# A Gaussian Process approach for IMU to Pose Spatiotemporal Calibration

**Cedric Le Gentil and Teresa Vidal-Calleja**
Robotics Institute at the University of Technology Sydney, Australia
cedric.legentil@uts.edu.au, teresa.vidalcalleja@uts.edu.au

## Abstract

This paper presents a method to perform Inertial Measurement Unit (IMU)-to-pose extrinsic calibration. Considering the pose information collected with a motion-capture system or from the pose of a robotic arm's end effector, the goal is to estimate the geometric transformation between the reference frame of an IMU and the one of a marker tracked by the motion-capture system (with the marker being rigidly mounted with the IMU) or the arm gripper. The method models the pose data in a continuous manner using Gaussian Processes (GP). With the use of linear operators, the GP models are utilised to predict the dynamics of the system in terms of velocities and accelerations without relying on any explicit motion model. Accordingly, the extrinsic parameters as well as the inter-sensor time-shift are estimated with a non-linear optimisation, minimising the difference between the IMU data and the predicted dynamics. Note that the method is not tailored to a specific motion-capture system or robotic arm and can be used with any exteroceptive sensor from which data can be preprocessed into poses through time. We demonstrate the soundness of our approach with real-world data throughout multiple experiments testing the robustness and repeatability of the results. We will release our open-source implementation[1].

## 1 Introduction

Over the years, the performance of robotics state estimation has relentlessly been improving. A challenging part of robotics research is the ability to measure the actual accuracy/performance of an estimation framework. A common approach is the use of motion-capture systems to obtain the 6-Degree-of-Freedom (DoF) pose of a marker through time. Attaching the marker to a sensor suite, it is possible to recover the accurate pose of the sensor suite. However, as illustrated in Fig. 1, the reference frame of the marker and the reference frame used in the robotic system (e.g., camera frame) are not the same. There is generally an unknown geometric transformation between the marker's frame and the one of the state estimation pipeline. The work presented in this paper originates from the task of obtaining the ground truth trajectories of Inertial Measurement Unit (IMU)-only sensor suites as needed in [McDonald *et al.*, 2023] or [Le Gentil and Vidal-Calleja, 2023]. The scope of this work is not limited to motion-capture to IMU calibration as the method is agnostic to the way the pose of the marker is obtained. Accordingly, the method could be applied to gripper-IMU calibration using a robotic arm's end effector pose. Similarly, it could perform the calibration between an exteroceptive sensor, such as lidar or camera, and an IMU given a preprocessing step that would convert the exteroceptive data into pose estimates through time.

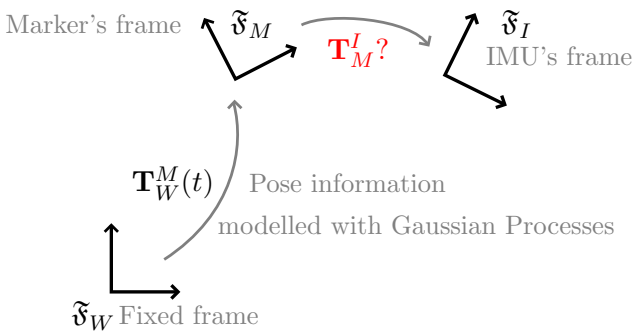



Figure 1: The proposed method aims at performing pose-IMU extrinsic calibration using Gaussian Processes to represent continuously the pose through time. An application example is the generation of an IMU ground truth trajectory using motion-capture systems or robotic arms.

[1] https://github.com/UTS-CAS/imu_pose_calib

In the context of multi-sensor perception, the extrinsic calibration (geometric transformation) between the different sensors is a crucial step for the overall performance of the system. As cameras are ubiquitous, a large number of works addressed the problem of vision-based extrinsic calibration. The works in [Alves *et al.*, 2003] and [Lobo and Dias, 2007] perform camera-IMU calibration relying on complex rigs with actuators and additional sensors. Later, [Kelly and Sukhatme, 2009] and [Furgale *et al.*, 2013] address the same problem by simply moving a handheld visual-inertial sensor suite in front of a simple checkerboard. Additionally, the continuous pose state representation based on basis functions in [Furgale *et al.*, 2013] allows for the estimation of an inter-sensor time-shift, therefore, accounting for unsynchronised data collection. Targeting the lidar-camera transformation estimation, [Nowicki, 2020] also leverages a continuous model of the systems pose to perform spatiotemporal calibration via the use of B-Splines. Similarly to camera-IMU, the lidar-IMU extrinsic calibration problem has been extensively studied starting with [Le Gentil *et al.*, 2018] where Gaussian Processes (GPs) are used to continuously model the inertial data before performing preintegration [Forster *et al.*, 2015]. Later, works like [Lv *et al.*, 2020] and [Mishra *et al.*, 2021] enabled such calibration without the need for any calibration target.

All the aforementioned works involve an exteroceptive sensor (camera or lidar) and use modality-dependent data. Other works address the calibration as a pose-to-pose or motion problem like [Taylor and Nieto, 2015] where the exteroceptive data is first preprocessed into pose estimates. This is related to the classic hand-in-eye calibration problem where the original purpose was to estimate the transformation between the reference frame of a camera and the one of a robotic arm [Tsai and Lenz, 1989]. Throughout the years, the hand-in-eye problem has been solved in multiple ways like [Park and Martin, 1994], [Andreff *et al.*, 1999], [Daniilidis, 1999]. More recently, [Zhang *et al.*, 2022] addresses the issue of camera-IMU-arm calibration using IMU preintegration in a first step to calibrate the camera with the IMU before performing the hand-in-eye calibration with the arm based on the previously estimated IMU poses.

In this paper, we directly perform the pose-IMU extrinsic calibration in a single formulation. Our method is based on GPs to continuously represent the pose measurements. GP regression is a non-parametric approach for probabilistic interpolation. With the use of linear operators [Särkkä, 2011], it is possible to analytically query the system's velocities and accelerations at any timestamp without relying on any explicit motion model (e.g., constant acceleration). This representation is inspired by our previous works on inertial preintegration for multi-modal state estimation [Le Gentil and Vidal-Calleja, 2021] and [Le Gentil and Vidal-Calleja, 2023]. These works demonstrate the ability of GP models to accurately accommodate unsynchronised sensor measurements in state estimation pipelines. While our previous works represent the pose derivatives and infer their integrals to get the actual pose, here we model the pose and infer the derivatives. Thus, similarly to [Furgale *et al.*, 2013], the residuals of our non-linear optimisation consist of the discrepancy between the inertial measurements and the dynamics predicted from the continuous pose models. This alleviates the need for constant bias assumptions between pose estimates which is common to any integration-based frameworks. The contributions of this work are:

- A novel GP-based formulation for spatiotemporal pose-IMU extrinsic calibration.
- The open-source implementation of the proposed method.
- Real-world experiments to demonstrate the soundness of the proposed method.

## 2 Gaussian Process background

### 2.1 Gaussian Process regression

GP regression is a kernel-based probabilistic method for non-parametric interpolation [Rasmussen and Williams, 2006]. Let us consider a signal $h(t) \in \mathbb{R}$ as a function of time $t$. Modelling $h$ with a GP as $h(t) \sim \mathcal{GP}\left(0, k_h(t,t')\right)$, the kernel function $k_h(t,t')$ corresponds to the covariance between any two instances of $h$: $k_h(t,t') = \text{cov}(h(t), h(t'))$. Given noisy measurements $y_i$ of $h(t_i)$

$$y_i = h(t_i) + \eta \text{ with } \eta \sim \mathcal{N}(0, \sigma_y), \tag{1}$$

and $i = 1, \cdots, N$, it is possible to express our GP model as a multivariate Gaussian distribution

$$\begin{bmatrix} \mathbf{y} \\ h^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I} & \mathbf{k}_h(\mathbf{t},t) \\ \mathbf{k}_h(t,\mathbf{t}) & k_h(t,t) \end{bmatrix}\right), \tag{2}$$

where $\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix}^\top$, $\mathbf{t} = \begin{bmatrix} t_1 & \cdots & t_N \end{bmatrix}^\top$, $\mathbf{k}_h(t,\mathbf{t}) = \begin{bmatrix} k_h(t,t_1) & \cdots & k_h(t,t_N) \end{bmatrix}$, $\mathbf{k}_h(\mathbf{t},t) = \mathbf{k}_h(t,\mathbf{t})^\top$, and

$$\mathbf{K}_h(\mathbf{t},\mathbf{t}) = \begin{bmatrix} k_h(t_1,t_1) & k_h(t_1,t_2) & \cdots & k_h(t_1,t_N) \\ k_h(t_2,t_1) & k_h(t_2,t_2) & \cdots & k_h(t_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ k_h(t_N,t_1) & k_h(t_N,t_2) & \cdots & k_h(t_N,t_N) \end{bmatrix},$$

with $h^*(t)$ a novel instance of the function $h$ at any arbitrary time $t$. By conditioning (2) with respect to the

noisy observations, the value at $t$ is inferred as

$$h^*(t) = \mathbf{k}_h(t,\mathbf{t})\big[\mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I}\big]^{-1}\mathbf{y} \quad (3)$$

$$\mathrm{var}\big(h^*(t)\big) = k_h(t,t) - \mathbf{k}_h(t,\mathbf{t})\big[\mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I}\big]^{-1}\mathbf{k}_h(\mathbf{t},t). \quad (4)$$

Consequently, (3) allows for the continuous probabilistic interpolation of a signal $h$ provided a set of noisy measurements $y_i$

### 2.2 Linear operators and Gaussian Process regression

Atop their probabilistic and non-parametric nature, another interesting property of GPs lies in the fact that applying a linear operator to a GP results in another GP [Särkkä, 2011]. In the context of GP regression, it means that given noisy measurement of a signal $h$, it is possible to directly infer a linear operation $g(t) = \mathcal{L}h(t)$ of the signal where $\mathcal{L}$ can be for example the differentiation operator $\mathcal{L}_d = \frac{\partial}{\partial t} \rightarrow g(t) = \mathcal{L}_d h(t) = \frac{\partial h(t)}{\partial t}$. Accordingly, if we consider a position signal, the use of GPs and linear operators allows for the direct inference of the velocity and acceleration without resorting to numerical differentiation methods.

Rewriting the multivariate Gaussian distribution (2) as

$$\begin{bmatrix}\mathbf{y}\\ g^*\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I} & \mathbf{k}_h(\mathbf{t},t)\mathcal{L}\\ \mathcal{L}\mathbf{k}_h(t,\mathbf{t}) & \mathcal{L}k_h(t,t)\mathcal{L}\end{bmatrix}\right), \quad (5)$$

the linearly operated signal is inferred as

$$g^*(t) = \mathcal{L}\mathbf{k}_h(t,\mathbf{t})\big[\mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I}\big]^{-1}\mathbf{y} \quad (6)$$

$$\mathrm{var}\big(g^*(t)\big) = \mathcal{L}k_h(t,t)\mathcal{L} - \mathcal{L}\mathbf{k}_h(t,\mathbf{t})\big[\mathbf{K}_h(\mathbf{t},\mathbf{t}) + \sigma_y^2\mathbf{I}\big]^{-1}\mathbf{k}_h(\mathbf{t},t)\mathcal{L}. \quad (7)$$

Note that the left application of the operator on the kernel function $k(t,t')$ corresponds to the operation being applied with respect to the first argument of the function, and the right application to the second argument.

## 3 Problem statement and system overview

Let us consider a motion-capture 3D marker and 6-DoF IMU composed of a 3-axis gyroscope and a 3-axis accelerometer. The IMU and the marker are rigidly mounted together. Their reference frames at time $t_i$ are denoted $\mathfrak{F}_{I_{t_i}}$ and $\mathfrak{F}_{M_{t_i}}$, respectively. The IMU provides gyroscope measurements $\tilde{\boldsymbol{\omega}}_{\mathfrak{t}_i^\omega}$ at times $\mathfrak{t}_i^\omega$ $(i = 1, \cdots, N_\omega)$ and accelerometer measurements $\tilde{\mathbf{f}}_{\mathfrak{t}_i^f}$ at times $\mathfrak{t}_i^f$ $(i = 1, \cdots, N_f)$. Note that accelerometer and gyroscope measurements can be provided asynchronously by the IMU. The pose of the motion-capture marker at time $\mathfrak{t}_i^m$ is collected as a rotation matrix $\mathbf{R}_W^{M_{\mathfrak{t}_i^m}}$ and a position vector $\mathbf{p}_W^{M_{\mathfrak{t}_i^m}}$ both expressed in a fixed reference frame $\mathfrak{F}_W$. The rigid transformation between the IMU and the marker is denoted as $\mathbf{R}_M^I$ for the rotation and $\mathbf{p}_M^I$ for the translation, corresponding to the pose of $\mathfrak{F}_I$ in the marker's frame $\mathfrak{F}_M$.

The goal of the proposed framework is to estimate the transformation between the IMU and the marker as well as a constant timeshift $\delta_t$ to account for the potential latency between the data collection of the motion-capture system and the IMU. The calibration parameters are estimated via a non-linear least-square optimisation that minimises the discrepancy between the continuous models of the motion-capture data and the IMU measurements. Before introducing our approach, we need to provide the reader with an overview of the IMU and system's dynamics.

### 3.1 Inertial dynamics and measurement model

In this subsection, we define the IMU dynamics and the inertial measurement model as in [Haug, 1989] and [Forster *et al.*, 2015]. The IMU pose $\mathbf{R}_W^I(t)$ and $\mathbf{p}_W^I(t)$ in the motion-capture system's fixed frame are ruled by the following differential equations:

$$\dot{\mathbf{R}}_W^I(t) = \mathbf{R}_W^I(t)\boldsymbol{\omega}_I(t)^\wedge, \quad (8)$$

$$\ddot{\mathbf{p}}_W^I(t) = \mathbf{a}_W^I(t), \quad (9)$$

where $\ddot{}$ is the double differentiation operator with respect to time $t$, $\mathbf{a}$ the linear acceleration of the sensor in $\mathfrak{F}_W$, $\boldsymbol{\omega}_I$ the angular velocity of the inertial frame relative to $\mathfrak{F}_W$ expressed in the IMU frame, and $^\wedge$ the operator that transforms a 3-by-1 vector into a skew-symmetric matrix as follows

$$\boldsymbol{\omega}^\wedge = \begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\end{bmatrix}^\wedge = \begin{bmatrix}0 & -\omega_3 & \omega_2\\ \omega_3 & 0 & -\omega_1\\ -\omega_2 & \omega_1 & 0\end{bmatrix}. \quad (10)$$

While the gyroscope directly measures the system's instantaneous angular velocity (subject to noise and additive biases), the accelerometer does not measure the linear acceleration in a fixed frame, but the *proper acceleration* $\tilde{\mathbf{f}}(t)$. The proper acceleration is a combination of the acceleration in a fixed frame and the gravitational acceleration. Overall, the inertial measurement model can be formalised as

$$\begin{aligned}\tilde{\mathbf{f}}(t) &= \mathbf{R}_W^I(t)^\top(\mathbf{a}_W^I(t) - \mathbf{g}_W) + \mathbf{b}_f(t) + \eta_f(t),\\ \tilde{\boldsymbol{\omega}}(t) &= \boldsymbol{\omega}_I(t) + \mathbf{b}_\omega(t) + \eta_\omega(t),\end{aligned} \quad (11)$$

where $\mathbf{g}$ is the gravity vector in $\mathfrak{F}_W$, $\mathbf{b}_f$ and $\mathbf{b}_\omega$ slowly varying sensor biases (modelled with Brownian motion),

and $\eta_f$ and $\eta_\omega$ the zero-mean Gaussian noises of variances $\sigma_f$ and $\sigma_\omega$ for the linear accelerations and angular velocities respectively.

### 3.2 Dynamics in rigid bodies

This subsection provides the dynamics of rigid bodies ([Kane and Levinson, 1985]) applied to our two reference frames (IMU and marker) linked via a rigid transformation. First, let us consider the angular velocities as observed from the IMU frame, $\boldsymbol{\omega}_I(t)$ and from the marker's frame, $\boldsymbol{\omega}_M(t)$. The angular velocity is constant throughout a rigid body (when expressed in a single reference frame: $\boldsymbol{\omega}_W^M = \boldsymbol{\omega}_W^I$). Accordingly, the relationship between the instantaneous angular velocities expressed in $\mathfrak{F}_{I_t}$ and $\mathfrak{F}_{M_t}$ solely need to account for the change of orientation $\mathbf{R}_M^I$

$$\boldsymbol{\omega}_M(t) = \mathbf{R}_M^I \boldsymbol{\omega}_I(t). \tag{12}$$

Regarding the acceleration, we need to account for the distance between the frames as well as the body rotational velocities and accelerations:

$$\mathbf{a}_W^M = \mathbf{a}_W^I + \boldsymbol{\alpha}_W^M \times \mathbf{p}_M^I + \boldsymbol{\omega}_W^M \times (\boldsymbol{\omega}_W^M \times \mathbf{p}_M^I), \tag{13}$$

with $\alpha_W^M = \dot{\boldsymbol{\omega}}_W^M$ the angular acceleration expressed in the fixed frame $\mathfrak{F}_W$.

## 4 IMU-pose extrinsic calibration

### 4.1 Continuous models

As motivated in the Introduction, we use GPs to model the pose of the motion-capture marker through time in a continuous manner. This accommodates asynchronous measurements from the motion-capture system, the accelerometer, and the gyroscope. It also enables analytical differentiation to infer velocities and accelerations (c.f. Section 2.2) and the estimation of a time-shift between the timestamps of the motion-capture system and the IMU. The three components of the position are models as

$$\mathbf{p}_{W\bullet}^M \sim \mathcal{GP}(0, k_{p_\bullet}(t,t)), \tag{14}$$

with $\bullet$ representing one of the 3 different axis $x$, $y$, or $z$. For the orientation, we first need to convert the rotation matrix into a minimal representation $\mathbf{r}_W^M = \mathrm{Log}(\mathbf{R}_W^M)$, with Log, the reverse operation of the SO(3) exponential mapping [Barfoot, 2017]. In order to ensure the continuity of the orientation through time we use the initialisation algorithm introduced in [Le Gentil and Vidal-Calleja, 2023]. The rotation vector $\mathbf{r}_W^M$ is modelled with three independent GPs

$$\mathbf{r}_{W\bullet}^M \sim \mathcal{GP}(0, k_{r_\bullet}(t,t)). \tag{15}$$

Leveraging (3) with the pose GPs (14) and (15), it is possible to infer the pose of the motion-capture marker, $\mathbf{R}_W^{M^*}(t) = \mathrm{Exp}(\mathbf{r}_W^{M^*}(t))$ and $\mathbf{p}_W^{M^*}(t)$ at any timestamp $t$. Additionally, using linear operators $\mathcal{L}_d = \frac{\partial}{\partial t}$, $\mathcal{L}_{dd} = \frac{\partial^2}{\partial t^2}$ and (6), we can estimate the linear acceleration, angular velocity and acceleration in the fixed frame as

$$\begin{aligned} \mathbf{a}_W^{M^*}(t) &= \mathcal{L}_{dd}\mathbf{p}_W^{M^*}(t), \\ \boldsymbol{\omega}_W^{M^*}(t) &= \mathcal{L}_d \mathbf{r}_W^{M^*}(t), \\ \text{and} \quad \boldsymbol{\alpha}_W^{M^*}(t) &= \mathcal{L}_{dd}\mathbf{r}_W^{M^*}(t), \end{aligned} \tag{16}$$

respectively.

### 4.2 Cost function

Given the GP models of the marker's pose and its derivatives, residuals can be derived to quantify the discrepancy between the motion-capture data and the IMU measurements. Atop the motion-capture and inertial data, the residuals involves a set of unknown variables that are estimated in a non-linear least-square problem

$$\begin{aligned} \mathcal{S}^* = \underset{\mathcal{S}}{\mathrm{argmin}} \Bigg( & \sum_{i=2}^{N^f} \|\mathbf{e}_{\mathbf{b}_f}^i\|^2_{\Sigma_{\mathbf{b}_f}} + \sum_{i=2}^{N^\omega} \|\mathbf{e}_{\mathbf{b}_\omega}^i\|^2_{\Sigma_{\mathbf{b}_\omega}} \\ & + \sum_{i=1}^{N^f} \|\mathbf{e}_f^i\|^2_{\Sigma_f} + \sum_{i=1}^{N^\omega} \|\mathbf{e}_\omega^i\|^2_{\Sigma_\omega} \Bigg) \end{aligned} \tag{17}$$

with $\mathcal{S}$ the estimated state $\mathcal{S} = [\mathbf{R}_E^W, \mathbf{R}_M^I, \mathbf{p}_M^I, \mathbf{b}_f^1, \cdots, \mathbf{b}_f^{N_f}, \mathbf{b}_\omega^1, \cdots, \mathbf{b}_\omega^{N_\omega}, \delta_t]$, $\mathbf{e}_{\mathbf{b}_f}^i$ and $\mathbf{e}_{\mathbf{b}_\omega}^i$ the accelerometer and gyroscope bias residuals, $\mathbf{e}_f^i$ the accelerometer measurement residuals, and $\mathbf{e}_\omega^i$ the gyroscope measureents residuals. The rest of this section will detail the different residuals in the cost function.

#### Gyroscope residuals

From (12), we can link the instantaneous angular velocities of the IMU and marker frames. While the instantaneous angular velocity of the IMU corresponds to its measurement minus the estimated biases, the one of the marker is not readily available. Fortunately, as derived in [Barfoot, 2017], the instantaneous angular velocity can be deduced from the marker's orientation and its derivative with respect to time

$$\mathbf{J}_r(\mathbf{r}_W^M(t))\dot{\mathbf{r}}_W^M(t) = \boldsymbol{\omega}_M(t).$$

Accordingly, the gyroscope residual is given by

$$\mathbf{e}_\omega^i = \mathbf{J}_r(\mathbf{r}_W^{M*}(\mathrm{t}_i^\omega + \delta_t))\dot{\mathbf{r}}_W^{M*}(\mathrm{t}_i^\omega + \delta_t) - \mathbf{R}_M^I\big(\tilde{\boldsymbol{\omega}}_{\mathrm{t}_i^\omega} - \mathbf{b}_\omega^i\big). \tag{18}$$

#### Accelerometer residuals

The residuals for the accelerometer are derived from (11) and (13). While most components of these equations are available as measurements, inferred from the GP models, or directly estimated, the gravity in the world frame

is not known. To address this issue, we included in the state $\mathcal{S}$ the orientation $\mathbf{R}_E^W$ of the fixed frame with respect to an Earth-gravity aligned frame $\mathfrak{F}_E$ in which the gravity is known as $\mathbf{g}_E = \begin{bmatrix}0 & 0 & g\end{bmatrix}^\top$, leading to $\mathbf{g}_W = {\mathbf{R}_E^W}^\top \mathbf{g}_E$. Thus, the residuals $\mathbf{e}_f^i$ are defined as

$$\begin{aligned}\mathbf{e}_f^i =& \mathbf{R}_W^{M*}(\mathrm{t}_i^f + \delta_t)\mathbf{R}_M^I(\tilde{\mathbf{f}}_{\mathrm{t}_i^f} - \mathbf{b}_f^i) - \mathbf{a}_W^{M*}(\mathrm{t}_i^f \\ &+ \delta_t) + \boldsymbol{\alpha}_W^{M*}(\mathrm{t}_i^f + \delta_t) \times \mathbf{p}_M^I \\ &+ \boldsymbol{\omega}_W^{M*}(\mathrm{t}_i^f + \delta_t) \times (\boldsymbol{\omega}_W^{M*}(\mathrm{t}_i^f + \delta_t) \times \mathbf{p}_M^I) \\ &+ {\mathbf{R}_E^W}^\top \mathbf{g}_E.\end{aligned}$$

**Bias residuals**

The bias residuals are used to enforce the Brownian motion model of the gyroscope and accelerometer biases. Intuitively it prevents two consecutive bias estimates from being drastically different. Formally, the residuals are defined as

$$\mathbf{e}_{\mathbf{b}_f}^i = \mathbf{b}_f^i - \mathbf{b}_f^{i-1}, \tag{19}$$

$$\mathbf{e}_{\mathbf{b}_\omega}^i = \mathbf{b}_\omega^i - \mathbf{b}_\omega^{i-1}. \tag{20}$$

# 5 Experiments

## 5.1 Implementation and hardware

The formulation presented above considers the simultaneous extrinsic calibration of the rotation and position between the IMU and marker. However, the rotational component can be estimated using only the gyroscope-related residuals (18) and (20). Thus, we are first optimising a rotation-only version of (17) with a reduced state that solely comprises $\mathbf{R}_M^I$, the gyroscope biases $\mathbf{b}_\omega^i$ and the time-shift $\delta_t$. This provides good initial conditions for solving the full calibration as formulated in Section 4.

The proposed method has been implemented in C++ using *Ceres* [Agarwal *et al.*, 2022] for the non-linear optimisation, and Python for data pre-processing and result visualisation. The square exponential kernel has been used for the GP models. Our open-source code has been made available publicly[2].

## 5.2 Datasets

The rest of this section provides empirical evidence of the soundness of our method. To do so, we use a sensor suite consisting of an Intel Realsense depth camera D455 (with an internal 6-DoF IMU) and a Microstrain 3DM-GX5 IMU as shown in Fig. 2. The sensor suite is equipped with IR-reflective spheres creating a 6-DoF marker rigidly attached to the sensing plate. A Vicon system is used to perform motion capture with sub-millimetre accuracy. We have collected two sequences of around 40 s moving the sensing suite exciting the 6-DoF. The data from both IMUs as well as the point clouds from the depth camera have been recorded. For the sake of simplicity, let us denote *Seq1I* the dataset corresponding to the first sequence with the internal IMU and *Seq1E* with the external IMU. Similarly, we use *Seq2I* and *Seq2E* for the second sequence. We also recorded a series of static image-pose pairs in front of a chessboard (camera and Vicon's marker pose) to perform the pose-camera extrinsic calibration mentioned in the next subsection.

(a) Sensor suite (b) Lab environment

Figure 2: Illustration of our real-world data collection. The sensor suite (a) consists of a depth camera Intel Realsense D455 (with an internal IMU), a Microstrain 3DM-GX5 IMU, and IR-reflective spheres (constituting the motion-capture marker). The datasets are collected in a lab environment (b) equipped with a Vicon motion-capture system (IR cameras circled in red).

## 5.3 Quantitative

Evaluating the performance of extrinsic calibration methods is always a challenging task as it is not possible to obtain the actual ground truth in real-world scenarios. It is because the ground truth cannot be perfectly obtained with other means in the first place that we need calibration methods. This subsection provides the reader with a number of quantitative results aiming at analysing the performance of the proposed approach.

**Comparison with chained calibration**

First, let us compare the results of the proposed pose-IMU extrinsic calibration against the chained calibration that consists of pose-camera and camera-IMU. The pose-camera component corresponds to the classic hand-in-eye calibration problem in which a camera is attached to the end effector of a robotic arm and the goal is to estimate the effector's pose-to-camera. Regarding the camera-IMU link, we have used the provided proprietary

[2] `https://github.com/UTS-CAS/imu_pose_calib`

| Sequence → IMU ↓ | Seq. 1 | Seq. 2 |
|---|---|---|
| Internal | 0.52°<br>15.6 mm | 0.32°<br>13.3 mm |
| External | 0.41°<br>32.3 mm | 0.51°<br>33.4 mm |

Table 1: Difference between the proposed pose-IMU calibration and a chained pose-camera/camera-IMU calibration pipeline.

calibration method for the internal IMU and *Kalibr* [Furgale *et al.*, 2013] for the external IMU.

Table 1 shows the norm of the difference of estimated calibration parameters with both our method and the chained calibration pipeline. Overall, the orientation difference is consistent throughout all the datasets. However, the use of the internal IMU leads to smaller translation differences. While it is not possible to quantify the exact error of the proposed method, one can see the soundness of the proposed method compared to well-established methods. Nevertheless, we would like to emphasise that the chained calibration can easily be subject to errors by compounding multiple sources of inaccuracy (e.g., lack of precision of the target for camera intrinsic and hand-eye calibration, or motion blur of the camera for the camera-IMU link). We would also like to stress the repeatability of the proposed method as the results between Seq1 and Seq2 are very similar.

**Robustness to initial conditions**

The set-up presented in this paragraph aims at analysing the sensitivity of our method with respect to the initial guess of the extrinsic parameters $\mathbf{R}_M^I$ and $\mathbf{p}_M^I$. Using *Seq1I*, we have performed a first calibration given a good initial guess. Then, provided the optimal parameters, the calibration have been run 200 times using a noisy initial guess. Please note that the biases and time-shift have been initialised to zero all for all the runs. Fig. 3 shows the error of the estimated extrinsic parameters with respect to the initial guess error. The plot shows that the initial guess has little to no impact on the method's accuracy.

**Robustness to noise**

In this set of experiments, we analyse the impact of the input data noise on the method's output. Similarly to the previous set-up, we first obtained the optimal extrinsic parameters with *Seq1I*. Next, the calibration has been performed by adding various combinations of noise to the input data. It is important to remember that the original data come from a real sensor suite and therefore already contain some noise (standard deviations of $\approx 0.02\,\mathrm{m.s^{-2}}$ and $\approx 0.005\,\mathrm{rad.s^{-1}}$). In the following, the

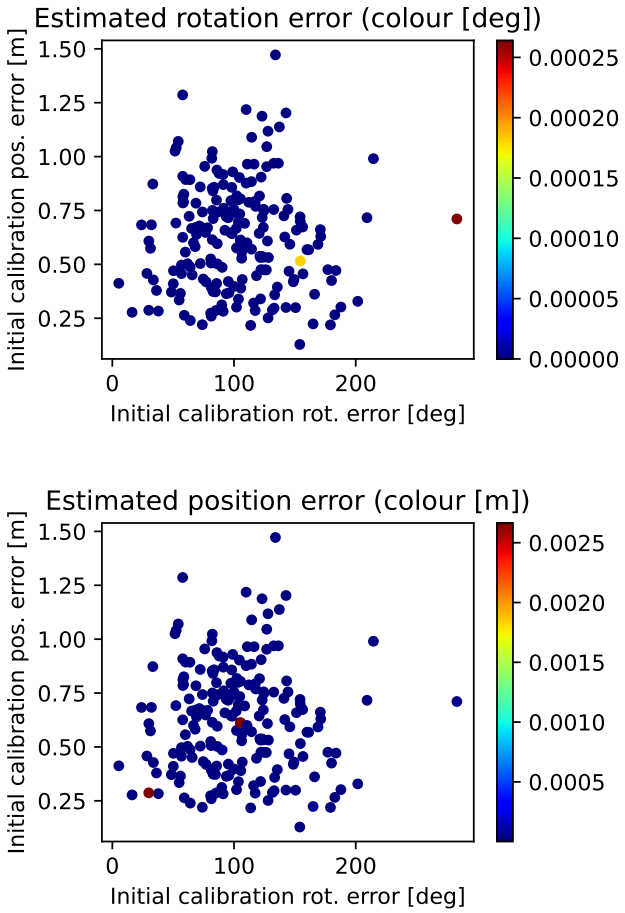



Figure 3: Analysis of the initial guess value for the proposed calibration accuracy. The error is shown with the colour of the scatter points as a function of the error on the initial guess (rotation and position).

| Gyr std → Acc std ↓ | 0.001 | 0.005 | 0.01 |
|---|---|---|---|
| 0.001 | 0.0012 | 0.0053 | 0.0129 |
| 0.01 | 0.0011 | 0.0060 | 0.0102 |
| 0.1 | 0.0012 | 0.0055 | 0.0115 |

Table 2: Analysis of the IMU noise on the calibration accuracy. Rotation RMSE [°] computed over 30 trials.

| Gyr std → Acc std ↓ | 0.001 | 0.005 | 0.01 |
|---|---|---|---|
| 0.001 | 0.050 | 0.043 | 0.049 |
| 0.01 | 0.498 | 0.654 | 0.638 |
| 0.1 | 1.776 | 1.865 | 1.850 |

Table 3: Analysis of the IMU noise on the calibration accuracy. Position RMSE [mm] computed over 30 trials.

mention of the standard deviation corresponds to the standard deviation of the additional noise.

Tables 2 and 3 present the Root Mean Squared Error (RMSE) [°] over 30 trials for additional noise to both the gyroscope and accelerometer measurements. As expected, the accuracy of $\mathbf{R}_M^I$ is mostly impacted by the noise of the gyroscope data while the accuracy of $\mathbf{p}_M^I$ is correlated with the accelerometer noise. In Tables 4 and 5, we perform a similar analysis for the motion-capture noise. It is interesting to see that the noise on the pose measurements have significantly more impact on the calibration accuracy than the IMU noise. This is sensible as the poses are differentiated once or twice in the various residuals of the proposed method, leading to a larger impact on the cost function and, therefore, the optimisation result.

### 5.4 Qualitative

In this subsection, we provide qualitative results by aggregating together point clouds from the depth camera using the pose from the Vicon system and the different transformation previously estimated: (a) uses the proposed pose-IMU calibration, combined with IMU-camera extrinsics, while (b) leverages directly the pose-camera eye-in-hand calibration. Fig. 4 shows both reconstructions of the scene as well as the difference between them after an ICP alignment [Segal *et al.*, 2009]. The

| Rot std → Pos std ↓ | 0.001 | 0.005 | 0.01 |
|---|---|---|---|
| 0.001 | 0.011 | 0.065 | 0.118 |
| 0.005 | 0.011 | 0.056 | 0.135 |
| 0.01 | 0.011 | 0.064 | 0.125 |

Table 4: Analysis of the pose noise on the calibration accuracy. Rotation RMSE [°] computed over 30 trials.

| Rot std → Pos std ↓ | 0.001 | 0.005 | 0.01 |
|---|---|---|---|
| 0.001 | 3.38 | 3.07 | 4.08 |
| 0.005 | 12.4 | 11.7 | 13.8 |
| 0.01 | 16.7 | 21.7 | 23.5 |

Table 5: Analysis of the pose noise on the calibration accuracy. Position RMSE [mm] computed over 30 trials.

camera range has been limited to 2 m to prevent highly noisy points further from the sensor. The transformation estimated with the ICP consists of a rotation of 0.4° and translation of 19 mm which is close to the difference between the two calibration pipelines discussed in Section 5.3. According to the histogram on the right of Fig. 4(c), the difference between the point clouds is mostly sub-millimetric, demonstrating the soundness of our calibration approach.

## 6 Conclusion

In this paper, we presented a novel approach to pose-IMU using GPs to represent the pose measurements in a continuous manner. The method optimises the geometric transformation between the IMU and the pose reference frame as well as a time-shift between the two. We have demonstrated the soundness of the method with real-world data by comparing our approach with a chained calibration pipeline using a camera. We have also performed various experiments to evaluate the robustness of our framework with respect to measurement noise and initial guesses of the parameters. Future work includes the multi-modal calibration with other sensors like cameras or lidars.

## 7 Acknowledgment

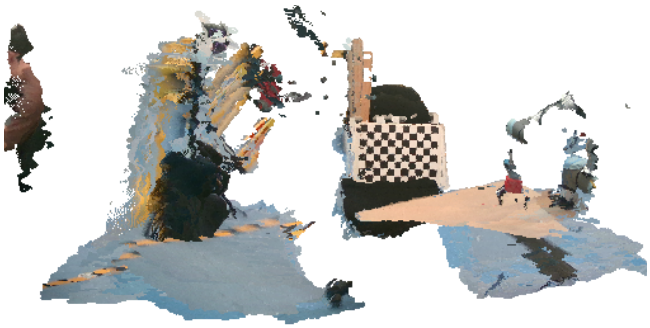
This work is supported by the Australian Research Council Discovery Project under Grant DP210101336. We would also like to thank Victor Hernandez Moreno for his help setting up the Vicon system used in our experiments.

## References

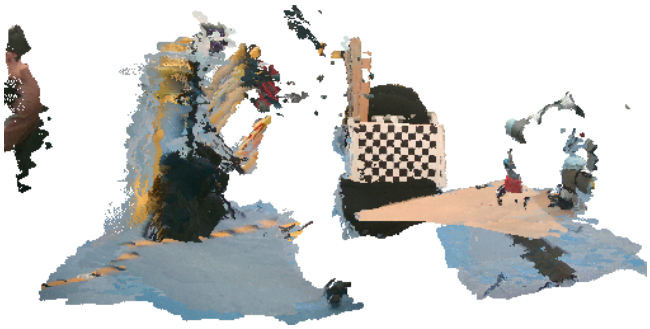
[Agarwal *et al.*, 2022] Sameer Agarwal, Keir Mierle, and The Ceres Solver Team. Ceres Solver. *https://github.com/ceres-solver/ceres-solver*, 3 2022.

[Alves *et al.*, 2003] João Alves, Jorge Lobo, and Jorge Dias. Camera-inertial sensor modelling and alignment for visual navigation. *International Conference on Advanced Robotics*, 5(3):103–111, 2003.

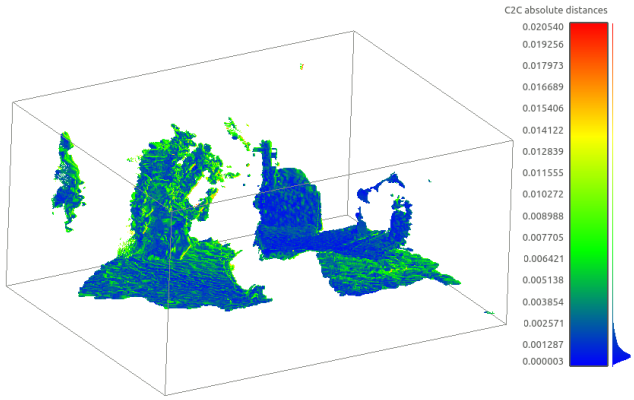
[Andreff *et al.*, 1999] Nicolas Andreff, Radu Horaud, and Bernard Espiau. On-line hand-eye calibration. In *Proceedings of the 2nd International Conference*



(a) Reconstruction using pose-IMU and IMU-camera extrinsics



(b) Reconstruction using pose-camera extrinsics




(c) Difference between point clouds (a) and (b)

Figure 4: Qualitative results aggregating the point clouds of the depth camera.

*on 3-D Digital Imaging and Modeling*, 3DIM'99, page 430–436, USA, 1999. IEEE Computer Society.

[Barfoot, 2017] Timothy D. Barfoot. *State Estimation for Robotics*. 2017.

[Daniilidis, 1999] Konstantinos Daniilidis. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18(3):286–298, 1999.

[Forster *et al.*, 2015] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. *Robotics: Science and Systems*, pages 6–15, 2015.

[Furgale *et al.*, 2013] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. *IEEE International Conference on Intelligent Robots and Systems*, pages 1280–1286, 2013.

[Haug, 1989] Edward J Haug. *Computer aided kinematics and dynamics of mechanical systems*, volume 1. Allyn and Bacon Boston, 1989.

[Kane and Levinson, 1985] Thomas R Kane and David A Levinson. *Dynamics, theory and applications*. McGraw Hill, 1985.

[Kelly and Sukhatme, 2009] Jonathan Kelly and Gaurav S. Sukhatme. Fast Relative Pose Calibration for Visual and Inertial Sensors. *Springer Tracts in Advanced Robotics*, 54:515–524, 2009.

[Le Gentil and Vidal-Calleja, 2021] Cedric Le Gentil and Teresa Vidal-Calleja. Continuous Integration over SO(3) for IMU Preintegration. *Robotics: Science and Systems XVII*, 2021.

[Le Gentil and Vidal-Calleja, 2023] Cedric Le Gentil and Teresa Vidal-Calleja. Continuous latent state preintegration for inertial-aided systems. *The International Journal of Robotics Research*, 0(0):02783649231199537, 2023.

[Le Gentil *et al.*, 2018] Cedric Le Gentil, Teresa Vidal-Calleja, and Shoudong Huang. 3D Lidar-IMU Calibration based on Upsampled Preintegrated Measurements for Motion Distortion Correction. *IEEE International Conference on Robotics and Automation*, 2018.

[Lobo and Dias, 2007] J. Lobo and J. Dias. Relative pose calibration between visual and inertial sensors. *The International Journal of Robotics Research*, 26:561–575, 2007.

[Lv *et al.*, 2020] Jiajun Lv, Jinhong Xu, Kewei Hu, Yong Liu, and Xingxing Zuo. Targetless calibration of lidar-imu system based on continuous-time batch

estimation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9968–9975, 2020.

[McDonald *et al.*, 2023] William McDonald, Cedric Le Gentil, and Teresa Vidal-Calleja. Global localisation in continuous magnetic vector fields using gaussian processes. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

[Mishra *et al.*, 2021] Subodh Mishra, Gaurav Pandey, and Srikanth Saripalli. Target-free extrinsic calibration of a 3d-lidar and an imu. In *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 1–7, 2021.

[Nowicki, 2020] Michal R. Nowicki. Spatiotemporal calibration of camera and 3d laser scanner. *IEEE Robotics and Automation Letters*, 5(4):6451–6458, 2020.

[Park and Martin, 1994] F.C. Park and B.J. Martin. Robot sensor calibration: solving ax=xb on the euclidean group. *IEEE Transactions on Robotics and Automation*, 10(5):717–721, 1994.

[Rasmussen and Williams, 2006] C E Rasmussen and C K I Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[Särkkä, 2011] Simo Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 151–158, 2011.

[Segal *et al.*, 2009] A Segal, D Haehnel, and S Thrun. Generalized-ICP. *Robotics: Science and Systems*, 5:168–176, 2009.

[Taylor and Nieto, 2015] Zachary Taylor and Juan Nieto. Motion-based calibration of multimodal sensor arrays. *IEEE International Conference on Robotics and Automation*, pages 4843–4850, 2015.

[Tsai and Lenz, 1989] R.Y. Tsai and R.K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358, 1989.

[Zhang *et al.*, 2022] Yinlong Zhang, Wei Liang, Mingze Yuan, Hongsheng He, Jindong Tan, and Zhibo Pang. Monocular visual-inertial and robotic-arm calibration in a unifying framework. *IEEE/CAA Journal of Automatica Sinica*, 9(1):146–159, 2022.