Lehrstuhl für Datenverarbeitung
Technische Universität München

Tracking Human Motion in Real-World Environments

Peter Hausamann

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung eines

Doktors der Ingenieurswissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Oliver Hayden

Prüfer der Dissertation:

1. Hon.-Prof. Dr. Martin Daumer
2. Assistant Prof. Paul McNeilage, Ph.D.
3. Prof. Dr.-Ing. Klaus Diepold

Die Dissertation wurde am 16.06.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 06.12.2021 angenommen.

Peter Hausamann. *Tracking Human Motion in Real-World Environments.* Dissertation, Technische Universität München, Munich, Germany, 2022.

Chair of Data Processing, Technische Universität München, 80290 München, Germany, http://www.ldv.ei.tum.de/.

Acknowledgments

Even in ideal conditions, writing a doctoral thesis is no easy endeavor. But sometimes, life decides to throw in a few extra obstacles. I'm extremely grateful that I was able to complete this thesis despite all the challenges and I wouldn't have been able to do so without the support of the people around me.

First and foremost, I want to thank PD Dr. Martin Kleinsteuber for being my supervisor for the first part of this journey and – after the untimely demise of our research group – Hon.-Prof. Dr. Martin Daumer for taking me over the finish line. A huge thanks also to Assistant Prof. Paul MacNeilage, Ph.D. for his support and guidance no matter the physical distance. I also want to thank Prof. Dr.-Ing. Klaus Diepold and Prof. Dr. Stefan Glasauer for their support and valuable advice.

During these past couple of years, I was blessed to be able to work with a (quite impressive) number of remarkable colleagues and just all around amazing human beings. Thanks to Rory, Vince, Alex, Wei, Clemens, Dominik, Hauser, Hao, Stefan, Martin, Matthias, Sven, Philipp, Michael M., Marcello, Timur, Gerhard, Astrid, Michael A., Mark, Brian, Bharath, Savvy and most importantly, Christian for the great times. Everyone else at GOL, LDV, Trium/SLC and UNR, thank you as well.

And while my work environment contributed a lot, the support I received from my friends and family was at least equally important. Thanks to my parents, my sister and above all, Karina - I would not be the person I am today without you.

Abstract

Accurate measurement of human movement is an essential requirement in many scientific disciplines. It is necessary for investigating motor control and other fundamental mechanisms in neuroscience and biomechanics. It also plays a crucial role in clinical research as physical activity is a key indicator for wellbeing and many diseases are associated with distinct motor patterns. In addition, emerging technologies such as augmented and virtual reality also require precise tracking of human movements.

While the quantitative study of human motion dates back hundreds of years, technology that permits accurate tracking outside of controlled environments has become available only recently. All of the applications described above can benefit from the ability to track movements in the real world: it enables sampling of natural and unbiased behavior in scientific and clinical studies and allows development of commercial products for everyday use.

This thesis presents several novel approaches to real-world tracking of human motion. First, I propose a combined head and eye tracking system based on existing commercial tracking devices. The system is mobile and lightweight, the software developed in this context is open source and the additional hardware necessary to combine the devices can be 3D printed. Secondly, the accuracy of the head tracker used in this system is evaluated. To this end, the device's measurements are compared against gold standard tracking methods across several real-world environments and locomotion tasks. Lastly, a study is conducted that assesses the validity of previously described parameters of head stabilization during everyday behavior. These head stability measures have been shown to have diagnostic value for diseases such as multiple sclerosis and Parkinson's disease. However, so far they have only been obtained in laboratory settings which makes it important to investigate how well previous findings generalize to the real world.

My results show that precise real-world motion tracking is achievable with today's technology. The combined head-eye tracking system's ability to accurately capture movements is demonstrated with different motion tasks. The performance of the head tracker by itself is promising, although some measurements such as the exact position of the head in space can still be subject to large inaccuracies. Finally, parameters of head stability seem to be reproducible in a real-life setting which suggests that they could be suitable for a potential clinical application.

Contents

List of Figures and Tables

List of Figures

List of Tables

1. Introduction

1.1. Motivation

The ability to move and be physically active is one of the most important factors in human quality of life (QoL) [44]. At the time of writing, the COVID-19 pandemic is still severely limiting mobility of individuals across the globe due to lockdown and quarantine measures aimed at limiting the spread of the virus. Studies show that individuals who engage in physical activity during quarantine report higher levels of QoL than those who don't [39].

This capacity for movement and physical activity can be severely impaired in diseases such as multiple sclerosis (MS) [91] and Parkinson's disease (PD) [26]. Patients diagnosed with MS rate the ability to walk as the most important bodily function in terms of contribution to overall wellbeing [49]. Oftentimes, illnesses are associated with distinctive movement patterns, for example in the case of autism spectrum disorder (ASD) [9]. These examples show that the ability to quantitatively measure human motion can be a key instrument for diagnosing and monitoring diseases as well as designing therapies that maximize QoL for patients.

On the other hand, the quantitative study of human movement can also be used to answer fundamental neuroscientific questions of sensory processing and motor control. Studying movement patterns of healthy individuals can reveal insights into the postural control system [89, 97] and the strategies of the nervous system when dealing with challenging tasks such as walking through rough terrain [83]. An important area of research is also how the mammalian vestibular system encodes information about self motion in neuronal representations [7].

Finally, emerging technologies such as augmented and virtual reality (AR and VR) also require faithful tracking of human motion. In VR, an artificial three-dimensional (3D) visual scene is shown on a head-mounted display and needs to be rendered correctly as the user's head moves through the virtual environment [8]. AR overlays visual and other sensory information onto the real world and often requires tracking the user's gaze or extremities to interact with virtual objects [27].

The examples of these technologies highlight the fact that the measurement of head and eye movements is of particular importance. In the human nervous system, head motion is measured by the vestibular organs and drives several important reflexes. The

vestibulo-ocular reflex (VOR) for example directly links eye and head movements by compensating for head rotation and translation while fixating environmental features [7]. This process is important to provide a stable image of the environment to the visual system. Due to this reflex, vestibular disorders can manifest themselves in involuntary eye movements, called nystagmus [107].

The next section will introduce the theoretical background for motion tracking and present several key technologies as well as their history. In many cases, the presented methods are constrained to controlled environments such as specially equipped laboratories and only recent technological advances make it possible to employ some of them in the real world.

The ability to track human movement in such real-world contexts is however important for various reasons. From a clinical point of view, recent studies have questioned how well measurements obtained in controlled environments generalize to real world conditions [20, 120]. Conducting studies in natural environments is also crucial for neuroscientific research because it provides a straightforward way of presenting the nervous system with natural stimuli. Research shows that the brain represents sensory information in a way that is adapted to naturally occurring stimulus distributions [40, 115, 118]. Finally, AR is an example of a commercial application that requires real-world motion tracking as users will want to utilize the technology in any kind of environment. This broad need for real-world tracking methods necessitates research efforts regarding their development, evaluation and application, which will be the focus of this thesis.

1.2. History and technical background

1.2.1. Motion of rigid bodies

Motion of an object is always relative to an observer. Formally, we can unambiguously describe this by the position of a reference frame (i.e., a uniquely located and oriented coordinate system) attached the object with respect to (w.r.t.) the reference frame of the observer over time. The position consist of a 3-degree of freedom (DOF) linear position $\mathbf{p}$, often referred to as just *position* and a 3-DOF angular position $\mathbf{o}$, often referred to as *orientation*. A solid object in which deformation is negligibly small is also called a *rigid body*. The combined position and orientation of a body is also referred to as its *pose*.

Position can be represented as a 3D vector from the origin O of the observer's reference frame O to the origin B of the body-fixed frame B:

$$\mathbf{p} = \overline{\mathrm{OB}} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} \tag{1.1}$$

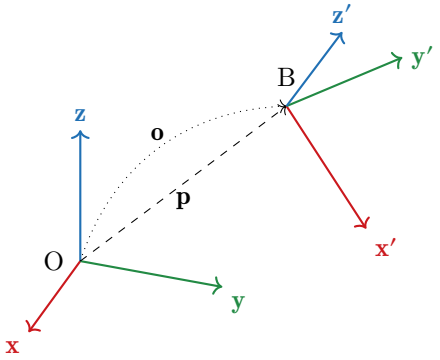

Figure 1.1.: Relationship between observer and body frame. The position $\mathbf{p}$ describes the location of the origin B of the body frame w.r.t. the origin O of the observer frame. The orientation $\mathbf{o}$ describes the relationship between the body-fixed coordinate vectors ($\mathbf{x}'$, $\mathbf{y}'$, $\mathbf{z}'$) and the observer-fixed coordinate vectors ($\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$).

Orientation can be represented with different conventions and is generally formalized as the rotation required to position the observer frame in the same direction as the body frame. Common representations of orientation include Euler angles (three consecutive rotations about the body-fixed coordinate axes), rotation vectors (describing a single rotation about an arbitrary axis) and rotation matrices (representing the rotation as a single matrix-vector product). A less intuitive, but compact and unambiguous way of describing orientation is with *quaternions*:

$$\mathbf{o} = \begin{pmatrix} o_w \\ o_x \\ o_y \\ o_z \end{pmatrix} \tag{1.2}$$

A quaternion representing orientation is also called a *versor* and can be divided into two parts:

- An imaginary part $\mathrm{Im}(\mathbf{o}) = (o_x, o_y, o_z)^\top$ that can be interpreted as a 3D vector whose direction defines an axis of rotation and whose length corresponds to the angle of rotation around that axis.
- A real part $\mathrm{Re}(\mathbf{o}) = o_w = \sqrt{1 - o_x^2 - o_y^2 - o_z^2}$ that ensures that $\mathbf{o}$ is a unit quaternion, i.e., $\|\mathbf{o}\| = 1$.

A unit quaternion $\mathbf{o}$ that describes the orientation of B w.r.t. O defines the relationship between the coordinate vectors of both frames.

Specifically, it satisfies:

$$\mathbf{x}' = \mathrm{rot}\,(\mathbf{o}, \mathbf{x}) \tag{1.3a}$$
$$\mathbf{y}' = \mathrm{rot}\,(\mathbf{o}, \mathbf{y}) \tag{1.3b}$$
$$\mathbf{z}' = \mathrm{rot}\,(\mathbf{o}, \mathbf{z}) \tag{1.3c}$$

where $\mathrm{rot}(\mathbf{o}, \mathbf{v})$ denotes the rotation of a vector $\mathbf{v}$ by a quaternion $\mathbf{o}$. The exact calculation as well as an overview of the quaternion algebra necessary for describing rotations can be found in appendix A.1.

Linear velocity is the rate of change of position over time:

$$\mathbf{v}(t) = \frac{\partial \mathbf{p}(t)}{\partial t} \tag{1.4}$$

Angular velocity is the rate of change of orientation over time and can be computed from the quaternion representation as follows:

$$\boldsymbol{\omega}(t) = \mathrm{Im}\left(2\mathbf{o}^* \frac{\partial \mathbf{o}(t)}{\partial t}\right) \tag{1.5}$$

Here, $\mathbf{o}^* = (o_w, -o_x, -o_y, -o_z)^T$ denotes the complex conjugate of $\mathbf{o}$.

Linear acceleration is the rate of change of linear velocity over time:

$$\mathbf{a}(t) = \frac{\partial \mathbf{v}(t)}{\partial t} \tag{1.6}$$

Angular acceleration is the rate of change of angular velocity over time:

$$\boldsymbol{\alpha}(t) = \frac{\partial \boldsymbol{\omega}(t)}{\partial t} \tag{1.7}$$

In many cases it is necessary to describe the simultaneous motion of multiple rigid bodies, e.g., when the observer itself is moving w.r.t. some world-fixed reference frame W. Often, this requires certain measurements of motion to be transformed between different reference frames. The necessary equations for transforming positions, orientations, velocities and accelerations across reference frames are detailed in appendix A.2.

1.2.2. Motion tracking

Motion tracking is the process of estimating certain aspects of the motion of either a single body or a set of bodies over time. For a single body, we can generally distinguish two different approaches:

- A stationary observer estimating the motion of the body from its point of view, referred to as *outside-in* tracking.
- An observer attached to the tracked body itself, estimating the motion of the the body w.r.t. its surroundings, referred to as *inside-out* tracking.

The first approach is usually limited to a so-called *tracking volume*, i.e., a subset of 3D space close to the observer within which it can estimate the motion with sufficient accuracy. This limitation does not apply to the second approach, although the tracking performance can greatly depend on the surroundings of the body (see chapter 3). Because of this, it should be clear that for most use cases where motion is tracked in real-world environments, inside-out methods are the only viable solution. This is especially true when the aim is to record human motion in an ecologically valid context, i.e., during natural everyday behavior.

Outside-in tracking generally scales well to multiple rigid bodies, although visual methods may need to deal with the problem of occlusions. On the other hand, for the inside-out approach, it is usually necessary to track each body individually which can lead to issues with scalability.

Another criteria to distinguish tracking methods is by the kind of motion than can be directly or indirectly tracked with sufficient accuracy. Some of the methods described in the following can estimate linear acceleration or angular velocity while others can track position and orientation.

Manual methods

Human motion has been scientifically studied for over 2000 years. As early as 500 BC, the Greek philosophers described their observations of human movement [6]. The physical principles of motion discovered by Galileo Galilei, Isaac Newton, and others between the 16th and 18th century made it possible to formalize the physical phenomena associated with human motion [87]. The quantitative study of human locomotion, i.e., movement from one place to another such as walking or running, dates back to the early 19th century. In 1836, the Weber brothers conducted one of the first studies that determined gait speed from measurements of time and distance [128].

Since then, automated techniques using electronic devices such as cameras an inertial sensors have become the norm for quantitative motion tracking. Nevertheless, manual tracking methods are still relevant today. For example, the six-minute walking test (6MWT) is a standardized clinical tests that can serve as an diagnostic measure for various health conditions such as MS and PD [125]. The test measures the distance a person can walk in six minutes and requires only a stop watch, a course of known length demarcated with cones and pen and paper to record the number of completed laps.

Inertial measurement units

Inertial measurement units (IMUs) are sensor devices that measure inertial motion, usually linear acceleration with an accelerometer and angular velocity with a gyroscope. Historically, IMUs have been employed for navigation of manned and unmanned vehicles such as aircraft or missiles since the 1960s [122]. In recent years, low-cost and compact microelectromechanical system (MEMS)-based IMUs have become commercially available, leading to widespread use of wearable sensors for human motion tracking [55].

Accelerometers estimate linear acceleration by measuring the compression or extension of a spring attached to an object of known mass (also called a *proof mass*) [61]. Newton's second law of motion states that the force $\mathbf{F}$ applied to a body of mass m is proportional to its acceleration $\mathbf{a}$:

$$\mathbf{F} = m \cdot \mathbf{a} \tag{1.8}$$

In the mass-spring system, this force is equal and opposite to the force $\mathbf{F}_s$ needed to extend or compress the spring with a constant stiffness k by the distance $\Delta\mathbf{x}$ (see fig. 1.2a):

$$\mathbf{F}_s = -k \cdot \Delta\mathbf{x} \tag{1.9}$$

In commercially available MEMS accelerometers $\Delta\mathbf{x}$ is usually transduced into an electrical signal through piezoelectric, piezoresistive, capacitive or similar effects [61].

Gyroscopes measure angular velocity based on a mass-spring system as well. However, the force measured by these systems is the *Coriolis force* $\mathbf{F}_c$, a fictitious force that acts upon objects moving at a velocity $\mathbf{v}$ w.r.t. a reference frame rotating with an angular velocity $\boldsymbol{\omega}$ [5]:

$$\mathbf{F}_c = -2m \cdot \mathbf{v} \times \boldsymbol{\omega} \tag{1.10}$$

The most common MEMS gyroscopes are so called *vibrating structure gyroscopes*. Here, the velocity $\mathbf{v}$ is imposed on the proof mass in the form of an oscillation of fixed frequency [10]. The resulting Coriolis force is then transformed into an electrical signal in a similar manner as in the case of the accelerometer (see fig. 1.2b).

It should be noted that the sensor systems described above can generally only measure motion in a single spatial direction. To determine all three DOF of a body's motion, commercial accelerometers and gyroscopes usually consist of three sensors oriented in a perpendicular manner.

Some IMUs also include a tri-axial magnetometer. This type of sensor measures the direction and strength of a magnetic field. While this is not a direct measurement of inertial motion it can be used to determine the direction of the earth's local magnetic field which in turn permits to estimate the orientation of the device w.r.t a local earth-fixed reference frame [5]. IMUs that combine a 3-DOF accelerometer, gyroscope, and magnetometer are commercially available in small form factors at low cost (fig. 1.3).

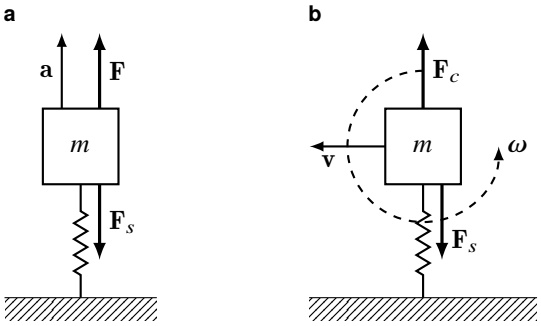

Figure 1.2.: Mass-spring system models of accelerometer (**a**) and gyroscope (**b**). The resulting force of the spring $\mathbf{F}_s$ is equal and opposite to the force applied to the body. In (**a**) the force $\mathbf{F}$ is due to the acceleration $\mathbf{a}$ according to Newton's second law. In (**b**) the angular velocity $\boldsymbol{\omega}$ and the imposed linear velocity $\mathbf{v}$ result in the fictitious Coriolis force $\mathbf{F}_c$.

Figure 1.3.: The SparkFun OpenLog Artemis[1](image licensed under CC BY 2.0[2]) features a 3-DOF accelerometer, gyroscope, and magnetometer as well as a programmable microprocessor and a microSD card allowing direct recording of motion data at a sampling rate of 250 Hz.

Apart from their use in navigation, IMUs have been extensively employed in clinical research, most notably in gait analysis [61, 68, 117], investigation of movement patterns related to conditions such as MS [100] and PD [21] as well as fall risk assessment [5, 67]. Accelerometers and gyroscopes are also standard components of today's smartphones where they can be used for activity and fitness tracking [65, 95, 127].

[1] https://www.sparkfun.com/products/16832
[2] https://creativecommons.org/licenses/by/2.0

Attitude estimation from inertial measurements

Since accelerometers measure linear acceleration, the second derivative of position, and gyroscopes angular velocity, the first derivative of orientation, a naive approach for estimating 6-DOF pose from inertial data would be integrating linear acceleration twice and angular velocity once. Unfortunately, sensor data is inevitably noisy, which leads to integration drifts that accumulate to position errors of several meters and orientation errors of several degrees within seconds [63].

Additionally, position is the double integral of inertial linear acceleration. Due to Einstein's equivalence principle, accelerometers cannot distiguish between gravitational force and force due to acceleration. Because of this, they can only measure overall gravito-inertial acceleration, the sum of inertial acceleration $\mathbf{i}$ and acceleration due to gravity $\mathbf{g}$ in the sensor's reference frame S:

$$^{S}\mathbf{a} = {}^{S}\mathbf{g} + {}^{S}\mathbf{i} \tag{1.11}$$

$^{S}\mathbf{g}$ can be calculated by rotating the acceleration due to gravity in the world frame $^{W}\mathbf{g}$ (which is constant and pointing in the upward vertical direction) by the quaternion representation of the sensor orientation $\mathbf{o}$:

$$^{S}\mathbf{g} = \mathrm{rot}\left(\mathbf{o}, {}^{W}\mathbf{g}\right) \tag{1.12}$$

To determine the inertial acceleration it is therefore necessary to estimate the orientation of the sensor, a process referred to as *attitude estimation*. Since both accelerometer and gyroscope measurements provide information about the orientation, data from both sensors can be combined for this estimate with a technique called *sensor fusion*.

The following will outline a sensor fusion-based attitude estimation approach in the context of *Kalman filtering*, an algorithm that uses multiple noisy measurements over time in combination with a dynamical system model to infer the unobservable state of the system [102]. This is by no means intended to be comprehensive nor rigorous; rather, it will provide an exemplary framework that will be further developed in the following section.

The filter assumes that the state $\mathbf{x}_k$ can be predicted from the previous state $\mathbf{x}_{k-1}$ by the state transition model:

$$\mathbf{x}_k = f\left(\mathbf{x}_{k-1}, \mathbf{u}_k\right) + \mathbf{w}_k \tag{1.13}$$

where f is some differentiable function called the state transition function, $\mathbf{u}_k$ denotes the control input[3] and $\mathbf{w}_k$ the process noise. The second component of the filter is the observation model that associates the measurements $\mathbf{z}_k$ with the current state:

$$\mathbf{z}_k = h\left(\mathbf{x}_k\right) + \mathbf{v}_k \tag{1.14}$$

[3] The name "control input" refers to the original problem formulation in the context of trajectory estimation in spaceflight. In fact, $\mathbf{u}_k$ can be any kind of known input signal, including a measurement from a sensor that is not part of the measurement model.

Here, h is the differentiable observation function and $\mathbf{v}_k$ denotes the observation noise. This formulation is called the extended Kalman filter (EKF) that generalizes the basic Kalman filter for non-linear functions f and h.

In addition to an estimate of the unobservable state $\mathbf{x}$, the EKF algorithm also keeps track of the uncertainty of the state estimate. This is reflected in the covariance matrix $\mathbf{P}$ that represents the joint variability of all state variables.

The EKF algorithm consists of two steps (see fig. 1.4):

1. In the *prediction* step, the *a priori* state estimate $\mathbf{x}_{k|k-1}$ is updated according to the state transition model:

$$\mathbf{x}_{k|k-1} = f\left(\mathbf{x}_{k-1|k-1}, \mathbf{u}_k\right) \tag{1.15}$$

   The a priori covariance matrix $\mathbf{P}_{k|k-1}$ is updated as well.

2. In the *update* step, the *a posteriori* state estimate $\mathbf{x}_{k|k}$ is updated using the observation model and the Kalman gain $\mathbf{K}_k$:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\left(\mathbf{z}_k - h\left(\mathbf{x}_{k|k-1}\right)\right) \tag{1.16}$$

   The a posteriori covariance matrix $\mathbf{P}_{k|k}$ is updated as well.

The Kalman gain $\mathbf{K}_k$ plays a central role in this algorithm as it determines the weight of the measurements vs. the prediction in the final state estimate. The calculations for $\mathbf{K}_k$, $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ are detailed in appendix A.3.

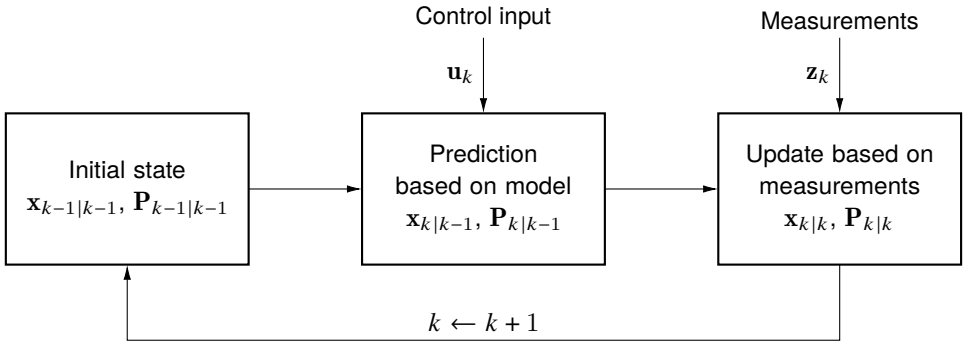

Figure 1.4.: The discrete-time EKF algorithm. In the prediction step, the state $\mathbf{x}$ and the covariance matrix $\mathbf{P}$ are estimated according to the state transition model. In the update step, the estimates are corrected according to the observer model.

Applying this algorithm to the problem of attitude estimation from linear acceleration and angular velocity, a quaternion-based state transition model can be written as [63]:

$$\underbrace{\mathbf{o}_k}_{\mathbf{x}_k} = \underbrace{\mathbf{o}_{k-1}}_{\mathbf{x}_{k-1}} \cdot q(\underbrace{\boldsymbol{\omega}_k}_{\mathbf{u}_k} \Delta t) \tag{1.17}$$

where $\boldsymbol{\omega}_k \, \Delta t$ is the estimated change of orientation from the angular velocity measurement $\boldsymbol{\omega}_k$ – represented as a rotation vector – and $q(\cdot)$ denotes the equivalent quaternion representation of this vector (see appendix A.1). Note that the measured angular velocity is interpreted as a control input in this model.

The corresponding observation model can be written as [63]:

$$\underbrace{\mathbf{a}_k}_{\mathbf{z}_k} = \mathrm{rot}(\underbrace{\mathbf{o}_k}_{\mathbf{x}_k}, {}^W\mathbf{g}) + \underbrace{\mathbf{i}_k}_{\mathbf{v}_k} \tag{1.18}$$

where ${}^W\mathbf{g}$ is the acceleration due to gravity in world coordinates. Note that the unobservable inertial acceleration is interpreted as a measurement error in this model.

As previously stated, the above state transition and observer models are developed merely for illustrative purposes. The problem of attitude estimation from IMU data is still an active research topic and the EKF algorithm is just one of the possible approaches to solving it. A survey of different estimation methods can be found in Crassidis et al. [33] and an EKF-based approach specifically designed for measuring orientation of human body segments in Luinge and Veltink [75].

Most recently, novel approaches make it possible to estimate earth-horizontal orientation and even 3D position and from purely inertial measurements by integrating machine learning models trained on large data sets of real-world human locomotion into the EKF architecture [72]. Another example where machine learning has been successfully applied to inertial data is for estimation of real-world walking speed based on acceleration measurements from a waist-worn device [110].

Visual-inertial simultaneous localization and mapping

Simultaneous localization and mapping (SLAM) refers to a class of algorithms originally developed in the 1980s for autonomous robots to solve the problem of an agent's navigation in an unknown environment [36]. The process has two goals:

- Constructing a map of the agent's surroundings (*mapping*).
- Estimating the position and orientation of the agent within that map (*localization*).

The general SLAM framework can incorporate various types of sensors, including inertial measurements, light detection and ranging (LIDAR) or global positioning system (GPS) data. One approach that has been extensively researched in recent years combines one or more video cameras with one or more IMUs and is referred to as visual-inertial simultaneous localization and mapping (VI-SLAM).

Many SLAM systems are built around an EKF and can be seen as an extension of the IMU-based approach described in the previous section. In addition to the orientation,

the state of the filter also incorporates the system's position and the location of so-called *landmarks* that the system tracks as it moves through the environment. A landmark can be thought of very much in its colloquial sense: a stationary feature of the environment that stands out from its surroundings and that can be easily detected from multiple locations.

The state of the EKF-SLAM algorithm consists of the agent's position and orientation w.r.t. the world frame:

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{p}_k \\ \mathbf{o}_k \end{pmatrix} \tag{1.19}$$

as well as the estimated positions of all currently tracked landmarks, also in the world frame:

$$\mathbf{m}_k = \begin{pmatrix} \mathbf{m}_k^{(1)} \\ \vdots \\ \mathbf{m}_k^{(n)} \end{pmatrix} \tag{1.20}$$

In the classic EKF-based VI-SLAM approach the inputs to the state transition model are the inertial measurements while the inputs to the observation model are the locations of visible landmarks in the camera image. The state transition model might look like this:

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{p}_{k-1} + \mathbf{t}_k \\ \mathbf{o}_{k-1} \Delta\mathbf{o}_k \end{pmatrix} \tag{1.21}$$

where

$$\Delta\mathbf{o}_k = q\left(\boldsymbol{\omega}_k \, \Delta t\right) \tag{1.22}$$

is the estimated change of orientation (see eq. 1.17) and

$$\mathbf{t}_k = {}^{W}\mathbf{i}_k \Delta t^2 = \left(\mathrm{rot}\left(\left(\mathbf{o}_{k-1}\Delta\mathbf{o}_k\right)^{-1}, \mathbf{a}_k\right) - {}^{W}\mathbf{g}\right)\Delta t^2 \tag{1.23}$$

is the change of position estimated by double-integrating the inertial acceleration in world coordinates ${}^{W}\mathbf{i}_k$.

In a monocular (i.e., single-camera) system the observer model describes the measurement of landmark locations by the camera. The position of a single landmark $\mathbf{m}^{(i)}$ w.r.t. the sensor frame S can be expressed as:

$${}^{S}\mathbf{m}_k^{(i)} = \begin{pmatrix} x_k^{(i)} \\ y_k^{(i)} \\ z_k^{(i)} \end{pmatrix} = \mathrm{rot}\left(\mathbf{o}_k^{-1}, \mathbf{m}^{(i)} - \mathbf{p}_k\right) \tag{1.24}$$

From this, the observation of a landmark is calculated as its projection onto the camera image plane [92]:

$$\mathbf{z}_k^{(i)} = \frac{1}{z_k^{(i)}} \begin{pmatrix} x_k^{(i)} \\ y_k^{(i)} \end{pmatrix} \tag{1.25}$$

Note that this simplified model assumes that measurements from camera and IMU are represented in the same reference frame S. For a rigorous discussion that takes into account the differences in reference frames as well as noise terms see Mourikis and Roumeliotis [92].

In order to track the landmark observations $\mathbf{z}_k^{(i)}$ across multiple camera images it is necessary to extract robust image features $\mathbf{f}_k^{(j)}$ that are invariant to position and orientation of the landmark w.r.t. the observer. This ensures that multiple observations of the same landmark can be properly associated. For this, the algorithm stores the corresponding features of all previously observed landmarks in a database.

A high-level schematic of this VI-SLAM algorithm is shown in fig. 1.5. It should be noted that the IMU usually provides measurements at a much higher sampling rate than the camera. This means that multiple prediction steps are carried out between update steps.

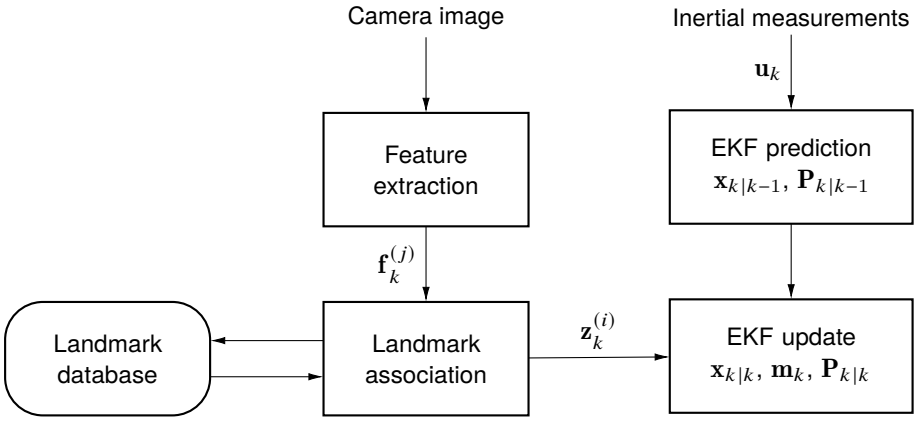

Figure 1.5.: High-level data flow of an EKF-based VI-SLAM system. The IMU measurements are interpreted as a control input $\mathbf{u}$ and used in the prediction step. Landmarks are extracted from the camera images and matched to previous observations. The perspective projections of the visible landmarks are the measurements $\mathbf{z}^{(i)}$ used in the update step.

One advantage of VI-SLAM is its suitability for tracking any kind of rigid body in both indoor and outdoor environments. Additionally, cameras and IMUs are cheap compared to other types of sensors (e.g., laser scanners). A drawback of using purely visual data for landmark tracking is that landmark depth can be difficult to estimate, even with stereo cameras.

The general SLAM framework also supports combining landmark observations from multiple agents or multiple runs across the environment by the same agent. Repeated observations of the same landmark reduce its uncertainty and drifts in the agent's

pose can be corrected when the agent returns to a previously visited location through a mechanism called *loop closure* [111].

In recent years, SLAM has seen interest especially as a key component of autonomous driving car systems [116]. Some VR systems such as the Oculus Quest also rely on SLAM for accurate head tracking because it eliminates the necessity for a stationary optical tracking system [50]. Nevertheless, applications for human motion tracking are still in their infancy, in part because commercial, self-contained SLAM systems have not been widely available. Novel devices such as the Intel® RealSense™ T265 tracking camera (fig. 1.6) could be used in the future for tracking human movement in the context of clinical or neuroscientific research (for a thorough discussion see chapters 2 and 3).

Figure 1.6.: The Intel® RealSense™ T265 tracking camera. The device features two fish eye cameras, an accelerometer, and a gyroscope and runs a VI-SLAM algorithm in real time on a dedicated chipset to produce a pose estimate at 200 Hz.

Optical tracking systems

An optical tracking system (OTS) is a system that estimates motion based purely on sequences of images of the moving body or bodies. This type of technique was used as early as 1887 by Eadward Muybridge to characterize equine gait based on photographic images [94]. Since then, video cameras with high frame rates have drastically improved the temporal resolution of image-based tracking, although estimating the pose e.g., of human limbs required the manual annotation of joint locations in the images [31].

The current state of the art in optical tracking are automatic marker-based systems. This technique uses multiple cameras that emit invisible infrared light which is reflected by specialized markers attached to the tracked body. Each rigid body must be outfitted with at least three markers to permit reconstructing its pose [31]. This reconstruction is

automatically performed by most commercial systems based on the observation of the markers by different cameras set up in a known geometry. For human motion tracking, placing markers on head, torso and limbs enables fitting of a skeleton model from which 3D joint positions and angles can be inferred (fig. 1.7a).

The capacity of these marker-based systems to track real-world human motion is however limited by the fact that pose estimation is only possible within a tracking volume observable by at least two cameras. Additionally, the markers can restrict movement and incorrect placement can lead to tracking errors [31]. Because of these limitations, markerless motion tracking is an active field of research in recent years. While certain methods still require multi-camera setups [93], novel machine-learning based approaches such as OpenPose [22] permit reconstructing 3D skeleton and face models from single camera images (fig. 1.7b).

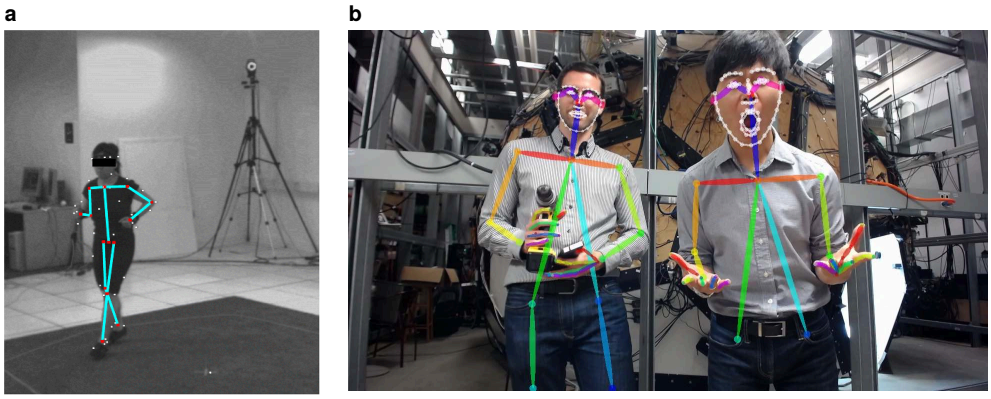

Figure 1.7.: Marker-based and markerless optical tracking systems. (**a**) Locations of optical tracking markers (white), keypoints (red) and skeleton model (cyan) detected by a marker-based multi-camera tracking system (from Colyer et al. [31], licensed under CC BY 4.0[4]). (**b**) Keypoints, skeleton and face model detected by the OpenPose markerless single camera tracking framework (from Cao et al. [22], copyright © 2021 IEEE).

One of the primary applications of OTSs is in the entertainment industry, notably for generating animations for movies, television shows and video games [88]. Many commercial VR systems also rely on optical tracking of the head-mounted display device and handheld controllers [8]. Finally, optical motion tracking is frequently used to investigate fundamental aspects of human locomotion [51, 97], sports bio-mechanics [31] and in clinical research [98, 119].

[4] https://creativecommons.org/licenses/by/4.0/

Eye tracking

Like with other kinds of human motion, the study of eye movements dates back more than two millennia; the Greek philosopher Aristotle already described the tight coupling between the motion of the left and right eye [126]. Precise quantitative automated eye tracking was first achieved in the 1960s with so called search coil systems. In this method, miniature metallic coils are placed in a contact lens which in turn is worn by subjects in their eyes. By applying an exterior magnetic field of constant direction and strength and measuring the resulting voltage induced in the coil, it is possible to determine the orientation of the coil relative to the field [103].

Nowadays, most commercially available eye trackers are video-based, i.e., they use one or more video cameras directed at the eyes. Most commonly, an infrared light source is used to illuminate the eyes whose reflections are picked up by the cameras. Processing the video image with specialized algorithms yields the location of the pupil (or pupils) and other distinct visual features of the eye within the image [52].

Video-based eye trackers can be broadly categorized into two categories: stationary and head-mounted. The former are placed in a fixed location in front of the subject to be tracked. In contrast, the latter are worn by the subject on their head and often feature a scene camera that captures a portion of the subject's visual field [52]. One example of a commercial head-mounted system is the Pupil eye tracking system [58] (fig. 1.8a).

Eye tracking is a central technique in researching the human visual system. Oftentimes, experiments involve presenting a stimulus on a screen viewed by a test subject while simultaneously tracking their eyes with a stationary system. Here, the goal of the tracking is to reconstruct the location of the participants' gaze in screen coordinates, a process called *gaze mapping*. This mapping – from eye locations in the eye camera image to gaze locations on the screen – has to be calculated for each session based on a calibration procedure. In a head-mounted setup, the same approach can be employed to locate the user's gaze within the scene camera image (fig 1.8b).

While this gaze mapping method allows to reconstruct two-dimensional (2D) gaze position in screen coordinates, it does not yield any information about the 3D pose of the eyes. Recently, model-based approaches of reconstructing eye pose w.r.t. the camera reference frame have been developed [121][5]. For a stationary system this directly corresponds to eye-in-world pose; a head-mounted system can be combined with a head tracker to obtain this (for a thorough discussion see chapter 2).

Eye tracking systems are used in neuroscientific investigation as eye movements are strongly related to processing of visual and vestibular stimuli by the brain. Measuring

[5] More precisely, eye orientation is usually restricted to 2 DOF because the eyes' capacity for torsional motion is limited and most tracking methods do not take it into account.

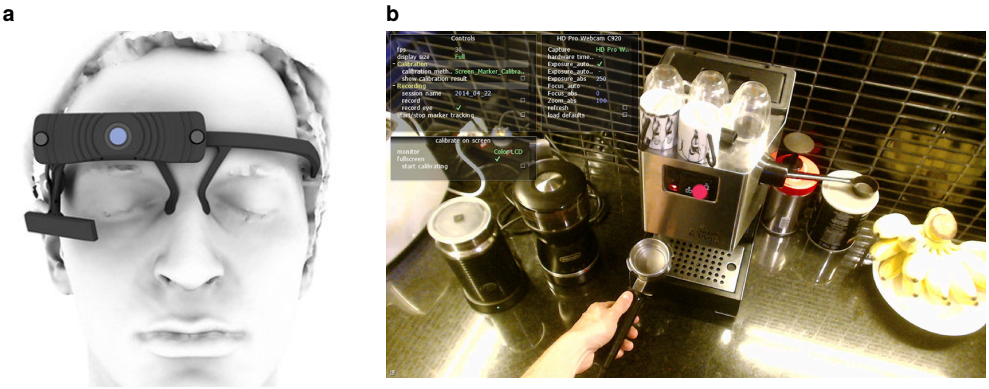

Figure 1.8.: The Pupil mobile eye tracking system (from Kassner et al. [59], licensed under CC BY-NC-SA 3.0[6]). (**a**) Render of the Pupil Pro headset showing frame, scene and eye camera. (**b**) Screen capture of the Pupil Capture software showing the scene camera image with the estimated gaze position of the user as a red circle.

eye motion is also important in clinical research as several diseases can manifest themselves in distinct motor patterns [107]. Commercial applications include marketing research and gaze-guided computer interfaces [52]. Novel VR systems based on multi-focal displays that decompose the rendered scene into multiple focus planes also require eye tracking [90].

1.3. Scope of this thesis

The goal of this thesis is to develop, validate and apply novel approaches of tracking human motion in real-world environments. To this end, I will address the following research questions:

RQ 1 *How can human motion be accurately tracked outside controlled environments?*

This question is addressed across the entire thesis. Chapter 2 introduces a novel concept for simultaneous real-world head and eye tracking. The system consists of a head-mounted eye tracker, the Pupil Core (see fig. 1.8) and a VI-SLAM-based head tracker, the Intel RealSense T265 (see fig. 1.6). The former allows tracking and recording position and orientation of both eyes w.r.t. its scene camera through a software application (Pupil Capture) that runs on a laptop connected to the tracker. The latter provides its pose w.r.t. a world-fixed reference frame which can be recorded with a plugin for the Pupil Capture software. The tracking devices are worn on the user's head and the laptop placed in a backpack which makes the system completely portable and suitable for use in any environment.

[6] https://creativecommons.org/licenses/by-nc-sa/3.0/

Beside the question how human motion can be tracked is it also important to assess how accurate this tracking is. Both the Pupil Core and the T265 are off-the-shelf products whose manufacturers provide claims of accuracy about their measurements without rigorous scientific support for these claims. Chapter 3 evaluates the tracking performance of the T265 for real-world human head tracking. This is especially relevant as the device was designed for general purpose motion tracking and not specifically for the use case of human movement.

Finally, chapter 4 discusses the use of IMUs for simultaneous measurement of head and trunk motion in daily life. In the study presented in this chapter, participants wear two sensors over a period of about 12 hours. From the measured data, periods of locomotion are extracted and derivative measures related to head stabilization are computed.

A drawback of the inside-out tracking approaches employed here is that they measure the motion of the tracking device which may differ from the movement of the body part it is attached to. As an example, sensors can be attached to the head at different locations and these locations can vary between subjects or trials. This problem can be solved by estimating the position and/or orientation of the device w.r.t. some anatomical reference based on a series of calibration movements (discussed in chapters 3 and 4).

RQ 2 *How can motion of different body parts (e.g., head, eye, trunk) be integrated?*

Tracking multiple body parts simultaneously with inside-out methods usually requires a dedicated tracker to be attached to each. These devices can be completely independent (e.g., when using multiple self-contained IMUs) which can result in offsets and drifts of the timestamps corresponding to each measurement. The same problem arises when the same body part is being tracked by different system, e.g., when evaluating the accuracy of a tracking method against a gold standard. Solving this issue requires time synchronization procedures that are discussed in chapters 3 and 4.

Another issue that arises when tracking multiple parts of the body (or the same parts with different methods) is the calculation of transformations between reference frames (see appendix A.2). This can be addressed in a similar manner to the problem of anatomical alignment described in RQ 1. A method to estimate reference frame transforms based on calibration movements is discussed in chapter 3. Chapter 2 describes an approach to determine the transformation between head and eye tracker by means of an image registration method. This allows to construct a full reference frame tree from the world frame to an eye-fixed frame and thus to calculate eye-in-world pose for both eyes.

RQ 3 *Does human motion, when tracked this way, differ from a controlled environment and if so, how?*

This question is addressed in two parts. First, it is important to investigate whether the reliability of the tracking system itself is influenced by real-world conditions. Chapter 3 examines the influence of the visual environment on the tracking quality of the T265's VI-SLAM algorithm. As discussed in section 1.2.2, the accuracy of a vision-based SLAM system considerably depends on its ability to estimate the 3D position of visual landmarks. This is especially important as the measurement of 6 DOF pose has historically been limited to stationary tracking systems such as OTSs. Portable IMU-based systems are subject to limitations in this regard (see section 1.2.2, p. 8) that could be overcome by VI-SLAM.

The second question is whether measurements of human movement obtained in a controlled environment are representative of real-world motion, i.e., whether movements are somehow biased due to the environment. Chapter 4 focuses on this issue by investigating whether measures of head stabilization reported by previous studies in a laboratory setting can be reproduced in the real world.

RQ 4 *What implications does this have for clinical and other applications?*

Last but not least, the research questions considered in this thesis should also inform clinical, commercial and other applications of human motion tracking. Chapter 3 explores the real-world tracking capabilities of the T265's VI-SLAM algorithm which could potentially be used as part of an AR system.

For clinical applications, sampling day-to-day behavior is important as any kind of clinical intervention will inevitably affect the patient's everyday life. Chapter 4 examines whether head stability parameters described in the literature could serve as diagnostic tools or outcome measures for clinical trials.

1.4. Related work

To the best of my knowledge, the portable head-eye tracking system proposed in chapter 2 is the first of its kind developed specifically for scientific research. The closest commercial solutions are the Tobii Pro Glasses 2[7] eye tracker and the Magic Leap 1[8] AR device, although none of these devices are particularly designed to measure 6-DOF head and eye pose.

Portable solutions for real-world eye tracking however have been described quite extensively in the literature; one of the first systems was presented by Babcock and Pelz

[7] https://www.tobiipro.com/de/produkte/tobii-pro-glasses-2/
[8] https://www.magicleap.com/en-us/magic-leap-1

[11]. Kassner et al. [58] introduced the first iteration of the Pupil eye tracking platform whose second generation is used in the system described in chapter 2. Several studies have also combined head and eye tracking. Kasneci et al. [57] investigated the visual exploration ability of patients with homonymous visual field defects during driving using a head-mounted eye tracker and a stationary (car-mounted) head tracker. Larsson et al. [66] described an eye tracking approach that uses an IMU to compensate head movements.

The suitability of SLAM for head tracking has also been discussed previously. Mayol et al. [85] reviewed the general requirements that a wearable SLAM system should fulfill. Cinaz and Kenn [29] proposed a SLAM-based head tracker consisting of an IMU and a laser scanner and Williamson et al. [130] evaluated an omni-directional approach combining multiple optical and depth cameras to address tracking issues caused by agile head movements. A novel VI-SLAM method specifically tailored to AR/VR applications is proposed by Fang et al. [38]. Multiple studies have also evaluated the tracking accuracy of the T265 [1, 3, 96]; for a detailed discussion of these papers refer to chapter 3.

There have also been numerous studies that investigated human motion in real-world settings. MacDougall [76] examined head, hip, wrist and ankle motion with accelerometers during every-day activities. 20 healthy subjects wore the sensors for a period of 10 hours during a typical day and annotated their activities. A frequency analysis revealed that the dominant frequency of vertical head acceleration was narrowly distributed around 2 Hz and did not depend on age, height or weight of the subject.

Carriot et al. [23] recorded head and foot motion with IMUs during predetermined activities. 8 subjects performed different active activities (e.g., running) as well as passive ones (e.g., riding a bus). The authors performed a frequency analysis and found that the spectral power of head motion stimuli does not follow a power law (i.e., it does not decay with $1/f^{\alpha}$ as is observed in other sensory modalities). Carriot et al. hypothesize that this is due to bio-mechanical filtering of passive motion stimuli by the body as well as motor control. In a follow-up study [24] the same authors also examined the power spectra of envelope signals of head movements and found that these were not well fit by a power law either. By modeling neuronal responses to the recorded head motion stimuli, they conclude that the unique envelope statistics of vestibular stimuli may drive the coding strategies of vestibular neurons.

MacNeilage and Glasauer [78] also measured head motion with IMUs during a fixed set of locomotor activities. They performed an analysis of variability across the stride cycle in order to quantify predictability of head movements. The authors found that predictability was lowest during mid-stance and higher during heel strike and toe off. For a maximum-likelihood model of cue integration by the brain, they propose a method to quantify the weight of vestibular sensory signals versus motor signals based on this predictability.

Stellmann et al. [120] investigated the ecological validity of standardized tests for clinical gait assessment in MS, i.e., how strongly the performance in these tests is associated with real-life mobility. For this, they recorded linear acceleration at the waist of 30 patients with MS performing 2 and 6 minute walking tests (2MWT and 6MWT), ten-meter walking tests (10mWTs), and during a 7 day period afterwards. They found that the ecological validity of the 10mWT was limited and longer uninterrupted walking occurred rarely during daily life in their population of subjects.

This example highlights the fact that measuring human motion in the real world is especially important for clinical applications. The approach that parameters relevant for diagnosis or monitoring of diseases should be sampled during everyday life is known as ecological momentary assessment (EMA) [112]. The application of EMA to measures of head stabilization is discussed in chapter 4. Previous studies that examined head stability in a laboratory setting [21, 51, 67, 86, 97, 98, 100] are reviewed there.

The next chapters tie into this related work as follows: chapter 2 proposes the first mobile open-source head/eye-tracking system, chapter 3 presents the first evaluation study of the Intel RealSense T265 specifically addressing natural human head motion and chapter 4 presents a study that discusses the validity of various head stability parameters in the real world for the first time.

2. An Open-Source Solution for Head-Eye Tracking Outside the Lab

*Originally published on June 2nd, 2020 in Proceedings of the ACM Symposium on Eye Tracking Research and Applications as **"Positional head-eye tracking outside the lab: an open-source solution"**.*

Original authors: Peter Hausamann, Christian Sinnott and Paul R. MacNeilage

DOI: `https://doi.org/10.1145/3379156.3391365`

Abstract

Simultaneous head and eye tracking has traditionally been confined to a laboratory setting and real-world motion tracking limited to measuring linear acceleration and angular velocity. Recently available mobile devices such as the Pupil Core eye tracker and the Intel RealSense T265 motion tracker promise to deliver accurate measurements outside the lab. Here, the researchers propose a hard- and software framework that combines both devices into a robust, usable, low-cost head and eye tracking system. The developed software is open source and the required hardware modifications can be 3D printed. The researchers demonstrate the system's ability to measure head and eye movements in two tasks: an eyes-fixed head rotation task eliciting the vestibulo-ocular reflex inside the laboratory, and a natural locomotion task where a subject walks around a building outside of the laboratory. The resultant head and eye movements are discussed, as well as future implementations of this system.

Summary and author contributions

This chapter addresses RQ 1 and RQ 2 by describing a lightweight, portable system for simultaneous tracking of real-world human head and eye motion. The system consists of the Pupil Core mobile eye tracker onto which a RealSense T265 is mounted by means of a custom-designed 3D-printed bracket. Data from the eye tracker can be acquired through the open-source Pupil Capture software. We developed a plugin for this software that allows simultaneous acquisition of head tracking data from the T265. A second plugin can be used to estimate the transformation between the Pupil Core's scene camera and the T265's stereo cameras (the so-called extrinsics) by means of an image registration method. Additionally, another plugin is presented that permits exporting recorded data to an analysis format through Pupil Player, an open-source application for offline visualization and analysis of Pupil Core recordings.

The ability of the system to simultaneously track head and eye movements is demonstrated with two experiments:

1. An eyes-fixed head rotation task eliciting the vestibulo-ocular reflex inside the laboratory.
2. A natural locomotion task where a subject walks around a building outside of the laboratory.

The experimental design was conceived by all authors and all recordings were piloted and carried out by Christian Sinnott who also designed the 3D-printed bracket for mounting the RealSense T265 onto the Pupil Core eye tracker and prepared figure 1. I implemented the plugins for recording, extrinsics estimation and data export, performed all analyses and generated the plots.

The manuscript was written by all authors; my primary contribution is the methods and results section while Christian Sinnott and Paul MacNeilage primarily wrote the introduction and discussion.

2.1. Introduction

Mobile observers tend to fixate environmental features that are stationary in a world-fixed reference frame. However, this behavior is usually impossible to characterize because mobile eye trackers measure eye movements in a head-fixed reference frame (eye-in-head) [62, 64]. Information about head movement relative to the environment (head-in-world) would allow characterizing eye movements in a world-fixed reference frame (eye-in-world), but positional head tracking has historically been difficult to achieve outside the lab. To comprehensively characterize natural interaction with the visual environment (eye-in-world), a convenient mobile solution is needed that allows for simultaneous positional tracking of head-in-world and eye-in-head movements outside the lab.

Measurement of head movement outside the laboratory setting has traditionally relied on microelectromechanical system (MEMS)-based inertial measurement units (IMUs). IMUs are advantageous due to their portability and affordability, but they do not measure 6 degree of freedom (DOF) head-in-world position. They are limited to measurement of angular velocity and sum total gravitoinertial acceleration. Angular position may be estimated by integrating angular velocity, but these estimates are subject to drift [62, 70]. Estimation of linear position is even more problematic due to the ambiguity of gravitoinertial acceleration. Distinct estimates of gravitational and inertial acceleration may be obtained, e.g. via Kalman filtering [71], but double-integration of the resulting inertial acceleration to obtain linear position yields highly unreliable estimates. Nevertheless, several recent studies have investigated the use of IMU data to compensate for head movements during eye tracking [64, 66].

More reliable positional estimates of head movement outside the lab may be obtained by fusing IMU data with other data streams. One recent study used a motion capture suit composed of multiple IMUs to track full-body kinematics along with eye-in-head movement [83]. However, use of multiple IMUs does not fully solve the problem of drift. Perhaps the most promising solution involves joint analysis of IMU data along with video recorded from a rigidly mounted camera, known as visual-inertial simultaneous localization and mapping (VI-SLAM). This is a computer vision-based method that relies on frame-by-frame video analysis to create a 3D representation of the environment and simultaneously localize the agent's position relative to that environment [28].

Here, we combine two devices to create a highly mobile system for simultaneous tracking of head and eye position. To track 6 DOF head position we use a ready-made VI-SLAM system, the Intel RealSense T265 tracking camera, which fuses inertial data from an IMU with two global shutter fisheye cameras to generate a 6 DOF pose estimate at 200Hz. The system is relatively compact and affordable, and the on-board VI-SLAM solution has the benefit of greatly reducing computational load. To track eye movements, we used the Pupil Core head-mounted eye tracker which is also relatively inexpensive

and lightweight. In addition, the software suite supporting the Pupil Core is entirely open source facilitating integration with other devices.

In this paper, we present software that allows the Pupil Core and Intel RealSense T265 to be used together for simultaneous positional head-eye tracking outside the lab. We demonstrate the functionality of the system by presenting two example datasets. Tasks were chosen to demonstrate capture of basic physiological head-eye behaviors in both a laboratory and a real-world setting.

2.2. Methods

2.2.1. Hardware

Pupil Core

The Pupil Core system is a mobile eye tracking device from Pupil Labs [59]. Our configuration of the Pupil Core eye-tracker features three cameras. Two of these film the eyes at 200 frames per second (FPS) with a resolution of 192x192 pixels. The last camera is an outward-facing world camera and records at 30 FPS with a 1280x720 pixel resolution and a diagonal field of view (FOV) of 100°. After consulting with other researchers we modified the Pupil Core to add a nose cushion (for ergonomics) and a IR-reflective film around the user's eyes (to reduce errant IR reflections during outdoor recording [18]).

RealSense T265

Head pose was measured using the Intel RealSense T265 tracking camera. The device consists of two global shutter fisheye world cameras (173° diagonal FOV; 30 Hz frame rate; 848x800 pixel resolution), a 3 DOF accelerometer (± 4g range; 62.5 Hz sampling rate), and a 3 DOF gyroscope ($\pm 2000 \frac{^\circ}{s}$ range; 200 Hz sampling rate). The T265 uses VI-SLAM to fuse data from these streams to estimate 6 DOF position and velocity of the camera relative to the environment at 200 Hz.

Camera mount

Our solution requires recording simultaneously from both the T265 and the native Pupil world camera, and the two must be rigidly attached to each other. To this end, we designed and 3D-printed a mounting bracket that holds the T265 and can be clipped and glued to the world camera. The CAD model for this bracket is freely available for download[1].

[1] `https://github.com/vedb/pupil-t265`

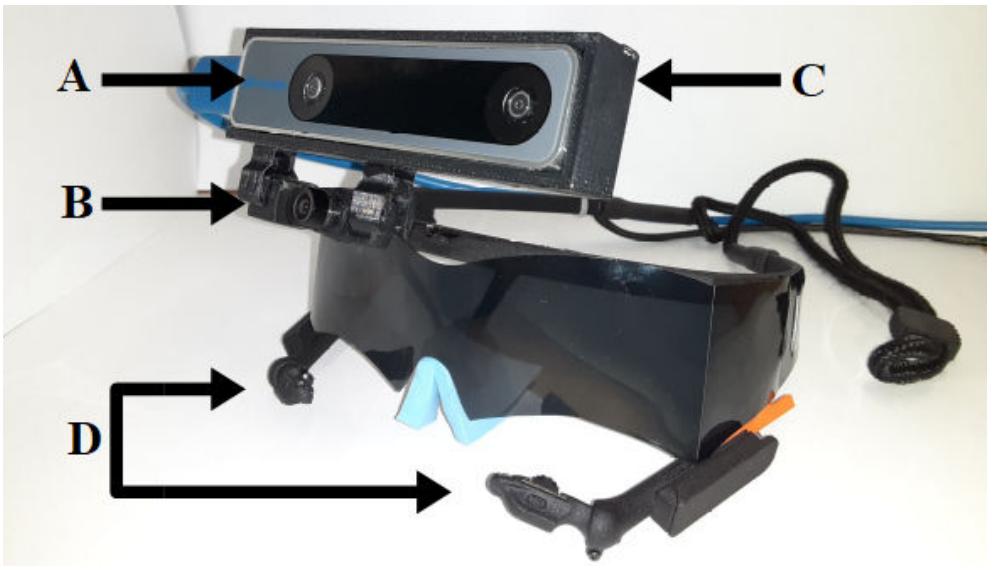

Figure 2.1.: Pupil and RealSense combined head and eye movement tracker. *The RealSense T265 (A) is mounted rigidly above the Pupil Core world camera (B) via a custom 3D-printed bracket (C). The eye cameras (D) are also visible.*

2.2.2. Software

Overview

Data acquisition and export was implemented as a set of Pupil Core software plugins. These plugins are written in Python and can be used by copying the plugin source file to a folder specified by the Pupil Core software. The software suite developed by Pupil Labs for use with Pupil Core consists of three programs: Pupil Capture, a data acquisition program with GUI; Pupil Player, a data visualizer and exporter; and Pupil Service, a debugging and acquisition (from terminal) program. We use Pupil Capture and Player for purposes of this report.

All data recorded by the plugins is stored in the same format as the data streams recorded by Pupil Capture (such as gaze positions) in order to to streamline the integration with the software. The plugins along with example data and analysis code are open source and freely available for download[1]. In the code, head tracking data is referred to as odometry as it includes positions as well as velocities.

Recording

Eye tracking data is recorded via Pupil Capture. In its original configuration, Pupil Capture allows the user to record from a world-facing camera and up to two eye cameras simultaneously. The user is able to modulate sampling rate, resolution, exposure time,

and other image parameters before recording. The software also allows for the user to map gaze in real time (online) or later, after a recording has been made (offline).

Recording of head tracking data is achieved through a plugin for Pupil Capture. The user can start the T265 device by clicking a button, and if the device is successfully started, they will see a live readout of the current tracker sampling rate, confidence, linear and angular position, and linear and angular velocity. When a recording is initiated through Pupil Capture, head tracking data is continuously written to disk along with eye tracking data. It is also possible to show the video stream recorded by the two T265 fisheye cameras in a separate window.

Camera extrinsics estimation

Both the T265 and the Pupil Core tracker define reference frames with respect to their world cameras. However, for analyses directly relating head and eye movements, the exact transformation between the two devices needs to be calculated in order to transform head and eye tracking data into a common reference frame. For this purpose, a second plugin was developed that estimates the position and orientation of the RealSense T265 with respect to the Pupil world camera. This is achieved by capturing a calibration pattern presented on the screen by all cameras from multiple angles and performing a stereo camera calibration for pairs of cameras using the `stereoCalibrate` function from the OpenCV computer vision library [19] (version 3.2, included in the Pupil software suite). The plugin also takes into account the different coordinate system of the T265 pose (x-axis leftwards, y-axis upwards, z-axis backwards) compared to the standard camera coordinates used by Pupil Core (x-axis leftwards, y-axis downwards, z-axis forwards).

As with the recording plugin, the user can inspect the fisheye video stream from the T265 which can be useful to verify that the calibration target is captured by all three cameras. The target, consisting of an asymmetric grid with 44 circle markers, can be shown in a window or in fullscreen mode. Calibration images (10 in total) can be acquired with a key or button press in the world camera window. The plugin will also draw the extent of the previously recorded targets over the current world camera image so that the user can make sure to cover the entire FOV of the camera.

Data export

Export is implemented as a plugin for Pupil Player. When the plugin is activated, head tracking data and camera extrinsics are saved as `.csv` files each time an export is triggered through Pupil Player. The export format is consistent with other exports that can be made with Pupil Player such as gaze data and includes timestamps and the indices of corresponding world camera frames.

2.2.3. Data acquisition

To illustrate the solution, we recorded data from two subjects and analyzed this data to demonstrate the workflow. Each subject put on the modified Pupil Core tracker which was attached to a laptop carried in a backpack. A 9-point calibration routine was performed in which the experimenter presented a specialized marker at nine locations configured in a 3x3 grid across the FOV of the world camera. The subject fixated the center of the marker for several seconds at each location, viewing the calibration target grid from a distance of about 1.2 m. This array corresponds to a 72° by 42° displacement of visual angle. Data from this calibration procedure was later used to perform post-hoc gaze mapping in the Pupil Player software.

After the calibration, the first subject was presented with the same kind of marker at the center of the visual field and was instructed to fixate on the marker while nodding and shaking their head several times to elicit a rotational vestibulo-ocular reflex (rVOR) response. Then, while still fixating on the target, they performed several vertical full-body squatting movements to elicit a translational VOR (tVOR) response. This was done at two viewing distances of approximately 1.2 and 2.1 m. The second subject performed the same calibration routine and then walked around the outside of a building for roughly five minutes before returning to the lab.

2.2.4. Analysis

Reference frame transformations

The T265 measures position and velocity of the device with respect to a world-fixed reference frame whose origin is at the point where the device was started. The Pupil Core measures the 3D position of the gaze point with respect to a reference frame fixed to the world camera. The transformation between the T265 and Pupil Core coordinate systems is computed with the calibration procedure outlined in section 2.2.2. The orientation of the eye with respect to the Pupil Core world camera frame can be estimated by computing the shortest arc rotation between a unit vector perpendicular to the image plane and the gaze point.

Finally, we define a coordinate system for analysis where the x-axis points forward, the y-axis to the left and the z-axis upwards. All measurements in our analysis are transformed from their original reference frame to this coordinate system. While this coordinate system is easier to interpret from an anatomical perspective, it is still relative to the Pupil Core world camera rather than an anatomical reference such as Reid's plane (see discussion). Directions and orientations are also expressed in a spherical coordinate system where the azimuth angle is measured from the positive x-axis.

Data pre-processing

The 3D gaze position was transformed to the spherical coordinate system defined in the previous section and smoothed with a boxcar filter with a window length of 500 ms. Afterwards, gaze velocity was computed by differentiating the gaze position with respect to time. Pupil Core provides a confidence score of the gaze estimate between 0 and 1. All gaze data with a confidence below 0.8 was excluded from the analysis. Similarly, the T265 reports a tracking confidence from 0 to 3, with a score of 3 corresponding to "high" tracking confidence. All tracking data below the highest score was also excluded. Additionally, gaze data whose velocity had a magnitude above $1000\frac{^\circ}{s}$ was excluded from analysis [52].

Vestibulo-ocular reflex

The angular and linear movements of the subject while fixating the calibration target were used to demonstrate the reflexive eye movements induced by the rVOR and tVOR. Both rVOR and tVOR are combined head and eye movements performed constantly in normally functioning vertebrates for image stabilization [37], and represent an ambiguous case for gaze classifiers when head movement data is unavailable. For the rVOR, we compared eye yaw and pitch velocity with negative head yaw and pitch velocity, respectively. For the tVOR, we compared eye pitch velocity with the angular velocity of the head relative to the marker with $\omega = \arctan(v_z/d)$ where v_z is the vertical head velocity and d is the distance of the marker from the head. VOR velocity gain was calculated by computing the median ratio of eye velocity to head velocity along the corresponding axis for each segment.

Eye-in-world velocity

Motion at the retina is driven predominantly by how the eye is moving relative to the world-fixed visual environment. Therefore, characterization of eye-in-world velocity can provide insight into typical visual motion stimuli during natural behavior [79, 82]. As described above, reconstruction of eye-in-world position and velocity is only possible by combining measures of head-in-world and eye-in-head movement.

To illustrate one application of positional head-eye tracking, we analyze the direction of linear velocity (heading) in eye coordinates (eye-in-world) during outdoor walking. Linear head velocity measured in the world frame by the T265 was transformed into spherical head coordinates and then further transformed into spherical eye coordinates using eye-in-head position (see section 2.2.4). We plot 2D histograms of heading as well as the eye-in-head position between $\pm 45^\circ$ azimuth and elevation.

2.3. Results

2.3.1. Vestibulo-ocular reflex

During the rVOR task, eye velocity and negative head velocity were closely aligned and exhibited a gain of 1.01 and 0.83 for yaw and pitch, respectively (figures 2.2A and B). Similar alignment was observed for eye velocity and angular velocity of the head relative to the fixation target during the tVOR task, although eye movement seemed to over-compensate for head movement (VOR velocity gains 1.32 and 1.52 for near and far fixation, respectively, figures 2.2C and D). This can be explained by the fact that the subject simultaneously performed small involuntary head pitch rotations that also need to be compensated.

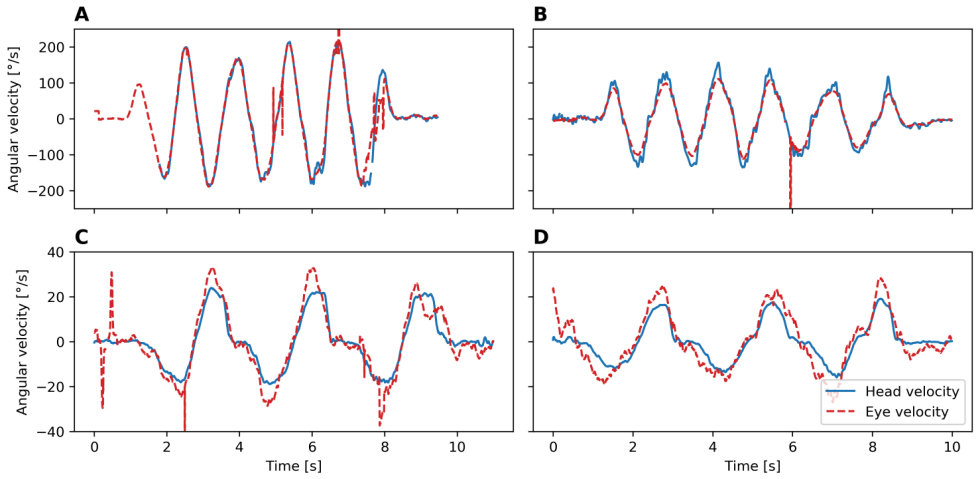

Figure 2.2.: Comparison of eye and head movements during fixation task. *Eye yaw velocity and negative head yaw velocity (A). Eye pitch velocity and negative head pitch velocity (B). Eye pitch velocity and angular velocity of the head relative to the fixation target (1.219 m distance, C). Eye pitch velocity and angular velocity of the head relative to the fixation target (2.143 m distance, D).*

2.3.2. Eye-in-world velocity

Direction of linear head-in-world velocity during walking for this subject and environment showed an elongated distribution along the elevation axis (figure 2.3A). The distribution for eye-in-head position was more circular (figure 2.3B). The offsets of these distributions relative to the origin are likely observed because the respective coordinate systems are not aligned with an anatomical reference frame such as Reid's plane. In contrast, eye-in-world velocity exhibited a Gaussian-like distribution, centered in both azimuth and elevation direction (figure 2.3C) reflecting a tendency to look in the direction of linear motion. This result shows that although head velocity and gaze position are not

aligned with an anatomical reference, the resulting eye-in-world velocity is aligned with an anatomical eye-fixed reference frame.

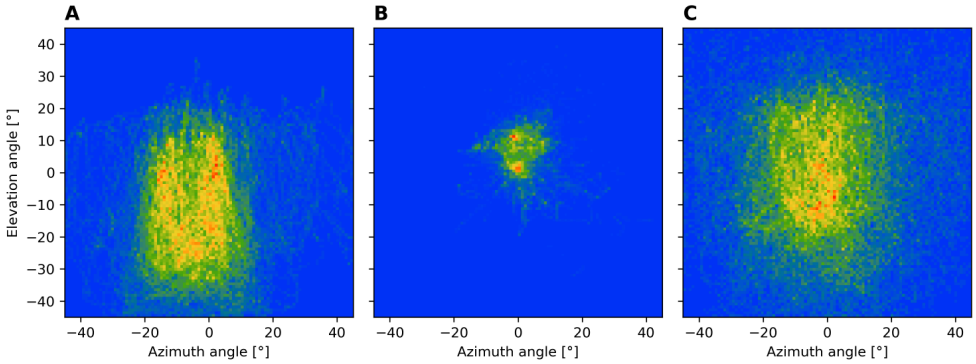

Figure 2.3.: Distributions of head/eye velocities and gaze direction during walking task. *Direction of linear head-in-world velocity (A). Eye-in-head gaze direction (B). Direction of linear eye-in-world velocity (C).*

2.4. Discussion

We demonstrate a mobile head- and eye-tracking platform capable of robust positional measurement and present analysis of two paradigms to validate the system, a fixation and a walking task. Simultaneous head and eye measurements during the fixation task revealed the expected compensatory eye movements (figure 2.2). Joint analysis of head and eye measurements during walking yielded a novel characterization of linear eye-in-world velocity that is a strong determinant of retinal image motion (figure 2.3).

While our head-eye tracker is relatively light on the user's head, it still requires use of a dedicated computer. We used a high-performance laptop carried in a backpack for data collection that was relatively heavy. Additionally, online data monitoring was impossible with this hardware configuration. This configuration is similar to another off-the-shelf product, Magic Leap. Our solution has the advantage of higher gaze sampling rates and open access to software and hardware collecting these data.

Additionally, weight of future hardware configurations is likely to decrease as personal computing improves. However, online data monitoring must be implemented by modifying how the laptop is carried to make the screen visible. One promising solution uses a lightweight hydration backpack with a window cutout to display the screen of the tablet being used for data collection [106].

Another concern is minimizing headset slippage. We attempted to control slippage using a sport strap fixed to the end-pieces of the Pupil Core eye tracker glasses. With

this solution, headset slippage did not prove to be an issue in the collection of this pilot data, but collection over longer periods must be tested.

To address slippage as well as conversion of measurements into an anatomically defined reference frame such as Reid's plane, it would be useful to measure the 3D position and orientation of the device relative to the head. This would be possible, for example, by scanning the participant's head with headset just before data collection. One possible method, originally developed to create custom head stabilization devices [43] employs a stereo camera attached to a cell phone. To quantify the amount of headset slippage experienced by any single participant one could compare scans acquired before and after a recording session.

Using VI-SLAM to measure head position is promising, but VI-SLAM-based solutions can deteriorate when visual features of the environment move, for example in a snow-storm, when walking in a crowd or when surrounded by foliage that is blowing in the wind. This is because VI-SLAM works under the assumption that environmental features are world-fixed. Roboticists have tried solving this problem by orienting cameras towards the ceiling in indoor environments [56], but this is not feasible with our system outdoors.

The Intel RealSense T265 was developed primarily to facilitate navigation in autonomous robots such as drones or self-driving vehicles [28]. Future human-centric development of a VI-SLAM pipeline leveraging regularities of human movement should lead to better estimation of head position. Higher quality head movement data can then be used to improve gaze classification performance [62, 64].

Several recent studies have characterized natural head and eye movements during everyday activities in natural environments [18, 79, 82, 83]. Novel, versatile, and accessible methods, like the ones presented here, will help advance these research endeavors by enabling longer recording sessions with data collection across a wider range of activities, environments, and subjects. Ultimately, this knowledge will lead to better understanding of human visual exploration and head-eye coordination, and this knowledge can be applied in the development of interactive technologies that rely on tracking and prediction of head and eye movements, such as augmented and virtual reality.

3. Evaluation of the Intel RealSense T265 for Tracking Head Motion

Originally published on June 14th, 2021 in Nature Scientific Reports as ***"Evaluation of the Intel RealSense T265 for Tracking Natural Human Head Motion"****.*

Original authors: Peter Hausamann, Christian Sinnott, Martin Daumer and Paul R. MacNeilage

DOI: `https://doi.org/10.1038/s41598-021-91861-5`

Abstract

Accurate and robust tracking of natural human head motion in natural environments is important for a number of applications including virtual and augmented reality, clinical diagnostics, as well as basic scientific research. Inertial measurement units (IMUs) provide a versatile solution for recording inertial data including linear acceleration and angular velocity, but reconstructing head position is difficult or impossible. This problem can be solved by incorporating visual data using a technique known as visual-inertial simultaneous localization and mapping (VI-SLAM). A recently released commercial solution, the Intel RealSense T265, uses a proprietary VI-SLAM algorithm to estimate linear and angular position and velocity, but the performance of this device for tracking of natural human head motion in natural environments has not yet been comprehensively evaluated against gold-standard methods. In this study, we used a wide range of metrics to evaluate the performance of the T265 with different walking speeds in different environments, both indoor and outdoor, against two gold-standard methods, an optical tracking system and a so-called perambulator. Overall, we find that performance of the T265 relative to these gold-standard methods is most accurate for slow to normal walking speeds in small- to medium-sized environments. The suitability of this device for future scientific studies depends on the application; data presented here can be useful in making that determination.

Summary and author contributions

This chapter is a significant contribution to RQ 1 as it evaluates a novel approach for measuring human movement in a real-world setting. The aim of this chapter is to evaluate the performance of an off-the-shelf VI-SLAM device, the Intel RealSense T265, specifically in the context of real-world head motion tracking. For this purpose, two studies are conducted comparing the data from the T265 against different gold standard tracking systems:

1. In an indoor environment with an optical tracking system (OTS).
2. In three different in- and outdoor environments with a perambulator, a device capable of measuring walking speed and distance.

The accuracy of the T265 with respect to the gold standard methods is assessed using a variety of metrics that quantify the errors in position and orientation as well as linear and angular velocity.

The primary result of this chapter is that the tracking performance of the T265 depends considerably on the environment and the movement speed. Overall, the lowest errors in both studies were measured during normal walking in small indoor environments. Additionally, the accuracy of the tracker seems to be acceptable for measures of direction such as the orientation of the device with respect to gravity and the direction of linear velocity. On the other hand, measures of magnitudes such as linear speed and displacement can be severely under-estimated, especially in large outdoor environments.

The OTS study was designed by myself, Christian Sinnott and Paul MacNeilage and all recordings were piloted and carried out by Christian Sinnott. The perambulator study was designed by myself, Martin Daumer and four students of the lecture “Clinical Applications of Computational Medicine” at TUM, Tobias Allgeier, Jana Daubmeier, Stefan Haupt and Mario Wenning. These students also performed extensive piloting for this study. The final recordings for this paper were however exclusively carried out by myself. I also developed the method for the reference frame transformations, carried out all data processing and analysis and generated all figures.

The paper itself was written by all authors; my primary contribution is the methods and results section as well as the supplementary material (appendix B) while Christian Sinnott and Paul MacNeilage primarily wrote the introduction and discussion and Martin Daumer revised the manuscript prior to submission.

3.1. Introduction

Tracking of human head motion is important across several domains. It is important for investigating basic scientific questions about reflexive control of posture, as well as reflexive stabilization of both head and eye movement [99]. It is also important in applied areas. For example, virtual and augmented reality (VR and AR) rely on tracking of human head motion to render the appropriate visual scene motion in head-mounted displays. And in a clinical setting, one can compare measures of head movement between normal and patient populations to assist in diagnosis and treatment of sensory, motor, and neurological disorders [48].

Historically, observation of how the head moves in space has been constrained to laboratory settings [77]. In early research, accurate, precise head tracking demanded that the participant wear bulky equipment to track the head mechanically or via magnetic search coil [14, 15, 32]. Advances in technology allowed robust head tracking to be conducted with optical tracking systems on humans and other mammals [80, 99]. This has been referred to as outside-in head tracking because stationary cameras "outside" the participant are used to track the moving head [101]. This method was more versatile, but robust performance was still confined to the laboratory. More recently, microelectromechanical system (MEMS)-based inertial measurement units (IMUs) have become accessible and affordable enough for widespread use, which in turn has allowed measurement of head movements outside the laboratory [24, 48, 76, 84].

MEMS IMUs typically consist of a tri-axial accelerometer and gyroscope, and sometimes a magnetometer, all built into a single small device. These allow estimating linear acceleration, angular velocity, and direction and strength of the local magnetic field, respectively. These estimates may be further processed to estimate orientation relative to gravity, linear velocity and position. Through each of these steps, error is introduced, particularly when integrating and double integrating to estimate linear velocity and position. Estimating orientation is less error-prone because accelerometer, gyroscopes and magnetometers all incorporate information about their orientation with respect to a local reference frame and model-based approaches such as the extended Kalman filter can fuse measurements from all three sensors [105].

One possibility to address these problems is to incorporate visual data. Visual-inertial simultaneous localization and mapping (VI-SLAM) is a method developed primarily for use in autonomous robots [42]. The method generally assumes that the IMU and camera(s) are rigidly attached to one another and relies on tracking of visual features of the stationary environment to augment the estimate of linear and angular position derived from IMU data. In the context of virtual and augmented reality, this type of tracking is referred to as inside-out (rather than outside-in) because the sensors mounted on the moving observer are used to track the stationary environment. Optimal algorithms for VI-SLAM are an area of active research. However, a commercially available VI-SLAM

device, the Intel RealSense T265, was recently released and represents a promising tool for versatile tracking of natural human head motion outside the lab. If the T265 device is going to be adopted as a standard tool, its performance must be evaluated. This is especially necessary because only a rough description of the T265's tracking method is provided [46]; no details about the proprietary closed-source VI-SLAM implementation are available.

Previous studies have compared estimates of position and orientation from the T265 against an optical tracking system (OTS). Alapetite et al. [3] mounted the device on a wheeled robot and investigated the influence of movement speed as well as the quantity of visual features and moving objects in the environment on the tracking quality. Their results show that tracking performance decreases with higher motion speeds and lower feature density. Ouerghi et al. [96] evaluated the tracking performance of a hand-held T265 in an industrial environment and measured positional errors below 2 % of the overall length of motion trajectories. Agarwal et al. [1] evaluated the device for indoor navigation of an unmanned aerial system (UAS) and report heading errors of around 3 degrees. Bayer and Faigl [16] proposed an approach combining the T265 with the RealSense D435 depth camera as a navigation system for a hexapod walking robot and report positional errors of around 10 cm in a laboraty environment. One major gap in the current literature is the evaluation of the T265 for tracking head motion. The device is lightweight (33 g) and affordable (~$ 200), and it is advertised as solution for head tracking for AR and VR. However, to our knowledge, and to date, there are no publicly available studies evaluating its performance in this context.

3.2. Methods

3.2.1. Evaluation in optical tracking space

A convenience sample of nine subjects (five female, four male; aged 20-46 years, mean age of 27.8 years) with no known history of vestibular or gait disorders were recruited. All procedures were approved by the Institutional Review Board of the University of Nevada Reno and carried out in accordance with relevant guidelines and regulations. In this study, the pose estimated by the T265 was compared with a gold-standard pose estimate generated by an OTS. A marker was attached to the T265 so that it could be tracked by the OTS. The device and marker were worn by participants on their heads using an elastic headband designed for mounting cameras on the head or helmet during sports activities (Fig. 3.1**a**, informed consent to publish the image in an online open-access publication was obtained from the participant).

Participants first performed a synchronization motion by nodding and shaking their head slowly five times each. This data was used to temporally align the T265 and the OTS recordings. Subjects then completed ten laps around the tracking space (Fig.

3.1**d**) at three self-chosen speeds: “at a leisurely walking pace”, “at a brisk walking pace”, and “at a jogging pace”. The first five laps for each pace were in a clockwise direction and the last five laps in a counterclockwise direction.

3.2.2. Evaluation with perambulator

Eight different subjects (three female, five male; aged 26-31 years, mean age of 28 years) with no history of vestibular or gait disorders were recruited for the second part of the study that investigated the speed estimated by the T265 in real-world environments at the main campus of Technical University of Munich (TUM). Here, a so-called perambulator was used as the gold standard measurement device. The perambulator is a surveyor’s wheel (see Fig. 3.1**c**) - a device generally used for measuring distances , e.g., in civil engineering - that was modified such that it was also capable of measuring speed [2]. Such a perambulator device has been used in previous studies [109] in order to measure real-world walking speed of participants. The device is lightweight and can be pushed in an ergonomic manner and thus did not considerably inhibit the normal walking and jogging movements of the participants.

Subjects performed the same set of tasks as in the first study (walk, slow walk, jog) in three different environments while wearing the T265 on the head and pushing the perambulator. The first environment was a hallway ($47 \times 4 \times 5$ m, about 80 m circuit length, Fig. 3.1**e**), the second a large lobby ($37 \times 12 - 30 \times 6$ m, about 80 m circuit length, Fig. 3.1**f**) and the third a large courtyard (60×60 m, surrounded by 5-6 story buildings, about 160 m circuit length, Fig. 3.1**g**). Participants were instructed to move along a pre-defined path in each environment. An experimenter took note of the distance measured by the perambulator in each task and environment. All subjects signed an informed consent form compliant with the European General Data Protection Regulation. The study protocol was approved by the institutional review board of the Sylvia Lawry Center for Multiple Sclerosis Research and procedures were carried out in accordance with relevant guidelines and regulations.

3.2.3. Hardware

The Intel Realsense T265 tracking camera has a diverse suite of sensors which all feed into a VI-SLAM pipeline, which fuses them into a 6 degree of freedom (DOF) estimation of position and velocity of the camera relative to the environment at 200 Hz. The sensors consist of two global shutter fisheye world cameras (173 ° diagonal field of view (FOV); 848×800 pixel resolution; 30 Hz sampling rate), a 3 DOF gyroscope ($\pm 2000 \ \frac{°}{s}$ range; 200 Hz sampling rate), and a 3 DOF accelerometer (± 4 g range; 62.5 Hz sampling rate). The 6 DOF estimation of camera position and velocity is computed

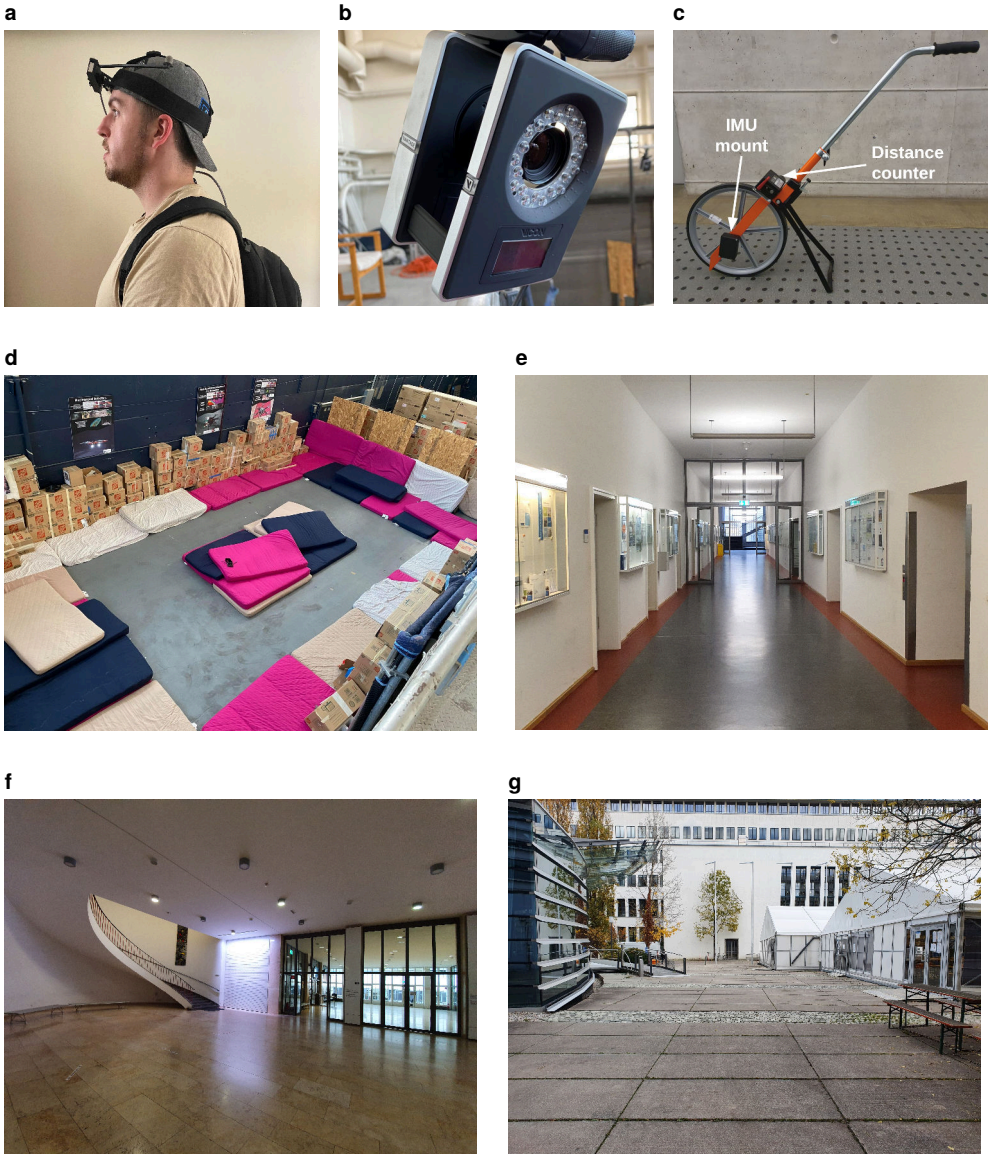

Figure 3.1.: Equipment used for the two studies and snapshots of the four environments where data was recorded. (**a**) Subject wearing head mount with T265 and tracking marker. (**b**) One of the Vicon Vantage cameras used in the optical tracking space. (**c**) Perambulator with distance counter and IMU mount. (**d**) Optical tracking space. (**e**) Hallway environment. (**f**) Lobby environment. (**g**) Courtyard environment.

in real-time onboard the T265 on a dedicated chipset. In both studies, data from the device was recorded via USB with a laptop carried by the participant in a slim backpack.

In the first study, twelve Vicon Vantage 8 cameras were used to perform optical tracking of a rigid body attached to the T265 worn on the head by participants. The Vantage 8 (Fig. 3.1**b**) is a purpose-built optical tracking camera produced by Vicon Motion Systems Ltd, UK, capable of recording at 260 Hz with an 8 megapixel resolution. By decreasing the resolution of the camera, the sampling rate can increase to a maximum of 2000 Hz. Each camera has an FOV of 61.7 ° horizontal by 47 ° vertical. These cameras created an optical tracking volume measuring $15 \times 8.5 \times 5$ meters, and yielded a 6 DOF pose estimate at 50 Hz.

The perambulator (Fig. 3.1**c**) is a modified surveyor's wheel (Nestle 12006001, Gottlieb Nestle GmbH, Germany) featuring a centimeter-precision distance counter. A housing for an IMU (actibelt RCT3, Trium Analysis Online GmbH, Germany) containing a tri-axial accelerometer (± 8 g range; 100 Hz sampling rate) and gyroscope ($\pm 2000 \ \frac{^\circ}{s}$ range; 100 Hz sampling rate) was attached to the axle of the wheel. The gyroscope recorded the instantaneous angular velocity of the wheel which directly corresponds to the speed of the device when being pushed across a surface.

The T265 was worn on the head via a custom-designed 3D-printed mount that holds the T265 securely through two M2.5 screws that thread into the back of the T265. The mount then fastens to an AmazonBasics camera head-strap system available on Amazon (ASIN B00R4YCKIK). In turn, this strap was worn on the head of the participant either directly or over a baseball cap (see Fig. 3.1**a**).

3.2.4. Software

Data from the T265 was recorded with custom software written in Python making use of the `pyrealsense2` library developed by Intel (`https://github.com/IntelRealSense/librealsense`, version 2.36.0). The software recorded the accelerometer and gyroscope streams as well the VI-SLAM position and velocity estimates to disk in a binary format. Information sampled by the OTS cameras was first sent to Vicon Blade software, where a rigid body was fit to the infrared optical marker data. The positional data of this rigid body was then published through the robot operating system (ROS) middleware via a custom wrapper. Data recorded on the IMU inside the perambulator was read out and processed by a custom software suite written in Julia developed by Trium Analysis Online GmbH.

3.2.5. Velocity and speed estimation

Linear and angular velocity ($\mathbf{v}$ and $\boldsymbol{\omega}$) were estimated from position ($\mathbf{p}$) and orientation ($\mathbf{q}$, in quaternions) provided by the OTS as $\mathbf{v} = \dot{\mathbf{p}}$ and $\boldsymbol{\omega} = \mathrm{Im}(2\mathbf{q}^{*}\dot{\mathbf{q}})$. Linear velocity recorded by the T265 as well as the estimate from the OTS (both measured in their respective world frames) were converted to speed by calculating the norm of the earth-horizontal components as $v = \|\mathbf{v}_{xy}\| = \sqrt{v_x^2 + v_y^2}$.

The angular velocity ω_P measured by the perambulator was filtered with a fourth-order Butterworth low-pass filter with a cutoff frequency of 10 Hz. Then it was transformed to linear speed by multiplying the angular velocity component in the direction of the axle with the circumference of the wheel ($C = 1\mathrm{m}$) as $v = C \cdot \omega_P$. Finally, samples with $v < 0.01$ m/s and segments shorter than 3 seconds were removed from the estimate.

3.2.6. Time synchronization

The timestamps of the data collected from the OTS were corrected by computing the cross-correlation function of the angular velocity $\boldsymbol{\omega}$ with that measured by the T265 ($\hat{\boldsymbol{\omega}}$) during the calibration segment (see *Evaluation in optical tracking space*). The temporal lag Δt of the maximum of this function was determined with $K = \mathrm{argmax}_k \sum_i^n \|\hat{\boldsymbol{\omega}}_i\| \cdot \|\boldsymbol{\omega}_{i+k}\|$ and $\Delta t = \hat{t}_K - t_K$ and the timestamps of the T265's measurements were shifted by this amount.

In the second study, the perambulator's IMU was tapped against the T265 at the beginning of each recording. This created visually distinguishable peaks in the accelerometer measurements of both devices. The timestamps of these peaks were used to manually correct the time offset.

For both studies, data recorded from the T265 was interpolated to match the timestamps of the respective gold standard (perambulator or OTS) after temporal alignment. A simple linear interpolation was used for position as well as linear and angular velocity. Orientation, expressed in quaternions, was interpolated using the spherical quadrangle method [113].

3.2.7. Reference frame transformations

The OTS provides position and orientation of the tracked rigid body with respect to its world frame W (denoted ${}^{W}\mathbf{p}$ and ${}^{W}\mathbf{q}$). The T265 provides its own position and orientation as well as linear and angular velocity with respect to a different world frame $\hat{W}$ (denoted ${}^{\hat{W}}\hat{\mathbf{p}}$, ${}^{\hat{W}}\hat{\mathbf{q}}$, ${}^{\hat{W}}\hat{\mathbf{v}}$ and ${}^{\hat{W}}\hat{\boldsymbol{\omega}}$).

The transformation between the world frames W and $\hat{W}$ was estimated using a basic point set registration (PSR) method [96] which was used to transform position and

orientation of the T265 from its own to the OTS world frame. A rotation-only PSR method was used to estimate the transformation from the body frames B and $\hat{B}$ to calibrated frames C and $\hat{C}$ that are independent of the orientation of the head mount on the subject's head. For the T265, we achieved this calibration by calculating the rotation that simultaneously centers heading direction (i.e. the direction of instantaneous linear velocity) along the longitudinal axis (x) and gravity direction along the vertical axis (z). The OTS was calibrated by aligning linear velocity and gravity direction to those measured by the T265 in its calibrated frame with the same rotation-only PSR method. Details on these estimations and transformations can be found in the supplementary material.

The above estimations (T265 world frame and calibrated frames) were performed for each subject and each task during the first 30 seconds of each task. The transformations obtained from these estimations were then applied to all measurements recorded during the task. This ensured that enough data was available for a robust estimation while at the same time reducing the possibility of drifts in position and orientation influencing the result. The complete reference frame tree with all transformations between frames is shown in fig. 3.2. Unless specified otherwise, positions and orientations reported below are represented in the world frame W (e.g. $\hat{\mathbf{p}}$ as a shorthand for ${}^{W}\hat{\mathbf{p}}$) while velocities and accelerations are represented in the respective calibrated body frames (e.g. $\hat{\mathbf{v}}$ as a shorthand for ${}^{\hat{C}}\hat{\mathbf{v}}$).

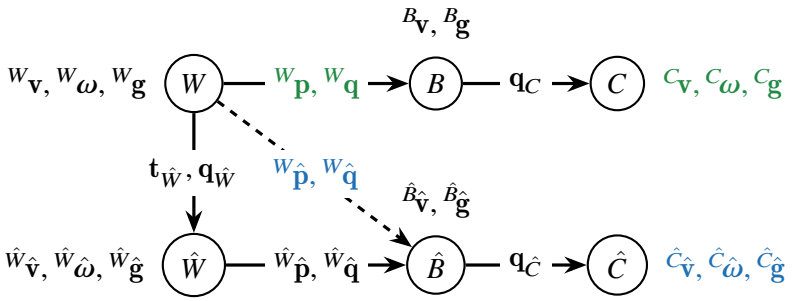

Figure 3.2.: Reference frame tree with corresponding measurements and transformations. The top row shows the relationship between world (W), body (B) and calibrated (C) frame of the OTS as well as the measurements used in the following analysis highlighted in green. The second row shows the relationship between world ($\hat{W}$), body ($\hat{B}$) and calibrated ($\hat{C}$) frame of the T265 as well as the measurements used in the following analysis highlighted in blue. The dashed arrow denotes the estimate of the T265's position and orientation with respect to the OTS world frame.

3.2.8. Performance metrics

The primary aim of this study was to compare the position and velocity estimates provided by the T265 with those provided by the OTS and perambulator. This comparison

was conducted using a number of metrics to quantify specific aspects of performance. An overview of these performance metrics is shown in table 3.1.

Metric	Unit	Description	Formula
Trajectory length error (TLE) *	%	Relative difference between trajectory length estimated by T265 and gold standard length	$\frac{\hat{L}-L}{L}$
Absolute translation error (ATE)	m	Root mean square (RMS) of distances between position estimated by T265 and OTS across the full trajectory [72]	$\sqrt{\frac{1}{n}\sum_i^n \lVert\hat{\mathbf{p}}_i - \mathbf{p}_i\rVert^2}$
Relative translation error (RTE)	m	RMS of relative distance between position estimated by T265 and OTS over a window of k samples [72]	$\sqrt{\frac{1}{n}\sum_i^n \lVert\hat{\mathbf{p}}_{i+k} - \hat{\mathbf{p}}_i - \mathrm{rot}(\hat{\mathbf{q}}_{\gamma,i} \cdot \mathbf{q}_{\gamma,i}^{-1}, \mathbf{p}_{i+k} - \mathbf{p}_i)\rVert^2}$
Translation drift (TDr)	%	Distance between final position estimates of a trajectory relative to trajectory length [72]	$\frac{\lVert\hat{\mathbf{p}}_n - \mathbf{p}_n\rVert}{L}$
Gravity direction error (GDE)	°	Mean angle between representation of gravity vectors in the respective body frames	$\frac{1}{n}\sum_i^n \arccos \frac{\langle\hat{\mathbf{g}}_i, \mathbf{g}_i\rangle}{\lVert\hat{\mathbf{g}}_i\rVert \cdot \lVert\mathbf{g}_i\rVert}$
GDE-α, GDE-β		Roll and pitch angle difference	$\hat{\alpha} - \alpha$, $\hat{\beta} - \beta$
Absolute yaw error (AYE)	°	RMS of yaw angle difference across full trajectory [72]	$\sqrt{\frac{1}{n}\sum_i^n \Delta\gamma_i^2}$
Relative yaw error (RYE)	°	RMS of yaw angle difference over a window of k samples [72]	$\sqrt{\frac{1}{n}\sum_i^n (\Delta\gamma_{i+k} - \Delta\gamma_i)^2}$
Yaw drift (YDr)	°/h	Final yaw angle difference relative to trajectory duration T [72]	$\frac{\Delta\gamma_n}{T}$
Speed error (SpE) *,†	%	Mean relative difference between earth-horizontal speed measured by the T265 and gold standard speed	$\frac{1}{n}\sum_i^n \frac{\hat{v}_i - v_i}{v_i}$
Heading direction error (HDE) †	°	Mean angle between linear velocity vectors estimated by T265 and OTS	$\frac{1}{n}\sum_i^n \arccos \frac{\langle\hat{\mathbf{v}}_i, \mathbf{v}_i\rangle}{\lVert\hat{\mathbf{v}}_i\rVert \cdot \lVert\mathbf{v}_i\rVert}$
HDE-θ, HDE-ϕ †		Azimuth and elevation angle diff.	$\hat{\theta}_v - \theta_v$, $\hat{\phi}_v - \phi_v$

Metric	Unit	Description	Formula
Linear velocity magnitude error (LVME) †	m/s	Mean difference between magnitudes of linear velocity estimated by T265 and OTS	$\frac{1}{n}\sum_i^n \|\hat{\mathbf{v}}_i\| - \|\mathbf{v}_i\|$
Angular velocity direction error (AVDE) †	°	Mean angle between angular velocity vectors estimated by T265 and OTS	$\frac{1}{n}\sum_i^n \arccos \frac{\langle\hat{\boldsymbol{\omega}}_i, \boldsymbol{\omega}_i\rangle}{\|\hat{\boldsymbol{\omega}}_i\| \cdot \|\boldsymbol{\omega}_i\|}$
AVDE-θ, AVDE-ϕ †		Azimuth and elevation angle diff.	$\hat{\theta}_\omega - \theta_\omega$, $\hat{\phi}_\omega - \phi_\omega$
Angular velocity magnitude error (AVME) †	°/s	Mean difference between magnitudes of angular velocity estimated by T265 and OTS	$\frac{1}{n}\sum_i^n \|\hat{\boldsymbol{\omega}}_i\| - \|\boldsymbol{\omega}_i\|$

Table 3.1.: Description of performance metrics. Metrics denoted with a star (∗) were computed for both studies, all other metrics only for the optical tracking space study. For metrics denoted with a dagger (†) we excluded samples where the gold-standard motion speed was below 0.1 m/s.

The trajectory length L was computed from position data provided by T265 and OTS as the sum of earth-horizontal displacements, i.e. the norm of the difference in x and y direction between two consecutive samples: $L = \sum_i^n \left\|\mathbf{p}_{xy,i} - \mathbf{p}_{xy,i-1}\right\|$. Since the position estimate by the T265 is occasionally subject to re-localization jumps that would result in an over-estimation of the trajectory length, samples where the instantaneous speed was above 5 m/s were considered artifacts and excluded from this computation. Trajectory length measured by the perambulator was directly provided by the distance counter of the device.

For the calculation of the relative translation error (RTE), yaw drift at the beginning of the window was removed by rotating the position estimate of the OTS with $\mathrm{rot}(\hat{\mathbf{q}}_{\gamma,i} \cdot \mathbf{q}_{\gamma,i}^{-1}, \mathbf{p}_{i+k} - \mathbf{p}_i)$. Here, $\mathrm{rot}(\mathbf{q}, \mathbf{v}) = \mathbf{q}\mathbf{v}\mathbf{q}^{-1}$ denotes the rotation of a vector $\mathbf{v}$ by the quaternion $\mathbf{q}$ and $\mathbf{q}_\gamma = [\sqrt{1-q_z^2}, 0, 0, q_z]^\top$ denotes the quaternion representing the yaw component of $\mathbf{q}$. Since the OTS data was recorded at a sampling rate of 50 Hz, we used a window length of $k = 50$ to obtain windows of approximately 1 second. The same window length was used for the calculation of the relative yaw error (RYE).

The roll (α) and pitch (β) angle with respect to gravity were computed as $\alpha = \arctan {}^{g_y}/_{g_z}$ and $\beta = -\arcsin {}^{g_x}/_{\|\mathbf{g}\|}$. Difference in yaw angle was computed as the geodesic distance between the yaw components of the orientations from T265 and OTS as $\Delta\gamma = \arccos\left(2 < \hat{\mathbf{q}}_\gamma, \mathbf{q}_\gamma >^2 -1\right)$ [54]. We excluded outliers in the first and 99th percentile of $\Delta\gamma$ from further analysis.

Heading is the instantaneous direction of linear velocity in head coordinates. Heading elevation (ϕ_v) and azimuth (θ_v) angle were computed as $\phi_v = \arcsin {}^{v_z}/_{\|\mathbf{v}\|}$ and $\theta_v = -\arctan {}^{v_y}/_{v_x}$. We also computed angular velocity elevation ($\phi_\omega = \arcsin {}^{\omega_z}/_{\|\boldsymbol{\omega}\|}$)

and azimuth angle ($\theta_\omega = -\arctan {}^{\omega_y}/_{\omega_x}$). This corresponds to the direction of the instantaneous axis of rotation.

3.2.9. Statistical analysis

Distributions of motion speeds v where estimated using a kernel density estimate (KDE) with $\hat{f}_h(v) = {}^1/_{nh} \sum_i^n K({}^{v-v_i}/_h)$[114]. We used a Gaussian kernel $K(x) = {}^1/_{\sqrt{2\pi}}\, e^{-x^2/2}$ and a bandwidth of $h = 0.2$.

We used a one-way repeated measures analysis of variance (ANOVA) to determine whether the means of error metrics were significantly different across tasks. We reported the F-statistic as well as p-values and considered effects significant if $p < 0.05$, in which case we performed a two-sided paired t-test between all pairs of tasks as a post-hoc test. For this test, we reported p-values corrected with the Bonferroni method and divided by 2 (to obtain a one-sided result dependent on the t-statistic, see below), considering differences significant when $p < 0.05$. Additionally, we deemed metrics to be larger in the first task of the pair if the t-statistic was positive and smaller if the t-statistic was negative. In the perambulator study, we performed a two-way repeated measures ANOVA in the same manner to determine the influence of task and environment on the error metrics, combined with a post-hoc paired t-test across tasks and environments in the case of significant effects. We did not compare metrics between the two studies since the difference in gold standard as well as environment and test subjects were confounding factors we could not control for.

All statistical analyses were performed in Python 3.6. We used the `statsmodels` library (version 0.12.1) for the ANOVA and the `scipy` library (version 1.5.3) for the KDE and t-tests. Error metrics were plotted for different tasks and environments using boxplots. Boxes were plotted from the first to the third quartile with the band indicating the median. Whiskers indicated the range from the lowest sample within 1.5 times the interquartile range (IQR) of the lower quartile to the highest sample within 1.5 times the IQR of the upper quartile.

3.3. Results

A comparison of trajectories recorded from T265 and OTS in the optical tracking space is shown in Fig. 3.3a-d. The figure demonstrates a case of successful tracking during walking (panels a and c) and a failure case during running (panels c and d). While there is an overestimation of displacement in the first case, the second case exhibits significant drift both in yaw angle (b) and vertical direction (d). Jumps in the T265 trajectory in Fig. 3.3b are due to re-localization of the device by means of loop closure,

i.e., a correction of the current pose estimate based on re-observation of previously observed landmarks.

Fig. 3.3e-j compares the earth-horizontal components of trajectories reported by the T265 in the perambulator study. The left column shows successful tracking cases in the hallway (panel e), lobby (panel g) and courtyard (panel h) environment while the right column (panels f, h and i) shows failure cases. The unsuccessful cases are characterized by yaw drift and, in the courtyard example (j), a task-dependent under-estimation of displacement.

Trajectory length errors (TLEs) are shown in Fig. 3.4a. Median TLEs in the OTS study are positive, indicating a task-dependent over-estimation of trajectory length ($F(2, 16) = 10.84, p = 0.001$) that is higher during slow walking ($p = 0.013$) and running ($p = 0.005$) compared to walking. TLEs in the perambulator study are dependent on environment ($F(2, 14) = 10.08, p = 0.002$) and task ($F(2, 14) = 7.75, p = 0.005$). Median values are negative and under-estimation is significantly smaller in the hallway environment when compared with the lobby ($p = 0.013$) and courtyard environments ($p = 0.015$). Additionally, we observe more under-estimation during running ($p = 0.012$) and walking ($p = 0.007$) compared to slow walking.

Positional data was also used to quantify absolute and relative translation error (ATEs, RTEs) and yaw error (AYEs, RYEs) as well as drift in translation (TDrs) and yaw (YDrs, Fig. 3.5). Median ATEs are around 0.4 m and values do not depend on task ($F(2, 16) = 1.77, p = 0.202$, Fig. 3.5a). RTEs are dependent on task ($F(2, 16) = 26.57, p < 0.001$) and significantly higher in the running task than in the walking ($p = 0.002$) and slow walking task ($p < 0.001$, Fig. 3.5b). Median TDrs are between 0.2 and 0.4 m/h and values are not task-dependent ($F(2, 16) = 2.35, p = 0.128$, Fig. 3.5c). Median AYEs are between 3 and 5° and are not dependent on task ($F(2, 16) = 3.09, p = 0.073$, Fig. 3.5d). Median RYEs increased from 0.9 to 1.4°from slow walking to running, although the overall effect is not statistically significant ($F(2, 16) = 0.46, p = 0.637$, Fig. 3.5e). YDrs depend on task ($F(2, 16) = 6.98, p = 0.007$) and are higher during running than slow walking ($p = 0.034$, medians between 1 and 3.5 °/h, Fig. 3.5f).

Orientation relative to gravity as tracked by both T265 and OTS shows an elongated distribution around the pitch axis (Fig. 3.6a and b). The gravity direction errors (GDEs) are relatively small and similar for both pitch and roll axes (Fig. 3.6c). Median values are between 1.6 and 3° and depend on task ($F(2, 16) = 8.49, p = 0.003$). Specifically, they are smaller in the walking task in comparison with the slow walking ($p = 0.026$) and running task ($p = 0.014$, Fig. 3.6d).

Linear and angular velocity measures were also compared. Example traces from both studies are shown in Fig. 3.7. 3-DOF linear and angular velocities of a participant measured by the T265 while walking in the optical tracking space are displayed in Fig. 3.7a and b. Panels c and d compare earth-horizontal movement speed calculated

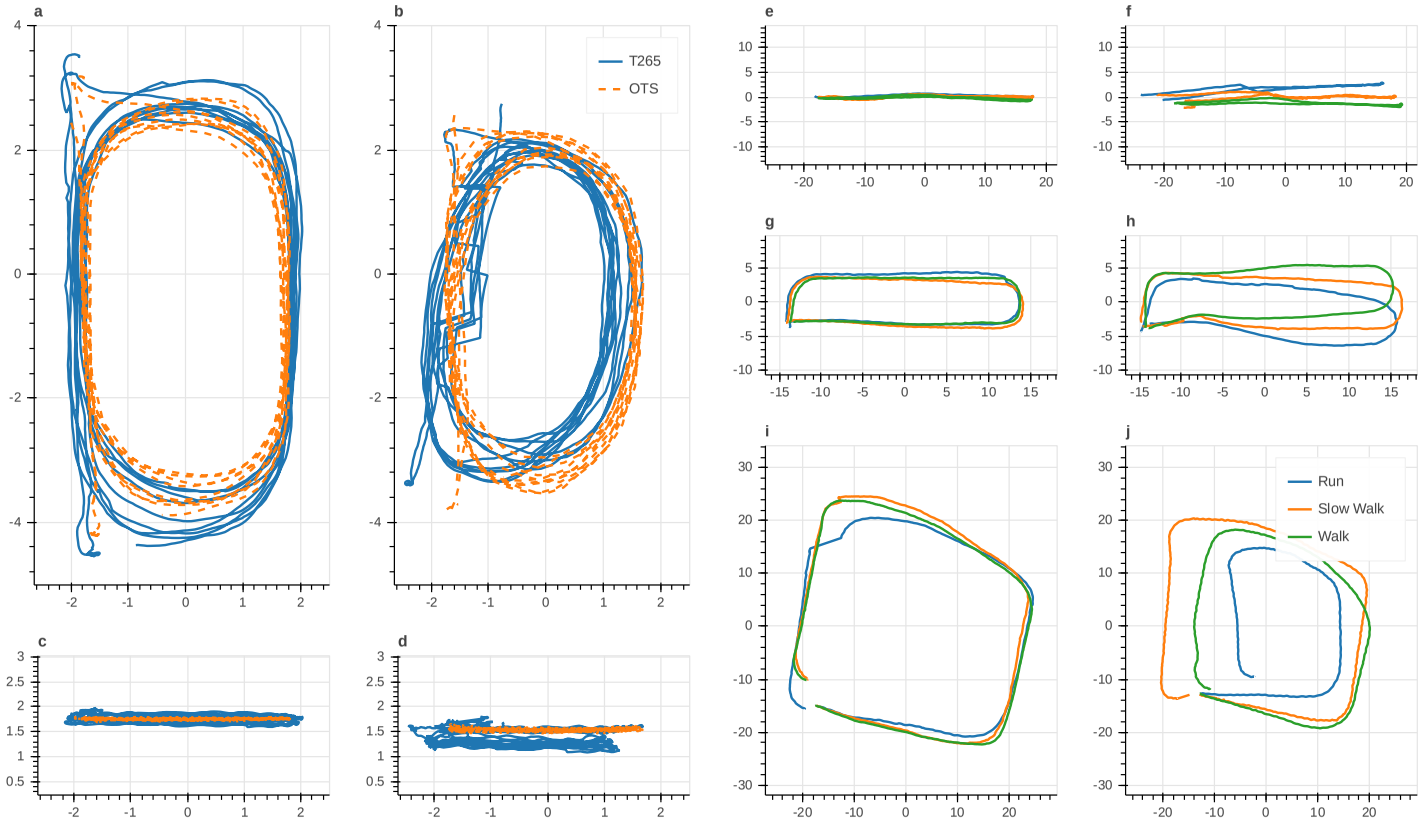

Figure 3.3.: Example trajectories recorded in all environments. All axes indicate position in m; axis labels were omitted to avoid cluttering. (**a-d**) Top and side view of trajectories recorded in the optical tracking space comparing position estimated by T265 (blue) and OTS (orange, dashed). (**a, c**) Successful tracking during walking. (**b, d**) Unsuccessful tracking exhibiting re-localizations and drift in both vertical direction and yaw during running. (**e-j**) Top view of example trajectories recorded in the real-world environments comparing position estimated by T265 across different tasks (blue: running, orange: slow walking, green: walking). (**e**) Successful tracking in hallway environment. (**f**) Unsuccessful tracking in hallway environment exhibiting yaw drift and re-localization. (**g**) Successful tracking in lobby environment. (**h**) Unsuccessful tracking in lobby environment exhibiting yaw drift. (**i**) Successful tracking in courtyard environment. (**j**) Unsuccessful tracking in courtyard environment exhibiting yaw drift and task-dependent under-estimation of displacement.

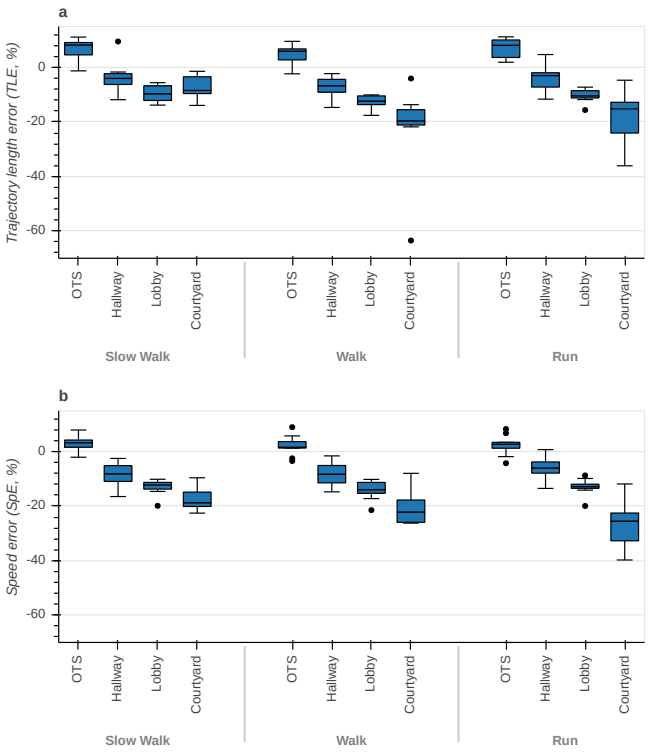

Figure 3.4.: Boxplots of trajectory length and speed errors between T265 and gold standard across different environments and tasks. (**a**) Trajectory length errors. (**b**) Speed errors.

from the T265 data with the gold standard perambulator measurements. Fig. 3.7c demonstrates a case of successful tracking in the hallway environment during walking, corresponding to the trajectory shown in Fig. 3.3e. In contrast, Fig. 3.7d illustrates a failure case characterized by a severe under-estimation of motion speed during running in the courtyard environment that corresponds to the trajectory in Fig. 3.3j.

Movement speeds are strongly dependent on task, both in the optical tracking space ($F(2, 16) = 219.98, p < 0.001$) and in the real-world environments ($F(2, 14) = 129.62, p < 0.001$, all post-hoc tests yielded $p < 0.001$, Fig. 3.8). Median speed during slow walking ranges from 0.7 m/s in the optical tracking space (Fig. 3.8a) to 1.2 m/s in the courtyard environment (Fig. 3.8d). Similarly, median values of walking and running speed range from 1.0 to 1.5 m/s and from 1.7 to 2.6 m/s, respectively. Running speed, especially in the hallway and lobby environments, shows a bi-modal distribution (Fig. 3.8b and c). However, there is no significant effect of environment on movement speed in the perambulator study ($F(2, 14) = 3.72, p = 0.051$).

Median speed errors (SpEs) are close to 0% in the optical tracking space and do not

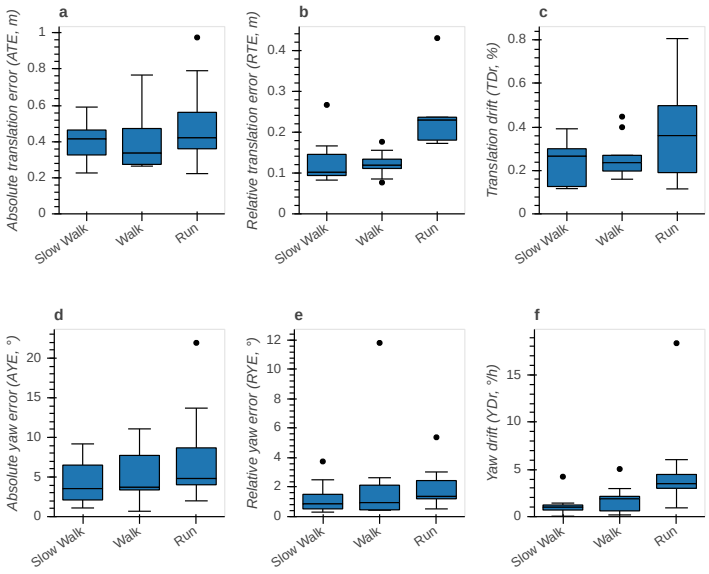

Figure 3.5.: Boxplots of translation and yaw errors between T265 and OTS across different tasks. (**a**) Absolute translation error. (**b**) Relative translation error. (**c**) Translation drift. (**d**) Absolute yaw error. (**e**) Relative yaw error. (**f**) Yaw drift.

depend on task ($F(2, 16) = 2.10, p = 0.155$, Fig. 3.4b). In the perambulator study, SpEs are negative and decrease with increasing size of the environment ($F(2, 14) = 27.22, p < 0.001$), with significantly lower values in the courtyard environment when compared to the hallway ($p = 0.001$) and lobby environments ($p = 0.009$). Additionally, we observe significantly lower values in the lobby in comparison with the hallway ($p = 0.024$). This indicates a tendency of the T265 to under-estimate speed in larger, more complex environments and is consistent with the results for TLEs. The error is also significantly dependent on task ($F(2, 14) = 4.52, p = 0.031$), where we observe more under-estimation during running compared to slow walking ($p = 0.028$).

Heading directions are centered around 0° elevation and azimuth angles (Fig. 3.9a and b). Error in heading direction (HDEs) shows a centered distribution, with similar extents in elevation and azimuth (Fig. 3.9c) and is dependent on task ($F(2, 16) = 8.32, p = 0.003$, Fig. 3.9e). Errors are lowest in the walking task with a median value of about 4°and significantly smaller compared to the slow walking ($p = 0.005$) and running task ($p = 0.012$). Errors in the magnitude of the linear velocity vector (LVMEs) are centered close to 0 °/s and do not depend on task ($F(2, 16) = 2.39, p = 0.123$, Fig. 3.9d).

Angular velocity directions (i.e., the axes of rotation) are distributed towards ±90° elevation and azimuth angles (Fig. 3.9f and g). This indicates that the instantaneous axis of head rotation is more frequently aligned with the pitch and yaw axis than with the

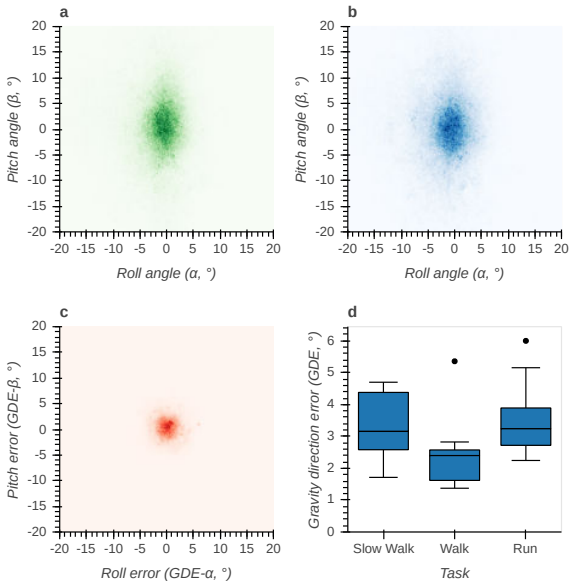

Figure 3.6.: Gravity direction measured by T265 and OTS and corresponding errors. Positive pitch angles correspond to forward pitch, positive roll angles to rightward roll. (**a**) Bi-variate histogram of pitch and roll angles measured by the OTS. (**b**) Bi-variate histogram of pitch and roll angles measured by the T265. (**c**) Bi-variate histogram of pitch and roll errors between T265 and OTS. (**d**) Boxplot of gravity direction errors between T265 and OTS across different tasks.

roll axis (which corresponds to zero azimuth and elevation). Error in angular velocity direction (AVDEs) shows a centered distribution that is elongated in the azimuth direction (Fig. 3.9h). Overall, errors are dependent on task ($F(2, 16) = 5.55, p = 0.015$), but the post-hoc analysis revealed no significant differences (Fig. 3.9j). Errors in the magnitude of angular velocity (AVMEs) are centered around -3 °/s and are not task-dependent ($F(2, 16) = 1.41, p = 0.273$, Fig. 3.9i).

3.4. Discussion

Measurement of natural human head motion in natural environments is important for a range of applications including VR/AR technology, clinical diagnostics, as well as basic scientific investigation of sensorimotor function. If VI-SLAM devices such as the T265 are going to be used for these applications, their accuracy must be evaluated. That is the primary aim of this study. Which measures of human head position and motion are most important varies greatly across applications. Therefore, we have evaluated accuracy using a wide range of metrics. Performance was evaluated relative to two gold-standard methods, the OTS because it can estimate all 6-DOF of head position and the perambulator because it can be used in any environment, including

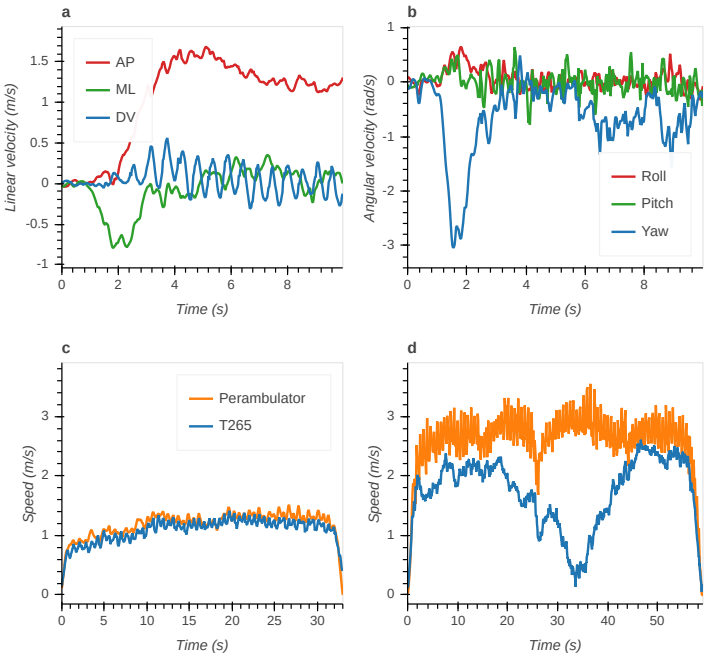

Figure 3.7.: Example time series of velocities measured by T265. (**a**) Linear velocity in anterior/posterior (AP), medial/lateral (ML) and dorsoventral (DV) directions in optical tracking space during walking. (**b**) Angular velocity around roll, pitch and yaw axes in optical tracking space during walking. (**c**) Comparison between speed measured by T265 and perambulator during walking in hallway environment showing successful tracking, see Fig. 3.3e. (**d**) Comparison between speed measured by T265 and perambulator during running in courtyard environment showing unsuccessful tracking with considerable under-estimation of speed, see Fig. 3.3j.

outdoors. We measured performance for a range of locomotor speeds because speed impacts both IMU data and visual data and may also impact how the VI-SLAM algorithm estimates linear and angular position. We also measured performance for a range of environments because environmental features are known to impact the reliability of visual data used for VI-SLAM.

Regarding the effect of environment, both trajectory length and movement speed are underestimated by the T265 relative to the perambulator and this error increases with the size of the environment (Fig. 3.4). This effect is likely to reflect underestimation of the physical scale or size of the visual scene and thus underestimation of the distance and speed of human movement. Large environment size can pose difficulties for many VI-SLAM algorithms [42, 108]. As the environment increases in size, landmarks used by the T265's VI-SLAM algorithm may increasingly get further from the cameras. In turn, the resultant landmark movement used to estimate camera motion may be underestimated due to the decreased stereo disparity of these landmarks sensed by the T265's cameras. Future studies using the T265 to measure ground speed and

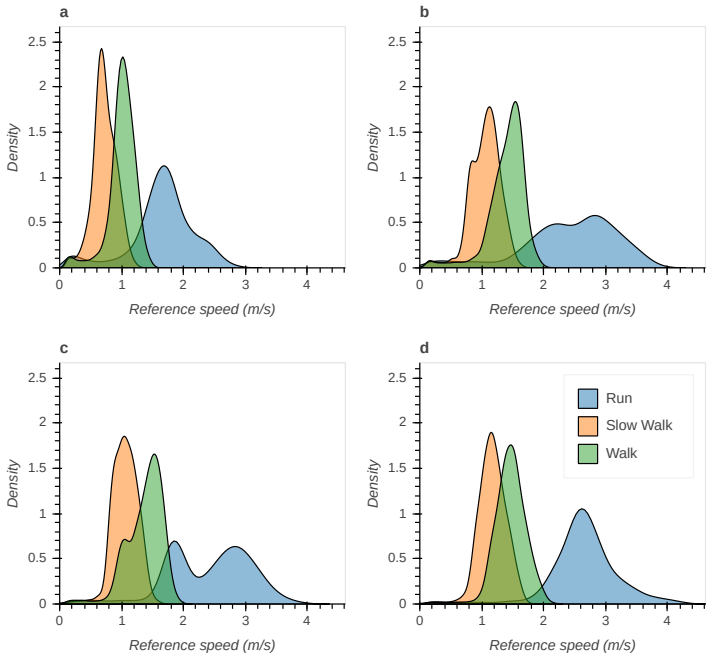

Figure 3.8.: Kernel density estimates of gold standard speed across different environments and tasks (blue: running, orange: slow walking, green: walking). (**a**) Optical tracking space. (**b**) Hallway environment. (**c**) Lobby environment. (**d**) Courtyard environment.

distance of linear head motion in diverse environments should be aware of this potential source of inaccuracy; other devices, such as the perambulator may be preferable in this context.

Some metrics are also affected by locomotor speed. Several of these show a monotonic increase in error with increasing locomotor speed. With the perambulator as gold-standard, underestimation of trajectory length is greater during running and walking than during slow-walking (Fig. 3.4). With the OTS as gold-standard, relative translation error is greater during running compared to walking and slow-walking; Yaw drift error is also higher during running than slow-walking (Fig. 3.5). These monotonic effects of speed are likely due to noise on VI-SLAM signals that increases with locomotor speed. For example, increased speed may lead to motion blur in the visual data which may hinder landmark localization. Regarding IMU data, noise may also increase with increased power at higher frequencies. Another factor is the relatively low camera frame rate of 30 Hz. With higher movement speed, tracked landmarks can move considerably within the camera image between consecutive frames, which in turn might degrade the tracking performance. Finally, the VI-SLAM algorithm itself may be optimized to operate best during slower, smoother motions.

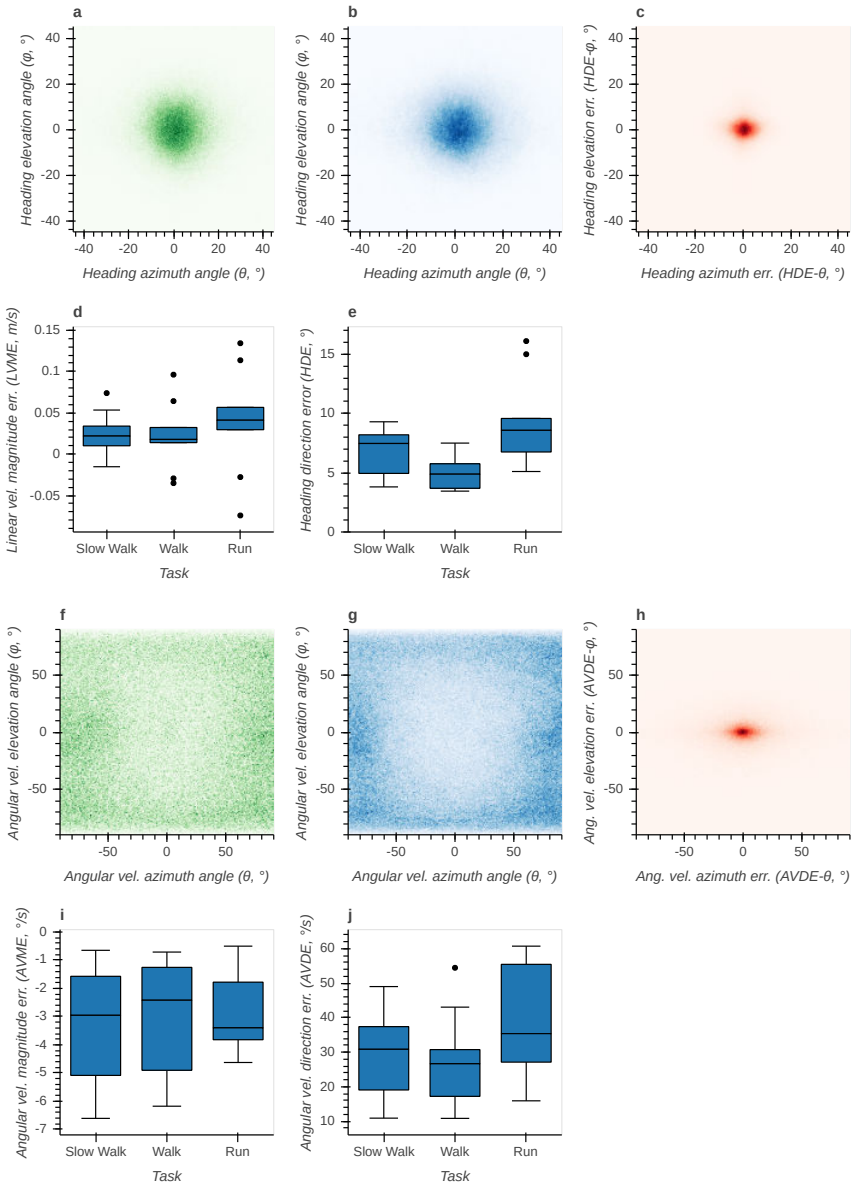

Figure 3.9.: Linear and angular velocity measured by T265 and OTS and corresponding errors. (**a**) Bi-variate histogram of heading azimuth and elevation angles measured by the OTS. Positive elevation angles correspond to upward heading, positive azimuth angles to rightward heading. (**b**) Bi-variate histogram of heading azimuth and elevation angles measured by the T265. (**c**) Bi-variate histogram of heading azimuth and elevation errors. (**d**) Boxplot of linear velocity magnitude errors across different tasks. (**e**) Boxplot of heading direction errors across different tasks. (**f**) Bi-variate histogram of angular velocity azimuth and elevation angles measured by the OTS. Positive elevation corresponds to a leftward rotation around the yaw axis, positive azimuth to an upward rotation around the pitch axis. (**g**) Bi-variate histogram of azimuth and elevation angles measured by the T265. (**h**) Bi-variate histogram of azimuth and elevation errors. (**i**) Boxplot of angular velocity magnitude errors across different tasks. (**j**) Boxplot of angular velocity direction errors across different tasks.

In contrast, other metrics show a non-monotonic effect of locomotor speed. Specifically, gravity direction error (Fig. 3.6) and heading direction error (Fig. 3.9) are both smallest during normal walking and greater for slow walking and running. The similar pattern of results for these two metrics may be because they both depend on how sum total linear acceleration, sensed by the accelerometer, is partitioned into gravitational and inertial components. The fact that performance is best for normal walking may be because the T265 algorithm has been specifically tuned to perform best during normal walking to facilitate its use in VR/AR applications. Unfortunately, it is not possible to verify these speculations because the T265 VI-SLAM algorithm is proprietary. Nevertheless, this pattern of results suggests that the T265 is a suitable choice for applications that require estimation of heading and gravity direction during normal walking.

These results are in line with those reported by Alapetite et al. [3] that suggest that motion speed and density of visual features in the environment have the greatest effect on the T265's performance while the presence of moving objects has less impact. We did not explicitly investigate the effect of feature richness as our focus was on real-world environments of varying size, although it could be argued that visual features are sparser in larger environments. In a future study, it would be interesting to evaluate the tracking performance in a feature-poor real-world setting such as a meadow. Our results regarding accuracies in position and heading angle also seem to confirm previous reports by Agarwal et al. [1], Bayer and Faigl [16], and Ouerghi et al. [96]. In the former, the authors note that the tracking performance of the T265 increases throughout multiple runs across the same environment which suggests that the device stores the features of a number of recently observed visual landmarks.

One limitation of the current study is the inability to calculate most performance metrics in more realistic, outdoor environments. To accomplish this, it would be necessary to use a gold-standard method that can measure linear and angular position in a large, naturalistic outdoor space. Unfortunately we did not have access to a large-scale, outdoor OTS or other method that would be suitable for this purpose. As a consequence, we were not able to evaluate how many of our metrics are affected by tracking in larger, outdoor environments. For example, we might expect yaw drift to be greater in larger environments due to the inability of the VI-SLAM algorithm to achieve loop closure (see, e.g., Fig. 3.3b and d). This was most likely not a problem in the small optical tracking space. Yaw drift and possibly other metrics measured in this space may not reflect performance in larger spaces.

Another limitation is the inability to evaluate what amount of measured error should be attributed to inaccuracy of the gold-standard. In particular, the T265 uses a gyroscope to measure angular velocity and this inertial measure of angular velocity may be more reliable than the angular velocity estimate provided by the OTS. Also, the measures derived from the perambulator are known to be less accurate when the walking trajectory is curved. We tried to mitigate this by instructing subjects to make curves as large

as possible and, in the hallway environment, lifting up the perambulator and turning around in place at the end of the hallway. It is also possible that the additional cognitive load of pushing the perambulator while walking or jogging caused participants to bias or change their own locomotion in some unforeseen way. Despite these possible limitations, the gold-standard methods used here are generally much more accurate than the T265 meaning that our metrics generally reflect performance of the T265 more than performance of the gold-standard method.

Moving forward, it would be important to further investigate in more detail which environmental features are most likely to cause tracking failures by the T265 or by VI-SLAM systems generally. Relevant environmental features include variation in light level of the environment, presence of independently moving objects, and environmental size or scale. These can all impact the ability of the VI-SLAM algorithm to identify, track, and use visual landmark features. Conventional SLAM systems often mitigate these issues through incorporation of additional sensors such as global positioning system (GPS), light detection and ranging (LIDAR), and wheel odometry [108]. While many early SLAM algorithms are highly redundant, using many landmarks, modern VI-SLAM methods limit the amount of landmarks used in order to increase computational efficiency [13], requiring greater emphasis on landmark selection and removal of erroneous or unusable landmarks. Recent efforts to combine machine learning techniques with VI-SLAM largely serve these needs, and decrease error driven by dynamic features [12]. The T265 does not include any sensors other than those previously described, and at time of writing it is unknown how the T265 performs feature selection.

An alternative to VI-SLAM that avoids environmental dependence is positional tracking based on purely non-visual data. The simplest form of non-visual tracking is based on inertial data only. Such tracking usually yields more noisy positional estimates, although certain biomechanical constraints such as the assumption of zero foot velocity during the stance phase can be used to correct drifts and integration errors [41]. Ongoing research is aimed at improving inertial tracking for the specific application of tracking human head position [72]. It would be possible to compare the tracking performance of the T265's VI-SLAM algorithm to an estimate computed only from its raw IMU data. However, this is subject to algorithm selection and parameter tuning which is why we deemed this comparison to be out of scope for the current study.

In principle, it would also be possible to improve the estimation yielded by the T265 by performing VI-SLAM post-hoc. While one of the larger selling points of the T265 is real-time VI-SLAM, it is not necessary for scientific investigation of human head motion. Measurements taken from the sensors of the T265 could be saved and subsequently passed through customized VI-SLAM algorithms that are optimized for the specific application of tracking natural human head motion in natural environments.

In summary, the T265 appears to be best suited for tracking human head position during normal walking in small- to medium-sized environments with limited dynamic features.

Increases (and sometimes decreases) in locomotor speed tend to increase observed error, as does use of the system in larger and more dynamic outdoor environments. The acceptability of the default, factory-set performance of the T265 depends on the application. In future, customized VI-SLAM algorithms may be applied to data collected by the T265 or other devices post-hoc such that estimation is optimized for tracking natural human head movement in natural environments. Additional evaluation studies in larger, dynamic, and outdoor environments would be very helpful for tuning of such custom VI-SLAM algorithms.

4. Ecological Momentary Assessment of Head Motion

*Originally published on June 4th, 2019 in Frontiers in Human Neuroscience as **"Ecological Momentary Assessment of Head Motion: Toward Normative Data of Head Stabilization"**.*

Original authors: Peter Hausamann, Martin Daumer, Paul R. MacNeilage and Stefan Glasauer

DOI: https://doi.org/10.3389/fnhum.2019.00179

Abstract

Head stabilization is fundamental for balance during locomotion but can be impaired in elderly or diseased populations. Previous studies have identified several parameters of head stability with possible diagnostic value in a laboratory setting. Recently, the ecological validity of measures obtained in such controlled contexts has been called into question. The aim of this study was to investigate the ecological validity of previously described parameters of head stabilization in a real-world setting. Ten healthy subjects participated in the study. Head and trunk movements of each subject were recorded with inertial measurement units (IMUs) for a period of at least ten hours. Periods of locomotion were extracted from the measurements and predominant frequencies, root mean squares (RMSs) and bout lengths were estimated. As parameters of head stabilization, attenuation coefficients (ACs), harmonic ratios (HRs), coherences and phase differences were computed. Predominant frequencies were distributed tightly around 2 Hz and ACs, HRs and coherences exhibited the highest values in this frequency range. All head stability parameters exhibited characteristics consistent with previous reports, although higher variances were observed. These results suggest that head stabilization is tuned to the 2 Hz fundamental frequency of locomotion and that previously described measures of head stability could generalize to a real-world setting. This is the first study to address the ecological validity of these measures, highlighting the potential

use of head stability parameters as diagnostic tools or outcome measures for clinical trials. The low cost and ease of use of the IMU technology used in this study could additionally be of benefit for a clinical application.

Summary and author contributions

This chapter investigates the validity of previously described parameters of head stability in a real-world setting, addressing RQ 3 and RQ 4. To this end, a study is conducted where head and trunk movement data of healthy human subjects is collected with IMUs for at least ten consecutive hours. From this data, head stabilization measures are computed and compared with previous reports where the same measures were obtained in a laboratory setting. The examined measures include:

1. attenuation coefficients, that quantify the attenuation of accelerations from trunk to head,
2. harmonic ratios, measuring the regularity of movement,
3. coherences, describing the similarity between head and trunk motion, and
4. phase differences, that can be used to quantify the temporal alignment of head and trunk movements.

These parameters are calculated for periods that are identified as locomotion. Additionally, the predominant frequency of trunk motion is computed as a proxy measure for walking speed.

The results of the study show that previous reports of head stability parameters can generally be replicated in a real-world context. This is important because several studies have shown that measures of head stabilization could be suitable tools for diagnosing diseases or serve as outcome measures for clinical trials.

The study design was developed between all four authors and my primary contribution is the identification of previously published parameters of head stability that could be evaluated with the study. Additionally, I carried out all motion recordings as well as data pre-processing and analysis. I also performed the validation experiments for the estimation of the direction of gravity from IMU data and the step detection described in appendix C. All figures were generated by me.

The manuscript was chiefly written by myself. Paul MacNeilage, Martin Daumer and Stefan Glasauer provided guidance and revised the manuscript at various points during the process.

4.1. Introduction

During locomotion, reflexive head movements operate to minimize horizontal head translation [35, 86] and simultaneously compensate for vertical translation by pitching the head [51, 97]. These stabilization behaviors are thought to be crucial for effective control of both balance and locomotion because they reduce undesired variability of vestibular and visual sensory inputs [97]. In elderly individuals, head stabilization is compromised during both steady-state walking [34] and gait initiation [69, 81]. Impaired head stabilization has also been associated with disorders such as Parkinson's disease (PD) [21, 67], multiple sclerosis (MS) [100] and bilateral vestibular defects [98].

Motion capture and accelerometry are widely used in the analysis of head stabilization during human locomotion [51, 61, 97]. However, studies using motion capture systems are usually constrained to a laboratory setting by design. Similarly, previous studies using wearable sensors have been limited by the need to instruct and supervise subjects and faithfully annotate periods of locomotion. Several recent studies have questioned the ecological validity of measurements obtained in such controlled contexts, i.e. how well these measurements generalize to real world conditions [20, 120]. An alternative approach, known as ecological momentary assessment (EMA) [112], advocates the sampling of clinically relevant parameters in a subject's natural environment rather than a clinical setting.

In support of EMA, researchers have observed that clinical measures such as 10-meter walk test times do not significantly correlate with more objective outcomes such as fall risk, raising doubts concerning the clinical relevance of these measures [20]. The frequently used 6-minute walking test has been challenged by the fact that in many diseased or elderly populations, 6 minutes of uninterrupted walking rarely occur during daily life [120]. While there are some clinical tests whose results correlate with objective outcomes (such as clinical assessment of gait speed, [4]), these examples highlight the need to validate standardized measures in a real-world context.

Wearable accelerometry devices have been suggested for sampling human motion during daily life [91] and can be used as a way to assess head stabilization performance in the spirit of EMA. Compared with clinical tests, they provide a cost-effective and straightforward method of recording ecologically valid measures. Previous studies of vestibular stimulation have used these kinds of sensors to address head and whole body motion in more realistic contexts, but were either constrained to pre-defined activities [23, 24] or lacked measurements of angular velocity [76].

In order to assess whether they are indicative of real-life locomotor function, previously established measures of head stability [17, 51, 86] need to be evaluated with respect to their ecological validity. Results obtained from a sample of healthy individuals could then be used as a normative baseline for future studies involving populations with balance, gait or neurological disorders.

Therefore, the aims of this study were: (i) to record a dataset of real-world human motion of trunk and head with wearable sensors, (ii) to compute previously described parameters of head stabilization from this data, and (iii) to compare the computed parameters with previous results obtained in controlled environments.

4.2. Methods

4.2.1. Subjects

A convenience sample of ten healthy human subjects (five male, five female, age 21-28, most of them students participating in lecture “Clinical Applications of Computational Medicine” at the Technical University of Munich) with no history of balance or gait disorders participated in the experiment. All subjects signed an informed consent form compliant with the European General Data Protection Regulation and gave explicit consent to the publication of the recorded data. The study protocol was approved by the institutional review board of the Sylvia Lawry Center for Multiple Sclerosis Research.

4.2.2. Sensor devices

We used a small, self-contained inertial measurement unit (IMU) to record both linear acceleration and angular velocity of the human head and trunk. The device (Actigraph GT9X Link) was chosen for its ability to continuously record accelerometer and gyroscope data at a sampling rate of 100 Hz for 24 hours. To record head motion, the sensor unit was firmly attached to the inside of a baseball cap that was worn by the subjects. To record trunk motion, an IMU of the same model was attached to a specialized neoprene belt (actibelt flex-belt, Trium GmbH, Munich, Germany) worn at the waist under the clothing. The actibelt system itself is frequently used in clinical accelerometry studies, but was not used in this study because in its current version it is not equipped with a gyroscope.

4.2.3. Data acquisition

Subjects were outfitted with the recording equipment in the morning of a typical work/university day and instructed to wear the equipment for at least ten hours. They were instructed to take note of periods during which they took off either sensor unit and these periods were subsequently excluded from analysis. The recording equipment was returned the next morning.

The IMUs were synchronized by knocking both devices against each other at the beginning and the end of each recording. This created clearly visible peaks in the accelerometer measurement that were used to correct timing offsets and drifts between

the devices. All subjects performed a calibration routine for both sensor units in order to align the sensor coordinates with head- and trunk-fixed reference frames. For the head device, they first held their heads in a slightly forward-pitched position that aligned Reid's plane [78] with an earth-horizontal plane. Afterwards, they nodded their heads five times around the pitch axis. This yields a unique transformation that rotates the acceleration due to gravity to be purely vertical and rotates the angular velocity to be purely around the medial/lateral axis for this calibration routine (resulting in a head-fixed reference frame as shown in fig. 4.1A). A similar routine was performed for the trunk device which was calibrated such that the acceleration due to gravity was purely vertical when the subjects stood up straight.

4.2.4. Coordinate frame transformations

The IMUs used for this study record linear acceleration and angular velocity, but provide no direct information about the orientation of the device in world coordinates. The calibration approach outlined in the previous section yields a head/trunk-fixed coordinate system (fig. 4.1A). However, for comparability with previously reported results obtained in laboratory settings [51, 86, 89] it is necessary to transform the measurements into a frame of reference whose vertical axis remains parallel to the direction of gravity. Reference frames in these studies are defined as right-handed coordinate systems with the vertical axis pointing upwards in the direction of gravity, the anterior/posterior axis pointing in the direction of the subjects' motion and the medial/lateral axis pointing to the left of the motion direction.

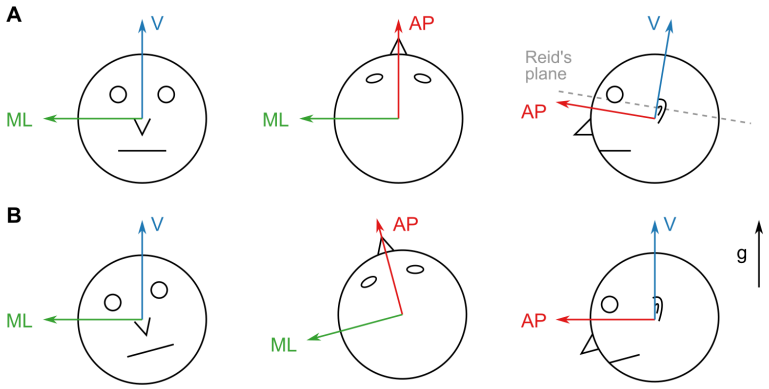

Figure 4.1.: (A) Axes of the head sensor coordinate system. The vertical axis is adjusted to be perpendicular to Reid's plane. This coordinate system remains head-fixed during all translations and rotations. **(B)** Axes of the aligned coordinate system. The vertical axis points in the direction of gravity (g). The system remains world-fixed during roll and pitch rotations (left and right figure) but remains head-fixed during yaw rotations (middle figure) and translations.

4.2.5. Estimation of the direction of gravity from IMU data

Gravitational acceleration g is linked to linear acceleration a and angular velocity through the following equations [45]:

$$g = a - i \tag{4.1}$$

$$\frac{\partial g}{\partial t} = \omega \times g \tag{4.2}$$

where i denotes the inertial acceleration of the device. Various filters are described in the literature that combine the linear acceleration and angular velocity measurements to produce an estimate of orientation. We propose a basic sensor fusion approach (table 4.1) that we show to be sufficiently accurate for typical trajectories occurring during human locomotion.

Inputs	Linear acceleration in sensor coordinates $a(t)$
	Angular velocity in sensor coordinates $\omega(t)$
	Correction factor α
	Acceleration due to gravity in world coordinates g_W
Outputs	Estimate of acceleration due to gravity in sensor coordinates $g(t)$
	Quaternion representing sensor orientation in aligned coordinates $q(t)$
Description	Start with initial estimate $g(0) = g_W$, for all $t \in 1..T$:
	1. Compute estimate of angular displacement: $\phi(t) = \omega(t)\Delta t$
	2. Compute estimate of gravity: $g(t) = \phi(t) \times g(t-1)$
	3. Update estimate with linear acceleration: $g(t) = g(t) + (1-\alpha)a(t)$
	4. Normalize estimate $g(t) = g(t)/\lvert g(t)\rvert$
	5. Compute $q(t)$ as the quaternion transforming $g(t)$ into g_W:
	a) Normalized axis of rotation: $n = g_W \times g(t)$, $n = n/\lvert n\rvert$
	b) Angle of rotation: $= \cos^{-1}(< g_W, g(t) >)$
	c) Quaternion from axis-angle representation: $q(t) = R(n, \theta)$

Table 4.1.: Description of the gravity filter algorithm for estimating gravity direction from IMU data. The correction factor $\alpha \in [0, 1]$ determines the weight of ω in the final estimate; an α of 0 means that ω is not used at all, while an α of 1 means that the linear acceleration is ignored. T is the number of samples and Δt is the time difference between two samples, corresponding to 10 ms at a sampling rate of 100 Hz. $< g_W, g(t) >$ denotes the inner product between g_W and $g(t)$ and $R(n, \theta)$ computes the quaternion from the axis-angle representation of the rotation:
$R(n, \theta) = \big(\cos(\theta/2), n_x \sin(\theta/2), n_y \sin(\theta/2), n_z \sin(\theta/2)\big)$

The angular velocity was high-passed at 0.1 Hz with a 5th-order Butterworth filter to remove errors due to gyroscope drift. The linear acceleration was low-passed with the same type of filter to reduce the influence of transient accelerations on the estimate. The estimates of orientation and acceleration due to gravity from the filter could then be used to transform the raw acceleration measured by the sensor into net inertial acceleration in aligned coordinates:

$$i_A = rot(q^{-1}, a - g) \tag{4.3}$$

where $rot(q, v)$ denotes the rotation of the vector v by the quaternion q. It should be noted that step 5 of table 4.1 ensures that the transformation has no yaw rotation component since $q(t)$ is computed from a rotation around the axis n which is always perpendicular to g_W. For consistency with previously reported results [51] where translations were described in a world-fixed, but rotations were described in a head/trunk-fixed frame, we did not transform the angular velocity into the aligned coordinate system.

We recorded a short dataset of one subject wearing one IMU attached to a baseball cap on the head. The sensor was mounted facing upwards on a plastic plate equipped with four optical markers for a motion capture system (8 Qualisys Oqus 100 cameras and Qualisys Track Manager software, version 2.9, Qualisys AB, Göteborg, Sweden). The subject performed different locomotor activities (walking, running) as well as spontaneous head movements while sitting for about 8 minutes. Afterwards, the sensor apparatus was removed from the baseball cap and rapidly swung around, creating high accelerations and rapid orientation changes of the device for about 1 minute. The motion capture data was used as a gold standard for evaluating the accuracy of the orientation estimate as well as finding the optimal parametrization of the algorithm.

We investigated the influence of the low-pass cut-off frequency of the linear acceleration (f_{LP}) as well as the correction factor α on the estimate quality and compared our approach with a previously described complementary filter method [129]. The accuracy was measured with the geodesic distance from the estimated quaternion q to the gold standard quaternion q_{GS} (corresponding to the angle of the shortest arc between the two orientations, [54]):

$$d = \cos^{-1}\left(2 < q, q_{GS} >^2 -1\right) \tag{4.4}$$

Both filter algorithms were implemented in Python 3.6 using the just-in-time compilation tools of the `numba` library (version 0.42) to greatly enhance execution speed. Run times were compared on an Intel Core i7-7700K CPU in single-threaded execution at a clock rate of 4.2 GHz. Based on the results of this analysis (see supplementary material), accelerometer and gyroscope data were transformed to the respective reference frames before further processing.

4.2.6. Step detection

In order to isolate periods of locomotion for analysis, we used a step detection method based on the inertial acceleration of the trunk sensor in aligned coordinates. We recorded a dataset of one subject wearing the trunk sensor, performing different locomotor activities at different speeds, including walking, running, stair walking and cycling. This data was used to parametrize a peak detector for extracting possible steps as well as to determine discriminative features that distinguish cycling from other types of motion.

Peaks were detected in the vertical axis component with a minimum height of 0.2 g, prominence of 0.4 g and distance of 20 ms (corresponding to a maximum detectable step frequency of 5 Hz, [110]). For each peak, we computed the short-time power spectrum $S(f)$ of the linear acceleration in all three spatial axes with a segment length of 1024 samples centered around the peak, weighted with a Blackman window function. The power spectrum was used to determine predominant frequency in each axis, i.e. the frequency with the highest spectral power. We investigated the distribution of root mean square (RMS) vertical accelerations as well as the difference between predominant frequencies in the vertical (V) and medial/lateral (ML) direction and used the results as criteria for the exclusion of cycling periods (see supplementary material).

The step detection method was applied to the trunk sensor data for each of the 10 subjects. Since we limited our analysis to frequencies above 1 Hz (see results), detected steps were grouped together as bouts if the time difference between two consecutive steps was smaller than one second. Bouts of single steps, i.e. where no other steps where detected within one second before and afterwards, were subsequently excluded from further analysis.

4.2.7. Predominant frequency as a proxy for walking speed

We determined the predominant frequencies of head and trunk accelerations for each step in all three spatial axes using the same short-time power spectrum approach as described above, albeit with a segment length of 512 samples. We used a shorter segment length than in the step detection procedure as it increased the temporal resolution at the expense of frequency resolution, yielding more accurate results for short bouts. We also calculated the magnitude of accelerations using the RMS for each step segment in all three directions. Furthermore, means and standard deviations of trunk predominant frequency in the V direction were calculated for each bout.

In [51], the authors showed a strong link between walking velocity and predominant frequency of vertical head translation. While we did not validate the exact correspondence for our data, predominant frequency of vertical head acceleration was used as a proxy measure for gait speed, allowing qualitative comparisons between previously

published results and ours. In the following, we use the term “predominant frequency” as a shorthand for predominant frequency of vertical head acceleration.

4.2.8. Assessment of head stabilization during locomotion

Attenuation coefficient

The reduction of linear accelerations through the upper body was quantified for each step segment using the attenuation coefficient (AC) between trunk and head. Segments consisted of 512 samples centered around the peak and were weighted using a Blackman window function in order to decrease the influence of non-locomotor accelerations for short bouts. ACs were calculated in the anterior/posterior (AP), ML and V directions using the RMS values of head (A_H) and trunk acceleration (A_T) [86] as:

$$AC = 1 - \frac{A_H}{A_T} \tag{4.5}$$

Positive values indicate an attenuation of head accelerations with respect to trunk accelerations whereas negative values correspond to increased accelerations at the head when compared to the trunk.

Harmonic ratio

Regularity and smoothness of motion was quantified using the harmonic ratio (HR) for both head and trunk accelerations. In the AP and V directions, the HR was calculated as the total spectral power of the even harmonics divided by the total spectral power of the odd harmonics of the predominant frequency:

$$HR = \frac{\sum_k^N S\left(2k f_{dom}\right)}{\sum_k^N S\left(2(k+1) f_{dom}\right)} \tag{4.6}$$

where f_{dom} denotes the predominant frequency of the segment in the respective direction and $N = 10$ is the number of harmonics we considered. Because of the biphasic nature of accelerations within strides (two steps), high values indicate that acceleration patterns remain in phase across stride cycles and are associated with stable gait [89]. In the ML direction, the HR was calculated inversely due to the fact that lateral motion is monophasic within one stride (left and right step, [73]):

$$HR = \frac{\sum_k^N S\left(2(k+1) f_{dom}\right)}{\sum_k^N S\left(2k f_{dom}\right)} \tag{4.7}$$

Coherence

We quantified head-trunk coordination and compensatory head motion during locomotion using the coherence [51]:

$$K_{xy}^2(f) = \frac{S_{xy}(f)^2}{S_{xx}(f)S_{yy}(f)} \tag{4.8}$$

where $S_{xy}(f)$ denotes the cross-power spectrum of signals x and y, $S_{xx}(f)$ is the power spectrum of signal x, and $S_{yy}(f)$ is the power spectrum of signal y. Coherence values were computed between head pitch velocity and vertical head acceleration and between head pitch velocity and trunk pitch velocity.

As the coherence for the power spectrum of a single segment is ill-defined, we used an extended segment length of 1024 samples centered around every step. Each segment was divided into 5 sub-segments of 512 samples with an overlap of 128 samples. This approach guaranteed a well-defined coherence measure for each segment with the same frequency resolution as in the rest of the experiments.

Phase difference

As another measure of head stabilization we used the phase difference between two signals x and y [51]. This was calculated by determining the peak of the cross-correlation between x and y, in segments of 512 samples centered around each detected step. The time-lag of this peak was then transformed into a phase difference by dividing by the period length of signal x, estimated via auto-correlation. Phases differences were calculated between vertical head acceleration and head pitch velocity and between vertical head acceleration and trunk pitch velocity.

Since we computed phases differences between acceleration and pitch velocity, we corrected the resulting differences to be comparable with previously reported results that compared vertical displacement and pitch angle [51]. Pitch angle is obtained from pitch velocity by integrating once (taking into account some initial value) and translation is obtained from acceleration by integrating twice. Since the integration of a sinusoidal signal introduces a phase shift of $-\frac{\pi}{2}$, the overall phase correction for the difference is $2\left(-\frac{\pi}{2}\right) - \left(-\frac{\pi}{2}\right) = -\frac{\pi}{2}$.

4.2.9. Statistical analysis

The influence of the predominant frequency on the calculated measures was estimated with a Kruskal-Wallis test by calculating an effect size as follows [123]:

$$\eta^2 = \frac{H - k + 1}{n - k} \tag{4.9}$$

where H is the Kruskal-Wallis statistic, k is the number of predominant frequency groups and n is the number of samples. Effect sizes were considered small for $\eta^2 < 0.04$, intermediate for $0.04 < \eta^2 < 0.11$ and large for $\eta^2 > 0.11$ [30]. For pairwise comparisons between independent samples (e.g. between previously reported results and ours), Welch's two-sample t-test was used. Pairwise comparisons between dependent samples (e.g. between different spatial directions) were performed with a paired t-test. For each test, we reported p-values and considered results to be significant if $p < 0.01$. However, since this was an exploratory study, statistical power of these tests might be limited.

Statistical analysis was performed with the `stats` module of the `scipy` library (version 1.2.0) in Python 3.6. Results of our analyses were plotted as a function of predominant frequency using boxplots. Boxes indicated the range from the first to the third quartile and the band indicated the median. Whiskers were plotted from the lowest sample within 1.5 times the interquartile range (IQR) of the lower quartile to the highest sample within 1.5 times the IQR of the upper quartile. Due to the large amount of samples, outliers were not plotted. The number of samples was $n = 34455$, the number of steps that fell within the analyzed predominant frequency range (93.74% of all detected steps, see results and supplementary material).

4.3. Results

Predominant frequency of vertical trunk acceleration was strongly correlated with predominant frequency of vertical head acceleration between 1 and 2.6 Hz ($\eta^2 = 0.887$, $p < 0.001$, fig. 4.2A). Figure 4.2B shows a re-plot of figure 8B from [51], showing the relationship between walking velocity and predominant frequency of vertical head acceleration. In order to make our results comparable to previously published results, we limited our analysis to segments with head predominant frequencies between 1 and 2.6 Hz, corresponding to the range of frequencies associated with walking speeds between 0.6 and 2.2 m/s determined in [51].

Predominant frequency of vertical head acceleration was approximately normally distributed around 1.86 Hz with a standard deviation of 0.23 Hz (fig. 4.3A). RMS vertical accelerations exhibited a distribution skewed towards higher RMS values with a peak at 0.3 g for both head and trunk (fig. 4.3C & E). RMS accelerations increased with predominant frequency for both head ($\eta^2 = 0.375$, $p < 0.001$, fig. 4.3B) and trunk ($\eta^2 = 0.377$, $p < 0.001$, fig. 5D) and exhibited broader distributions with higher frequencies. This indicated a strong preference of subjects to move with a fundamental frequency close to 2 Hz and maintaining moderate accelerations of both head and trunk.

Distribution of bout lengths decreased logarithmically with the logarithm of bout length

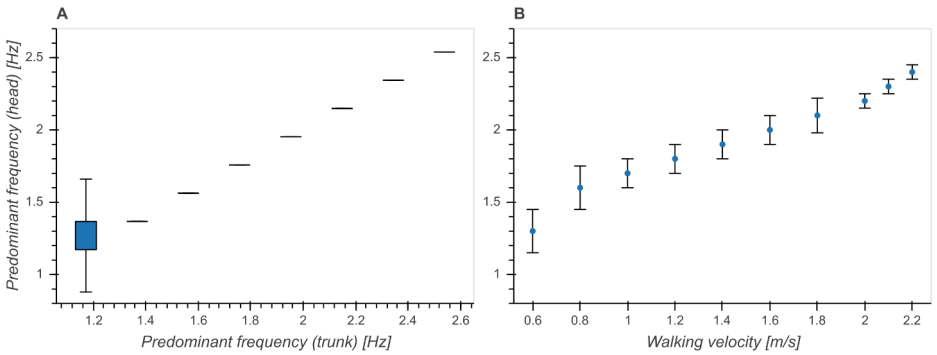

Figure 4.2.: Relationship between predominant frequency of vertical head acceleration and **(A)** predominant frequency of vertical trunk acceleration; **(B)** walking velocity (fig. 8B from [51]). Boxes above 1.2 Hz in (A) are not visible because all of the samples between the first and third quartile had the same value.

(fig. 4.4A). The effect of bout length on per-bout mean predominant frequencies was small ($\eta^2 = 0.022$, $p < 0.001$), although the median seemed to increase with larger bout lengths and they exhibited broader distributions for shorter bouts (fig. 4.4B). Standard deviations of predominant frequencies showed an intermediate dependence on bout length ($\eta^2 = 0.101$, $p < 0.001$) and exhibited smaller variances above 100 steps (fig. 4.4C). This showed a clear preference of subjects towards walking short bouts while longer bouts seemed to be connected to an increase of predominant frequency and a simultaneous decrease of variability.

The effect of predominant frequency on ACs in V direction was small ($\eta^2 = 0.039$, $p < 0.001$, fig. 4.5A). However, ACs increased with predominant frequency up to 2 Hz and afterwards decreased with higher frequencies in both AP ($\eta^2 = 0.165$, $p < 0.001$) and ML ($\eta^2 = 0.144$, $p < 0.001$) directions (fig. 4.5A). Pairwise comparisons between directions revealed significant differences between each pair of directions ($p < 0.001$), with ACs in V direction being lower than those in AP and ML directions. These differences were especially evident around 2 Hz, corresponding to the frequency range containing the highest number of samples (see also fig. 4.3A). ACs in V and AP direction differed significantly ($p < 0.001$) from those reported by [86], but not in the ML ($p = 0.043$) direction (fig. 4.5B). We found the most substantial difference in the V direction where we observed higher values, indicating that real-world vertical accelerations of the head are more strongly attenuated than previously reported.

The influence of predominant frequency on HRs was small across all directions for both head and trunk ($\eta^2 < 0.04$, $p < 0.001$), although we observed higher standard deviations between 2 and 2.4 Hz, especially in the AP and V directions (fig. 4.6A & C). Distributions differed significantly between each pair of directions ($p < 0.001$). Statistical testing revealed no significant differences between our results and those reported by [89] except for the head in the ML direction ($p < 0.001$), but we saw higher

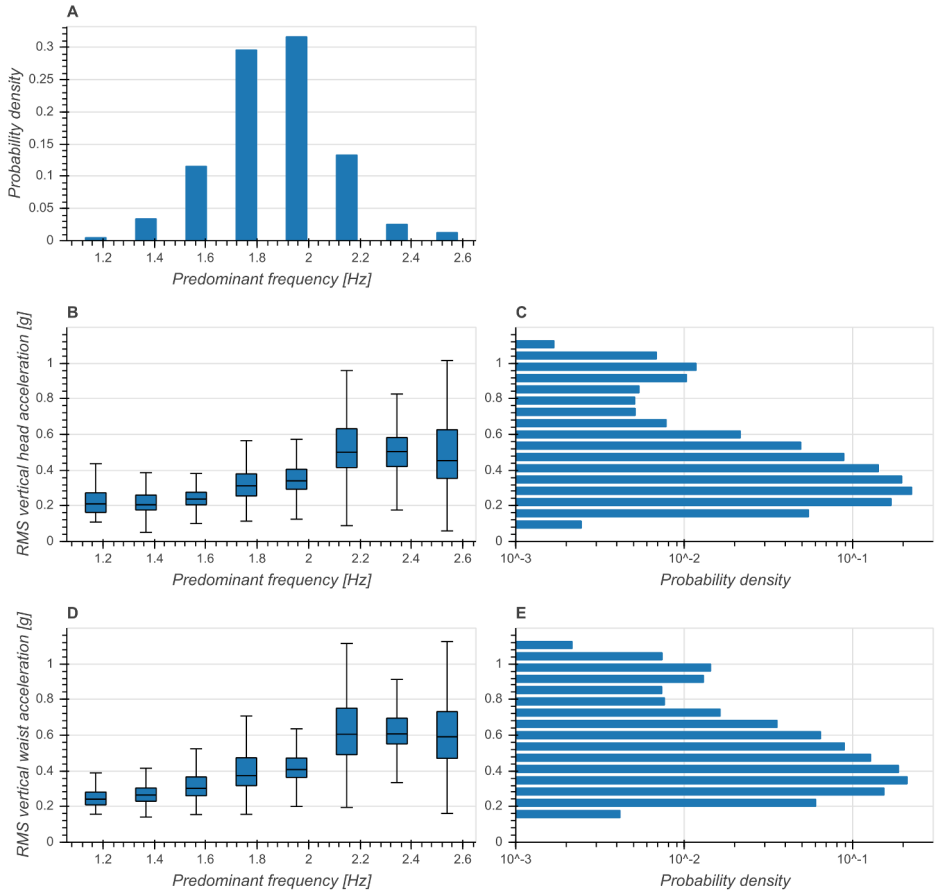

Figure 4.3.: (A) Distribution of predominant frequency of vertical head acceleration; **(B)** Boxplot of RMS vertical head accelerations as a function of predominant frequency. **(C)** Distribution of RMS vertical head accelerations (logarithmic scale). **(D)** Boxplot of RMS vertical trunk accelerations as a function of predominant frequency. **(E)** Distribution of RMS vertical trunk accelerations (logarithmic scale).

standard deviations for all axes and both sensor locations (fig. 4.6B & D). The high values of HRs measured around 2 Hz are an indication of highly regular and stable gait in this frequency range.

There was an intermediate effect of predominant frequency on coherence both between vertical head acceleration and head pitch velocity ($\eta^2 = 0.109$, $p < 0.001$, fig. 4.7A) and between head and trunk pitch velocity ($\eta^2 = 0.084$, $p < 0.001$, fig 4.7C). We observed an increase of mean coherence value around 2.15 Hz as well as a decrease of standard deviation. Coherence values differed significantly between head and trunk in the predominant frequency range from 1.37 to 2.34 Hz. These results are consistent with those reported in [51] (fig. 4.7B & D), although it should be noted that they obtained values for vertical displacement and pitch angle instead of vertical acceleration and

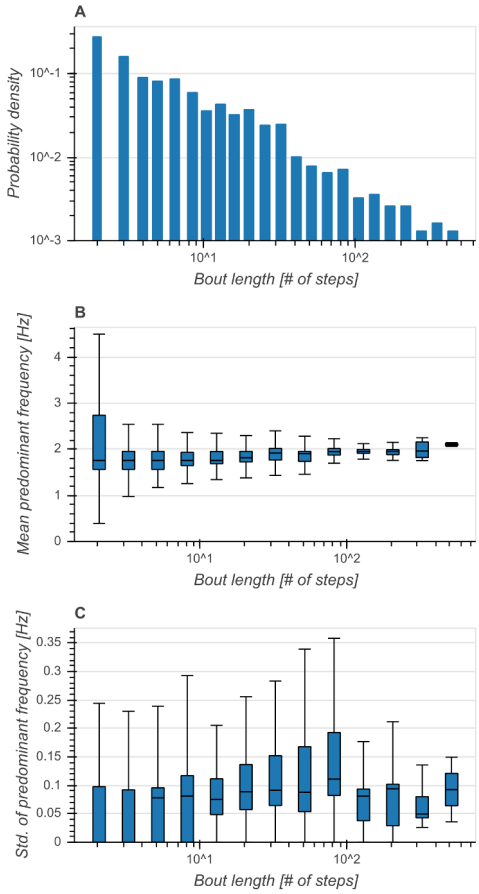

Figure 4.4.: (A) Distribution of bout lengths (logarithmic scale). **(B)** Boxplot of mean predominant frequency for each bout as a function of bout length. Broader distributions indicated a higher variance of predominant frequencies between bouts. **(C)** Boxplot of standard deviation of predominant frequency for each bout as a function of bout length. Higher values indicated a higher variance of predominant frequencies within bouts.

pitch velocity. However, since the coherence measures the similarity between signals at the predominant frequency, a mere phase shift as introduced by the integration of a sinusoidal signal component should not alter the value of the coherence function. These results demonstrate a tight coupling between both head pitch and vertical head translation as well as head and trunk pitch around the preferred predominant frequency of 2 Hz.

Predominant frequency had a small effect on phase differences for both head ($\eta^2 = 0.006$, $p < 0.001$, fig. 4.8A) and trunk ($\eta^2 = 0.022$, $p < 0.001$, fig. 4.8C). There was a significant difference between head and trunk for the whole analyzed range of predominant frequencies except for 1.17, 1.37 and 2.15 Hz. While the overall mean phase differences were comparable to those reported in [51], we did not observe a dependence on predominant frequency (fig. 4.8B & D). This indicates a phase lock between vertical head displacement and head/trunk pitch angle, independent of predominant frequency.

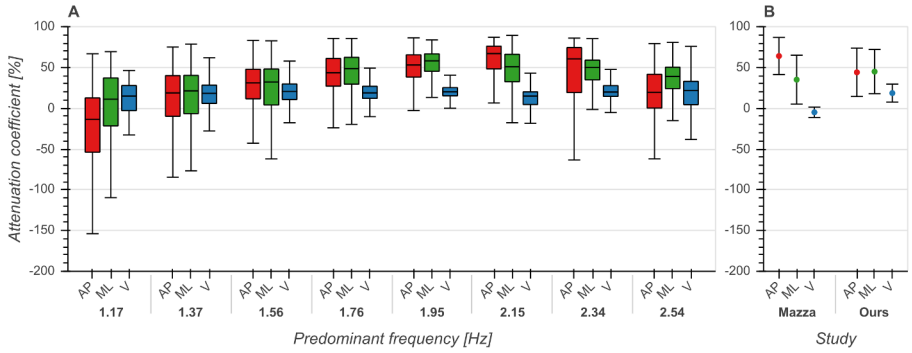

Figure 4.5.: Attenuation coefficients of accelerations between trunk and head in anterior/posterior (AP), medial/lateral (ML) and vertical (V) directions. **(A)** Attenuation coefficients as a function of predominant frequency. **(B)** Comparison between mean +/- std attenuation coefficients from [86] and our data. Means for [86] were computed as the average of the means of the two groups (male, female). Standard deviations were estimated by multiplying the reported standard error of the mean by the square root of the sample size and then computing the square root of the sum of squares of the groups. See also first row of fig. 2 from [86] for comparison.

4.4. Discussion

Due to the limited ecological validity of measurements obtained in a controlled laboratory setting [20, 91], there is a need for methods to measure and analyze head stabilization and head-trunk coordination in real-world scenarios. For clinical applications, it is first necessary to obtain normative data from healthy individuals as a baseline for possible diagnostic use. In this study, we measured head and trunk motion in an ecologically valid context and calculated several derivative measures of head stabilization performance. These measures were chosen based on those reported in the literature, and they evaluate horizontal head stabilization as well as head motion that compensates for vertical translation. Overall, our measures based on real-world accelerometry data agree quite well with similar measures derived from laboratory-based data, suggesting that these methods for quantifying head stabilization performance could generalize. However, we noticed some important differences and in general we observed larger variances in the distribution of these measures.

Predominant frequencies of motion were tightly coupled between trunk and head (fig. 4.2) and exhibited a narrow distribution around 2 Hz (fig. 4.3). Incidence of bout lengths decreased strongly towards longer bouts, but means and standard deviations of predominant frequencies did not strongly depend on bout length, showing only a small increase of means and simultaneous decrease of standard deviations towards longer bouts (fig. 4.4). These findings seem to confirm previous reports [76] which identified 2 Hz as the fundamental frequency of human locomotion across a wide range of activities. The observed changes in predominant frequency distribution as

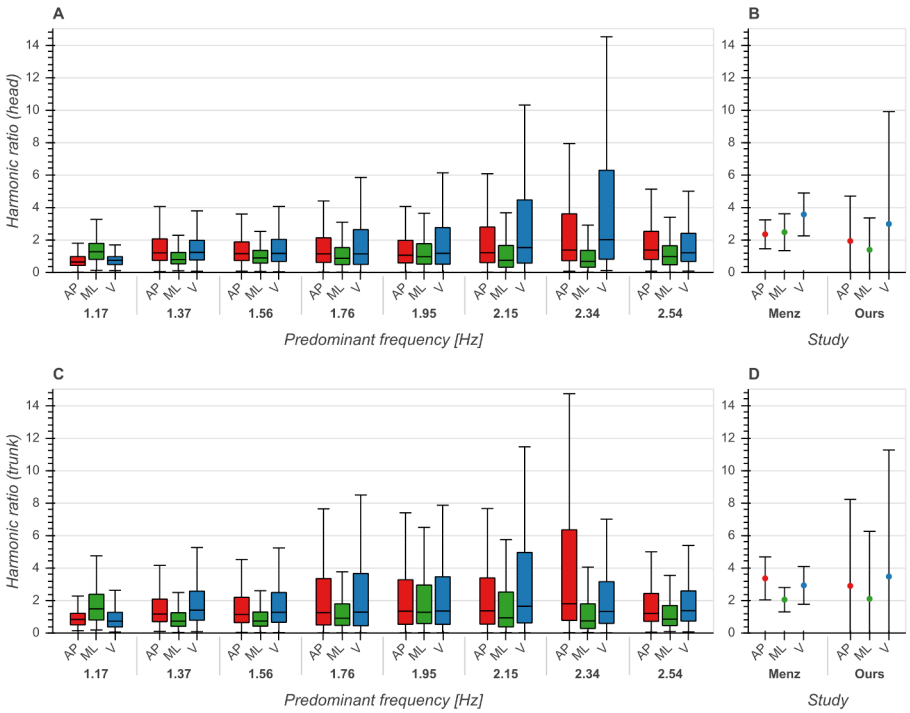

Figure 4.6.: Harmonic ratios of accelerations in anterior/posterior (AP), medial/lateral (ML) and vertical (V) directions. **(A)** Boxplot of harmonic ratios of head accelerations as a function of predominant frequency. **(B)** Comparison between mean +/- std harmonic ratios (head) from [89] and our data. **(C)** Boxplot of harmonic ratios of trunk accelerations as a function of predominant frequency. **(D)** Comparison between mean +/- std harmonic ratios (trunk) from [89] and our data. See also fig. 6 from [89] for comparison.

a function of bout length indicate a tendency of subjects towards more goal-directed and stable walking for longer distances. However, the observed differences for short bouts could also have other causes: On the one hand, these bouts could consist of false positive steps detected during cycling. With a larger annotated dataset it should be possible to develop a more refined step detection approach, possibly involving machine learning techniques or GPS data. Special care needs to be taken in order to faithfully detect slow or asymmetric gaits if the goal is to develop a diagnostic tool. On the other hand, it is possible that this in an artifact of the spectral analysis used for determining predominant frequency, which analyses segments of 5 seconds length in order to achieve the desirable frequency resolution. This choice arguably influenced the analysis of very short bouts as non-locomotion data was included in the transform window. Yet, for the analysis of elderly people and pathological gaits, short bouts are of paramount importance, as they make up most of the daily walking activity [110]. Special frequency analysis techniques for non-stationary data such as the empirical mode decomposition [53] could help circumvent this issue.

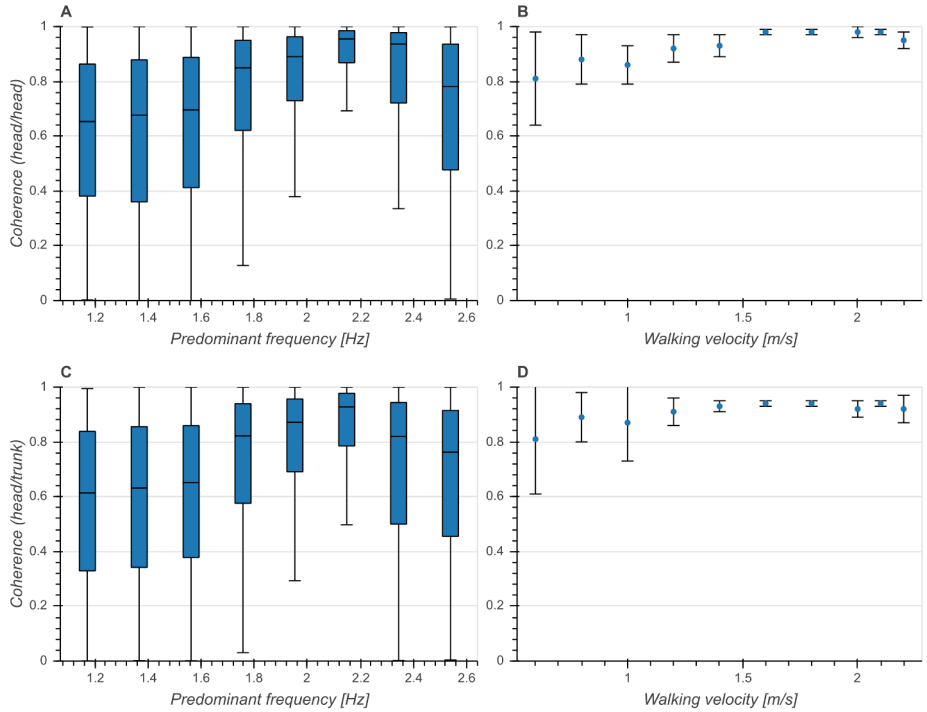

Figure 4.7.: Coherence at predominant frequency. **(A)** Boxplot of coherence between vertical head acceleration and head pitch velocity as a function of predominant frequency. **(B)** Coherence between vertical head displacement and head pitch angle as a function of walking velocity (fig. 9A from [51]). **(C)** Boxplot of coherence between head pitch velocity and trunk pitch velocity as a function of predominant frequency. **(D)** Coherence between head pitch angle and trunk pitch angle as a function of walking velocity (fig. 9B from [51]).

Attenuation of accelerations from trunk to head was stronger in AP and ML directions than in the V direction (fig. 7), consistent with previous reports [60, 86]. The reason for this is that the kinematic chain of the upper body aims at minimizing horizontal accelerations in order to stabilize the head in space. Compared with the results of [86] we observed stronger attenuation in the V direction; this could be due to characteristics of our uncontrolled environment such as inclusion of stair walking. Buckley et al. [21] observed that attenuation of accelerations in the ML direction was significantly lower in patients with Parkinson's disease when compared with healthy controls. This deterioration in patients seems to indicate that attenuation of lateral accelerations is due to active stabilization and not simply biomechanical constraints of the head-trunk chain. Attenuation strengths in AP and ML directions also showed a dependence on predominant frequency, exhibiting the highest values around 2 Hz. To the best of our knowledge, this is the first time that ACs were characterized as a function of predominant frequency. These results suggest that the attenuation of horizontal head accelerations

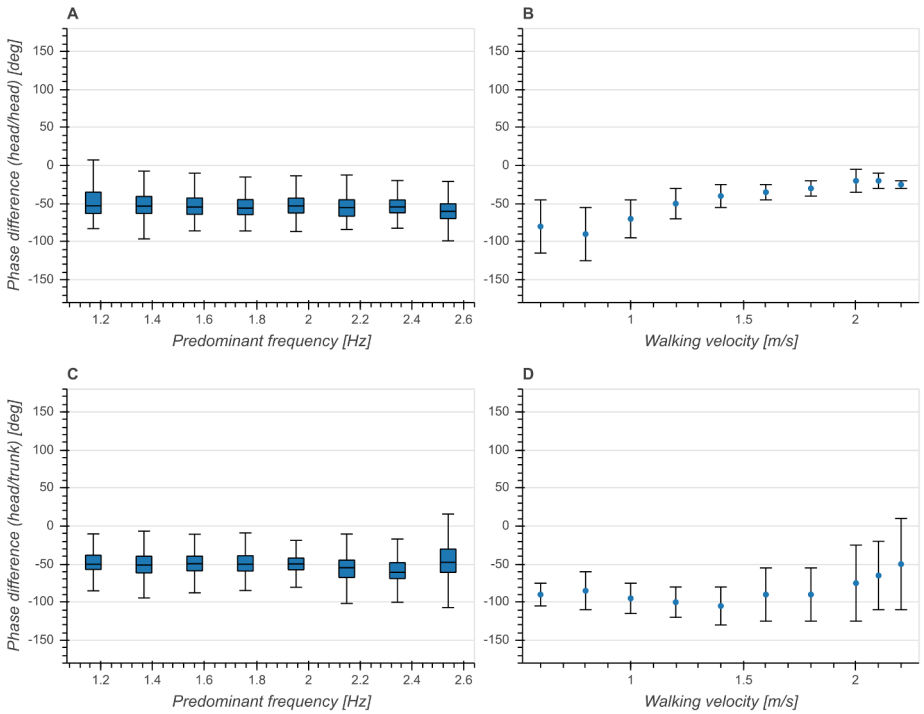

Figure 4.8.: (A) Boxplot of corrected phase difference between vertical head acceleration and head pitch velocity as a function of predominant frequency. **(B)** Phase difference between vertical head displacement and head pitch angle as a function of walking velocity (fig. 5C from [51]). **(C)** Boxplot of corrected phase difference between vertical head acceleration and trunk pitch velocity as a function of predominant frequency. **(D)** Phase difference between vertical head displacement and trunk pitch angle as a function of walking velocity (fig. 5B from [51]).

is tuned to the fundamental frequency of locomotion and that the quantification of this attenuation could be used as an ecologically valid objective measure of head stability.

Regularity of motion as measured by the HR was consistent with previous reports [89], although we found higher variances in all directions of motion (fig. 8). This could be explained by the fact that a significant effect of environmental factors such as walking on uneven surfaces [89] or unilateral limb loading [17] on the measured HRs has been observed. Previous studies found significantly lower HRs at both trunk and head between patients with MS [100] or PD [67, 74] and healthy controls, although there have been differing reports in the case of PD [21]. We observed an increase in HRs with predominant frequencies above 2 Hz, most prominently in the AP and V directions, in accordance with earlier reports [89]. Based on these findings, we conclude that the HR might be a suitable measure of head stabilization in a real-world context.

The similarity between vertical head acceleration and head pitch and between head pitch and trunk pitch as measured by the coherence was maximal at predominant

frequencies around 2.2 Hz (fig. 9). This is in line with previous reports [51] which observed the highest coherence values at walking speeds above the most common gait velocity of 1.4 m/s. Compared to their results, we measured lower means and higher standard deviations of coherence values across the entire range of analyzed predominant frequencies. These differences can be explained by the fact that Hirasaki et al. [51] analyzed steady-state walking on a treadmill with a target for gaze fixation. High coherences are associated with compensatory head motion aimed at maintaining gaze stability [51]. In a real-world setting, often characterized by intermittent walking and frequent gaze shifts, it is not surprising that overall lower coherence values are observed. Lower coherences have also been linked to vestibular deficits [98], suggesting a possible applicability of this measure in a clinical context.

Phase differences between vertical head acceleration and head/trunk pitch were distributed around -50° across the entire analyzed range of predominant frequencies (fig. 10). This is partly consistent with previous studies [51], however these studies reported an effect of walking velocity on the phase difference which we did not observe. Similar to the coherence, we hypothesize that the observed differences are due to our measurement scenario lacking a target for gaze fixation. We are not aware of any studies investigating phase differences of subjects with gait, balance or neurological disorders.

With the exception of phase differences, all analyzed metrics indicated strongest head stabilization around 2 Hz, corresponding to the preferred walking speed of the participants. We also observed the lowest variances of these measures in this range, in line with previous reports by Wuehr et al. [131] who showed that coefficients of variation of gait parameters such as stride time and stride length are lowest at self-selected walking speeds. Additionally, they measured higher variances in patients with cerebellar ataxia, especially outside of the range of preferred speeds, raising the question whether similar effects could occur for parameters of head stability.

Another disorder characterized by movement deficits is autism spectrum disorder (ASD) [124]. Children diagnosed with ASD exhibit atypical motor patterns that can be identified using machine learning techniques with great accuracy [9]. Computer-vision based tracking of head motion revealed that magnitude and velocity of head turning as well as velocity of head inclination are greater in children with ASD than in healthy controls [25]. This difference was especially evident when subjects watched video of social stimuli. Therefore, assessment of head motion during real-world social interactions could be a valuable tool for ASD diagnosis and research.

It should be noted that the size and makeup of our sample of participants is a possible source for bias. The sample included exclusively young subjects which facilitated comparison with previously reported results. In contrast, a normative dataset for comparison with diseased populations will likely have to include older subjects. A longer measurement period (at least one week) could also be helpful in increasing

the significance of findings. Furthermore, neither the gravity estimation nor the step detection algorithm have been independently validated and we did not control for movement of the sensors relative to head or trunk. However, all analyses performed in the aligned coordinate system are largely robust to small shifts in sensor position. The other concerns can be addressed in the study design of future studies.

In conclusion, we have shown that several previously described head stability parameters, when measured in an ecologically valid context, exhibited characteristics similar to those obtained in a laboratory setting. We have also characterized these parameters in function of predominant frequency as a proxy for walking speed (figs. 4.5-4.8). Nevertheless, we found some critical differences that could be attributed to features unique to the real-world context. Real-world measurements of attenuation coefficients were comparable to those previously obtained in a laboratory setting [86], as were measurements of harmonic ratios [89]. We could also replicate previously reported characteristics of coherences and phase differences [51]. Most of these measures have been shown to have value for diagnostic purposes or as endpoints for clinical trials. Our results indicate that the evaluated parameters are largely robust to characteristics that are usually absent in a laboratory context, such as frequent and large shifts of gaze and attention, dual tasking or walking with a companion. The data recorded in this study could serve as a model for collecting normative reference data of healthy individuals. Future studies will have to address the direct comparison of ecologically valid head stabilization parameters between healthy controls and patients with gait, balance, or neurological disorders. This way, mobile accelerometry could serve as a cheap and easy method to gain clinically relevant insights.

5. Conclusion

In this thesis, I have presented several novel approaches for real-world human motion tracking. The aim was to present various key stages of the scientific inquiry of motion tracking, from development and evaluation of tracking systems to their application in studies with human subjects. An emphasis was put on tracking methods specifically tailored to measuring head movements.

Chapter 2 introduced a mobile system for simultaneous measurement of head and eye movements based on two off-the-shelf components, the Pupil Core and the Intel RealSense T265. Two experiments were performed that demonstrated the system's ability to track head and eye movements at the same time. The primary advantages of this system are that the components are relatively inexpensive, all software is free and open-source and the necessary modifications to the existing hardware can be 3D-printed. To my knowlegde, this is the first open-source system for joint head-eye tracking. A future iteration of this system could be made even more lightweight and inobtrusive by using smaller devices and integrating them into a single rigid frame.

The suitability of the T265 for the specific use case of tracking human head motion was evaluated in chapter 3. Although previous studies had assessed the accuracy of the tracker as part of autonomous robotic systems [1, 3, 16] and hand-held devices [96], this chapter describes the first evaluation study explicitly addressing head tracking performance. Additionally, a primary aim of the study was to investigate the tracking accuracy in real-world environments. The results showed that certain measurements such as the orientation of the device with respect to gravity are quite accurate while others such as overall movement speed can be considerably under-estimated in large, open environments. These shortcomings could be addressed in the future by using the raw sensor data from the device and running a tracking algorithm designed specifically for human motion and open environments post-hoc.

Finally, chapter 4 discussed a study where head and trunk motion was recorded simultaneously with the objective of replicating previous studies on head stability in a real-world setting. The head stabilization parameters examined in these studies [51, 86, 89] had been collected in a laboratory setting which raised the question whether the same results could be obtained from real-life measurements. Our study mostly confirmed the previous reports, although some differences such as higher variances of the head stability measures were found. These results suggest that the investigated

stability parameters – when measured in a real-world context – could be useful as diagnostic tools or outcome measures for clinical trials.

In summary, this work has shown that recent technological developments have made it possible to accurately track human motion outside the lab. This kind of motion tracking has a wide range of applications, from the investigation of fundamental neuroscientific questions to clinical research and commercial applications such as augmented reality. One of the central open challenges identified in this thesis is the accuracy of fully positional tracking methods independent of environment. This motivates the development of novel devices and algorithms that address the specific use case of human motion tracking in the real world. Furthermore, miniaturization of tracking devices will be a key requirement for user acceptance, especially for studies that aim to measure unbiased real-world motion as well as for commercial products. Finally, the issue of privacy lies at the heart of any technology capable of quantifying human behavior. Camera-based tracking systems are a concern not only for the users themselves, but also for everyone around them. One possibility to avoid this problem is to perform all calculations on the device without recording any camera images. However, this makes it impossible to perform post-hoc motion estimation and thus imposes even higher requirements on the robustness of the tracking algorithm. Nevertheless, if all of these challenges can be addressed in the future, human motion tracking during everyday life could become a truly pervasive technology.

A. Mathematical appendix

A.1. Quaternion algebra

The quaternion number system was first described by William Rowan Hamilton in 1843 as an extension of the complex numbers [47]. Quaternions can be used to unambiguously describe rotations in three-dimensional (3D) space and are used throughout this thesis as the primary representation for orientations and rotations.

A quaternion $\mathbf{q}$ can be written as:

$$\mathbf{q} = q_w + q_x\mathbf{i} + q_y\mathbf{j} + q_z\mathbf{k} \tag{A.1}$$

with the real numbers q_w, q_x, q_y and q_z, and the *basic quaternions* $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. More compactly, quaternions can also be written as four-dimensional vectors:

$$\mathbf{q} = \begin{pmatrix} q_w \\ q_x \\ q_y \\ q_z \end{pmatrix} \tag{A.2}$$

Scalar and vector part

The scalar or real part of a quaternion is:

$$\mathrm{Re}(\mathbf{q}) = q_w \tag{A.3}$$

and its vector or imaginary part is:

$$\mathrm{Im}(\mathbf{q}) = \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} \tag{A.4}$$

Conjugate

The conjugate of a quaternion is defined as:

$$\mathbf{q}^* = \begin{pmatrix} \mathrm{Re}(\mathbf{q}) \\ -\mathrm{Im}(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} q_w \\ -q_x \\ -q_y \\ -q_z \end{pmatrix} \tag{A.5}$$

Norm

The norm of a quaternion is computed as:

$$\|\mathbf{q}\| = \sqrt{q_w^2 + q_x^2 + q_y^2 + q_z^2} \tag{A.6}$$

Quaternions with norm 1 are called unit quaternions or versors and can be used to represent rotations in 3D space.

Inverse

The inverse of a quaternion can be calculated as:

$$\mathbf{q}^{-1} = \frac{\mathbf{q}^*}{\|\mathbf{q}\|^2} \tag{A.7}$$

For unit quaternions, this is the same as the conjugate and corresponds to the inverse rotation.

Hamilton product

The Hamilton product is the non-commutative product of two quaternions $\mathbf{q}$ and $\mathbf{o}$. For unit quaternions, this corresponds to chaining two rotations.

$$\mathbf{q} \cdot \mathbf{o} = \begin{pmatrix} q_w r_w - q_x r_x - q_y r_y - q_z r_z \\ q_w r_x + q_x r_w + q_y r_z - q_z r_y \\ q_w r_y - q_x r_z + q_y r_w + q_z r_x \\ q_w r_z + q_x r_y - q_y r_x + q_z r_w \end{pmatrix} \tag{A.8}$$

Quaternion rotation

A vector $\mathbf{v}$ can be rotated by a unit quaternion $\mathbf{q}$ with:

$$\mathrm{rot}\,(\mathbf{q}, \mathbf{v}) = \mathrm{Im}\left(\mathbf{q}\,\tilde{\mathbf{v}}\,\mathbf{q}^{-1}\right) \tag{A.9}$$

where $\tilde{\mathbf{v}} = (0, v_x, v_y, v_z)^\top$ is a "pure" quaternion with real part 0, constructed from $\mathbf{v}$.

Construction from axis-angle representation

A unit quaternion can be constructed from a vector $\mathbf{v}$ whose length $\alpha = \|\mathbf{v}\|$ represents the angle of a rotation and whose direction $\mathbf{v}' = {}^{\mathbf{v}}/_{\alpha}$ represents the axis of the rotation:

$$q(\mathbf{v}) = \begin{pmatrix} \cos\left({}^{\alpha}/_{2}\right) \\ \sin\left({}^{\alpha}/_{2}\right) v'_x \\ \sin\left({}^{\alpha}/_{2}\right) v'_y \\ \sin\left({}^{\alpha}/_{2}\right) v'_z \end{pmatrix} \tag{A.10}$$

A.2. Reference frame transformations

This section will show how motion can be transformed between different reference frames. To demonstrate this, assume that a body B is moving with respect to (w.r.t.) an observer O who in turn is moving w.r.t. a world-fixed frame W (see fig. A.1).

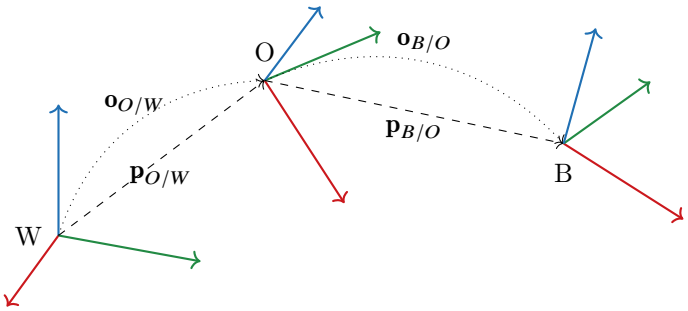

Figure A.1.: A reference frame tree with three different frames: world, observer and body frame.

Position

Assume the observer measures the position $\mathbf{p}_{B/O}$ of the body w.r.t. itself as well as its own position $\mathbf{p}_{B/O}$ and orientation $\mathbf{o}_{O/W}$ w.r.t. the world frame. The position $\mathbf{p}_{B/W}$ of

the body w.r.t. the world frame can then be calculated as:

$$\mathbf{p}_{B/W} = \mathrm{rot}\left(\mathbf{o}_{O/W}, \mathbf{p}_{B/O}\right) + \mathbf{p}_{O/W} \tag{A.11}$$

Orientation

The orientation $\mathbf{o}_{B/W}$ of B w.r.t. is simply the Hamilton product of $\mathbf{o}_{O/W}$ and $\mathbf{o}_{B/O}$:

$$\mathbf{o}_{B/W} = \mathbf{o}_{O/W} \cdot \mathbf{o}_{B/O} \tag{A.12}$$

Linear velocity

The linear velocity $\mathbf{v}_{B/W}$ can be computed from $\mathbf{v}_{B/O}$ with the "three term" velocity formula [104, p.887]:

$$\mathbf{v}_{B/W} = \mathbf{v}_{B/O} + \mathbf{v}_{O/W} + \boldsymbol{\omega}_{O/W} \times \mathbf{p}_{B/O} \tag{A.13}$$

Angular velocity

The angular velocity $\boldsymbol{\omega}_{B/W}$ is simply the sum of $\boldsymbol{\omega}_{B/O}$ and $\boldsymbol{\omega}_{O/W}$ [104, p.902]:

$$\boldsymbol{\omega}_{B/W} = \boldsymbol{\omega}_{B/O} + \boldsymbol{\omega}_{O/W} \tag{A.14}$$

Linear acceleration

The linear acceleration $\mathbf{a}_{B/W}$ can be computed from $\mathbf{a}_{B/O}$ with the "five term" acceleration formula [104, p.890]:

$$\mathbf{a}_{B/W} = \mathbf{a}_{B/O} + \mathbf{a}_{O/W} + \underbrace{2\boldsymbol{\omega}_{O/W} \times \mathbf{v}_{B/O}}_{\text{Coriolis accel.}} + \underbrace{\boldsymbol{\omega}_{O/W} \times \left(\boldsymbol{\omega}_{O/W} \times \mathbf{p}_{B/O}\right)}_{\text{centrifugal accel.}} + \underbrace{\boldsymbol{\alpha}_{O/W} \times \mathbf{p}_{B/O}}_{\text{Euler accel.}} \tag{A.15}$$

Angular acceleration

Again, the calculation of the angular acceleration $\boldsymbol{\alpha}_{B/W}$ is a lot simpler [104, p.903]:

$$\boldsymbol{\alpha}_{B/W} = \boldsymbol{\alpha}_{B/O} + \boldsymbol{\alpha}_{O/W} \tag{A.16}$$

Vectors and representation frames

Assume the observer measures the linear velocity $\mathbf{v}_{B/O}$ of the body w.r.t. itself. This is a directed vector quantity and its coordinates are subject to a coordinate system. This coordinate system is called the representation frame and denoted by a leading superscript, i.e., ${}^{O}\mathbf{v}_{B/O}$ denotes $\mathbf{v}_{B/O}$ represented in the observer coordinate system O. This vector can be transformed into the representation frame of the body with:

$${}^{B}\mathbf{v}_{B/O} = \mathrm{rot}\left(\mathbf{o}_{B/O}, {}^{O}\mathbf{v}_{B/O}\right) \tag{A.17}$$

It should be noted that, in the above transformation equations, all directed vectors (i.e., velocities and accelerations, but not positions and orientations) have to be expressed in the same representation frame.

A.3. Extended Kalman filter

The extended Kalman filter (EKF) is an algorithm used to estimate the unobservable state $\mathbf{x}$ of a non-linear dynamical system that is modeled with the state transition model:

$$\mathbf{x}_k = f\left(\mathbf{x}_{k-1}, \mathbf{u}_k\right) + \mathbf{w}_k \tag{A.18}$$

and the measurement model:

$$\mathbf{z}_k = h\left(\mathbf{x}_k\right) + \mathbf{v}_k \tag{A.19}$$

Here, f and h are the differentiable state transition and measurement functions, $\mathbf{u}_k$ the control input, $\mathbf{z}_k$ the measurements and $\mathbf{w}_k$ and $\mathbf{v}_k$ the process and measurement noise [102].

State transition and observation matrices

For the EKF algorithm, the state transition function f and the measurement function h need to be linearized about the working point. This is achieved by computing the Jacobians:

$$\mathbf{F}_k = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}_{k-1|k-1}, \mathbf{u}_k} \tag{A.20}$$

and

$$\mathbf{H}_k = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\mathbf{x}_{k|k-1}} \tag{A.21}$$

Prediction step

In this step, the predicted state estimate $\mathbf{x}_{k|k-1}$ is calculated according to the state transition function:

$$\mathbf{x}_{k|k-1} = f(\mathbf{x}_{k-1|k-1}, \mathbf{u}_k) \tag{A.22}$$

as well as the predicted covariance estimate:

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^\top + \mathbf{Q}_k \tag{A.23}$$

Here, $\mathbf{Q}_k$ denotes the covariance of the process noise $\mathbf{w}_k$.

Update step

In this step, the measurement residual (or innovation) $\mathbf{y}_k$ is calculated based on the measurement $\mathbf{z}_k$ and the measurement function:

$$\mathbf{y}_k = \mathbf{z}_k - h(\mathbf{x}_{k|k-1}) \tag{A.24}$$

as well as the residual (or innovation) covariance:

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k \tag{A.25}$$

Here, $\mathbf{R}_k$ denotes the covariance of the measurement noise $\mathbf{v}_k$. From this, the Kalman gain $\mathbf{K}_k$ is calculated as:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top \mathbf{S}_k^{-1} \tag{A.26}$$

and finally the updated state estimate:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k \mathbf{y}_k \tag{A.27}$$

as well as the updated covariance estimate:

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{A.28}$$

B. Supplementary material for chapter 3

B.1. Details of reference frame transformations

The transformation between the world frame W of the optical tracking system (OTS) and that of the T265 ($\hat{W}$) was estimated using a basic point set registration (PSR) method. The method estimates a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ for two sets $\mathcal{X} = \{\mathbf{x}_i \mid i \in 1..n\}$ and $\mathcal{Y} = \{\mathbf{y}_i \mid i \in 1..n\}$ of corresponding points that minimizes $\sum_i^n \|\mathbf{y}_i - (\mathbf{R}\mathbf{x}_i + \mathbf{t})\|_2$.

$$\mathbf{H} = \sum_i^n (\mathbf{x}_i - \overline{\mathbf{x}})\,(\mathbf{y}_i - \overline{\mathbf{y}})^\top \tag{B.1a}$$

$$\mathbf{U}, \mathbf{S}, \mathbf{V} = \mathrm{SVD}(\mathbf{H}) \tag{B.1b}$$

$$\mathbf{R} = \mathbf{V}\mathbf{U}^\top \tag{B.1c}$$

$$\mathbf{t} = -\mathbf{R}\overline{\mathbf{y}} + \overline{\mathbf{x}} \tag{B.1d}$$

where $\overline{\mathbf{x}} = \frac{1}{n}\sum_i^n \mathbf{x}_i$ and $\overline{\mathbf{y}} = \frac{1}{n}\sum_i^n \mathbf{y}_i$ denote the centroids of $\mathcal{X}$ and $\mathcal{Y}$, respectively. A modified version this algorithm with $\mathbf{H} = \sum_i^n \mathbf{x}_i\,\mathbf{y}_i$ can be used to estimate a transform that consists only of a rotation and minimizes $\sum_i^n \|\mathbf{y}_i - \mathbf{R}\mathbf{x}_i\|_2$.

With $\mathcal{X} = \{{}^{\hat{W}}\hat{\mathbf{p}}_i \mid i \in 1..n\}$ and $\mathcal{Y} = \{{}^{W}\mathbf{p}_i \mid i \in 1..n\}$ this yields $\mathbf{R}_{\hat{W}}$ and $\mathbf{t}_{\hat{W}}$ which can be used to transform position and orientation of the T265 from its own to the OTS world frame:

$${}^{W}\hat{\mathbf{p}} = \mathrm{rot}\left(\mathbf{q}_{\hat{W}}, {}^{\hat{W}}\hat{\mathbf{p}}\right) + \mathbf{t}_{\hat{W}} \tag{B.2a}$$

$${}^{W}\hat{\mathbf{q}} = \mathrm{rot}\left(\mathbf{q}_{\hat{W}}, {}^{\hat{W}}\hat{\mathbf{q}}\right) \tag{B.2b}$$

Here, $\mathrm{rot}(\mathbf{q}, \mathbf{v}) = \mathbf{q}\mathbf{v}\mathbf{q}^{-1}$ denotes the rotation of a vector $\mathbf{v}$ by the quaternion $\mathbf{q}$ and $\mathbf{q}_{\hat{W}}$ the equivalent quaternion representation of $\mathbf{R}_{\hat{W}}$.

The modified PSR method was used to estimate the transformation from the body frames B and $\hat{B}$ to calibrated frames C and $\hat{C}$ that are independent of the orientation of the head mount on the subject's head. In the case of the T265, this calibration was achieved by calculating the rotation that centers heading direction along the longitudinal x-axis and gravity direction along the vertical z-axis. For this, we first computed the representation of the gravity vector ${}^{\hat{W}}\mathbf{g}$ - a unit length vector pointing upwards in the

vertical direction in world coordinates - as well as the linear velocity vector in the body frame $\hat{B}$:

$$^{\hat{B}}\mathbf{g} = \mathrm{rot}\left(^{\hat{W}}\hat{\mathbf{q}}, {}^{\hat{W}}\mathbf{g}\right) \tag{B.3a}$$

$$^{\hat{B}}\hat{\mathbf{v}} = \mathrm{rot}\left(^{\hat{W}}\hat{\mathbf{q}}, {}^{\hat{W}}\hat{\mathbf{v}}\right) \tag{B.3b}$$

The modified PSR algorithm on the sets

$$\mathcal{X} = \left\{ {}^{\hat{B}}\hat{\mathbf{v}}_i \;\middle|\; i \in 1..n \right\} \cup \left\{ {}^{\hat{B}}\mathbf{g}_i \;\middle|\; i \in 1..n \right\} \tag{B.4a}$$

$$\mathcal{Y} = \left\{ [\|\hat{\mathbf{v}}_i\|, 0, 0]^\top \;\middle|\; i \in 1..n \right\} \cup \left\{ [0, 0, \|\mathbf{g}_i\|]^\top \;\middle|\; i \in 1..n \right\} \tag{B.4b}$$

yields the rotation matrix $\mathbf{R}_{\hat{C}}$ and its equivalent quaternion representation $\mathbf{q}_{\hat{C}}$ which was used to transform the linear and angular velocity as well as the gravity vector measured by the T265 into $\hat{C}$:

$$^{\hat{C}}\hat{\mathbf{v}} = \mathrm{rot}\left(^{\hat{W}}\hat{\mathbf{q}} \cdot \mathbf{q}_{\hat{C}}, {}^{\hat{W}}\hat{\mathbf{v}}\right) \tag{B.5a}$$

$$^{\hat{C}}\hat{\boldsymbol{\omega}} = \mathrm{rot}\left(^{\hat{W}}\hat{\mathbf{q}} \cdot \mathbf{q}_{\hat{C}}, {}^{\hat{W}}\hat{\boldsymbol{\omega}}\right) \tag{B.5b}$$

$$^{\hat{C}}\hat{\mathbf{g}} = \mathrm{rot}\left(^{\hat{W}}\hat{\mathbf{q}} \cdot \mathbf{q}_{\hat{C}}, {}^{\hat{W}}\mathbf{g}\right) \tag{B.5c}$$

The OTS was calibrated by aligning linear velocity and gravity direction to those measured by the T265 in its calibrated frame by running the modified PSR on the sets

$$\mathcal{X} = \left\{ {}^{\hat{B}}\hat{\mathbf{v}}_i \;\middle|\; i \in 1..n \right\} \cup \left\{ {}^{\hat{B}}\mathbf{g}_i \;\middle|\; i \in 1..n \right\} \tag{B.6a}$$

$$\mathcal{Y} = \left\{ {}^{B}\mathbf{v}_i \;\middle|\; i \in 1..n \right\} \cup \left\{ {}^{B}\mathbf{g}_i \;\middle|\; i \in 1..n \right\} \tag{B.6b}$$

to obtain the rotation matrix $\mathbf{R}_C$ and its equivalent quaternion representation $\mathbf{q}_C$. With this, we transformed the linear and angular velocity of the marker tracked by the OTS as well as the gravity vector into C:

$$^{C}\mathbf{v} = \mathrm{rot}\left(^{W}\mathbf{q} \cdot \mathbf{q}_C, {}^{W}\mathbf{v}\right) \tag{B.7a}$$

$$^{C}\boldsymbol{\omega} = \mathrm{rot}\left(^{W}\mathbf{q} \cdot \mathbf{q}_C, {}^{W}\boldsymbol{\omega}\right) \tag{B.7b}$$

$$^{C}\mathbf{g} = \mathrm{rot}\left(^{W}\mathbf{q} \cdot \mathbf{q}_C, {}^{W}\mathbf{g}\right) \tag{B.7c}$$

All estimations and transformations were performed in Python 3.6 using the `rigid-body-motion` library (version 0.3.0, `https://github.com/phausamann/rigid-body-motion`).

C. Supplementary material for chapter 4

C.1. Estimation of the direction of gravity from IMU data

For the locomotion data, the proposed filter approach - termed gravity filter (GF) - showed a similar performance to the complementary filter (CF), with a minimum root mean square (RMS) orientation error of 7.2° at a low-pass cut-off frequency of f_{LP} = 0.75 Hz and a correction factor of $\alpha = 0.1$. The influence of α on the gravity filter's performance was small, although higher factors resulted in higher errors. The choice of f_{LP} had a strong effect; errors increased with higher cut-off frequencies up to 9° at 5 Hz (fig. C.1A and B). For the high acceleration data, errors were significantly higher, in the case of the GF with a minimum of 51.5° at f_{LP} = 1 Hz and α = 0.8. Errors decreased with higher values of and exhibited a flat minimum for values of f_{LP} between 1 and 2 Hz for both GF and CF (fig. C.1C and D). A notable exception is the case of α = 0.9 that resulted in higher overall errors for the GF but lower errors for the CF, even with increasing f_{LP}. Regarding execution speed, the GF implementation outperformed the CF implementation by a factor of 4, requiring about 3 s per hour of recording compared with 12 s in the case of the CF.

As a result, we used the GF with f_{LP} = 1 Hz and α = 0.8 to transform the raw accelerometer and gyroscope data before further analysis. This parametrization provided a trade-off between the two motion scenarios in our test data. It should be noted that high acceleration scenario contains data that is very unlikely to occur during natural human locomotion and the corresponding high orientation errors are not a cause for concern.

C.2. Step detection

The relationship between RMS vertical trunk acceleration $A_{T,V}$ and the difference between predominant frequencies in in vertical (V) and medial/lateral (ML) direction Δf_{dom} for the different activities in the dataset is shown in figure C.2. Peaks detected during cycling were clustered around Δf_{dom} = -2.5 Hz and $A_{T,V}$ = 0.2 g. Most locomotion segments exhibited a non-negative Δf_{dom} except for some stair walking data which in turn showed values of $A_{T,V}$ greater than 0.4 g. On the basis of these results we classified all segments with both $\Delta f_{dom} < 0$ Hz and $A_{T,V} < 0.4$ g as cycling

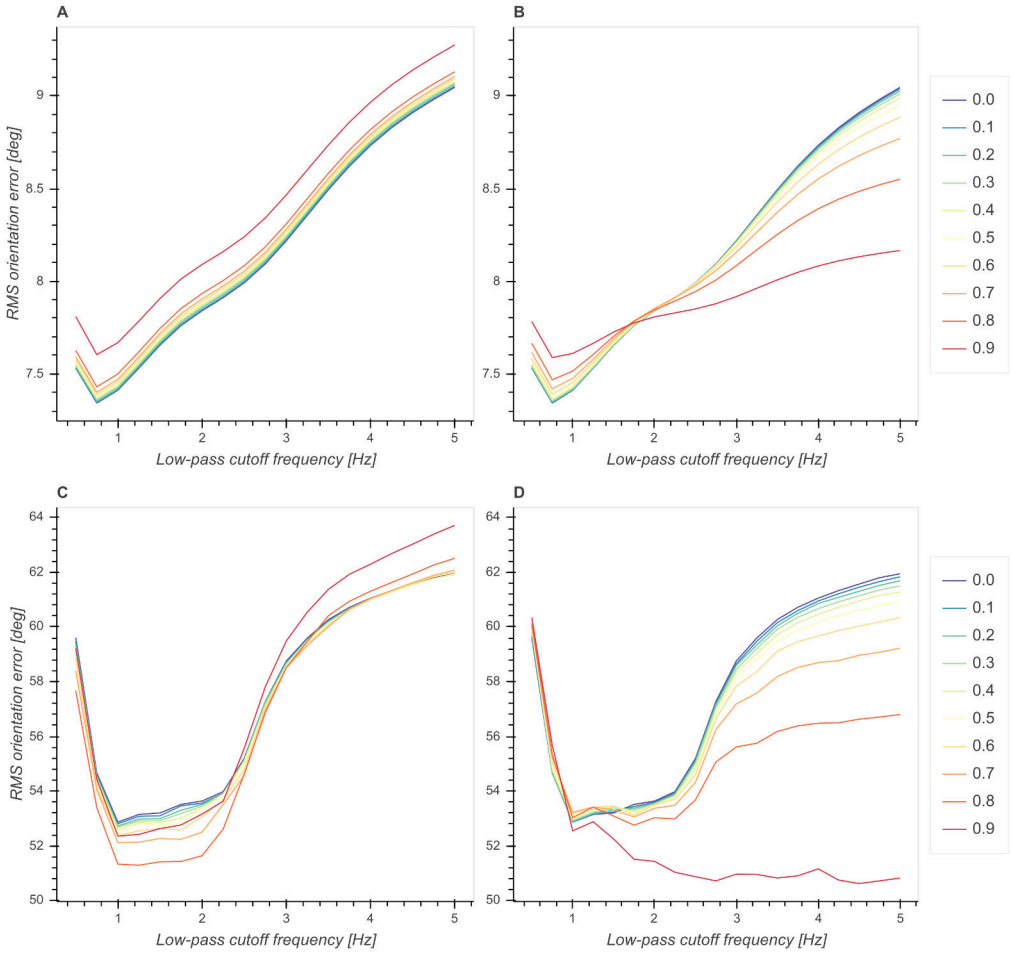

Figure C.1.: Effect of low-pass cut-off frequency f_{LP} and correction factor α (colored lines) on RMS orientation error of **(A)** gravity filter estimate for locomotion data; **(B)** complementary filter estimate for locomotion data; **(C)** gravity filter estimate for high acceleration data; **(D)** complementary filter estimate for high acceleration data

and excluded them from further analysis. This rule eliminated over 80 % of incorrectly identified steps in the test dataset while rejecting less than 2.5 % of actual walking and running steps. Since the number of peaks detected during cycling was low to begin with, this approach detected only 52 steps in a 5 minute period of cycling, less than 2 % of the number of steps occurring during a walking bout of the same length. In the analyzed dataset, we detected a total of 36755 steps.

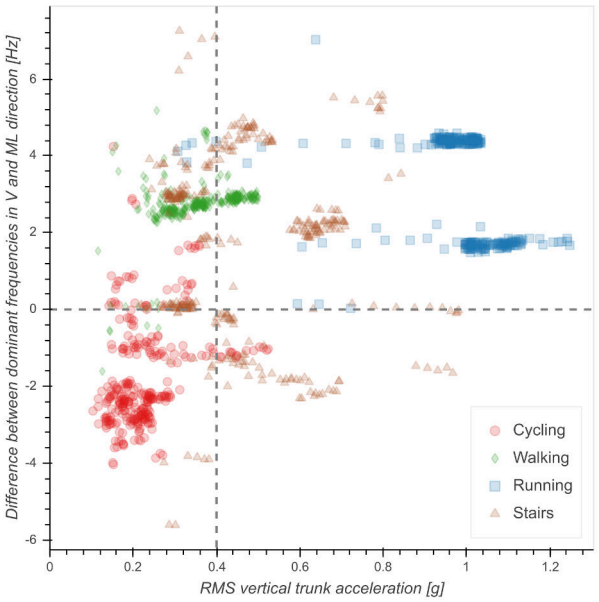

Figure C.2.: Relationship between difference of dominant frequency Δf_{dom} and RMS vertical trunk acceleration $A_{T,V}$ for the four types of annotated activities (cycling, walking, running, stair walking). A small random offset was added to each value of Δf_{dom} in order to make clusters in the data easier to distinguish.

Acronyms

2MWT two-minute walking test

2D two-dimensional

3D three-dimensional

6MWT six-minute walking test

10mWT ten-meter walking test

AC attenuation coefficient

ANOVA analysis of variance

AP anterior/posterior

AR augmented reality

ASD autism spectrum disorder

ATE absolute translation error

AVDE angular velocity direction error

AVME angular velocity magnitude error

AYE absolute yaw error

CF complementary filter

DOF degree of freedom

DV dorsoventral

EKF extended Kalman filter

EMA ecological momentary assessment

FOV field of view

FPS frames per second

GDE gravity direction error

GF	gravity filter
GPS	global positioning system
HDE	heading direction error
HR	harmonic ratio
IMU	inertial measurement unit
IQR	interquartile range
KDE	kernel density estimate
LIDAR	light detection and ranging
LVME	linear velocity magnitude error
MEMS	microelectromechanical system
ML	medial/lateral
MS	multiple sclerosis
OTS	optical tracking system
PD	Parkinson's disease
PSR	point set registration
QoL	quality of life
RMS	root mean square
ROS	robot operating system
RTE	relative translation error
RYE	relative yaw error
SLAM	simultaneous localization and mapping
SpE	speed error
TDr	translation drift
TLE	trajectory length error
TUM	Technical University of Munich

UAS unmanned aerial system

V vertical

VI-SLAM visual-inertial simultaneous localization and mapping

VOR vestibulo-ocular reflex

rVOR rotational vestibulo-ocular reflex

tVOR translational VOR

VR virtual reality

w.r.t. with respect to

YDr yaw drift

Bibliography

[1] A. Agarwal, J. R. Crouse, and E. N. Johnson. “Evaluation of a Commercially Available Autonomous Visual Inertial Odometry Solution for Indoor Navigation”. In: *2020 International Conference on Unmanned Aircraft Systems, ICUAS 2020* (2020), pp. 372–381. DOI: 10.1109/ICUAS48674.2020.9213962.

[2] G. Aigner, B. Grimm, C. Lederer, and M. Daumer. “Method to collect ground truth data for walking speed in real-world environments: description and validation”. In: *PeerJ Preprints* 7 (2019), e27558v1. ISSN: 2167-9843. DOI: 10.7287/peerj.preprints.27558. URL: https://doi.org/10.7287/peerj.preprints.27558v1.

[3] A. Alapetite, Z. Wang, J. P. Hansen, M. Zajaçzkowski, and M. Patalan. “Comparison of three off-the-shelf visual odometry systems”. In: *Robotics* 8(3) (July 2020), p. 56. ISSN: 22186581. DOI: 10.3390/ROBOTICS9030056. URL: https://www.mdpi.com/2218-6581/9/3/56.

[4] H. Albrecht, C. Wötzel, L. P. Erasmus, M. Kleinpeter, N. König, and W. Pöllmann. “Day-to-day variability of maximum walking distance in MS patients can mislead to relevant changes in the expanded disability status scale (EDSS): Average walking speed is a more constant parameter”. In: *Multiple Sclerosis* 7(2) (2001), pp. 105–109. ISSN: 13524585. DOI: 10.1191/135245801678227621.

[5] K. Aminian and B. Najafi. “Capturing human motion using body-fixed sensors: Outdoor measurement and clinical applications”. In: *Computer Animation and Virtual Worlds* 15(2) (2004), pp. 79–94. ISSN: 15464261. DOI: 10.1002/cav.2.

[6] T. P. Andriacchi and E. J. Alexander. “Studies of human locomotion: Past, present and future”. In: *Journal of Biomechanics* 33(10) (2000), pp. 1217–1224. ISSN: 00219290. DOI: 10.1016/S0021-9290(00)00061-0.

[7] D. E. Angelaki and K. E. Cullen. “Vestibular System: The Many Facets of a Multimodal Sense”. In: *Annual Review of Neuroscience* 31(1) (2008), pp. 125–150. ISSN: 0147-006X. DOI: 10.1146/annurev.neuro.31.060407.125555. URL: http://www.annualreviews.org/doi/10.1146/annurev.neuro.31.060407.125555.

[8] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller. “State of the art of virtual reality technology”. In: *IEEE Aerospace Conference Proceedings* 2016-June (2016). ISSN: 1095323X. DOI: 10.1109/AERO.2016.7500674.

[9] A. Anzulewicz, K. Sobota, and J. T. Delafield-Butt. “Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism”. In: *Scientific Reports* 6(July) (2016), pp. 1–13. ISSN: 20452322. DOI: 10.1038/srep31107. URL: http://dx.doi.org/10.1038/srep31107.

[10] V. Apostolyuk. “Theory and Design of Micromechanical Vibratory Gyroscopes”. In: *MEMS/NEMS*. 2007, pp. 173–195. DOI: 10.1007/0-387-25786-1_6.

[11] J. S. Babcock and J. B. Pelz. “Building a lightweight eyetracking headgear”. In: *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA’2004*. New York, New York, USA: ACM Press, 2004, pp. 109–114. ISBN: 1581138253. DOI: 10.1145/968363.968386. URL: http://portal.acm.org/citation.cfm?doid=968363.968386.

[12] M. S. Bahraini, A. B. Rad, and M. Bozorg. “SLAM in dynamic environments: A deep learning approach for moving object tracking using ML-RANSAC algorithm”. In: *Sensors (Switzerland)* 19(17) (Sept. 2019). ISSN: 14248220. DOI: 10.3390/s19173699.

[13] T. Bailey and H. Durrant-Whyte. “Simultaneous localization and mapping (SLAM): Part II”. In: *IEEE Robotics and Automation Magazine* 13(3) (2006), pp. 108–117. ISSN: 10709932. DOI: 10.1109/MRA.2006.1678144.

[14] G. R. Barnes. “Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets”. In: *Journal of Physiology* 287 (1979), pp. 121–147. URL: papers3://publication/uuid/BC7BF009-9E2A-4CF7-BA53-E73EBCB8C000.

[15] A. E. Bartz. “Eye and Head Movements in Peripheral Vision: Nature of Compensatory Eye Movements”. In: *Science* 152(3729) (1966), pp. 1644–1645.

[16] J. Bayer and J. Faigl. “On Autonomous Spatial Exploration with Small Hexapod Walking Robot using Tracking Camera Intel RealSense T265”. In: *2019 European Conference on Mobile Robots (ECMR)*. IEEE, Sept. 2019, pp. 1–6. ISBN: 978-1-7281-3605-9. DOI: 10.1109/ECMR.2019.8870968. URL: https://ieeexplore.ieee.org/document/8870968/.

[17] J. L. Bellanca, K. A. Lowry, J. M. VanSwearingen, J. S. Brach, and M. S. Redfern. “Harmonic ratios: A quantification of step to step symmetry”. In: *Journal of Biomechanics* 46(4) (2013), pp. 828–831. ISSN: 00219290. DOI: 10.1016/j.jbiomech.2012.12.008. arXiv: 15334406. URL: http://dx.doi.org/10.1016/j.jbiomech.2012.12.008.

[18] K. Bonnen, J. S. Matthis, A. Gibaldi, M. S. Banks, D. Levi, and M. Hayhoe. “A role for stereopsis in walking over complex terrains”. In: *Journal of Vision* 19(10) (178b 2019).

[19] G. Bradski. “The OpenCV Library”. In: *Dr. Dobb’s Journal of Software Tools* (2000).

[20] M. A. Brodie, M. J. Coppens, A. Ejupi, Y. J. Gschwind, J. Annegarn, D. Schoene, R. Wieching, S. R. Lord, and K. Delbaere. “Comparison between clinical gait and daily-life gait assessments of fall risk in older people”. In: *Geriatrics and Gerontology International* 17(11) (2017), pp. 2274–2282. ISSN: 14470594. DOI: 10.1111/ggi.12979.

[21] C. Buckley, B. Galna, L. Rochester, and C. Mazzà. “Attenuation of Upper Body Accelerations during Gait: Piloting an Innovative Assessment Tool for Parkinson’s Disease”. In: *BioMed Research International* 2015 (2015), pp. 1–6. ISSN: 2314-6133. DOI: 10.1155/2015/865873. URL: http://www.hindawi.com/journals/bmri/2015/865873/.

[22] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh. “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43(1) (2021), pp. 172–186. ISSN: 19393539. DOI: 10.1109/TPAMI.2019.2929257. arXiv: 1812.08008.

[23] J. Carriot, M. Jamali, M. J. Chacron, and K. E. Cullen. “Statistics of the Vestibular Input Experienced during Natural Self-Motion: Implications for Neural Processing”. In: *Journal of Neuroscience* 34(24) (June 2014), pp. 8347–8357. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.0692-14.2014. URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0692-14.2014.

[24] J. Carriot, M. Jamali, K. E. Cullen, and M. J. Chacron. “Envelope statistics of self-motion signals experienced by human subjects during everyday activities: Implications for vestibular processing”. In: *PLOS ONE* 12(6) (June 2017). Ed. by M. Lappe, e0178664. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0178664. URL: http://dx.plos.org/10.1371/journal.pone.0178664.

[25] J. Cassell, J. F. Cohn, J. C. Britton, K. B. Martin, G. Ren, Z. Hammal, D. S. Messinger, A. Gutierrez, and M. Ogihara. “Objective measurement of head movement differences in children with and without autism spectrum disorder”. In: *Molecular Autism* 9(1) (2018), pp. 1–10. DOI: 10.1186/s13229-018-0198-4.

[26] J. T. Cavanaugh, T. D. Ellis, G. M. Earhart, M. P. Ford, K. B. Foreman, and L. E. Dibble. “Capturing ambulatory activity decline in parkinson’s disease”. In: *Journal of Neurologic Physical Therapy* 36(2) (June 2012), pp. 51–57. ISSN: 15570584. DOI: 10.1097/NPT.0b013e318254ba7a. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf%20http://www.ncbi.nlm.nih.gov/pubmed/22592060%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3934648.

[27] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. “Mobile Augmented Reality Survey: From Where We Are to Where We Go”. In: *IEEE Access* 5 (2017), pp. 6917–6950. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2698164.

[28] C. Chen, H. Zhu, M. Li, and S. You. “A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives”. In: *Robotics* 7 (45 July 2018), pp. 1–20.

[29] B. Cinaz and H. Kenn. “HeadSLAM - simultaneous localization and mapping with head-mounted inertial and laser range sensors”. In: *2008 12th IEEE International Symposium on Wearable Computers*. (January 2008). IEEE, 2008, pp. 3–10. ISBN: 978-1-4244-2637-9. DOI: 10.1109/ISWC.2008.4911575. URL: http://ieeexplore.ieee.org/document/4911575/.

[30] J. Cohen. *Statistical power analysis for the behavioral sciences*. 2nd. Taylor & Francis Inc, 1988. ISBN: 9780805802832.

[31] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. Salo. “A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System”. In: *Sports Medicine - Open* 4(1) (2018). ISSN: 21989761. DOI: 10.1186/s40798-018-0139-y.

[32] B. T. Crane and J. L. Demer. “Human gaze stabilization during natural activities: Translation, rotation, magnification, and target distance effects”. In: *Journal of Neurophysiology* 78(4) (1997), pp. 2129–2144. ISSN: 00223077. DOI: 10.1152/jn.1997.78.4.2129.

[33] J. L. Crassidis, F. Landis Markley, and Y. Cheng. “Survey of nonlinear attitude estimation methods”. In: *Journal of Guidance, Control, and Dynamics* 30(1) (2007), pp. 12–28. ISSN: 15333884. DOI: 10.2514/1.22452.

[34] R. L. Cromwell, R. A. Newton, and G. Forrest. “Head stability in older adults during walking with and without visual input”. In: *Journal of Vestibular Research* 11(2) (2001), pp. 105–114.

[35] R. L. Cromwell, R. A. Newton, and L. G. Carlton. “Horizontal plane head stabilization during locomotor tasks”. In: *Journal of Motor Behavior* 33(1) (2001), pp. 49–58. ISSN: 00222895. DOI: 10.1080/00222890109601902.

[36] H. Durrant-Whyte and T. Bailey. “Simultaneous localization and mapping: Part I”. In: *IEEE Robotics and Automation Magazine* 13(2) (2006), pp. 99–108. ISSN: 10709932. DOI: 10.1109/MRA.2006.1638022.

[37] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König. “Human eye-head co-ordination in natural exploration”. In: *Network: Computation in Neural Systems* 18 (3 Sept. 2007), pp. 267–297.

[38] W. Fang, L. Zheng, H. Deng, and H. Zhang. “Real-Time Motion Tracking for Mobile Augmented/Virtual Reality Using Adaptive Visual-Inertial Fusion”. In: *Sensors* 17(5) (May 2017), p. 1037. ISSN: 1424-8220. DOI: 10.3390/s17051037. URL: http://www.mdpi.com/1424-8220/17/5/1037.

[39] L. N. Ferreira, L. N. Pereira, M. da Fé Brás, and K. Ilchuk. “Quality of life under the COVID-19 quarantine”. In: *Quality of Life Research* (November 2020) (2021). ISSN: 15732649. DOI: 10.1007/s11136-020-02724-x. URL: https://doi.org/10.1007/s11136-020-02724-x.

[40] D. J. Field. “What Is the Goal of Sensory Coding?” In: *Neural Computation* 6(4) (1994), pp. 559–601. ISSN: 0899-7667. DOI: 10.1162/neco.1994.6.4.559. URL: http://www.mitpressjournals.org/doi/10.1162/neco.1994.6.4.559.

[41] E. Foxlin. “Pedestrian Tracking with Shoe-Mounted Inertial Sensors”. In: *IEEE Computer Graphics and Applications* 25(6) (Nov. 2005), pp. 38–46. ISSN: 0272-1716. DOI: 10.1109/MCG.2005.140. URL: http://www.computer.org/portal/web/csdl/doi/10.1109/MCG.2005.140%20http://ieeexplore.ieee.org/document/1528431/.

[42] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha. “Visual simultaneous localization and mapping: a survey”. In: *Artificial Intelligence Review* 43(1) (2012), pp. 55–81. ISSN: 15737462. DOI: 10.1007/s10462-012-9365-8.

[43] J. Gao, A. Huth, and Y. Park. *Case Forge*. 2017. URL: https://caseforge.co/.

[44] D. L. Gill, C. C. Hammond, E. J. Reifsteck, C. M. Jehu, R. A. Williams, M. M. Adams, E. H. Lange, K. Becofsky, E. Rodriguez, and Y. T. Shang. "Physical activity and quality of life". In: *Journal of Preventive Medicine and Public Health* 46(SUPPL.1) (2013), pp. 28–34. ISSN: 19758375. DOI: 10.3961/jpmph.2013.46.S.S28.

[45] S. Glasauer. "Interaction of Semicircular Canals and Otoliths in the Processing Structure of the Subjective Zenith". In: *Annals of the New York Academy of Sciences* 656(1 Sensing and C) (May 1992), pp. 847–849. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.1992.tb25272.x. URL: http://doi.wiley.com/10.1111/j.1749-6632.1992.tb25272.x.

[46] A. Grunnet-Jepsen, M. Harville, B. Fulkerson, D. Piro, S. Brook, and J. Radford. *Introduction to Intel® RealSense™ Visual SLAM and the T265 Tracking Camera*. 2019. URL: https://dev.intelrealsense.com/docs/intel-realsensetm-visual-slam-and-the-t265-tracking-camera.

[47] W. R. Hamilton. *Elements of quaternions*. Longmans, Green, & Company, 1866. URL: https://archive.org/details/elementsofquater00hamiuoft.

[48] P. Hausamann, M. Daumer, P. R. MacNeilage, and S. Glasauer. "Ecological momentary assessment of head motion: Toward normative data of head stabilization". In: *Frontiers in Human Neuroscience* 13(June) (2019), pp. 1–13. ISSN: 16625161. DOI: 10.3389/fnhum.2019.00179. URL: https://www.frontiersin.org/article/10.3389/fnhum.2019.00179/full.

[49] C. Heesen, J. Böhm, C. Reich, J. Kasper, M. Goebel, and S. M. Gold. "Patient perception of bodily functions in multiple sclerosis: Gait and visual function are the most valuable". In: *Multiple Sclerosis* 14(7) (2008), pp. 988–991. ISSN: 13524585. DOI: 10.1177/1352458508088916.

[50] J. Hesch, A. Kozminksi, and O. Linde. *Powered by AI: Oculus Insight*. 2019. URL: https://ai.facebook.com/blog/powered-by-ai-oculus-insight/ (visited on 03/31/2021).

[51] E. Hirasaki, S. T. Moore, T. Raphan, and B. Cohen. "Effects of walking velocity on vertical head and body movements during locomotion". In: *Experimental Brain Research* 127(2) (July 1999), pp. 117–130. ISSN: 00144819. DOI: 10.1007/s002210050781. URL: http://link.springer.com/10.1007/s002210050781.

[52] K. Holmqvist and R. Andersson. *Eye-tracking: A comprehensive guide to methods, paradigms and measures*. Nov. 2017. ISBN: ISBN-13: 978-1979484893.

[53] N. E. Huang, Z. Shen, S. R. Long, M.-L. C. Wu, H. H. Snin, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 454(1971) (Mar. 1998), pp. 903–995. ISSN: 13645021. DOI: 10.1098/rspa.1998.0193. URL: http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1998.0193.

[54] D. Q. Huynh. “Metrics for 3D rotations: Comparison and analysis”. In: *Journal of Mathematical Imaging and Vision* 35(2) (2009), pp. 155–164. ISSN: 09249907. DOI: 10.1007/s10851-009-0161-2.

[55] M. Iosa, P. Picerno, S. Paolucci, and G. Morone. “Wearable inertial sensors for human movement analysis”. In: *Expert Review of Medical Devices* 13(7) (2016), pp. 641–659. ISSN: 17452422. DOI: 10.1080/17434440.2016.1198694.

[56] W. Jeong and K. M. Lee. “CV-SLAM: A new ceiling vision-based SLAM technique”. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2005), pp. 1–6.

[57] E. Kasneci, K. Sippel, K. Aehling, M. Heister, W. Rosenstiel, U. Schiefer, and E. Papageorgiou. “Driving with Binocular Visual Field Loss? A Study on a Supervised On-Road Parcours with Simultaneous Eye and Head Tracking”. In: *PLoS ONE* 9(2) (Feb. 2014). Ed. by P. Bex, e87470. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0087470. URL: https://dx.plos.org/10.1371/journal.pone.0087470.

[58] M. Kassner, W. Patera, and A. Bulling. “Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction”. In: *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2014), pp. 1151–1160. DOI: 10.1145/2638728.2641695. arXiv: 1405.0006.

[59] M. Kassner, W. Patera, and A. Bulling. “Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction”. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Sept. 2014), pp. 1151–1160.

[60] J. J. Kavanagh, S. Morrison, and R. S. Barrett. “Coordination of head and trunk accelerations during walking”. In: *European Journal of Applied Physiology* 94(4) (July 2005), pp. 468–475. ISSN: 1439-6319. DOI: 10.1007/s00421-005-1328-1. URL: http://link.springer.com/10.1007/s00421-005-1328-1.

[61] J. J. Kavanagh and H. B. Menz. “Accelerometry: A technique for quantifying movement patterns during walking”. In: *Gait and Posture* 28(1) (2008), pp. 1–15. ISSN: 09666362. DOI: 10.1016/j.gaitpost.2007.10.010.

[62] T. Kinsman, K. Evans, G. Sweeny, T. Keane, and J. Pelz. “Ego-motion compensation improves fixation detection in wearable eye tracking”. In: *Proceedings of the Symposium for Eye Tracking Research and Applications* (Mar. 2012), pp. 221–224.

[63] M. Kok, J. D. Hol, and T. B. Schön. “Using Inertial Sensors for Position and Orientation Estimation”. In: *Foundations and Trends® in Signal Processing* 11(1-2) (2017), pp. 1–153. ISSN: 1932-8346. DOI: 10.1561/2000000094. arXiv: 1704.06053. URL: http://arxiv.org/abs/1704.06053%7B%5C%%7D0Ahttp://dx.doi.org/10.1561/2000000094%20http://www.nowpublishers.com/article/Details/SIG-094.

[64] R. Kothari, Z. Yanh, C. Kanan, R. Bailey, J. Pelz, and G. Diaz. “Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities”. In: *arXiv* (May 2019).

[65] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. “Activity recognition using cell phone accelerometers”. In: *ACM SIGKDD Explorations Newsletter* 12(2) (2011), p. 74. ISSN: 19310145. DOI: 10.1145/1964897.1964918. arXiv: arXiv:1501.01199v1. URL: http://portal.acm.org/citation.cfm?doid=1964897.1964918.

[66] L. Larsson, M. Nyström, A. Schwaller, M. Stridh, and K. Holmqvist. “Compensation of head movements in mobile eye-tracking data using an inertial measurement unit”. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Sept. 2014), pp. 1161–1167.

[67] M. D. Latt, H. B. Menz, V. S. Fung, and S. R. Lord. “Acceleration patterns of the head and pelvis during gait in older people with Parkinson’s disease: A comparison of fallers and nonfallers”. In: *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 64(6) (2009), pp. 700–706. ISSN: 10795006. DOI: 10.1093/gerona/glp009.

[68] M. D. Latt, H. B. Menz, V. S. Fung, and S. R. Lord. “Walking speed, cadence and step length are selected to optimize the stability of head and pelvis accelerations”. In: *Experimental Brain Research* 184(2) (2008), pp. 201–209. ISSN: 00144819. DOI: 10.1007/s00221-007-1094-x.

[69] L. Laudani, A. Casabona, V. Perciavalle, and A. Macaluso. “Control of head stability during gait initiation in young and older women”. In: *Journal of Electromyography and Kinesiology* 16(6) (2006), pp. 603–610. ISSN: 10506411. DOI: 10.1016/j.jelekin.2006.08.001.

[70] F. Li, J. B. Pelz, and S. J. Daly. “Measuring hand, head and vehicle motions in commuting environments”. In: *Proceedings of SPIE* 7240 (2009), pp. 7240-1-7240–11.

[71] W. Li and J. Wang. “Effective adaptive Kalman filter for MEMS-IMU/Magnetometers integrated attitude and heading reference systems”. In: *The Journal of Navigation* 66 (1 2012), pp. 99–113.

[72] W. Liu, D. Caruso, E. Ilg, J. Dong, A. I. Mourikis, K. Daniilidis, V. Kumar, and J. Engel. “TLIO: Tight Learned Inertial Odometry”. In: *IEEE Robotics and Automation Letters* 5(4) (2020), pp. 5653–5660. ISSN: 23773766. DOI: 10.1109/LRA.2020.3007421. arXiv: 2007.01867.

[73] K. A. Lowry, N. Lokenvitz, and A. L. Smiley-Oyen. “Age- and speed-related differences in harmonic ratios during walking”. In: *Gait and Posture* 35(2) (2012), pp. 272–276. ISSN: 09666362. DOI: 10.1016/j.gaitpost.2011.09.019. URL: http://dx.doi.org/10.1016/j.gaitpost.2011.09.019.

[74] K. A. Lowry, A. L. Smiley-Oyen, A. J. Carrel, and J. P. Kerr. “Walking stability using harmonic ratios in Parkinson’s disease”. In: *Movement Disorders* 24(2) (2009), pp. 261–267. ISSN: 08853185. DOI: 10.1002/mds.22352.

[75] H. J. Luinge and P. H. Veltink. “Measuring orientation of human body segments using miniature gyroscopes and accelerometers”. In: *Medical and Biological Engineering and Computing* 43(2) (2005), pp. 273–282. ISSN: 01400118. DOI: 10.1007/BF02345966.

[76] H. G. MacDougall. “Marching to the beat of the same drummer: the spontaneous tempo of human locomotion”. In: *Journal of Applied Physiology* 99(3) (2005), pp. 1164–1173. ISSN: 8750-7587. DOI: 10.1152/japplphysiol.00138.2005. URL: http://jap.physiology.org/cgi/doi/10.1152/japplphysiol.00138.2005.

[77] P. MacNeilage. “Characterization of Natural Head Movements in Animals and Humans”. In: *The Senses: A Comprehensive Reference*. Ed. by B. Fritsch and H. Straka. Vol. 6. Elsevier, Academic Press, 2020, pp. 69–87.

[78] P. R. MacNeilage and S. Glasauer. “Quantification of Head Movement Predictability and Implications for Suppression of Vestibular Input during Locomotion”. In: *Frontiers in Computational Neuroscience* 11(0) (June 2017). ISSN: 1662-5188. DOI: 10.3389/fncom.2017.00047. URL: http://journal.

frontiersin.org/article/10.3389/fncom.2017.00047/full.

[79] P. R. MacNeilage, L. Nguyen, and C. Sinnott. "Characterization of natural head and eye movements driving retinal flow". In: *Journal of Vision* 19(10) (147d 2019).

[80] M. D. Malinzak, R. F. Kay, and T. E. Hullar. "Locomotor head movements and semicircular canal morphology in primates". In: *Proceedings of the National Academy of Sciences of the United States of America* 109(44) (2012), pp. 17914–17919. ISSN: 00278424. DOI: 10.1073/pnas.1206139109.

[81] A. Maslivec, T. M. Bampouras, S. Dewhurst, G. Vannozzi, A. Macaluso, and L. Laudani. "Mechanisms of head stability during gait initiation in young and older women: A neuro-mechanical analysis". In: *Journal of Electromyography and Kinesiology* 38(November 2017) (2018), pp. 103–110. ISSN: 18735711. DOI: 10.1016/j.jelekin.2017.11.010. URL: https://doi.org/10.1016/j.jelekin.2017.11.010.

[82] J. S. Matthis, K. S. Muller, and M. M. Hayhoe. "Retinal optic flow and the control of locomotion". In: *Journal of Vision* 19(10) (179 2019).

[83] J. S. Matthis, J. L. Yates, and M. M. Hayhoe. "Gaze and the control of foot placement when walking in natural terrain". In: *Current Biology* 28 (2018), pp. 1224–1233.

[84] A. Mayerhoffer and P. MacNeilage. "Natural Statistics of Vestibular Stimulation During Human Locomotion". In: *Neuroscience Meeting Planner* (580.20) (2011).

[85] W. W. Mayol, A. J. Davison, B. J. Tordoff, and D. W. Murray. "Applying Active Vision and SLAM to Wearables". In: *Springer Tracts in Advanced Robotics*. Vol. 15. 2005, pp. 325–334. DOI: 10.1007/11008941_35. URL: http://link.springer.com/10.1007/11008941%7B%5C_%7D35.

[86] C. Mazzà, M. Iosa, P. Picerno, and A. Cappozzo. "Gender differences in the control of the upper body accelerations during level walking". In: *Gait & Posture* 29(2) (Feb. 2009), pp. 300–303. ISSN: 09666362. DOI: 10.1016/j.gaitpost.2008.09.013. URL: http://link.springer.com/10.1007/978-3-540-89208-3%7B%5C_%7D12%20https://linkinghub.elsevier.com/retrieve/pii/S096663620800297X.

[87] V. Medved. *Measurement of human locomotion*. CRC press, 2000.

[88] A. Menache. *Understanding motion capture for computer animation*. Elsevier, 2011.

[89] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick. “Acceleration patterns of the head and pelvis when walking on level and irregular surfaces”. In: *Gait & Posture* 18(1) (Aug. 2003), pp. 35–46. ISSN: 09666362. DOI: 10.1016/S0966-6362(02)00159-5. URL: http://toronto.cmha.ca/%20http://linkinghub.elsevier.com/retrieve/pii/S0966636202001595.

[90] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman. “Fast gaze-contingent optimal decompositions for multifocal displays”. In: *ACM Transactions on Graphics*. 2017. DOI: 10.1145/3130800.3130846.

[91] R. W. Motl, B. M. Sandroff, and J. J. Sosnoff. “Commercially available accelerometry as an ecologically valid measure of ambulation in individuals with multiple sclerosis”. In: *Expert Review of Neurotherapeutics* 12(9) (Sept. 2012), pp. 1079–1088. ISSN: 1473-7175. DOI: 10.1586/ern.12.74. URL: http://www.tandfonline.com/doi/full/10.1586/ern.12.74.

[92] A. I. Mourikis and S. I. Roumeliotis. “A multi-state constraint Kalman filter for vision-aided inertial navigation”. In: *Proceedings - IEEE International Conference on Robotics and Automation* (2007), pp. 3565–3572. ISSN: 10504729. DOI: 10.1109/ROBOT.2007.364024.

[93] L. Mündermann, S. Corazza, and T. P. Andriacchi. “The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications”. In: *Journal of NeuroEngineering and Rehabilitation* 3 (2006), pp. 1–11. ISSN: 17430003. DOI: 10.1186/1743-0003-3-6.

[94] E. Muybridge. *Animal locomotion*. Da Capo Press, 1887.

[95] F. J. Ordóñez and D. Roggen. “Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition”. In: *Sensors (Switzerland)* 16(1) (2016). ISSN: 14248220. DOI: 10.3390/s16010115.

[96] S. Ouerghi, N. Ragot, R. Boutteau, and X. Savatier. “Comparative study of a commercial tracking camera and ORB-SLAM2 for person localization”. In: *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* 4(Visigrapp) (2020), pp. 357–364. DOI: 10.5220/0008980703570364.

[97] T. Pozzo, A. Berthoz, and L. Lefort. “Head stabilization during various locomotor tasks in humans. I. Normal subjects.” In: *Experimental brain research* 82(1) (1990), pp. 97–106. ISSN: 0014-4819. URL: http://www.ncbi.nlm.nih.gov/pubmed/2257917.

[98] T. Pozzo, A. Berthoz, L. Lefort, and E. Vitte. “Head stabilization during various locomotor tasks in humans. II. Patients with bilateral peripheral vestibular deficits.” In: *Experimental brain research* 85(1) (1991), pp. 208–17. ISSN: 0014-4819. URL: http://www.ncbi.nlm.nih.gov/pubmed/1884759.

[99] T. Pozzo, A. Berthoz, and L. Lefort. *Head kinematic during various motor tasks in humans*. 1989. DOI: 10.1016/S0079-6123(08)62233-5.

[100] M. Psarakis, D. A. Greene, M. H. Cole, S. R. Lord, P. Hoang, and M. Brodie. “Wearable technology reveals gait compensations, unstable walking patterns and fatigue in people with multiple sclerosis”. In: *Physiological Measurement* 39(7) (July 2018), p. 075004. ISSN: 1361-6579. DOI: 10.1088/1361-6579/aac0a3. URL: http://stacks.iop.org/0967-3334/39/i=7/a=075004?key=crossref.daed267e9c9e98e6ac3d546b98b74c1a.

[101] D. Pustka, J. P. Hülß, J. Willneff, F. Pankratz, M. Huber, and G. Klinker. “Optical outside-in tracking using unmodified mobile phones”. In: *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers* (November) (2012), pp. 81–89. DOI: 10.1109/ISMAR.2012.6402542.

[102] M. I. Ribeiro. “Kalman and extended kalman filters: Concept, derivation and properties”. In: *Institute for Systems and Robotics* 43 (2004), p. 46.

[103] D. A. Robinson. “A Method of Measuring Eye Movement Using a Scieral Search Coil in a Magnetic Field”. In: *IEEE transactions on bio-medical engineering* 10(4) (Oct. 1963), pp. 137–45. ISSN: 0018-9294. DOI: 10.1109/tbmel.1963.4322822. URL: http://www.ncbi.nlm.nih.gov/pubmed/14121113.

[104] A. Ruina and R. Pratap. *Introduction to Statics and Dynamics*. 2002.

[105] A. M. Sabatini. “Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing”. In: *IEEE Transactions on Biomedical Engineering* 53(7) (2006), pp. 1346–1356. ISSN: 00189294. DOI: 10.1109/TBME.2006.875664.

[106] T. Santini, H. Brinkmann, L. Reitstätter, H. Leder, R. Rosenberg, W. Rosenstiel, and E. Kasneci. “The art of pervasive eye tracking: Unconstrained eye tracking in the Austrian gallery Belvedere.” In: *PETMEI ’18: 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction* (2018), pp. 1–8.

[107] N. Sarvananthan, M. Surendran, E. O. Roberts, S. Jain, S. Thomas, N. Shah, F. A. Proudlock, J. R. Thompson, R. J. McLean, C. Degg, G. Woodruff, and I. Gottlob. “The Prevalence of Nystagmus: The Leicestershire Nystagmus Survey”. In: *Investigative Opthalmology & Visual Science* 50(11) (Nov. 2009), p. 5201. ISSN: 1552-5783. DOI: 10.1167/iovs.09-3486. URL: http://iovs.

arvojournals.org/article.aspx?doi=10.1167/iovs.09-3486.

[108] D. Scaramuzza and F. Fraundorfer. “Tutorial: Visual odometry”. In: *IEEE Robotics and Automation Magazine* 18(4) (Dec. 2011), pp. 80–92. ISSN: 10709932. DOI: 10.1109/MRA.2011.943233.

[109] M. Schimpl, C. Lederer, and M. Daumer. “Development and validation of a new method to measure walking speed in free-living environments using the actibelt® platform”. In: *PLoS ONE* 6(8) (2011). ISSN: 19326203. DOI: 10.1371/journal.pone.0023080.

[110] M. Schimpl, C. Moore, C. Lederer, A. Neuhaus, J. Sambrook, J. Danesh, W. Ouwehand, and M. Daumer. “Association between Walking Speed and Age in Healthy, Free-Living Individuals Using Mobile Accelerometry—A Cross-Sectional Study”. In: *PLoS ONE* 6(8) (Aug. 2011). Ed. by A. Lucia, e23299. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0023299. URL: https://dx.plos.org/10.1371/journal.pone.0023299.

[111] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart. “maplab: An Open Framework for Research in Visual-inertial Mapping and Localization”. In: (2017). ISSN: 2377-3766. DOI: 10.1109/LRA.2018.2800113. arXiv: 1711.10250. URL: http://arxiv.org/abs/1711.10250%7B%5C%%7D0Ahttp://dx.doi.org/10.1109/LRA.2018.2800113.

[112] S. Shiffman, A. A. Stone, and M. R. Hufford. “Ecological momentary assessment.” In: *Annual review of clinical psychology* 4 (2008), pp. 1–32. ISSN: 1548-5943. URL: http://www.ncbi.nlm.nih.gov/pubmed/18509902.

[113] K. Shoemake. “Quaternion Calculus and Fast Animation, Computer Animation: 3-D Motion Specification and Control”. In: *SIGGRAPH 1987 Tutorial*. Siggraph. 1987, pp. 101–121.

[114] B. W. Silverman. *Density estimation: For statistics and data analysis*. 2018, pp. 1–175. ISBN: 9781351456173. DOI: 10.1201/9781315140919.

[115] E. P. Simoncelli and B. A. Olshausen. “Natural image statistics and neural representation.” In: *Annu Rev Neurosci* 24 (2001), pp. 1193–1216. ISSN: 0147-006X. DOI: 10.1146/annurev.neuro.24.1.1193. URL: http://www.ncbi.nlm.nih.gov/pubmed/11520932.

[116] A. Singandhupe and H. La. “A Review of SLAM Techniques and Security in Autonomous Driving”. In: *Proceedings - 3rd IEEE International Conference on Robotic Computing, IRC 2019* (19) (2019), pp. 602–607. DOI: 10.1109/IRC.2019.00122.

[117] G. L. Smidt, J. S. Arora, and R. C. Johnston. “Accelerographic analysis of several types of walking.” In: *American journal of physical medicine* 50(6) (1971), pp. 285–300. ISSN: 0002-9491.

[118] E. C. Smith and M. S. Lewicki. “Efficient auditory coding”. In: *Nature* 439(7079) (Feb. 2006), pp. 978–982. ISSN: 14764687. DOI: 10.1038/nature04485. arXiv: arXiv:1011.1669v3. URL: http://www.nature.com/articles/nature04485.

[119] A. Sorriento, M. B. Porfido, S. Mazzoleni, G. Calvosa, M. Tenucci, G. Ciuti, and P. Dario. “Optical and Electromagnetic Tracking Systems for Biomedical Applications: A Critical Review on Potentialities and Limitations”. In: *IEEE Reviews in Biomedical Engineering* 13 (2020), pp. 212–232. ISSN: 19411189. DOI: 10.1109/RBME.2019.2939091.

[120] J. P. Stellmann, A. Neuhaus, N. Götze, S. Briken, C. Lederer, M. Schimpl, C. Heesen, and M. Daumer. “Ecological validity of walking capacity tests in multiple sclerosis”. In: *PLoS ONE* 10(4) (2015), pp. 1–11. ISSN: 19326203. DOI: 10.1371/journal.pone.0123822.

[121] L. Świrski and N. A. Dodgson. “A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting”. In: *Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI)* (January) (2013).

[122] D. Tazartes. “An historical perspective on inertial navigation systems”. In: *1st IEEE International Symposium on Inertial Sensors and Systems, ISISS 2014 - Proceedings* (2014), pp. 2–6. DOI: 10.1109/ISISS.2014.6782505.

[123] M. Tomczak and E. Tomczak. “The need to report effect size estimates revisited. An overview of some recommended measures of effect size”. In: *Trends in Sport Sciences* 1(21) (2014), pp. 19–25. ISSN: 22999590. DOI: 10.7326/0003-4819-159-3-201308060-00005.

[124] C. Trevarthen and J. T. Delafield-Butt. “Autism as a developmental disorder in intentional movement and affective engagement”. In: *Frontiers in Integrative Neuroscience* 7(July) (2013), pp. 1–16. ISSN: 1662-5145. DOI: 10.3389/fnint.2013.00049. URL: http://journal.frontiersin.org/article/10.3389/fnint.2013.00049/abstract.

[125] P. O. Ubuane, B. A. Animasahun, O. A. Ajiboye, M. O. Kayode-Awe, O. A. Ajayi, and F. O. Njokanma. “The historical evolution of the six-minute walk test as a measure of functional exercise capacity: a narrative review”. In: *Journal of Xiangya Medicine* 3(3) (2018), pp. 40–40. DOI: 10.21037/jxym.2018.11.01.

[126] N. J. Wade. “Pioneers of eye movement research”. In: *i-Perception* 1(2) (2010), pp. 33–68. ISSN: 20416695. DOI: 10.1068/i0389.

[127] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. “Deep Learning for Sensor-based Activity Recognition: A Survey”. In: (2017), pp. 1–10. ISSN: 01678655. DOI: 10.1016/j.patrec.2018.02.010. arXiv: 1707.03502. URL: http://arxiv.org/abs/1707.03502%7B%5C%%7D0Ahttp://dx.doi.org/10.1016/j.patrec.2018.02.010.

[128] W. E. Weber and E. Weber. *Mechanik der menschlichen Gehwerkzeuge: eine anatomisch-physiologische Untersuchung*. Vol. 1. Dieterich, 1836.

[129] G. Wetzstein. *EE 267 Virtual Reality Course Notes : Orientation Tracking with IMUs*. Tech. rep. (2). 2017.

[130] B. M. Williamson, A. Vargas, P. Garrity, R. Sottilare, and J. J. LaViola. “AgileSLAM: A Localization Approach for Agile Head Movements in Augmented Reality”. In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. (August 2019). IEEE, Oct. 2018, pp. 25–30. ISBN: 978-1-5386-7592-2. DOI: 10.1109/ISMAR-Adjunct.2018.00025. URL: https://ieeexplore.ieee.org/document/8699274/.

[131] M. Wuehr, R. Schniepp, J. Ilmberger, T. Brandt, and K. Jahn. “Speed-dependent temporospatial gait variability and long-range correlations in cerebellar ataxia”. In: *Gait and Posture* 37(2) (2013), pp. 214–218. ISSN: 09666362. DOI: 10.1016/j.gaitpost.2012.07.003. URL: http://dx.doi.org/10.1016/j.gaitpost.2012.07.003.