two authors contributed equally in this work.)

Conference papers

4. Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Maria Alejandra Sanchez and Didier Stricker, **"Development of High Rate Wearable MIMU Tracking System Robust to Magnetic Disturbances and Body Acceleration,"** In *Intelligent Systems Conference (IntelliSys), London, UK*, 2019. (**Best scientific paper award**)

5. Hammad Tanveer Butt, Mathias Musahl, Manthan Pancholi, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Intelligent Sensor Fusion with Online Distributed MIMU Calibration for Wearable Motion Capture (Poster),"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

6. Hammad Tanveer Butt, Manthan Pancholi, Mathias Musahl, Pramod Murthy, Maria Alejandra Sanchez and Didier Stricker, **"Inertial Motion Capture Using Adaptive Sensor Fusion and Joint Angle Drift Correction,"** In *22nd International Conference on Information Fusion, Ottawa, Canada*, 2019.

7. Pramod Murthy[†], Hammad Tanveer Butt[†], Sandesh Hiremath, Alireza Khoshhal and Didier Stricker, **"Learning 3D Joint Constraints from Vision based Motion Capture Datasets,"** In *International Conference on Machine Vision Application (MVA), Tokyo, Japan*, 2019. - the long version of Conference paper is peer reviewed and published as Journal paper too. ([†] First two authors contributed equally in this work.)