# On learning pseudo-sensors to improve visual egomotion estimation

by

Valentin Peretroukhin

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Institute for Aerospace Studies
University of Toronto

# Abstract

On learning pseudo-sensors to improve visual egomotion estimation

Valentin Peretroukhin
Doctor of Philosophy
Graduate Department of Institute for Aerospace Studies
University of Toronto
2019

The ability to estimate *egomotion* is at the heart of safe and reliable mobile autonomy. By inferring pose changes from sequential sensor measurements, egomotion estimation forms the basis of mapping and navigation pipelines, and permits mobile robots to self-localize within environments where external localization information may be intermittent or unavailable. Visual egomotion estimation, also known as *visual odometry*, has become ubiquitous in mobile robotics due to the availability of high-quality, compact, and inexpensive cameras that capture rich representations of the world. To remain computationally tractable, 'classical' visual odometry pipelines make simplifying assumptions that, while permitting reliable operation in ideal conditions, often lead to systematic error. In this dissertation, we present several data-driven *pseudo-sensors* that serve to augment conventional pipelines by inferring latent information from sensor data. Our approach retains many of the benefits of traditional pipelines, while leveraging high-capacity hyper-parametric models to extract complementary information that can be used to improve uncertainty quantification, correct for systematic bias, and improve robustness to difficult-to-model deleterious effects. We validate our pseudo-sensors on several kilometres of sensor data collected in sundry settings such as urban roads, indoor labs, and planetary analogue sites in the Canadian High Arctic.

# Epigraph

A little learning is a dangerous thing;
drink deep, or taste not the Pierian
spring: there shallow draughts
intoxicate the brain, and drinking
largely sobers us again.

---

Alexander Pope

The universe is no narrow thing and the order within it is not constrained by any latitude in its conception to repeat what exists in one part in any other part. Even in this world more things exist without our knowledge than with it and the order in creation which you see is that which you have put there, like a string in a maze, so that you shall not lose your way. For existence has its own order and that no man's mind can compass, that mind itself being but a fact among others.

---

Cormac McCarthy

Elephants don't play chess.

---

Rodney Brooks

To all those who encouraged (or, at least, *never discouraged*) my intellectual wanderlust.

# Acknowledgements

This document would not have been possible without the generous support and guidance of my supervisor[1], the perennial love of my family and friends[2], and the limitless patience of my lab mates[3]. Thank you all.

[1] as well as all of my collaborators and academic mentors (special thanks to Lee)
[2] especially the support and encouragement of Elyse
[3] in humouring my insatiable need for debate and banter (special thanks to Lee)

# Contents

# Notation

| | | |
|---|---|---|
| $a$ | : | Symbols in this font are real scalars. |
| $\mathbf{a}$ | : | Symbols in this font are real column vectors. |
| $\boldsymbol{a}$ | : | Symbols in this font are real column vectors in homogeneous coordinates. |
| $\mathbf{A}$ | : | Symbols in this font are real matrices. |
| $\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$ | : | Normally distributed with mean $\boldsymbol{\mu}$ and covariance $\mathbf{R}$. |
| $E[\cdot]$ | : | The expectation operator. |
| $\underrightarrow{\boldsymbol{\mathcal{F}}}_a$ | : | A reference frame in three dimensions. |
| $(\cdot)^\wedge$ | : | An operator associated with the Lie algebra for rotations and poses. It produces a matrix from a column vector. |
| $(\cdot)^\vee$ | : | The inverse operation of $(\cdot)^\wedge$. |
| $\mathbf{1}$ | : | The identity matrix. |
| $\mathbf{0}$ | : | The zero matrix. |
| $\mathbf{p}_a^{cb}$ | : | A vector (resp. homogenous coodrdinates) from point $b$ to point $c$ (denoted by the superscript) and expressed in $\underrightarrow{\boldsymbol{\mathcal{F}}}_a$ (denoted by the subscript). |
| $\mathbf{C}_{ab}$ | : | The $3 \times 3$ rotation matrix that transforms vectors from $\underrightarrow{\boldsymbol{\mathcal{F}}}_b$ to $\underrightarrow{\boldsymbol{\mathcal{F}}}_a$: $\mathbf{p}_a^{cb} = \mathbf{C}_{ab}\mathbf{p}_b^{cb}$. |
| $\mathbf{T}_{ab}$ | : | The $4 \times 4$ transformation matrix that transforms homogeneous points from $\underrightarrow{\boldsymbol{\mathcal{F}}}_b$ to $\underrightarrow{\boldsymbol{\mathcal{F}}}_a$: $\boldsymbol{p}_a^{ca} = \mathbf{T}_{ab}\boldsymbol{p}_b^{cb}$. |

# Chapter 1

# Introduction

> To be sure, a writer cannot begin with a thesis; he must rather use his writerly sensitivity to intuit what is going on, even if he cannot understand its implications.
>
> Gary Morson, *How the great truth dawned*

In this dissertation, we present a general approach to improve visual egomotion estimation for mobile autonomous platforms. Such mobile *automata* have been a part of human culture since antiquity. In ancient Greek mythology, Hephaestus–the 'patron of invention and technology' (Mayor, 2019)–was said to have forged autonomous handmaidens that assisted him in his workshop. In ancient India, the king Ajatashatru was said to use *bhuta vahana yanta* ('spirit movement machines') to protect the relics of Gautama Buddha after his death in the fourth century BCE. According to Burmese legend, the bhuta vahana yanta of Ajatashatru were made with stolen secrets from a group Greco-Roman 'roboticists' named the *yantakara*. The methods of the yantakara were closely guarded, and mechanical assassins were said to pursue those who attempted to disseminate them[1] (Mayor, 2019).

In the millennia since, mobile automata have been largely documented only in isolated demonstrations (e.g., the programmable cart of Hero of Alexandria or the 'autonomous' knight of Leonardo da Vinci) or relegated to the pages of literary work (e.g., Shelley's *Frankenstein*). Although the secrets of the yantakara may never be rediscovered, the centuries-long pursuit of true autonomy has finally started to yield techniques and machinery that show great promise in aiding humanity in the twenty-first century and beyond.

[1] Disseminate this document at your own risk.

## 1.1 Egomotion Estimation

One of said techniques is *egomotion* estimation: the process of computing changes in position and orientation of a moving platform from onboard measurements. The basic principle of egomotion estimation—the method of *dead reckoning*—has been used since ancient times to determine the position of a ship at sea. Although almanacs combined with star measurements can yield *latitude* estimates during the night, the problem of longitude estimation was not solved until the 18th century[2]. As a result, marine navigators could only rely on relative ship speed measurements paired with magnetic heading to infer east-west motion. In a similar process, early aviators computed egomotion through magnetic heading, airspeed, and an estimate of prevailing winds. Although all dead-reckoning-based egomotion estimates will exhibit unbounded error growth, these early methods were particularly inaccurate and required regular corrections through known landmarks.[3] In the twentieth century, the goals of inter-continental flight and space exploration necessitated the development of highly accurate egomotion sensors (e.g., gimballed inertial sensors like accelerometers and gyroscopes) and an associated set of estimation techniques that could compute egomotion without human intervention (e.g., the Kalman filter (Grewal and Andrews, 2010)).

Further, the development of extra-planetary rovers motivated the development of a new approach to egomotion estimation. Although ground vehicles can infer motion through *wheel odometry* (integrating wheel rate and orientation encoder measurements into a kinematic model), this approach can be highly inaccurate on surfaces that induce wheel slip (e.g., the rock and sand covered surface of Mars). To address this, a number of researchers in the 1980s developed the technique of *visual* odometry (Scaramuzza and Fraundorfer, 2011) (or VO), as a way to compute egomotion using images collected sequentially. VO is closely related to the techniques of *bundle adjustment* (Triggs et al., 2000) and *structure from motion* that were initially developed to automate the recon-

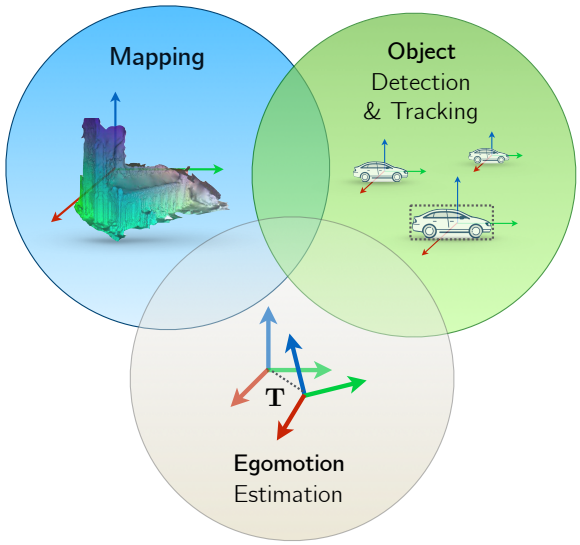


Figure 1.1: Visual egomotion estimation can be used in tandem with other visual estimation techniques.

[2]With the development of the marine chronometer by John Harrison.

[3]Perhaps the most famous example of dead-reckoning error was made by Christopher Columbus in 1492. He believed he had reached the *Indies* (modern Indonesia) but had really arrived in the Bahamas.

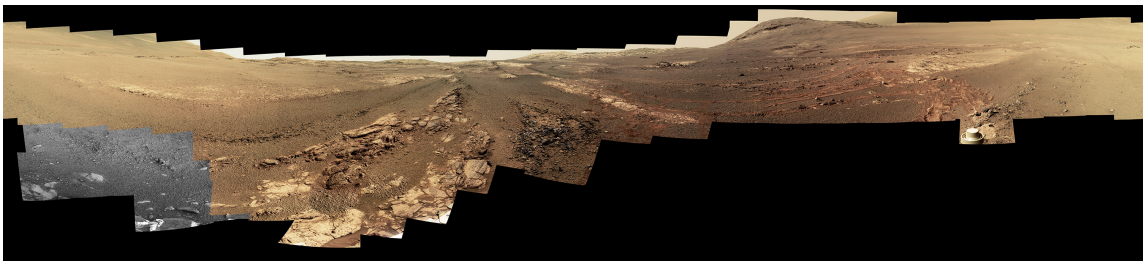

Figure 1.2: The last 360 degree panorama taken by the PanCam apparatus of the Mars Exploration Rover, *Opportunity*, at its final resting place on Mars, the western rim of the Endeavour Crater. Contact with *Opportunity* was lost shortly after this was captured, due to a severe dust storm (credit: NASA/JPL-Caltech/Cornell/ASU).

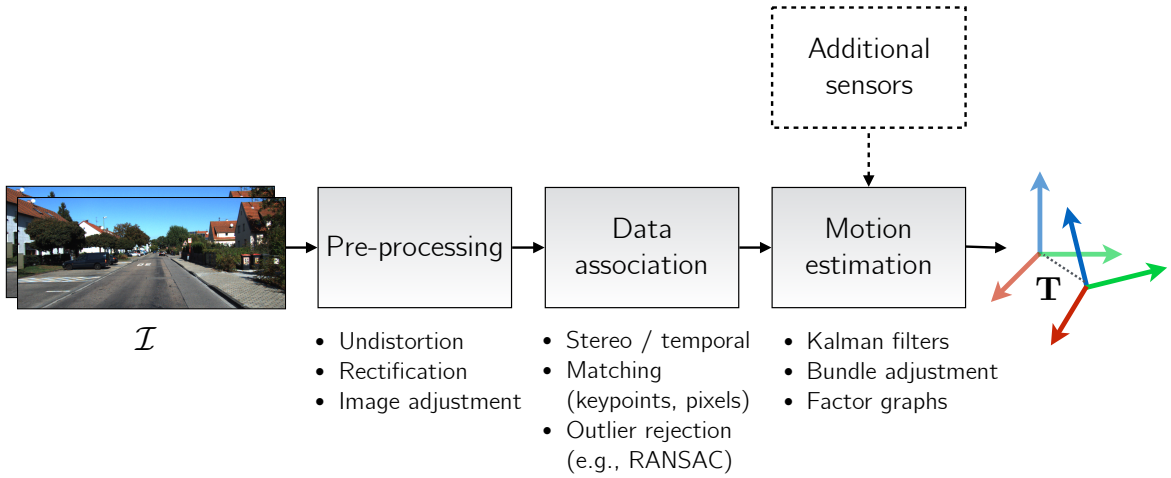


Figure 1.3: A 'classical' visual odometry pipeline consists of several distinct components that have interpretable inputs and outputs.

struction of cold-war-era reconnaissance imagery.

In the last decade, the development of compact, relatively-inexpensive, high resolution cameras has made vision-based estimation ubiquitous in mobile autonomous applications. In addition to computing egomotion, camera data can be used to build detailed maps of an environment (which can often be solved in tandem with egomotion estimation, resulting in a technique named simultaneous localization and mapping, or SLAM (Cadena et al., 2016)), as well as detect, track and avoid other objects (see Figure 1.1).

## 1.2 A Visual *Pipeline*

Central to *classical* visual odometry algorithms (which, in this context, refers to the bulk of VO research published during what Cadena et al. (2016) call the *classical* and *algorithmic-analysis* ages of VO and SLAM research between 1986 and 2015) is the idea of a pipeline.

A pipeline consists several distinguishable blocks that have interpretable inputs and outputs. By carefully processing information contained within sensor data, pipelines facilitate the construction of complex state estimation architectures that can fuse observations from sensors of varied modality to create rich models of the external world and infer the state of a mobile platform within it. In this dissertation, we will largely deal with the improvement of a baseline VO pipeline– we illustrate its major components in Figure 1.3.

Such Classical VO pipelines (e.g., Leutenegger et al. (2015); Cvišić and Petrović (2015); Tsotsos et al. (2015)) have achieved impressive localization accuracy on trajectories spanning several kilometres by carefully extracting and tracking sparse visual features (using *hand-crafted* algorithms) across consecutive images. Simultaneously, significant effort has gone to developing localization pipelines that eschew sparse features in favour of *dense* visual data (Alcantarilla and Woodford, 2016; Forster et al., 2014), typically relying on loss functions that use direct pixel intensities.

In the last five years, a significant part of the visual state estimation literature has also focused on the idea of replacing classical pipelines with parametric modelling through deep convolutional neural networks (CNNs) and data-driven learning. Although initially developed for image classification (LeCun et al., 2015), CNN-based measurement models have been applied to numerous problems in geometric state estimation (e.g., homography estimation (DeTone et al., 2016), single image depth reconstruction (Garg et al., 2016), camera relocalization (Kendall and Cipolla, 2016), place recognition (Sünderhauf et al., 2015)). A number of recent CNN-based approaches have also tackled the problem of egomotion estimation, often purporting to obviate the need for classical visual localization pipelines by learning pose changes *end-to-end*, directly from image data (e.g., Melekhov et al. (2017), Handa et al. (2016), Oliveira et al. (2017)).

Despite this surge of excitement, significant debate has emerged within the robotics and computer vision communities regarding the extent to which deep models should replace existing geometric state estimation algorithms. Owing to their representational power, deep models may move the onerous task of selecting 'good' (i.e., robust to environmental vagaries and sensor motion) visual features from the roboticist to the learned model. By design, deep models also provide a straight-forward formulation for using *dense* data while being flexible in their loss function, and taking full advantage of modern computing architecture to minimize run-time. Despite these potential benefits, end-to-end regression techniques for state estimation often generalize poorly to new environments, come with few analytical guarantees, and provide only point estimates of latent parameters (see Table 1.1 for more details). Indeed, the most accurate visual egomotion pipelines at the time of writing[4] remain largely those based

[4]Based on the KITTI Odometry benchmark leaderboard at http://www.cvlibs.net/datasets/kitti/

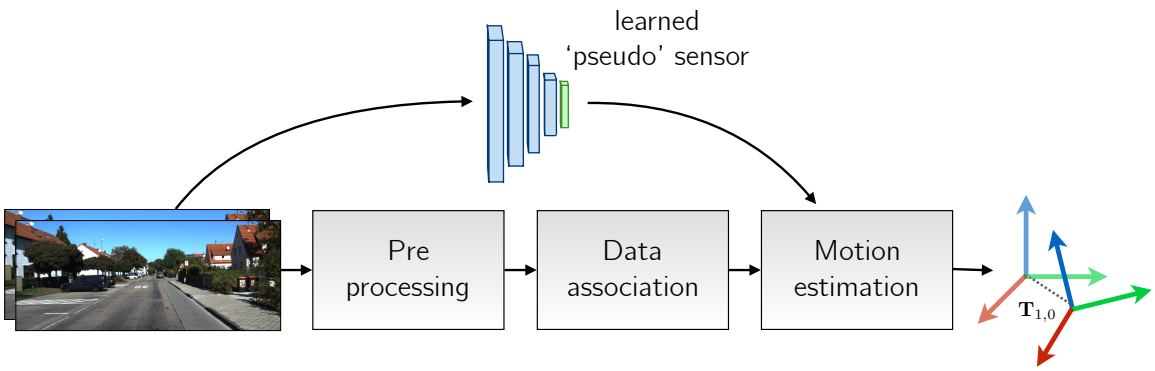


Figure 1.4: A learned *pseudo-sensor* extracts latent information from the same data stream.

on carefully selected sparse features. Furthermore, there is recent empirical evidence (Zhou et al., 2019) that suggests that designing a pipeline with interpretable components (e.g., optical flow, scene segmentation) is crucial to generalization on various visual tasks.

Table 1.1: A comaprison of pipelines and end-to-end deep models for visual egomotion estimation.

| | **Classical Pipelines** | **Deep Models** |
|---|---|---|
| *Maturity* | Decades of literature & domain knowledge | Nascent with few uses in mobile autonomy |
| *Interpretability* | Good, each component has interpretable input and output | Poor, often with no interpretable intermediate outputs |
| *Uncertainty* | Foundational to *probabilistic robotics* | Few nascent methods (Monte-carlo Dropout (Gal and Ghahramani, 2016b), Bootstrap (Osband et al., 2016)) |
| *Robustness* | Empirically generalizable (Zhou et al., 2019) | Highly dependant on training data |
| *Flexibility* | Limited by ingenuity of designer | Limited by training data |

`eval_odometry.php`.

## 1.3 The Learned Pseudo-Sensor

As state estimation enters the robust-perception age (Cadena et al., 2016), algorithms that work in limited contexts will need to be adapted and augmented to ensure they can operate over longer time-periods, and through sundry environments. Towards this end, we introduce the paradigm of the *learned pseudo-sensor*. Learned pseudo-sensors allow one to retain the benefits of classical state estimation pipelines while leveraging the representational power of modern data-driven learning techniques. Instead of completely replacing the classical pipeline, herein we present four ways in which machine learning can be used to train a hyper-parametric model that extracts latent information from an existing visual and inertial data stream. By fusing the output of these sensors with the output of the pipeline, we can make the final egomotion estimates more accurate and more robust to difficult-to-model effects (Figure 1.4). To accomplish this fusion, we rely on two approaches. The first, which we call Predictive Robust Estimation (PROBE, Chapter 4), uses the paradigm of a pseudo-sensor to build a heteroscedastic noise model from extant visual-inertial data. By predicting uncertainty information, PROBE effectively re-scales a robust loss function to better account for deleterious visual effects. The second approach (used by Sun-BCNN, DPC-Net, and HydraNet, Chapters 5 to 7 respectively) produces geometric quantities (probabilistic estimates of an illumination source, $\mathrm{SE}(3)$ corrections to existing egomotion estimates, and independent probabilistic rotation estimates, respectively), that can be fused with the original pipeline through a factor graph optimization routine.

## 1.4 Original Contributions

This dissertation consists of several published contributions under the umbrella of *learned pseudo-sensors* that improve a canonical visual egomotion pipeline. Before detailing each pseudo-sensor, we present some mathematical foundations (Chapter 2) and a common baseline for an indirect stereo visual odometry pipeline (Chapter 3) which all four methods build upon. In total, there are two journal papers and five conference papers associated with our work. Below, we briefly summarize each of the pseudo-sensors and list the publications that are associated with each.

1. **Predictive Robust Estimation (PROBE)**,
   Predictive Robust Estimation (Chapter 4, Appendix A) uses the method of Generalized Kernel (GK) estimation (Vega-Brown et al., 2014) to derive an efficient Bayesian model for stereo tracking uncertainty based a set of training reprojection errors. By setting a prior on covariance, we derive a robust loss that can be predictively scaled to improve

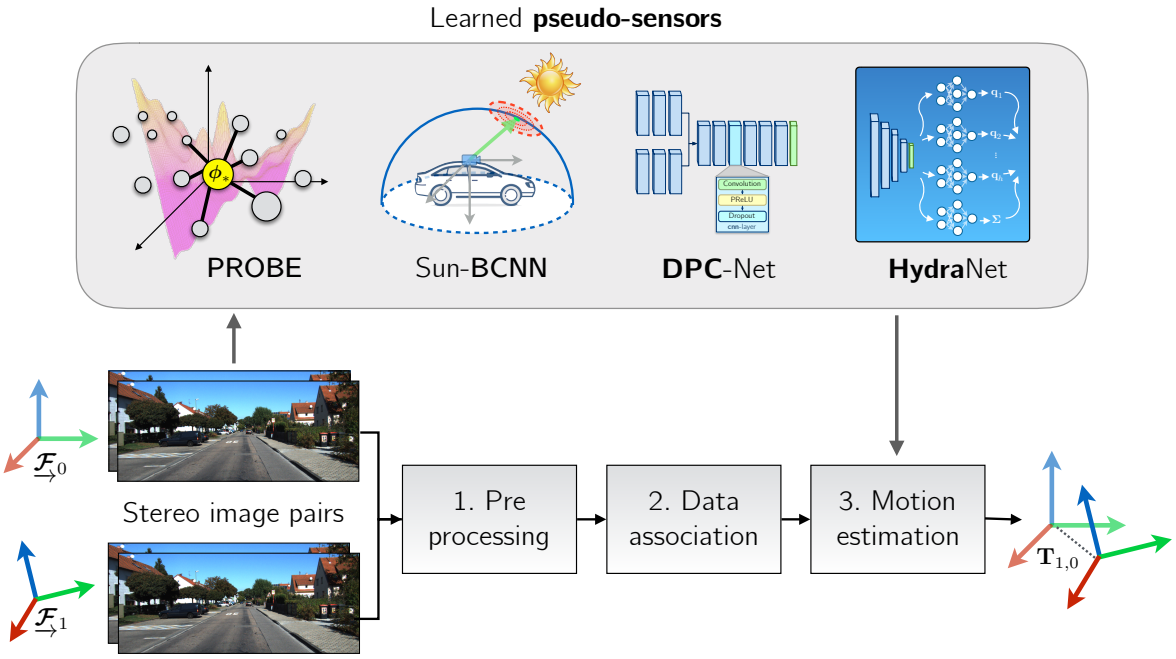


Figure 1.5: This dissertation details four examples of *pseudo-sensors* that improve 'classical' egomotion estimation through data-driven learning.

the accuracy and consistency of an indirect stereo visual odometry pipeline. PROBE is associated with three publications listed below. The first two publications (and Appendix A) explore useful predictors for uncertainty and build a non-Bayesian isotropic covariance model. The latter publication presents the Bayesian GK approach.

- Peretroukhin, V., Clement, L., Giamou, M., and Kelly, J. (2015a). PROBE: Predictive robust estimation for visual-inertial navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*, pages 3668–3675, Hamburg, Germany,
- Peretroukhin, V., Clement, L., and Kelly, J. (2015b). Get to the point: Active covariance scaling for feature tracking through motion blur. In *Proceedings of the IEEE International Conference on Robotics and Automation Workshop on Scaling Up Active Perception*, Seattle, Washington, USA,
- Peretroukhin, V., Vega-Brown, W., Roy, N., and Kelly, J. (2016). PROBE-GK: Predictive robust estimation using generalized kernels. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 817–824.

2. **Sun-BCNN: Learned sun sensor**
   Sun-BCNN (Chapter 5) is a technique to infer a probabilistic estimate of the direction

of the sun from a single RGB image using a Bayesian Convolutional Neural Network (BCNN). The method works much like hardware sun sensors (Lambert et al., 2012), but does not requires any additional sensing equipment, and can provide mean and covariance estimates that can be readily incorporated into existing visual odometry frameworks. It is associated with three publications listed below. Initial exploratory work was published at ISER, and the BCNN improvement was presented at ICRA. An additional journal paper summarizing the work of the prior two papers, adding data from the Canadian High Arctic and Oxford, and investigating the effect of cloud cover and transfer learning was published in the International Journal of Robotics Research, Special Issue on Experimental Robotics.

- Clement, L., Peretroukhin, V., and Kelly, J. (2017). Improving the accuracy of stereo visual odometry using visual illumination estimation. In Kulic, D., Nakamura, Y., Khatib, O., and Venture, G., editors, *2016 International Symposium on Experimental Robotics*, volume 1 of *Springer Proceedings in Advanced Robotics*, pages 409–419. Springer International Publishing, Berlin Heidelberg. Invited to Journal Special Issue,
- Peretroukhin, V., Clement, L., and Kelly, J. (2017). Reducing drift in visual odometry by inferring sun direction using a bayesian convolutional neural network. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'17)*, pages 2035–2042, Singapore,
- Peretroukhin, V., Clement, L., and Kelly, J. (2018). Inferring sun direction to improve visual odometry: A deep learning approach. *International Journal of Robotics Research*, 37(9):996–1016.

3. **DPC-Net: Learned pose corrections**
   Deep Pose Correction (Chapter 6) is an approach to improving egomotion estimates through $\mathrm{SE}(3)$ pose corrections learned through deep regression with a supervised loss based on Lie theory. The Deep Pose Correction Network (DPC-Net) learns low-rate, 'small' *corrections* from training data that are then fused with the original estimates from the canonical pipeline. DPC-Net does not require any modification to an existing pipeline, and can learn to correct multi-faceted errors from estimator bias, sensor miscalibration or environmental effects. It is associated with one journal publication listed below.

   - Peretroukhin, V. and Kelly, J. (2018). DPC-Net: Deep pose correction for visual localization. *IEEE Robotics and Automation Letters*, 3(3):2424–2431.

4. **HydraNet: Learned probabilistic rotation estimation**

   Finally, HydraNet (Chapter 7) is a multi-headed network structure that can regress probabilistic estimates of rotation (elements of the matrix Lie group, $\mathrm{SO}(3)$) that account for both aleatoric and epistemic uncertainty. HydraNet builds upon results from both Sun-BCNN and DPC that show that correcting rotation is critical to accurate egomotion estimation. Towards this end, HydraNet is designed to produce well-calibrated notions of uncertainty over $\mathrm{SO}(3)$ that facilitate fusion with classical egomotion pipelines through a probabilistic factor graph formulation. It is associated with one publication:

   - Peretroukhin, V., Wagstaff, B., and Kelly, J. (2019). Deep probabilistic regression of elements of SO(3) using quaternion averaging and uncertainty injection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'19) Workshop on Uncertainty and Robustness in Deep Visual Learning*, pages 83–86, Long Beach, California, USA.

# Chapter 2

# Mathematical Foundations

> By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems, and, in effect, increases the mental power of the race.
>
> — Alfred North Whitehead

## 2.1 Coordinate Frames

Before we can present the main contributions of this dissertation, it will be useful to first outline the notation and mathematical foundations that underly the work. Throughout this dissertation, we largely follow the notation of Barfoot (2017) when dealing with three-dimensional rigid-body kinematics.

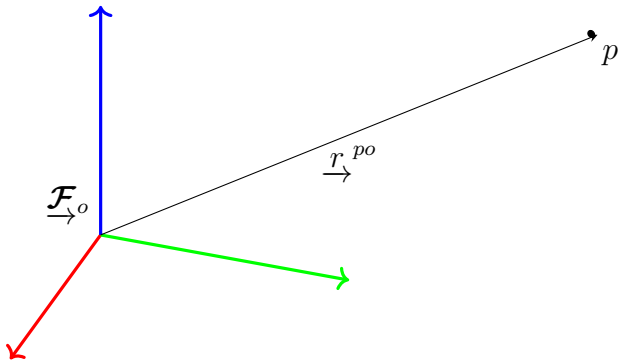


Figure 2.1: A position vector expressed in a coordinate frame.

We refer to a three-dimensional position vector, $\underrightarrow{r}^{po}$, as one that originates at the origin of a coordinate reference frame, $\underrightarrow{\mathcal{F}}_o$, and terminates at the point $p$. This geometric quantity has

the numerical coordinates $\mathbf{r}_o^{po}$ when expressed in $\underrightarrow{\mathcal{F}}_o$. Often, we will refer to two reference frames such as a world or *inertial* frame, $\underrightarrow{\mathcal{F}}_i$, and a vehicle frame, $\underrightarrow{\mathcal{F}}_v$. Rotation matrices or rigid-body transformations that convert coordinates from $\underrightarrow{\mathcal{F}}_i$ to $\underrightarrow{\mathcal{F}}_v$ will be represented as $\mathbf{T}_{vi}$, and $\mathbf{C}_{vi}$[1], respectively.

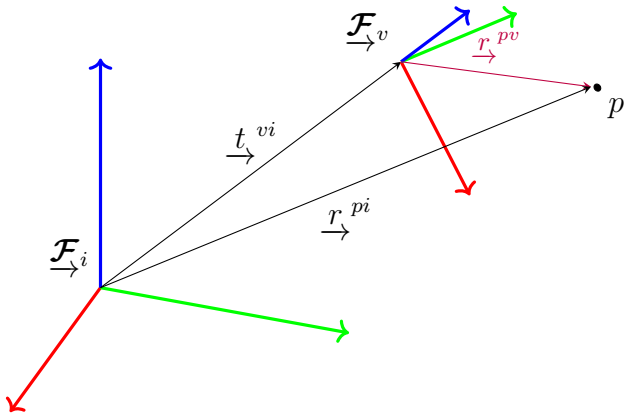


Figure 2.2: Two common references frames used throughout this thesis.

## 2.2 Rotations

The rotation matrix $\mathbf{C}$ is a member of the matrix Lie group[2] $\mathrm{SO}(3)$ (the Special Orthogonal group). We can define it as follows:

$$\mathrm{SO}(3) = \{\mathbf{C} \in \mathbb{R}^{3\times 3} | \ \mathbf{C}^T\mathbf{C} = \mathbf{1}, \det \mathbf{C} = 1\}. \tag{2.1}$$

### Active and Passive Representations

An active (or *alibi*) rotation changes the coordinates of a position directly while implicitly assuming that the reference frame is fixed. A passive (or *alias*) rotation rotates the reference frame. Following Barfoot (2017), all rotation matrices in this dissertation are passive unless otherwise noted.

### Exponential and Logarithmic Maps

Since rotations form a matrix Lie group (we refer the reader to Solà et al. (2018) and Barfoot (2017) for a thorough treatment of Lie groups for state estimation), we can define a surjective

[1]We use $\mathbf{C}$ and not $\mathbf{R}$ for rotation matrices to avoid confusion with common notation for measurement model covariance.

[2]A Lie group is a group that is also a differentiable manifold. See Barfoot (2017) for more details.

exponential map[3] from three axis-angle parameters, $\boldsymbol{\phi} = \phi\mathbf{a}, \quad \phi \in \mathbb{R}, \mathbf{a} \in S^2$, to a rotation matrix, $\mathbf{C}$:

$$\mathbf{C} = \mathrm{Exp}\left(\boldsymbol{\phi}\right) = \exp\left(\boldsymbol{\phi}^{\wedge}\right) = \sum_{n=0}^{\infty} \frac{1}{n!}(\boldsymbol{\phi}^{\wedge})^n \tag{2.2}$$

$$= \cos\phi\mathbf{1} + (1 - \cos\phi)\mathbf{a}\mathbf{a}^T + \sin\phi\mathbf{a}^{\wedge}, \tag{2.3}$$

where the wedge operator $(\cdot)^{\wedge}$[4] is defined as

$$\mathbf{a}^{\wedge} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}^{\wedge} = \begin{bmatrix} 0 & -a_2 & a_1 \\ a_2 & 0 & -a_0 \\ -a_1 & a_0 & 0 \end{bmatrix}. \tag{2.4}$$

Equation (2.3) is often referred to as the Euler-Rodriguez formula and it can also be derived geometrically, starting from Euler's theorem that any rotation can be expressed as an axis of rotation and an angle of rotation about that axis. Although the map in Equation (2.2) is surjective, we can define an inverse map if we restrict its domain to $0 \leq \phi < \pi$:

$$\boldsymbol{\phi} = \mathrm{Log}\left(\mathbf{C}\right) = \log\left(\mathbf{C}\right)^{\vee} = \frac{\phi(\mathbf{C} - \mathbf{C}^T)^{\vee}}{2\sin\phi}, \tag{2.5}$$

where $\phi = \arccos\left(\frac{\mathrm{tr}(\mathbf{C})-1}{2}\right)$ and the *vee* operator, $(\cdot)^{\vee} : \mathbb{R}^{3\times3} \to \mathbb{R}^3$, is defined as the unique inverse of the wedge operator $(\cdot)^{\wedge}$. Note Equation (2.5) is undefined at both $\phi = 0$ and at $\phi = \pi$. In the former case, we can use a small-angle approximation and define

$$\mathrm{Log}\left(\mathbf{C}\right) \approx (\mathbf{C} - \mathbf{1})^{\vee} \quad \text{when} \quad \phi \approx 0. \tag{2.6}$$

The latter case (when $\phi = \pi$) defines the *cut locus* of the space where $\mathrm{Exp}\left(\cdot\right)$ is not a covering map and both $+\boldsymbol{\phi}$ and $-\boldsymbol{\phi}$ map to the same $\mathbf{C}$. This *cut locus* is related to the idea that any three parameterization of $\mathrm{SO}(3)$ will have singularities associated with it.

[3] We follow Solà et al. (2018) and also define *capitalized* map for notational clarity.

[4] This operator is sometimes also expressed as $(\cdot)^{\times}$ or $[\cdot]_{\times}$ and is known as the skew-symmetric operator.

### 2.2.1 Unit Quaternions

We can also represent a rotation with unit quaternion, $\mathbf{q}$. A unit quaternion consists of a scalar value $q_\omega$, and a three-dimensional vector component, $\mathbf{q}_v$:

$$\mathbf{q} = \begin{bmatrix} q_\omega \\ \mathbf{q}_v \end{bmatrix} \in S^3, \quad (\|\mathbf{q}\| = 1). \tag{2.7}$$

Unit quaternions also form a Lie group (Solà et al., 2018) and and lie on a three-dimensional unit sphere within $\mathbb{R}^4$. As with rotation matrices, we can define a surjective map from three parameters to the group itself,

$$\mathbf{q} = \mathrm{Exp}\left(\boldsymbol{\phi}\right) = \begin{bmatrix} \cos\left(\phi/2\right) \\ \mathbf{a}\sin\left(\phi/2\right) \end{bmatrix}. \tag{2.8}$$

By inspection, we can see that both $\mathbf{q}$ and $-\mathbf{q}$ represent the same axis-angle pair, $\{\phi, \mathbf{a}\}$. As a result, unit quaternions represent a *double cover* of $\mathrm{SO}(3)$ and we must be careful to account for this when using them as a rotation parametrization. In particular, when computing the logarithmic map,

$$\boldsymbol{\phi} = \mathrm{Log}\left(\mathbf{q}\right) = 2\mathbf{q}_v \frac{\arctan\left(\|\mathbf{q}_v\|, q_\omega\right)}{\|\mathbf{q}_v\|}, \tag{2.9}$$

we can must account for the double cover by replacing $\mathbf{q}$ with $-\mathbf{q}$ if $q_\omega$ is negative. Also note that as with rotation matrices Equation (2.9) is undefined when $\phi = 0$, but, importantly, we do not face any issues when $\phi = \pi$ due to the half-angle. In the former case, we can again rely on small angle approximations to define:

$$\mathrm{Log}\left(\mathbf{q}\right) \approx \frac{\mathbf{q}_v}{q_\omega}\left(1 - \frac{\|\mathbf{q}_v\|^2}{3q_\omega^2}\right) \quad \text{when} \quad \phi \approx 0. \tag{2.10}$$

A fantastic summary of the history of rotation parameterizations, unit quaternions and the story of Hamilton and Rodriguez can be found in Altmann (1989).

### 2.2.2 Topology

Topologically, $\mathrm{SO}(3)$ is *diffeomorphic*[5] to the real projective space, $\mathbb{RP}^3$, the space of all lines passing through the origin in $\mathbb{R}^4$. As a result, any global $n$-parametrization of $\mathrm{SO}(3)$ will incur some cost. If we use rotation matrices ($n = 9$), we need to ensure orthonormality and

[5] A diffeomorphism is a smooth invertible function that maps one differentiable manifold to another.

that $\det \mathbf{C} = 1$. With unit quaternions ($n = 4$), we need to account for the unit-norm constraint and take note of the double cover. Parametrizations with $n = 3$ (like Euler angles and axis-angle parameters) will be bounded, but unconstrained. However, due to the topological structure of $\mathrm{SO}(3)$ all three-parameter parametrizations will not be invertible for certain rotations. With Euler angles, one has to be wary of *gimbal lock*, wherein two angles become indeterminate from each other. For axis-angle parameters, it is not possible to uniquely represent rotations whose angle is $\pi$.

To see why this is the case, consider that according to Euler's theorem, we can represent any element in $\mathrm{SO}(3)$ by the axis-angle pair $\{\mathbf{a}, \phi\}$, $\mathbf{a} \in S^2$, and $0 \leq \phi \leq \pi$. We can partially represent this space by the closed ball of radius $\pi$ in $\mathbb{R}^3$ using the combined axis-angle coordinates $\boldsymbol{\phi} = \phi\mathbf{a}$. However, at the boundary ($\phi = \pi$), we must account for the fact that rotations represented by $\{\mathbf{a}, \pi\}$ are identical to those represented by $\{-\mathbf{a}, \pi\}$. In other words, we must *identify* all antipodal points, $\boldsymbol{\phi}$ and $-\boldsymbol{\phi}$ when $\|\boldsymbol{\phi}\| = \pi$. This closed 3-ball with identified antipodal points on its boundary is topologically equivalent to the 3-sphere ($S^3$) with its antipodal points identified. In turn, this space is equivalent to $\mathbb{RP}^3$ since any line passing through the origin in $\mathbb{R}^4$ can be mapped to two unit normals, $\pm\mathbf{n} \in S^3$.

This *identification* makes rotation representation particularly tricky in $\mathbb{R}^3$ and clearly explains why unit quaternions, $\mathbf{q} \in S^3$, are a *double* cover of $\mathrm{SO}(3)$, since we must add the relation $\mathbf{q} = -\mathbf{q}$ to make these two spaces equivalent.

Accordingly, in this dissertation, we parametrize rotations as the constrained quantities, $\mathbf{q}$ or $\mathbf{C}$. When dealing with perturbations about a given rotation (e.g., to compute updates to a state, or to propagate uncertainty), we use *small* rotations, $\delta\mathbf{C}$ or $\delta\mathbf{q}$, parametrized using three unconstrained parameters (since, in this case, we can assume $\|\boldsymbol{\phi}\| << \pi$).

## 2.3 Spatial Transforms

The rigid body transform $\mathbf{T}$ is a also a member of the matrix Lie group, the Special Euclidian group $\mathrm{SE}(3)$ and can be defined as a $4 \times 4$ matrix as follows:

$$\mathrm{SE}(3) = \{\mathbf{T} = \begin{bmatrix} \mathbf{C} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \in \mathbb{R}^{4\times4} | \ \mathbf{C} \in \mathrm{SO}(3), \mathbf{t} \in \mathbb{R}^3\}. \tag{2.11}$$

As a member of a matrix Lie group, it also admits a surjective exponential map,

$$\mathbf{T} = \mathrm{Exp}\left(\boldsymbol{\xi}\right) = \exp\left(\boldsymbol{\xi}^\wedge\right) = \sum_{n=0}^{\infty} \frac{1}{n!} (\boldsymbol{\xi}^\wedge)^n \tag{2.12}$$

where $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\phi} \end{bmatrix} \in \mathbb{R}^6$ and the wedge operator is overloaded (following Barfoot (2017)) as follows:

$$\boldsymbol{\xi}^\wedge \triangleq \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\phi} \end{bmatrix}^\wedge = \begin{bmatrix} \boldsymbol{\phi}^\wedge & \boldsymbol{\rho} \\ \mathbf{0}^T & 0 \end{bmatrix}. \tag{2.13}$$

In practice, we can evaluate the exponential map through the Euler-Rodriguez formula (Equation (2.3)) and by computing the left-Jacobian of $\mathrm{SO}(3)$, $\mathbf{J}$,

$$\mathbf{T} = \mathrm{Exp}\left(\begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\phi} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{C}(\boldsymbol{\phi}) & \mathbf{J}(\boldsymbol{\phi})\boldsymbol{\rho} \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{2.14}$$

where

$$\mathbf{J}(\boldsymbol{\phi}) = \frac{\sin\phi}{\phi}\mathbf{1} + (1 - \frac{\sin\phi}{\phi})\mathbf{a}\mathbf{a}^T + \frac{1-\cos\phi}{\phi}\mathbf{a}^\wedge. \tag{2.15}$$

### 2.3.1 Applying Transforms

Applying our notation for coordinate frames (and referring back to Section 2.1), a transform, $\mathbf{T}_{vi}$ can be expressed as

$$\mathbf{T}_{vi} = \begin{bmatrix} \mathbf{C}_{vi} & \mathbf{t}_v^{iv} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{2.16}$$

This allows us to use the homogenous point representation for $\mathbf{r}_i^{pi}$ and express the following relation:

$$\boldsymbol{r}_v^{pv} = \mathbf{T}_{vi}\boldsymbol{r}_i^{pi}, \tag{2.17}$$

or

$$\begin{bmatrix} \mathbf{r}_v^{pv} \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{C}_{vi} & \mathbf{t}_v^{iv} \\ \mathbf{0}^T & 1 \end{bmatrix}}_{\mathbf{T}_{vi}} \begin{bmatrix} \mathbf{r}_i^{pi} \\ 1 \end{bmatrix}, \tag{2.18}$$

which is numerically equivalent to

$$\mathbf{r}_v^{pv} = \mathbf{C}_{vi}\mathbf{r}_i^{pi} + \mathbf{t}_v^{iv}. \tag{2.19}$$

## 2.4 Perturbations

When solving optimization problems that involve rotations or rigid-body transforms, it is often useful to consider a small *perturbation* about an operating point. By leveraging a core property of Lie groups (that they are locally 'Euclidian'), we can convert difficult non-linear problems into ones that have local linear approximations.

Using rotations as an example, we can perturb an operating point, $\overline{\mathbf{C}} \triangleq \operatorname{Exp}\left(\overline{\boldsymbol{\phi}}\right)$, in three different ways:

$$\mathbf{C} = \begin{cases} \operatorname{Exp}\left(\delta\boldsymbol{\phi}^{\ell}\right)\overline{\mathbf{C}} & \text{left perturbation,} \\ \operatorname{Exp}\left(\overline{\boldsymbol{\phi}} + \delta\boldsymbol{\phi}^{m}\right) & \text{middle perturbation,} \\ \overline{\mathbf{C}}\operatorname{Exp}\left(\delta\boldsymbol{\phi}^{r}\right) & \text{right perturbation.} \end{cases} \tag{2.20}$$

We can relate all the left and middle perturbations through the left Jacobian of $\mathrm{SO}(3)$ with the following useful identity,

$$\operatorname{Exp}\left(\left(\boldsymbol{\phi} + \delta\boldsymbol{\phi}^{m}\right)\right) \approx \operatorname{Exp}\left(\mathbf{J}(\boldsymbol{\phi})\delta\boldsymbol{\phi}^{m}\right)\operatorname{Exp}\left(\boldsymbol{\phi}\right). \tag{2.21}$$

From this it follows that $\delta\boldsymbol{\phi}^{\ell} \approx \mathbf{J}(\boldsymbol{\phi})\delta\boldsymbol{\phi}^{m}$ and elucidates why $\mathbf{J}$ is called the *left* Jacobian.

In this dissertation, we will use the left and middle perturbations when appropriate. Using small angle approximations, the Euler-Rodriguez formula (Equation (2.3)) yields $\operatorname{Exp}\left(\delta\boldsymbol{\phi}\right) \approx \mathbf{1} + \delta\boldsymbol{\phi}^{\wedge}$, which allows us to write the useful formula for the left perturbation:

$$\mathbf{C} = (\mathbf{1} + (\delta\boldsymbol{\phi}^{\ell})^{\wedge})\overline{\mathbf{C}}. \tag{2.22}$$

Similarly, we can write analogous expressions for a rigid body transform, $\mathbf{T} \in \mathrm{SE}(3)$, as composed of an operating point $\overline{\mathbf{T}} \triangleq \operatorname{Exp}\left(\overline{\boldsymbol{\xi}}\right)$, and a small perturbation about that operating point:

$$\mathbf{T} = \begin{cases} \operatorname{Exp}\left(\delta\boldsymbol{\xi}^{\ell}\right)\overline{\mathbf{T}} & \text{left perturbation,} \\ \operatorname{Exp}\left(\overline{\boldsymbol{\xi}} + \delta\boldsymbol{\xi}^{m}\right) & \text{middle perturbation,} \\ \overline{\mathbf{T}}\operatorname{Exp}\left(\delta\boldsymbol{\xi}^{r}\right) & \text{right perturbation.} \end{cases} \tag{2.23}$$

Now, we can also note a similar identity for $\mathrm{SE}(3)$,

$$\operatorname{Exp}\left(\left(\boldsymbol{\xi} + \delta\boldsymbol{\xi}^{m}\right)\right) \approx \operatorname{Exp}\left(\left(\boldsymbol{\mathcal{J}}(\boldsymbol{\xi})\delta\boldsymbol{\xi}^{m}\right)\right)\operatorname{Exp}\left(\boldsymbol{\xi}\right), \tag{2.24}$$

where $\boldsymbol{\mathcal{J}}$ is the left Jacobian of $\mathrm{SE}(3)$ and defined as

$$\boldsymbol{\mathcal{J}}(\boldsymbol{\xi}) \triangleq \begin{bmatrix} \mathbf{J}(\boldsymbol{\phi}) & \mathbf{Q}(\boldsymbol{\xi}) \\ \mathbf{0} & \mathbf{J}(\boldsymbol{\phi}) \end{bmatrix}, \tag{2.25}$$

where $\mathbf{Q}(\boldsymbol{\xi})$ can be evaluated analytically (see Barfoot (2017)). This again allows us to write

$\delta\boldsymbol{\xi}^{\ell} \approx \boldsymbol{\mathcal{J}}(\boldsymbol{\xi})\delta\boldsymbol{\xi}^{m}$ and form a similar expression,

$$\mathbf{T} = (\mathbf{1} + (\delta\boldsymbol{\xi}^{\ell})^{\wedge})\overline{\mathbf{T}}. \tag{2.26}$$

To derive locally linear systems from sets of rigid-body transforms, or 'poses', we can apply Equation (2.26). To update an operating point, we solve for $\delta\boldsymbol{\xi}^{\ell}$ and then use the constraint-sensitive update $\mathbf{T} \leftarrow \operatorname{Exp}\left(\delta\boldsymbol{\xi}^{\ell}\right)\overline{\mathbf{T}}$.

Finally, we note that we will often drop the perturbation superscripts $(\cdot)^{\ell}$ and $(\cdot)^{m}$ after defining the perturbation scheme.

## 2.5 Uncertainty

We can also use perturbation theory to implicitly define uncertainty on constrained manifolds (see Barfoot and Furgale (2014) for a thorough discussion).

Given a concentrated normal density, $\delta\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{6\times6})$, we can *inject* this unconstrained density onto the Lie group through left perturbations about some mean using

$$\mathbf{T} = \operatorname{Exp}\left(\delta\boldsymbol{\xi}\right)\overline{\mathbf{T}}. \tag{2.27}$$

This allows us to keep track of a random variable, $\mathbf{T}$, by keeping its mean in group form, $\overline{\mathbf{T}}$, while its second statistical moment is stored as a standard $6 \times 6$ covariance matrix, $\boldsymbol{\Sigma}$.

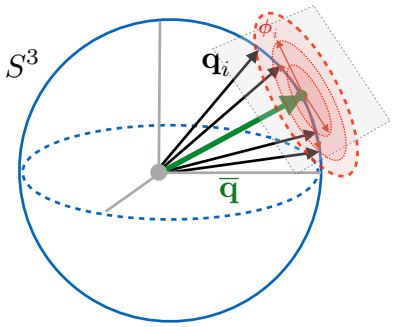


Figure 2.3: We can define uncertainty in the left tangent space of a mean element of a Lie group (here illustrated for unit quaternions).

We can define an analogous density for rotation matrices given normal densities over rotation perturbations $\delta\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{3\times3})$,

$$\mathbf{C} = \operatorname{Exp}\left(\delta\boldsymbol{\phi}\right)\overline{\mathbf{C}}, \tag{2.28}$$

and also, for unit quaternions,

$$\mathbf{q} = \operatorname{Exp}\left(\delta\boldsymbol{\phi}\right) \otimes \overline{\mathbf{q}} \tag{2.29}$$

where $\otimes$ refers to the standard quaternion product operator Sola (2017).

# Appendices

# Bibliography


Agarwal, S., Mierle, K., et al. (2016). Ceres solver.

Alcantarilla, P. F. and Woodford, O. J. (2016). Noise models in feature-based stereo visual odometry.

Altmann, S. L. (1989). Hamilton, rodrigues, and the quaternion scandal. *Math. Mag.*, 62(5):291–308.

Barfoot, T. D. (2017). *State Estimation for Robotics*. Cambridge University Press.

Barfoot, T. D. and Furgale, P. T. (2014). Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Trans. Rob.*, 30(3):679–693.

Brachmann, E. and Rother, C. (2018). Learning less is more-6d camera localization via 3d surface regression. In *Proc. CVPR*, volume 8.

Byravan, A. and Fox, D. (2017). SE3-nets: Learning rigid body motion using deep neural networks. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 173–180.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., and Leonard, J. J. (2016). Past, present, and future of simultaneous localization and mapping: Toward the Robust-Perception age. *IEEE Trans. Rob.*, 32(6):1309–1332.

Carlone, L., Rosen, D. M., Calafiore, G., Leonard, J. J., and Dellaert, F. (2015a). Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 125–132.

Carlone, L., Tron, R., Daniilidis, K., and Dellaert, F. (2015b). Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4597–4604.

Cheng, Y., Maimone, M. W., and Matthies, L. (2006). Visual odometry on the mars exploration rovers - a tool to ensure accurate driving and science imaging. *IEEE Robot. Automat. Mag.*, 13(2):54–62.

Clark, R., Wang, S., Wen, H., Markham, A., and Trigoni, N. (2017). Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem.

Clement, L. and Kelly, J. (2018). How to train a CAT: learning canonical appearance transformations for direct visual localization under illumination change. *IEEE Robotics and Automation Letters*, 3(3):2447–2454.

Clement, L., Peretroukhin, V., and Kelly, J. (2017). Improving the accuracy of stereo visual odometry using visual illumination estimation. In Kulic, D., Nakamura, Y., Khatib, O., and Venture, G., editors, *2016 International Symposium on Experimental Robotics*, volume 1 of *Springer Proceedings in Advanced Robotics*, pages 409–419. Springer International Publishing, Berlin Heidelberg. Invited to Journal Special Issue.

Costante, G., Mancini, M., Valigi, P., and Ciarfuglia, T. A. (2016). Exploring representation learning with CNNs for Frame-to-Frame Ego-Motion estimation. *IEEE Robotics and Automation Letters*, 1(1):18–25.

Crete, F., Dolmiere, T., Ladret, P., and Nicolas, M. (2007). The blur effect: perception and estimation with a new no-reference perceptual blur metric. In *Human vision and electronic imaging XII*, volume 6492, page 64920I. International Society for Optics and Photonics.

Cvišić, I. and Petrović, I. (2015). Stereo odometry based on careful feature selection and tracking. In *Proc. European Conf. on Mobile Robots (ECMR)*, pages 1–6.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition, (CVPR)*, pages 248–255.

DeTone, D., Malisiewicz, T., and Rabinovich, A. (2016). Deep image homography estimation.

Duan, Y., Chen, X., Houthooft, R., Schulman, J., and Abbeel, P. (2016). Benchmarking deep reinforcement learning for continuous control. In *Proc. Int. Conf. on Machine Learning*, ICML'16, pages 1329–1338.

Eisenman, A. R., Liebe, C. C., and Perez, R. (2002). Sun sensing on the mars exploration rovers. In *Aerosp. Conf. Proc.*, volume 5, pages 5–2249–5–2262 vol.5. IEEE.

Engel, J., Stuckler, J., and Cremers, D. (2015). Large-scale direct SLAM with stereo cameras. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst. (IROS)*, pages 1935–1942.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer.

Fischler, M. and Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395.

Fisher, R. (1953). Dispersion on a sphere. In *Proc. Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 217, pages 295–305. The Royal Society.

Fitzgibbon, A. W., Robertson, D. P., Criminisi, A., Ramalingam, S., and Blake, A. (2007). Learning priors for calibrating families of stereo cameras. In *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pages 1–8.

Florez, S. A. R. (2010). *Contributions by vision systems to multi-sensor object localization and tracking for intelligent vehicles*. PhD thesis.

Forster, C., Carlone, L., Dellaert, F., and Scaramuzza, D. (2015). IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation.

Forster, C., Pizzoli, M., and Scaramuzza, D. (2014). SVO: Fast semi-direct monocular visual odometry. In *Proc. IEEE Int. Conf. Robot. Automat.(ICRA)*, pages 15–22. IEEE.

Furgale, P. (2011). *Extensions to the Visual Odometry Pipeline for the Exploration of Planetary Surfaces*. PhD thesis.

Furgale, P. and Barfoot, T. D. (2010). Visual teach and repeat for long-range rover autonomy. *J. Field Robot.*, 27(5):534–560.

Furgale, P., Carle, P., Enright, J., and Barfoot, T. D. (2012). The devon island rover navigation dataset. *Int. J. Rob. Res.*, 31(6):707–713.

Furgale, P., Enright, J., and Barfoot, T. (2011). Sun sensor navigation for planetary rovers: Theory and field testing. *IEEE Trans. Aerosp. Electron. Syst.*, 47(3):1631–1647.

Furgale, P., Rehder, J., and Siegwart, R. (2013). Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286.

Gal, Y. (2016). *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge.

Gal, Y. and Ghahramani, Z. (2016a). Bayesian convolutional neural networks with Bernoulli approximate variational inference. In *Proc. Int. Conf. Learning Representations (ICLR), Workshop Track*.

Gal, Y. and Ghahramani, Z. (2016b). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proc. Int. Conf. Mach. Learning (ICML)*, pages 1050–1059.

Garg, R., Carneiro, G., and Reid, I. (2016). Unsupervised CNN for single view depth estimation: Geometry to the rescue. In *European Conf. on Comp. Vision*, pages 740–756. Springer.

Geiger, A., Lenz, P., Stiller, C., and Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *Int. J. Rob. Res.*, 32(11):1231–1237.

Geiger, A., Ziegler, J., and Stiller, C. (2011). StereoScan: Dense 3D reconstruction in real-time. In *Proc. IEEE Intelligent Vehicles Symp. (IV)*, pages 963–968.

Geman, S., McClure, D. E., and Geman, D. (1992). A nonlinear filter for film restoration and other problems in image processing. *CVGIP: Graphical models and image processing*, 54(4):281–289.

Glocker, B., Izadi, S., Shotton, J., and Criminisi, A. (2013). Real-time rgb-d camera relocalization. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 173–179.

Grewal, M. S. and Andrews, A. P. (2010). Applications of kalman filtering in aerospace 1960 to the present [historical perspectives]. *IEEE Control Syst. Mag.*, 30(3):69–78.

Haarnoja, T., Ajay, A., Levine, S., and Abbeel, P. (2016). Backprop KF: Learning discriminative deterministic state estimators. In *Proc. Advances in Neural Inform. Process. Syst. (NIPS)*.

Handa, A., Bloesch, M., Pătrăucean, V., Stent, S., McCormac, J., and Davison, A. (2016). gvnn: Neural network library for geometric computer vision. In *Computer Vision – ECCV 2016 Workshops*, pages 67–82. Springer, Cham.

Hartley, R., Trumpf, J., Dai, Y., and Li, H. (2013). Rotation averaging. *Int. J. Comput. Vis.*, 103(3):267–305.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6(9):813–827.

Hu, H. and Kantor, G. (2015). Parametric covariance prediction for heteroscedastic noise. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst. (IROS)*, pages 3052–3057.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proc. ACM Int. Conf. Multimedia (MM)*, pages 675–678.

Kelly, J., Saripalli, S., and Sukhatme, G. S. (2008). Combined visual and inertial navigation for an unmanned aerial vehicle. In *Proc. Field and Service Robot. (FSR)*, pages 255–264.

Kendall, A. and Cipolla, R. (2016). Modelling uncertainty in deep learning for camera relocalization. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 4762–4769.

Kendall, A. and Cipolla, R. (2017). Geometric loss functions for camera pose regression with deep learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6555–6564.

Kendall, A., Grimes, M., and Cipolla, R. (2015). PoseNet: A convolutional network for Real-Time 6-DOF camera relocalization. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2938–2946.

Kerl, C., Sturm, J., and Cremers, D. (2013). Robust odometry estimation for RGB-D cameras. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 3748–3754.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 6402–6413. Curran Associates, Inc.

Lalonde, J.-F., Efros, A. A., and Narasimhan, S. G. (2011). Estimating the natural illumination conditions from a single outdoor image. *Int. J. Comput. Vis.*, 98(2):123–145.

Lambert, A., Furgale, P., Barfoot, T. D., and Enright, J. (2012). Field testing of visual odometry aided by a sun sensor and inclinometer. *J. Field Robot.*, 29(3):426–444.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D., and Batra, D. (2015). Why M heads are better than one: Training a diverse ensemble of deep networks.

Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., and Furgale, P. (2015). Keyframe-based visual–inertial odometry using nonlinear optimization. *Int. J. Rob. Res.*, 34(3):314–334.

Levine, S., Finn, C., Darrell, T., and Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.*

Li, Q., Qian, J., Zhu, Z., Bao, X., Helwa, M. K., and Schoellig, A. P. (2017a). Deep neural networks for improved, impromptu trajectory tracking of quadrotors. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 5183–5189.

Li, R., Wang, S., Long, Z., and Gu, D. (2017b). UnDeepVO: Monocular visual odometry through unsupervised deep learning.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Ma, W.-C., Wang, S., Brubaker, M. A., Fidler, S., and Urtasun, R. (2016). Find your way by observing the sun and other semantic cues.

MacTavish, K. and Barfoot, T. D. (2015). At all costs: A comparison of robust cost functions for camera correspondence outliers. In *Proc. Conf. on Comp. and Robot Vision (CRV)*, pages 62–69.

Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2016). 1 year, 1000 km: The oxford RobotCar dataset. *Int. J. Rob. Res.*

Maimone, M., Cheng, Y., and Matthies, L. (2007). Two years of visual odometry on the mars exploration rovers. *J. Field Robot.*, 24(3):169–186.

Mayor, A. (2019). *Gods and Robots*. Princeton University Press.

McManus, C., Upcroft, B., and Newman, P. (2014). Scene signatures: Localised and point-less features for localisation. In *Proc. Robotics: Science and Systems X*.

Melekhov, I., Ylioinas, J., Kannala, J., and Rahtu, E. (2017). Relative camera pose estimation using convolutional neural networks. In *Proc. Int. Conf. on Advanced Concepts for Intel. Vision Syst.*, pages 675–687. Springer.

Oliveira, G. L., Radwan, N., Burgard, W., and Brox, T. (2017). Topometric localization with deep learning. *arXiv preprint arXiv:1706.08775*.

Olson, C. F., Matthies, L. H., Schoppers, M., and Maimone, M. W. (2003). Rover navigation using stereo ego-motion. *Robot. Auton. Syst.*, 43(4):215–229.

Osband, I., Blundell, C., Pritzel, A., and Van Roy, B. (2016). Deep exploration via bootstrapped DQN. In *Proc. Advances in Neural Inform. Process. Syst. (NIPS)*, pages 4026–4034.

Peretroukhin, V., Clement, L., Giamou, M., and Kelly, J. (2015a). PROBE: Predictive robust estimation for visual-inertial navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*, pages 3668–3675, Hamburg, Germany.

Peretroukhin, V., Clement, L., and Kelly, J. (2015b). Get to the point: Active covariance scaling for feature tracking through motion blur. In *Proceedings of the IEEE International Conference on Robotics and Automation Workshop on Scaling Up Active Perception*, Seattle, Washington, USA.

Peretroukhin, V., Clement, L., and Kelly, J. (2017). Reducing drift in visual odometry by inferring sun direction using a bayesian convolutional neural network. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'17)*, pages 2035–2042, Singapore.

Peretroukhin, V., Clement, L., and Kelly, J. (2018). Inferring sun direction to improve visual odometry: A deep learning approach. *International Journal of Robotics Research*, 37(9):996–1016.

Peretroukhin, V. and Kelly, J. (2018). DPC-Net: Deep pose correction for visual localization. *IEEE Robotics and Automation Letters*, 3(3):2424–2431.

Peretroukhin, V., Kelly, J., and Barfoot, T. D. (2014). Optimizing camera perspective for stereo visual odometry. In *Canadian Conference on Comp. and Robot Vision*, pages 1–7.

Peretroukhin, V., Vega-Brown, W., Roy, N., and Kelly, J. (2016). PROBE-GK: Predictive robust estimation using generalized kernels. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 817–824.

Peretroukhin, V., Wagstaff, B., and Kelly, J. (2019). Deep probabilistic regression of elements of SO(3) using quaternion averaging and uncertainty injection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'19) Workshop on Uncertainty and Robustness in Deep Visual Learning*, pages 83–86, Long Beach, California, USA.

Punjani, A. and Abbeel, P. (2015). Deep learning helicopter dynamics models. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 3223–3230.

Rosen, D. M., Carlone, L., Bandeira, A. S., and Leonard, J. J. (2019). SE-Sync: A certifiably correct algorithm for synchronization over the special euclidean group. *Int. J. Rob. Res.*, 38(2-3):95–125.

Scaramuzza, D. and Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE Robot. Autom. Mag.*, 18(4):80–92.

Sibley, G., Matthies, L., and Sukhatme, G. (2007). Bias reduction and filter convergence for long range stereo. In *Robotics Research*, pages 285–294. Springer Berlin Heidelberg.

Sola, J. (2017). Quaternion kinematics for the error-state kalman filter. *arXiv preprint arXiv:1711.02508*.

Solà, J., Deray, J., and Atchuthan, D. (2018). A micro lie theory for state estimation in robotics.

Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B., and Milford, M. (2015). On the performance of ConvNet features for place recognition. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst. (IROS)*, pages 4297–4304.

Sunderhauf, N., Shirazi, S., Jacobson, A., Dayoub, F., Pepperell, E., Upcroft, B., and Milford, M. (2015). Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free. In *Proc. Robotics: Science and Systems XII*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition, (CVPR)*, pages 1–9.

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle Adjustment — A Modern Synthesis. In Goos, G., Hartmanis, J., van Leeuwen, J., Triggs, B., Zisserman,

A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, volume 1883, pages 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg.

Tsotsos, K., Chiuso, A., and Soatto, S. (2015). Robust inference for visual-inertial sensor fusion. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 5203–5210.

Umeyama, S. (1991). Least-Squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(4):376–380.

Vega-Brown, W. and Roy, N. (2013). CELLO-EM: Adaptive sensor models without ground truth. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pages 1907–1914.

Vega-Brown, W. R., Doniec, M., and Roy, N. G. (2014). Nonparametric Bayesian inference on multivariate exponential families. In *Proc. Advances in Neural Information Proc. Syst. (NIPS) 27*, pages 2546–2554.

Wang, S., Clark, R., Wen, H., and Trigoni, N. (2017). DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2043–2050.

Yang, F., Choi, W., and Lin, Y. (2016). Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 2129–2137.

Yang, N., Wang, R., Stueckler, J., and Cremers, D. (2018). Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry. In *European Conference on Computer Vision (ECCV)*. accepted as oral presentation, arXiv 1807.02570.

Zhang, G. and Vela, P. (2015). Optimally observable and minimal cardinality monocular SLAM. In *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, pages 5211–5218.

Zhou, B., Krähenbühl, P., and Koltun, V. (2019). Does computer vision matter for action?

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in Neural Inform. Process. Syst. (NIPS)*, pages 487–495.

Zhou, T., Brown, M., Snavely, N., and Lowe, D. G. (2017). Unsupervised learning of depth and Ego-Motion from video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6612–6619.