Linköping studies in science and technology. Dissertations.
No. 1533

# Navigation and Mapping for Aerial Vehicles Based on Inertial and Imaging Sensors

**Zoran Sjanic**

Department of Electrical Engineering
Linköping University, SE–581 83 Linköping, Sweden

Linköping 2013

**Cover illustration:** Background constitutes of a composite image of Washington D.C. where an optical image from Google Maps is used in the upper-right and a SAR image from Sandia National Laboratories is used in the lower-left corner. For the foreground, a front view image of the rotor Unmanned Aerial Vehicle Saab V-200 Skeldar is used.

Linköping studies in science and technology. Dissertations.
No. 1533

**Navigation and Mapping for Aerial Vehicles Based on Inertial and Imaging Sensors**

Zoran Sjanic

`zoran@isy.liu.se`
`www.control.isy.liu.se`
*Division of Automatic Control*
*Department of Electrical Engineering*
*Linköping University*
*SE–581 83 Linköping*
*Sweden*

ISBN 978-91-7519-553-7 ISSN 0345-7524

Copyright © 2013 Zoran Sjanic

Printed by LiU-Tryck, Linköping, Sweden 2013

*Za Almu*

# Abstract

Small and medium sized Unmanned Aerial Vehicles (UAV) are today used in military missions, and will in the future find many new application areas such as surveillance for exploration and security. To enable all these foreseen applications, the UAV's have to be cheap and of low weight, which restrict the sensors that can be used for navigation and surveillance. This thesis investigates several aspects of how fusion of navigation and imaging sensors can improve both tasks at a level that would require much more expensive sensors with the traditional approach of separating the navigation system from the applications. The core idea is that vision sensors can support the navigation system by providing odometric information of the motion, while the navigation system can support the vision algorithms, used to map the surrounding environment, to be more efficient. The unified framework for this kind of approach is called Simultaneous Localisation and Mapping (SLAM) and it will be applied here to inertial sensors, radar and optical camera.

Synthetic Aperture Radar (SAR) uses a radar and the motion of the UAV to provide an image of the microwave reflectivity of the ground. SAR images are a good complement to optical images, giving an all-weather surveillance capability, but they require an accurate navigation system to be focused which is not the case with typical UAV sensors. However, by using the inertial sensors, measuring UAV's motion, and information from the SAR images, measuring how image quality depends on the UAV's motion, both higher navigation accuracy and, consequently, more focused images can be obtained. The fusion of these sensors can be performed in both batch and sequential form. For the first approach, we propose an optimisation formulation of the navigation and focusing problem while the second one results in a filtering approach. For the optimisation method the measurement of the focus in processed SAR images is performed with the image entropy and with an image matching approach, where SAR images are matched to the map of the area. In the proposed filtering method the motion information is estimated from the raw radar data and it corresponds to the time derivative of the range between UAV and the imaged scene, which can be related to the motion of the UAV.

Another imaging sensor that has been exploited in this framework is an ordinary optical camera. Similar to the SAR case, camera images and inertial sensors can also be used to support the navigation estimate and simultaneously build a three-dimensional map of the observed environment, so called inertial/visual SLAM. Also here, the problem is posed in optimisation framework leading to batch Maximum Likelihood (ML) estimate of the navigation parameters and the map. The ML problem is solved in both the straight-forward way, resulting in nonlinear least squares where both map and navigation parameters are considered as parameters, and with the Expectation-Maximisation (EM) approach. In the EM approach, all unknown variables are split into two sets, hidden variables and actual parameters, and in this case the map is considered as parameters and the navigation states are seen as hidden variables. This split enables the total problem to be solved computationally cheaper then the original ML formulation. Both optimisation problems mentioned above are nonlinear and non-convex requiring good initial solution in order to obtain good parameter estimate. For this purpose a method for initialisation of inertial/visual SLAM is devised where the conditional linear structure of the problem is used to obtain the initial estimate of the parameters. The benefits and performance improvements of the methods are illustrated on both simulated and real data.

# Populärvetenskaplig sammanfattning

Obemannade flygande farkoster, eller Unmanned Aerial Vehicles (UAV) på engelska, är en typ av farkoster där operatören inte sitter ombord utan styr den från marken. Dessa används idag, i huvudsak, av militären för olika typer av uppdrag, men det finns en stor potential för framtida civil användning, speciellt inom övervaknings- och kartläggningsuppdrag. Dessutom ger det faktum att ingen människa sitter i UAV:en en viktig fördel över bemannade farkoster, nämligen att även uppdrag som skulle kunna vara farliga nu kan utföras utan att liv riskeras. För att göra UAV-operatörer ännu effektivare i utförandet av uppdragen, måste farkosterna bli mer och mer autonoma, dvs. ha förmågan att själva utföra vissa uppgifter, som till exempel flyga en fördefinierad rutt. En kritisk komponent i detta är navigeringssystemet, som har som huvuduppgift att beräkna och förse olika andra system med farkostens position, höjd, fart och orientering. För att göra detta, används olika mätsensorer som mäter UAV:ens rörelse eller dess läge i förhållande till andra objekt med kända positioner. Kraven på navigeringssystemet för att kunna klara av att ge stöd till övriga system är att det måste ha en hög noggrannhet och tillförlitlighet.

Idag finns det navigeringshjälpmedel som ger möjlighet till en väldigt noggrann navigering, som det globala satellit navigeringssystemet, eller Global Positioning System (GPS) på engelska. Det systemet använder satelliter som kretsar runt jorden för att bestämma sin egen position. Ett problem som dessa har är att tillgången till dem inte alltid kan förutsättas beroende på olika omständigheter. Satelliter kanske inte syns från nuvarande position eller är det någon som avsiktligt stör ut dessa signaler. I sådana situationer är ett navigeringssystem som är oberoende av externa hjälpmedel högst önskvärt.

Denna avhandling presenterar olika metoder som, genom att använda information från olika sensorer som en typisk UAV bär med sig, skapar ett navigeringssystem som kan klara de krav som ställs på det. Sensorerna som används är främst bildalstrande sensorer, såsom optisk kamera och radar, men även tröghetssensor som mäter UAV:ens rörelse. Metoderna bygger på att de bildalstrande sensorerna observerar omgivningen UAV:en flyger i, och på så sätt räknar ut hur den har rört sig. Detta löses med ett matematisk ramverk som kallas optimering, där det gäller att hitta den bästa lösningen till det ställda problemet.

En annan fördel sensorer såsom kamera eller radar har är att med hjälp av dessa kan kartor över omgivningen skapas samtidigt som man löser navigeringsproblemet. Denna procedur kallas Simultaneous Localisation and Mapping, eller förkortat SLAM, på engelska och är huvudmetoden som utnyttjas i denna avhandling. De resulterande kartorna kan användas för olika syften, t.ex. att skapa en karta över ett landskap där förändringar av omgivningen kan ha skett som efter en naturkatastrof. De resultat som har erhållits visar att de använda metoderna har en betydande potential att användas i praktiken för att skapa ett robust navigeringssystem.

# Acknowledgments

First of all, I would like to thank all the people that make this little thing that I have done possible, Prof. Fredrik Gustafsson, my supervisor, and Dr. Thomas Schön, my co-supervisor, for all the guidance and help during this journey. Also, my thanks go to Prof. Lennart Ljung and people at Saab Aeronautics in Linköping for giving me this opportunity in the first place. All my former and present bosses; Dr. Predrag Pucar, Niklas Ferm, Jonas Palm, Björn Kullberg, Tobias Jansson and research responsible Dr. Gunnar Holmberg. Your support during this time was priceless, and don't worry, soon I'll be back to actually do something useful. Thanks go to the Control Theory group's boss Prof. Svante Gunnarsson for taking such a good care of the group and to all the administrators Ulla Salaneck, Åsa Karmelind and Ninna Stensgård for taking care of all the practicalities.

During this long journey, you can't survive alone, and that is why some other people deserve my thanks. Naturally, people at the Control Theory group with which I, in some way, worked with, Dr. Martin Enqvist, Lic. Roger Larsson, Dr. Christian Lundquist, Dr. David Törnqvist, Dr. Fredrik Gunnarsson, Dr. Carsten Fritsche, Dr. Umut Orguner and especially "my partner in crime" Lic. Martin "Morgan" Skoglund. I would also like to thank people from different parts of Saab, Dr. Patrik Dammert, Dr. Hans Hellsten, Dr. Per-Johan Nordlund, Dr. Ola Härkegård and Dr. Sören Mollander for the help and good discussions. It was a pleasure to cooperate with all of you.

And since "all work and no play makes Jack a dull boy", it is important to have fun besides all the work. Therefore it is nice that a whole bunch of people in the Control Theory group know how to have fun. So thanks to (in order of appearance) Dr. Jonas Callmer, Dr. Karl Granström, Dr. Gustaf Hendeby, Dr. Henrik Ohlsson, Dr. Christian Lundquist, Dr. Christian Lyzell, Dr. Emre Özkan, Dr. Henrik Tidefelt, Dr. Ragnar Wallin, Dr. Daniel Ankelhed, Lic. Patrik Axelsson, Lic. André Carvalho Bittencourt, Niclas Evestedt, Lic. Rikard Felkeborn, Ylva Jung, Lic. Sina "my name is to long to fit on the Amex card" Koshfetrat Pakazad, Lic. Roger Larsson, Jonas Linder, George Mathai, Isak Nielsen, Hanna Nyqvist, Michael Roth, Marek Syldatk, Lic. Niklas Wahlström. And Jonas, Kalle and Christian Lu., thanks for sharing so many nights out both in Sweden and elsewhere.

Special thanks go to people that shared the working room with me during all these years (and actually kept their mental sanity(?!)) Lic. Daniel Petersson, Lic. Fredrik "quiz" Lindsten and Manon Kok. Thanks for making our room a fun and nice place to work in.

All this work would be impossible without an infinite support from my family and friends, thanks for all the support. And Alma, I know that this took toooooo much time but it is soon finished $\ddot{\smile}$, so thanks for having patience and trust in me during all this time ♡.

Last, but not least, thanks to the LINK-SIC Industry Excellence Center and Saab for financing my studies.

*Linköping, September 2013*
*Zoran Sjanic*

# Contents

**Notation** **xv**

## I Background

**1 Introduction** **3**
- 1.1 Background and Motivation . . . 3
- 1.2 Contributions . . . 6
  - 1.2.1 Included Publications . . . 6
  - 1.2.2 Other Publications . . . 9
- 1.3 Thesis outline . . . 9

**2 Estimation and Sensor Fusion Framework** **11**
- 2.1 Estimation Theory for Dynamical Systems . . . 11
- 2.2 Simultaneous Localisation and Mapping . . . 15
  - 2.2.1 SLAM System Models . . . 15

**3 Navigation System and Sensors** **21**
- 3.1 UAV Navigation System . . . 21
- 3.2 Inertial Measurement Unit . . . 23
- 3.3 Synthetic Aperture Radar . . . 26
  - 3.3.1 Real and Synthetic Aperture Radar . . . 26
  - 3.3.2 Measuring Focus . . . 32
  - 3.3.3 Evaluation on Test Optical Image . . . 35
  - 3.3.4 Evaluation on Test SAR Images . . . 35
- 3.4 Optical Camera . . . 51
  - 3.4.1 Projective Geometry Model . . . 51
  - 3.4.2 Image Feature Extraction . . . 56

**4 Concluding Remarks** **59**
- 4.1 Conclusions . . . 59
- 4.2 Future Work . . . 60

**Bibliography** **61**

# II Publications

**A Simultaneous Navigation and Synthetic Aperture Radar Focusing** **67**
1 Introduction . . . . . 69
2 Notation and Problem Formulation . . . . . 72
3 Navigation Framework . . . . . 75
  3.1 Aircraft Model . . . . . 76
  3.2 Navigation Performance . . . . . 76
4 Focus Measures . . . . . 77
  4.1 Two Entropy Measures . . . . . 77
  4.2 Focus Measure Performance . . . . . 77
5 Search Methods . . . . . 81
  5.1 Joint Optimisation of Trajectory and Focus . . . . . 81
  5.2 Gradient Search . . . . . 81
  5.3 Calculating the Gradient . . . . . 82
6 Numerical Examples for Simulated Images . . . . . 85
  6.1 Two-Dimensional Optimisation . . . . . 85
  6.2 High-Dimensional Optimisation . . . . . 87
7 Example with Real SAR Image . . . . . 91
8 Conclusions . . . . . 95
Bibliography . . . . . 96

**B Navigation and SAR focusing with Map Aiding** **99**
1 Introduction . . . . . 101
2 SAR Imaging Principle . . . . . 104
3 Motion Models . . . . . 106
  3.1 SAR Geometry . . . . . 107
4 Image Matching Approach . . . . . 108
  4.1 Edge Detector . . . . . 109
  4.2 Chamfer Image Matching . . . . . 110
  4.3 Modified Matching Approach . . . . . 112
5 Kinematic Parameter Estimation . . . . . 113
6 Results . . . . . 113
  6.1 Results of the Image Matching Approach . . . . . 114
  6.2 Results for the Kinematic Parameters Estimation . . . . . 115
  6.3 Discussion . . . . . 120
7 Conclusions and Future Work . . . . . 120
Bibliography . . . . . 123

**C Navigation and SAR Auto-focusing Based on the Phase Gradient Approach** **127**
1 Introduction . . . . . 129
2 Sensor Fusion Framework . . . . . 133
3 Phase Gradient Auto-focusing Method . . . . . 135
  3.1 Basics of the PG method . . . . . 135

3.2 Estimating the Phase/Range Gradient . . . . . . . . . . . . . . . 136
4 EKF Auto-focusing and Evaluation of the Performance . . . . . . . . . 139
4.1 Extended Kalman Filter . . . . . . . . . . . . . . . . . . . . . 139
4.2 Measurement Model . . . . . . . . . . . . . . . . . . . . . . . 140
4.3 Numerical Results . . . . . . . . . . . . . . . . . . . . . . . 140
5 Conclusions and Future Work . . . . . . . . . . . . . . . . . . . . . 145
Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146

**D A Nonlinear Least-Squares Approach to the SLAM Problem 149**
1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 151
2 Problem Formulation . . . . . . . . . . . . . . . . . . . . . . . . . 152
3 Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 153
3.1 Dynamics . . . . . . . . . . . . . . . . . . . . . . . . . . . . 153
3.2 Landmark State Parametrisation . . . . . . . . . . . . . . . . . 154
3.3 Camera Measurements . . . . . . . . . . . . . . . . . . . . . . 155
4 Solution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
4.1 Initialisation . . . . . . . . . . . . . . . . . . . . . . . . . . 156
4.2 Nonlinear Least-Squares Smoothing . . . . . . . . . . . . . . . 156
5 Experiments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 158
5.1 Experimental Setup . . . . . . . . . . . . . . . . . . . . . . . 160
5.2 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 160
6 Conclusions and Future Work . . . . . . . . . . . . . . . . . . . . . 161
Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 164

**E Initialisation and Estimation Methods for Batch Optimisation of Inertial/Visual SLAM 167**
1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 169
2 Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 173
2.1 Position and Orientation . . . . . . . . . . . . . . . . . . . . 173
2.2 IMU Measurements . . . . . . . . . . . . . . . . . . . . . . . 173
2.3 Camera Measurements . . . . . . . . . . . . . . . . . . . . . . 174
3 SLAM Initialisation . . . . . . . . . . . . . . . . . . . . . . . . . 174
3.1 Feature Tracks . . . . . . . . . . . . . . . . . . . . . . . . . 175
3.2 Track Clustering . . . . . . . . . . . . . . . . . . . . . . . . 176
3.3 Rotation Initialisation . . . . . . . . . . . . . . . . . . . . . 177
3.4 Linear SLAM . . . . . . . . . . . . . . . . . . . . . . . . . . 178
3.5 Iterative Outlier Removal . . . . . . . . . . . . . . . . . . . . 181
4 Nonlinear Least-Squares SLAM . . . . . . . . . . . . . . . . . . . . 182
5 Heuristic Motivation of the Linear Initialisation . . . . . . . . . . . 182
6 Monte Carlo Simulations . . . . . . . . . . . . . . . . . . . . . . . 183
6.1 Efficiency of the Linear Initialisation . . . . . . . . . . . . . 184
6.2 Sensitivity to Initial Rotation Errors . . . . . . . . . . . . . . 184
6.3 Iterative Outlier Removal . . . . . . . . . . . . . . . . . . . . 185
7 Real Data Experiments . . . . . . . . . . . . . . . . . . . . . . . . 185
7.1 Clustering Results . . . . . . . . . . . . . . . . . . . . . . . 190
8 Conclusions and Future Work . . . . . . . . . . . . . . . . . . . . . 194

Bibliography . . . . . . . . . . 199

**F Expectation-Maximisation Maximum Likelihood Estimation for Inertial/Visual SLAM 203**
1 Introduction . . . . . . . . . . 205
2 Expectation Maximisation . . . . . . . . . . 207
3 EM-SLAM . . . . . . . . . . 208
3.1 E-step . . . . . . . . . . 209
3.2 M-step . . . . . . . . . . 209
4 Models . . . . . . . . . . 210
4.1 IMU Parametrisation . . . . . . . . . . 210
4.2 Camera Measurements . . . . . . . . . . 212
5 Nonlinear Least-Squares . . . . . . . . . . 213
6 Computation Complexity . . . . . . . . . . 214
7 Obtaining an Initial Estimate . . . . . . . . . . 215
8 Results . . . . . . . . . . 215
8.1 Simulations . . . . . . . . . . 215
8.2 Real Data Experiments . . . . . . . . . . 218
9 Conclusions and Future Work . . . . . . . . . . 218
Bibliography . . . . . . . . . . 223

**G Cellular Network Non-Line-of-Sight Reflector Localisation Based on Synthetic Aperture Radar Methods 227**
1 Introduction . . . . . . . . . . 229
2 SAR and Multistatic SAR . . . . . . . . . . 230
3 OFDM Signal and SAR Modeling . . . . . . . . . . 233
4 Radio Channel Measurements . . . . . . . . . . 234
5 Results . . . . . . . . . . 235
6 Conclusions and Future Work . . . . . . . . . . 238
Bibliography . . . . . . . . . . 239

# Notation

## ABBREVIATIONS

| Abbreviation | Meaning |
|---|---|
| 3GPP LTE | 3rd Generation Partnership Project Long Term Evaluation |
| ADC | Air Data Computer |
| BFGS | Broyden-Fletcher-Goldfarb-Shanno (quasi-Newton algorithm) |
| CARABAS | Coherent All RAdio BAnd Sensing |
| CML | Constrained Maximum Likelihood (estimate) |
| CMAP | Constrained *Maximum a Posteriori* (estimate) |
| DCT | Discrete Cosine Transform (focus measure) |
| DOF | Degrees Of Freedom |
| EKF | Extended Kalman Filter |
| EM | Expectation-Maximisation |
| E-UTRAN | Evolved Universal Terrestrial Radio Access Network |
| GNSS | Global Navigation Satellite System |
| GPS | Global Positioning System |
| IDP | Inverse Depth Parametrisation |
| IMU | Inertial Measurement Unit |
| INS | Inertial Navigation System |
| KF | Kalman Filter |
| LS | Least-Squares |
| MAP | *Maximum a Posteriori* (estimate) |
| MC | Monte Carlo (simulation) |
| MEMS | Micro Electro-Mechanical Sensors |
| ML | Maximum Likelihood (estimate) |

## ABBREVIATIONS

| Abbreviation | Meaning |
|---|---|
| MV | Minimum Variance (estimate) |
| NLS | Nonlinear Least-Squares |
| PG | Phase Gradient |
| RAR | Real Aperture Radar |
| RMSE | Root Mean Square Error |
| SAR | Synthetic Aperture Radar |
| SIFT | Scale-Invariant Feature Transform |
| SLAM | Simultaneous Localisation and Mapping |
| SML | Sum-Modified-Laplacian (focus measure) |
| SURF | Speeded Up Robust Features |
| TG | Tenengrad (focus measure) |
| UAV | Unmanned Aerial Vehicle |
| VGA | Video Graphics Array |

## SYMBOLS AND OPERATORS

| Notation | Meaning |
|---|---|
| $x_t$ | States vector |
| $y_t$ | Measurements vector |
| $u_t$ | Known inputs vector |
| $f_t(x_t, u_t)$ | System dynamics |
| $h_t(x_t, u_t)$ | Measurement function |
| $w_t$ | Process noise |
| $e_t$ | Measurement noise |
| $Q_t$ | Process noise variance |
| $R_t$ | Measurement noise variance or range |
| $\theta$ | Parameter vector |
| $\hat{x}$ | Estimate of $x$ |
| $\mathcal{N}(\mu, P)$ | Gaussian distribution with mean $\mu$ and variance $P$ |
| $x_{0:N}, \mathcal{X}$ | Short notation for $\{x_0,\ x_1,\ \ldots,\ x_N\}$ |
| $p^X, p^Y, p^Z$ | Position in Cartesian coordinates |
| $X, Y, Z$ | Position in Cartesian coordinates (alternative notation) |
| $v^X, v^Y, v^Z$ | Velocity in $X-$, $Y-$ and $Z-$direction, respectively |
| $a^X, a^Y, a^Z$ | Acceleration in $X-$, $Y-$ and $Z-$direction, respectively |
| $q_0, q_1, q_2, q_3$ | Quaternions |
| $\mathsf{R}$ | Rotation matrix |
| $\omega^X, \omega^Y, \omega^Z$ | Angular velocity around $X-$, $Y-$ and $Z-$direction, respectively |
| $I$ | Optical or (complex-valued) SAR image |
| $\mathbb{R}$ | Set of real numbers |
| $\mathcal{O}$ | Ordo, in order of |
| $\dot{x}_t$ | Time derivative of $x_t$ |
| $\sim$ | Is distributed according to |
| $\|x\|_{P^{-1}}$ | $P^{-1}-$weighted norm of $x$, $\sqrt{x^T P^{-1} x}$ |
| $x^*$ | Complex conjugate of $x$ |
| $x^T$ | Transpose of a vector or a matrix |
| $\arg\min_x$ | Minimising argument with respect to $x$ |
| $\arg\max_x$ | Maximising argument with respect to $x$ |
| **s. t.** | subject to |

## SYMBOLS AND OPERATORS

| Notation | Meaning |
|---|---|
| $\in$ | Is member of |
| $\notin$ | Is not member of |
| diag$\{a, b, c\}$ | $3 \times 3$ matrix with elements $a,\ b$ and $c$ on the diagonal and zeros otherwise |
| $\lvert x \rvert$ | Absolute value of $x$ or magnitude of a complex number $x$ |
| $\mathbf{dim}(x)$ | Dimension of $x$ |
| arg$\{x\}$ | Argument of a complex number $x$ |
| $\Re\{x\}$ | Real part of a complex number $x$ |
| $\Im\{x\}$ | Imaginary part of a complex number $x$ |
| $A * B$ | Convolution of matrices $A$ and $B$ |
| Tr$(A)$ | Trace of a matrix $A$ |
| $\det(A)$ | Determinant of a matrix $A$ |
| $\frac{\partial}{\partial x} f(x, y)$ | Partial derivative of $f(x, y)$ with respect to $x$ |

# Part I

# Background

# 1

# Introduction

In this introductory chapter a background to the work performed in this thesis and a motivation for it will be presented. Also, a brief introduction to Unmanned Aerial Vehicles is provided. Furthermore, all the contributions the author has been involved in are listed and briefly described. This list is divided between publications that are included in the Part II of the thesis and the ones that are not. The chapter is finalised with the outline of the thesis.

## 1.1 Background and Motivation

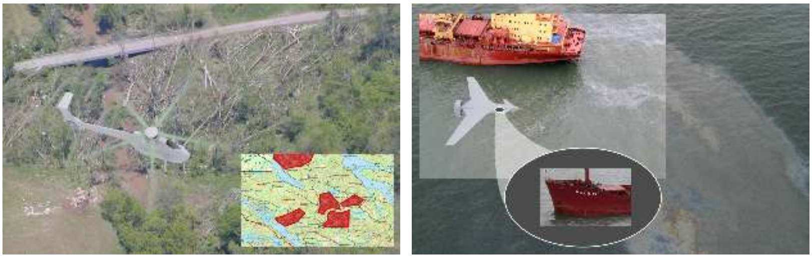
Throughout the history, the exploratory nature of the people has driven them to travel and explore their world. This has led to a need for the art of navigation, i.e., knowing where we are in relation to some known environment. Navigation is used in many different applications, like sailing with boats or flying with aircraft. Many of these applications should not be possible without navigation, like intercontinental flights for example. Navigation also allows for the creation of the systems that can assist human operators in performing different tasks (or even completely remove them), where operating the actual vehicle, like e.g., flying, is the most basic task that the operator can be assisted with. An example of aerial vehicles where a human operator's involvement is partially or completely removed is the Unmanned Aerial Vehicles (UAV). The most obvious thing in the case of UAVs is that the operator is not actually sitting in the UAV, but rather is controlling it remotely from the ground. In this way, the UAVs can carry out missions that might be dangerous and where a risk for a loss of an aircraft is high. Another case might be long reconnaissance missions lasting for 24 hours, where it is impossible to have one pilot or crew to operate for so long, like for Northrop-Grumman Global Hawk, Figure 1.1.

Today, most of the UAVs are used for military purposes, such as reconnaissance of the

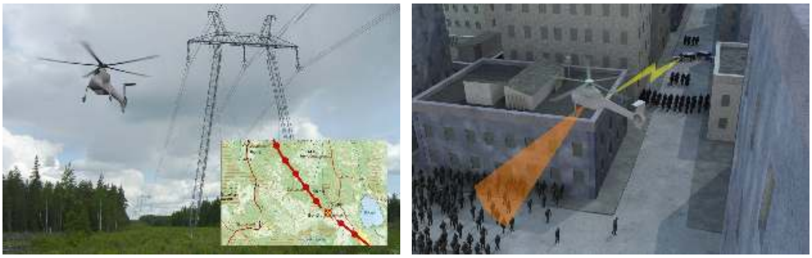
***Figure 1.1:*** *Northrop-Grumman Global Hawk fixed wing UAV. Image: Northrop-Grumman website.*

battlefield or enemy's forces. When it comes to the civilian applications, the UAVs are not present in a bigger scale today. However, there is a huge potential for using the UAVs in the civilian sector, such as police or fire departments or environment surveying missions, like assessing the damage after catastrophes like floods or earthquakes. In Figure 1.2 some of these concepts are illustrated. Actually, saying that UAVs are not used in the civilian sector is, from the technical point of view, not entirely correct. Auto-pilots are used in all commercial airplanes for about $95\%$ of the flight time implying that it is actually the "manned" UAVs that are used every day. The potential is reinforced by the fact that many of the tasks that the UAVs might perform are usually long, boring or, as mentioned, dangerous. Also, an advantage of using UAVs over an manned aircraft is that an operator should be able to concentrate on the mission instead of concentrating on the basic tasks, like flying an aircraft for example. It is therefore crucial to build UAVs with as much autonomy and support in the basic functions as possible. This implies that onboard systems must be able to handle different situations, that may arise during the mission, without the constant monitoring of the operator.

One of the most important parts of the total system to enable this is the navigation system of the UAV. The navigation system provides all the important quantities needed for the UAV to operate autonomously. In order to be able to do this with high reliability, the navigation system must have high accuracy and robustness. This could be realised with the navigation aids that give high quality navigation performance, like Global Navigation Satellite Systems (GNSS) of which Nav-Star Global Positioning System (GPS) (GPS, 2008) is the most famous one. However, the availability of these external navigation aids cannot be assumed in many applications. This can be caused by different things, like occlusions rendering the reception of the satellite signals impossible or even mali-



*(a) Disaster damage assessing.* *(b) Oil spill identification.*



*(c) Power lines surveying.* *(d) Riots monitoring.*

***Figure 1.2:*** *Some of the examples of UAV usage areas in the civilian sector. Images: © Saab AB.*

cious jamming of the signals by opposing military force. In all these cases, some kind of navigation system that is independent of the external systems, like satellites, is highly desirable. Such problem can, in general, be solved with expensive and usually big and heavy components, e.g., high-grade Inertial Navigation Systems (INS), like it is done in commercial or fighter aircraft. Since another reason for using UAVs is that they should be cheaper and smaller than manned aircraft, this is not a very good option. An example of this is Saab V-200 Skeldar rotor UAV, Figure 1.3, which weighs $200\ \mathrm{kg}$ and has a length of approximately $4\ \mathrm{m}$. In this case the less the components weigh the more fuel can be carried, which in turn gives longer mission time. The conclusion is that tactical performance, like endurance, puts the requirements on the choice of the components and the design of the UAV. The end result is that the used equipment, including the navigation system, must be cheap (in this context cheap means a couple of thousand USD class), small and lightweight. These constraints are often in conflict with the previously stated requirements on accuracy and robustness.

In the cases where UAVs are used for surveillance and reconnaissance missions, it is important to create an overview of the environment where the operation is carried out. For example, during a flood the configuration of the landscape changes and it is important to obtain the latest information about the status of the environment. In the more military oriented applications, an UAV can be used to create an up-to-date map of the interesting area for tactical evaluation of the mission or if the prior map was not available. Furthermore, the maps can also be used for navigation purposes, if the UAV carries sensors that can

***Figure 1.3:*** *Saab V-200 Skeldar rotor UAV. Image: © Saab AB. Photographer: Stefan Kalm.*

observe the environment and in that way relate these observations to the maps, serving as an alternative to the high-precision external aids, like GPS.

The work in this thesis is motivated by the above-mentioned aspects of the navigation systems in UAVs. Due to restrictions on weight and size of the equipment in small UAVs, high-performing inertial navigation systems, which are both large in size and heavy, are not a feasible option. Less accurate inertial navigation systems are prone to drift in navigation parameters making them dependent on external aids. Fortunately, UAVs are often equipped with imaging sensors like optical camera or imaging radar, usually due to the mission requirements, allowing us to use these and support the navigation system. Since these sensors primarily observe the surrounding environment, it is also possible to actually create the map of this environment at the same time as localising the UAV. This procedure is known as Simultaneous Localisation and Mapping, or SLAM (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006), and will be used as an unified framework in order to perform stable navigation and mapping of the environment. These maps can be used for different purposes, ranging from tactical evaluation of the terrain to creating high-fidelity maps or even three-dimensional models of the previously uncharted areas. They can in turn be later used to facilitate the navigation for other vehicles.

## 1.2 Contributions

### 1.2.1 Included Publications

Articles included in this thesis are listed below in chronological order together with a short description and contribution of each paper.

> Z. Sjanic and F. Gustafsson. Navigation and SAR Auto-focusing Based on the Phase Gradient Approach. In *Proceedings of 14th International Confer-*

> *ence on Information Fusion*, Chicago, USA, July 2011.

This article, **Paper C**, describes a method for iterative trajectory refinement and UHF SAR auto-focusing based on the filtering framework. Inertial measurements (accelerations) are used together with an estimate of the range time derivative, or phase gradient, in an Extended Kalman Filter (EKF) to obtain a stable estimate of the position. The main contribution is formulation and incorporation of the phase (or range) derivative into the filtering framework enabling sequential solution of the navigation and auto-focusing problem. Furthermore, the importance of using the exact expression for the range and its derivative in UHF SAR case is also shown.

> Z. Sjanic, M. A. Skoglund, T. B. Schön, and F. Gustafsson. A Nonlinear Least-Squares Approach to the SLAM Problem. In *Proceedings of 18th IFAC World Congress*, Milano, Italy, August/Septemeber 2011.

In this paper, which is **Paper D**, a *Maximum a posteriori* (MAP) batch formulation of the inertial/visual SLAM is proposed resulting in a nonlinear least squares optimisation problem. The inherent structure and sparsity of the resulting NLS problem is utilised to efficiently obtain the metrically correct solution to the 6-Degrees-of-Freedom (6-DoF) estimate of the platform's states, i.e., position, velocity and orientation, and a three dimensional map of the environment. Also, for the parametrisation of the landmark location an Inverse Depth Parametrisation (IDP) is used which also gives the special structure to the NLS problem. The benefit of IDP, which uses six parameters, over the regular three-parameter parametrisation is faster convergence of the landmark depth if an EKF is used to estimate the states and the map. Here the EKF is applied to the initialisation procedure of the estimated parameters and to obtain the data association.

> Z. Sjanic and F. Gustafsson. Simultaneous Navigation and Synthetic Aperture Radar Focusing. *Provisionally accepted to IEEE Transactions on Aerospace and Electronic Systems*, August 2013a.

An optimisation based method for simultaneous navigation and SAR auto-focusing is proposed in this contribution, **Paper A**, resulting in basically a Maximum Likelihood (ML) formulation. Here, the accelerations sensors are used together with SAR images, utilising image focus measures, image entropy in particular, to obtain the likelihood function. This likelihood is quite a complicated function of the optimisation variables, which are here the platform's states, making the direct gradient calculation difficult. However, as an important part of the article's contribution, an efficient way of calculating this gradient is devised, which is based on the chain rule for differentiation. The solution is made even more efficient by using quasi-Newton method where the Hessian matrix is estimated from gradient information during the iterations.

> Z. Sjanic and F. Gustafsson. Navigation and SAR focusing with Map Aiding. *Submitted to IEEE Transactions on Aerospace and Electronic Systems*, June 2013b.

This paper, **Paper B**, devises a method for solving the navigation and SAR focusing problem with an aid from prior information in form of optical maps. The basic idea is based on the image matching approach, called Chamfer matching, which is, in this contribution, reformulated in the nonlinear least squares optimisation framework. With this approach it is possible to solve for both the global position and direction of flight, since the map has known global coordinates, and for the flight trajectory, since the assumption is that the most focused image will give the matching cost with the smallest value.

> M. A. Skoglund, Z. Sjanic, and F. Gustafsson. Initialisation and Estimation Methods for Batch Optimisation of Inertial/Visual SLAM. *Submitted to IEEE Transactions on Robotics*, September 2013.

Inertial/visual SLAM problem formulation in optimisation form, as most optimisation problems, requires an initial value of the optimisation parameters. The nonlinear and non-convex nature of the posed problem may cause suboptimal solutions due to bad initialisation and consequently convergence to a local minimum. In Paper D EKF is used to initialise the parameters and data association, but that solution is not feasible for large data sets due to the cubic complexity for calculating the matrix inverse involved in the EKF iteration. This article, **Paper E**, contributes with a multi-stage procedure for the initialisation of the nonlinear inertial/visual SLAM optimisation problem. The approach is based on both pure visual methods, like image based rotation estimation and appearance based data association, and combined inertial/visual methods, where an almost linear formulation of the problem is utilised to efficiently obtain an initial solution and data association. It is also demonstrated that this estimate leads to a better initial values than straight forward initialisation with measurements only.

> Z. Sjanic, M. A. Skoglund, and F. Gustafsson. Expectation-Maximisation Maximum Likelihood Estimation for Inertial/Visual SLAM. *Submitted to IEEE Transactions on Robotics*, September 2013b.

Here, **Paper F**, an Expectation-Maximisation (EM) approach to the ML formulation of the SLAM problem is proposed. The EM framework is usually used when the ML problem is difficult to solve by introducing a new set of unknowns, so called hidden variables. The problem is then split into two subproblems, Expectation step (E-step) where hidden variables are estimated conditioned on the old parameters and Maximisation step (M-step) where parameters are estimated based on the hidden variables from the E-step by solving an optimisation problem. This framework is highly applicable to inertial/visual SLAM where the platform's states are seen as hidden variables and the map is represented as parameters. In this particular case, the E-step is approximately, but efficiently, solved with Extended Rauch-Tung-Striebel Smoother and the M-step is, also efficiently, solved with quasi-Newton method. Since the smoother step is computationally cheap and the optimisation problem in the M-step has fewer variables than the full ML formulation, the total computational time can be lower for the big problems.

> Z. Sjanic, F. Gunnarsson, C. Fritsche, and F. Gustafsson. Cellular Network Non-Line-of-Sight Reflector Localisation Based on Synthetic Aperture Radar Methods. *Submitted to IEEE Transactions on Antennas and Propagation*, September 2013a.

This short paper, **Paper G**, contributes with an application of a nonparametric multi-static SAR technique to cellular network data. The data consists of measured range to base stations which contains both line-of-sight and non-line-of-sight signals between the phone and the base stations, the latter ones caused by the signal reflections from the buildings in the urban environment. This data is used to estimate a map (or a multi-static SAR image) of the strongest reflectors. Furthermore, with this map, it is possible to find reflector-trajectory correspondence which can be used for outlier rejection in a cellular network localisation application.

### 1.2.2 Other Publications

Other articles which are not included in the thesis are listed below. Also here, the description and contribution are described.

> R. Larsson, Z. Sjanic, M. Enqvist, and L. Ljung. Direct Prediction-error Identification of Unstable Nonlinear Systems Applied to Flight Test Data. In *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

In this contribution a system identification method for unstable and nonlinear systems based on direct prediction-error approach is devised. The predictor is based on the nonlinear state space description of the system and calculated using EKF. The methods are demonstrated on estimation of aerodynamic parameters for the fighter aircraft.

> Z. Sjanic and F. Gustafsson. Simultaneous Navigation and SAR Auto-focusing. In *Proceedings of 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.

This article introduces an early concept of solving navigation and SAR focusing problem simultaneously and in the optimisation framework. The ideas presented here can be seen as a predecessor to the much more thorough work done in Paper A.

> Z. Sjanic and F. Gustafsson. Fusion of Information from SAR and Optical Map Images for Aided Navigation. In *Proceedings of 15th International Conference on Information Fusion*, Singapore, July 2012.

In this paper a method for matching of SAR images and optical map in order to aid the navigation is proposed. In this way a global position and direction of flight can be estimated. The method is based on the nonlinear least squares optimisation approach which also allows for the estimate of the parameter covariance. This contribution is the basis for Paper B.

## 1.3 Thesis outline

The thesis is divided into two parts, theoretical and subject background in the first part and edited versions of published papers in the second part. Chapter 1, this chapter, introduces the main problem and the motivation for solving it. It also briefly describes UAVs and lists all the included publications in Part II of the thesis and scientific contributions that are included in these articles. Chapter 2 briefly introduces the estimation theory that is used. UAV navigation systems and all the utilised sensors, i.e., inertial sensors, optical camera and SAR, and their properties are described in Chapter 3. Finally, the first part is concluded with Chapter 4 where some concluding remarks are discussed and suggestions about the future work are given.

# 2

# Estimation and Sensor Fusion Framework

In this chapter an overview the sensor fusion framework with short estimation theory and Simultaneous Localisation and Mapping (SLAM) framework is described, where only the main ideas and concepts are mentioned.

## 2.1 Estimation Theory for Dynamical Systems

The primary task of the sensor fusion function is to utilise the information from all the available sensors and fuse it together in order to improve certain properties of the output entities of interest. It can be that sensors have noisy measurements, imperfections due to biases or low signal fidelity. In some cases, the sensors do not measure quantities that we want directly, and they must be calculated or estimated in some way. The general approach to this problem is called estimation, where sensor fusion is one particular instance. In a way, we can say that the navigation system is an estimator of the navigational quantities, which in turn describe the motion state of the UAV, given some measurements of this motion. Systems in motion are usually denoted dynamical systems, which means that an input to such a system at a particular time will affect future behavior of the system. The opposite case is the static system where the input at the particular time affects the output at that time only.

In this framework, a particularly useful description of the dynamical systems is the state space description, usually in the form

$$x_{t+1} = f_t(x_t, u_t, w_t) \tag{2.1a}$$

$$y_t = h_t(x_t, u_t, e_t) \tag{2.1b}$$

where $x_t$ are the states of the system, usually the quantities that need to be estimated, $u_t$ are the known inputs, $w_t$ is the system noise, $y_t$ are the measurements and $e_t$ is the measurement noise. The function $f_t(\,\cdot\,)$ describes the dynamics of the system and $h_t(\,\cdot\,)$ is a function that relates the measurements and the states of the system. Note that these functions may vary in time. This is a rather general model of a system and one that is often used is

$$x_{t+1} = f_t(x_t, w_t) \tag{2.2a}$$

$$y_t = h_t(x_t) + e_t \tag{2.2b}$$

where noise term in the measurement equation appears in an additive fashion and the known input $u_t$ is omitted. This poses no practical problems since, in most cases, the additive measurement noise is a plausible model and a known input can be modelled through the time dependency of $f_t$ and $h_t$. In the rest of this thesis $u_t$ will mostly be omitted for notational convenience, and be included only when it is necessary. The descriptions of the dynamics and measurements (2.1) and (2.2) are both in discrete time, which is suitable for implementation in computers and because most of the modern sensors deliver data in sampled form. Usually, the system dynamics is dependent on the moving platform, so it can be fixed for each application. In our case the dynamics of the system will be modelled as aircraft dynamics which will be described below.

Another, and a bit more general, way of describing the system (2.2) is in the form of conditional probability density functions, or pdf, for state time transition and measurements

$$x_{t+1} \sim p(x_{t+1}|x_t) \tag{2.3a}$$

$$y_t \sim p(y_t|x_t) \tag{2.3b}$$

Equation (2.3b) is also known as likelihood function. From the system perspective these two descriptions are equivalent. The model above are referred to as Markov process', i.e., the state at time $t$ is only dependent of the state at time $t-1$. In a similar way the measurement at time $t$ is conditionally independent of the states in all times except the state at time $t$. As an example of model (2.3), take equation (2.2b) and suppose that the measurement noise, $e_t$, has Gaussian distribution with zero mean and variance $R_t$, $e_t \sim \mathcal{N}(0, R_t)$. This will, for one time instance, yield

$$p(y_t|x_t) = p_{e_t}(y_t - h(x_t)) = \frac{1}{\det\{2\pi R_t\}^{1/2}} e^{-\frac{1}{2}(y_t - h_t(x_t))^T R_t^{-1}(y_t - h_t(x_t))} \tag{2.4}$$

The estimation problem can now be posed as determining the states $x_{0:N} = \{x_0, \ldots, x_N\}$ given all the measurements $y_{1:N} = \{y_1, \ldots, y_N\}$ and the system model (2.2) or (2.3).

To define the solution to this problem, we can start with the simple method called *Maximum Likelihood* (ML) (Fisher, 1912) which is a well known statistical method where a model parameters, usually called $\theta$, are estimated from a set of measurements, assumed to be generated from this model as

$$y_t \sim \varphi(y_t|\theta) \tag{2.5}$$

where $\varphi(y_t|\theta)$ represents the likelihood of the measurements parametrised with $\theta$. The

ML solution is obtained by solving the maximisation problem

$$\hat{\theta}^{\mathrm{ML}} = \arg\max_{\theta} \varphi(y_{1:N}|\theta) \tag{2.6}$$

Using (2.3b) and identifying $\varphi = p$ and $\theta = x_{0:N}$, ML formulation is obtained and the method can be applied to the state estimation problem as

$$\hat{x}_{0:N}^{\mathrm{ML}} = \arg\max_{x_{0:N}} p(y_{1:N}|x_{0:N}) \tag{2.7}$$

where the joint likelihood for all the measurements, $p(y_{1:N}|x_{0:N})$, is used. In many practical applications the measurement noise is assumed to be independent in time, which is quite reasonable assumption in these cases. This implies that the joint likelihood can be written as

$$p(y_{1:N}|x_{0:N}) = p(x_0) \prod_{t=1}^{N} p(y_t|x_t) \tag{2.8}$$

where $p(x_0)$ is so called prior of the states, which acts as a belief about the state values at time 0, i.e., before any measurements have been obtained.

The ML approach above does not utilise the state transition pdf, (2.3a), which can actually be useful information. In order to include state dynamics, we can use a conceptually slightly different approach than ML. In ML the true parameter $\theta^0$ is considered as deterministic. If the parameter is seen as stochastic instead, i.e., its properties can be described with some pdf, $p(\theta)$, in the same way as for the measurements which are stochastic, we can “augment” the likelihood with this pdf simply by multiplying these together as $p(y|x)p(x)$. This is one way of incorporating the prior knowledge of the states and with help from the Bayes’ rule (Bayes, 1763)

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{2.9}$$

a so called posterior pdf of the states given measurements is obtained. Now $x_{0:N}$ that maximises this pdf can be obtained leading to so called *Maximum a Posteriori* estimate (MAP)

$$\hat{x}_{0:N}^{\mathrm{MAP}} = \arg\max_{x_{0:N}} p(x_{0:N}|y_{1:N}) = \arg\max_{x_{0:N}} p(y_{1:N}|x_{0:N})p(x_{0:N}) \tag{2.10}$$

Here, $p(y_{1:N})$ is omitted since it does not depend on $x_{0:N}$ and does not influence the maximisation solution. This formulation together with Markov and measurement model independency assumptions leads to the following MAP optimisation

$$\hat{x}_{0:N}^{\mathrm{MAP}} = \arg\max_{x_{0:N}} p(x_0) \prod_{t=1}^{N} p(y_t|x_t)p(x_t|x_{t-1}) \tag{2.11}$$

where $p(x_0)$ is as before the prior of the states before any measurements have arrived. Notice that if state prior is chosen as so called flat, i.e., $p(x) =$ constant, then the solution to MAP and ML will coincide. However, conceptually in MAP the parameters (or states) are still seen as stochastic while in ML they are not.

One powerful feature of the optimisation formulations (2.7) and (2.10) is that any kind of constraints on the states can be added, for example if states represent some physical entities which must be in certain bounds. Another example is if the prior distribution is degenerated, i.e., some of the states are deterministic. In these cases both maximisation problems can be augmented with constraints leading to *Constrained Maximum Likelihood*

$$\hat{x}_{0:N}^{\text{CML}} = \arg\max_{x_{0:N}} \; p(y_{1:N}|x_{0:N}) \qquad \textbf{s. t. } x_{0:N} \in \mathbb{X} \tag{2.12}$$

or *Constrained Maximum a Posteriori* formulation

$$\hat{x}_{0:N}^{\text{CMAP}} = \arg\max_{x_{0:N}} \; p(x_{0:N}|y_{1:N}) \qquad \textbf{s. t. } x_{0:N} \in \mathbb{X} \tag{2.13}$$

where $\mathbb{X}$ is a constraint set for the states. This formulation gives a great deal of flexibility in the problem description, allowing us to pose the problem in a form that is easy to solve. In most cases it will be assumed that involved probability functions have Gaussian form, i.e., the state and measurement noises are Gaussian. In that case it is appropriate to optimise a negative logarithm of the criterion resulting in a minimisation problem. Since the negative logarithm is a monotonously decreasing function the solutions will be the same.

One important special case of the system and measurement models is the case where both the dynamic and the measurement models are linear, the prior has Gaussian distribution and the process and the measurement noise are Gaussian and white, which gives

$$f_t(x_t, w_t) = F_t x_t + G_t w_t \tag{2.14a}$$

$$h_t(x_t) = H_t x_t \tag{2.14b}$$

$$x_0 \sim \mathcal{N}(\bar{x}, P_0), \; w_t \sim \mathcal{N}(0, Q_t), \; e_t \sim \mathcal{N}(0, R_t) \tag{2.14c}$$

Here, $P_0$ is the initial state covariance matrix, i.e., the uncertainty of $x_0$ and $\bar{x}$ is its prior value. The formulation of (2.11) in this case becomes (with negative logarithm applied)

$$\hat{x}_{0:N}^{\text{MAP}} = \arg\min_{x_{0:N}} \; \|x_0 - \bar{x}\|^2_{P_0^{-1}} + \sum_{t=1}^{N} \|y_t - H_t x_t\|^2_{R_t^{-1}} + \|x_t - F_{t-1}x_{t-1}\|^2_{(G_t Q_t G_t^T)^{-1}} \tag{2.15}$$

which is a weighted least squares problem. This example also shows a possible degeneracy issue in the problem formulation. To pose this problem it is required that the matrix $G_t Q_t G_t^T$ is actually invertible, and that is not always the case. In such problems, the formulations (2.12) and (2.13) are really useful.

All the methods presented above assume that all the measurements are available at the estimation time. If that is the case, the solution $\hat{x}_{0:N}$ is called smoothed estimate. Smoothing in this context is defined as an estimate in certain time that depends on the past and future measurements. If that is not the case, i.e., if estimate depends only on past measurements, we obtain a filtered estimate. One important advantage of the filtering is that it can be implemented in a sequential manner. The general sequential solution to (2.11) is given by recursive Bayesian filtering. The posterior distribution for each time instant can be obtained as, see e.g., Gustafsson (2010) for complete derivation,

$$p(x_1|y_0) = p(x_0) \qquad \text{(Initialisation)} \tag{2.16a}$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \tag{2.16b}$$

$$p(y_t|y_{1:t-1}) = \int\limits_{\mathbb{R}^n} p(y_t|x_t)p(x_t|y_{1:t-1})\,\mathrm{d}x_t \tag{2.16c}$$

$$p(x_t|y_{1:t-1}) = \int\limits_{\mathbb{R}^n} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})\,\mathrm{d}x_{t-1} \tag{2.16d}$$

The procedure (2.16) defines a very general nonlinear filter that can be applied to a large class of dynamic and measurements models. However, the closed form solution exists only in a few cases. One such case is (2.14) and it can be shown that procedure (2.16) becomes the Kalman Filter (KF) (Kalman, 1960) which is both the Minimum Variance (MV) and the Best Linear Unbiased Estimator (BLUE). If all the measurements are available at the estimation time, i.e., smoothing solution is possible, the solution to (2.15) can be obtained with the Kalman smoother instead of solving it directly. If the model is not linear, some approximate solutions to solve (2.16) must be applied, where Extended Kalman Filter (EKF) (Kailath et al., 2000) and Particle Filter (PF) (Gordon et al., 1993) are the most common approaches.

## 2.2 Simultaneous Localisation and Mapping

The main idea of Simultaneous Localisation and Mapping (SLAM) is to estimate a map of the surrounding environment from a moving platform, while simultaneously localising the platform (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006). Localisation in this context includes estimation of navigation states, which includes position, velocity and attitude. However depending on the sensor setup, some of the states cannot be estimated. This will be handled from case to case. The main prerequisite for mapping is that some of the available sensors actually observe the environment. Typical sensors that can do that are laser ranging sensors, radars and different kinds of cameras, e.g., infra-red or visible light cameras. In this work, radar (and in particular imaging radar or SAR) and optical camera will be used as the environment observing sensors. The maps generated by different sensors will also differ in their appearance. The advantage is that different maps may be used for different purposes depending on situation at hand. For example a SAR image can be seen as a map of the imaged environment where certain features of the environment are accentuated. In a similar way, an optical camera will produce another kind of map which might be useful in a different situation. How these maps are used is a question of tactical behaviour and will not be discussed in more detail here.

### 2.2.1 SLAM System Models

To put the SLAM problem in the sensor fusion framework described in Section 2.1 we have to define the dynamics and the measurement models or equivalently the state transition pdf and the likelihood function. To do that we need to introduce two coordinate systems which are important for the navigation. The first one is the local navigation frame, also called world frame, and it has a fixed position in the world. The second one is the body frame, which is aligned with the moving platform, and moves and rotates together with it. The navigation states basically describe body frame's position, velocity, acceleration and orientation relative the navigation frame. In SLAM the navigation frame is usually free to be placed arbitrarily, since most of the measurements are relative. However, the estimation results from SLAM can be combined with the prior information, like global maps, in order to fix the origin of the navigation frame to some global coordinate system.

In most cases the dynamics of the platform can be described by simple model using basic equations of motion. For linear motion, where states normally are position, linear velocity and acceleration, a double integrator model is used as (Farrell and Barth, 1999)

$$\dot{p}_t = v_t \tag{2.17a}$$
$$\dot{v}_t = a_t \tag{2.17b}$$
$$\dot{a}_t = w_t^a \tag{2.17c}$$

where $p_t$, is the position of the platform expressed in the local navigation frame, $v_t$ is its velocity and $a_t$ is the acceleration. $w_t^a$ is the unknown input to the system, here modelled as a jerk. The dimension of the states, i.e., $p$, $v$ and $a$ as well as the noise $w^a$, will depend on the application, but will usually be 2 or 3. This is continuous time description of the system, but for the implementation we need a discrete time description. The model above can be discretised exactly if a zero order hold assumption is made. In that case the discrete time model is

$$p_{t+1} = p_t + T_s v_t + \frac{T_s^2}{2} a_t + \frac{T_s^3}{6} w_t^a \tag{2.18a}$$
$$v_{t+1} = v_t + T_s a_t + \frac{T_s^2}{2} w_t^a \tag{2.18b}$$
$$a_{t+1} = a_t + T_s w_t^a \tag{2.18c}$$

where $T_s$ is the sampling time. This model can be compactly written in a matrix form as

$$x_{t+1} = F x_t + G w_t^a \tag{2.19a}$$
$$x_t = \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix}, \; F = \begin{bmatrix} I & T_s I & \frac{T_s^2}{2} I \\ 0 & I & T_s I \\ 0 & 0 & I \end{bmatrix}, \; G = \begin{bmatrix} \frac{T_s^3}{6} I \\ \frac{T_s^2}{2} I \\ T_s I \end{bmatrix} \tag{2.19b}$$

where $I$ and $0$ are the identity and zeros matrices of appropriate dimensions.

The rotational dynamics in $\mathbb{R}^3$ can be parametrised with so called Euler angles, denoted roll, $\phi$, pitch, $\theta$, and yaw, $\psi$, see Figure 2.1. These angles describe the rotation of one three-dimensional coordinate system relative another. In this case the coordinate systems are the local navigation and platform's body frames. The dynamics of these angles is

$$\dot{\phi}_t = \omega_t^X + \sin(\phi_t)\tan(\theta_t)\omega_t^Y - \cos(\phi_t)\tan(\theta_t)\omega_t^Z \tag{2.20a}$$
$$\dot{\theta}_t = \cos(\phi_t)\omega_t^Y + \sin(\phi_t)\omega_t^Z \tag{2.20b}$$
$$\dot{\psi}_t = -\frac{\sin(\phi_t)}{\cos(\theta_t)}\omega_t^Y + \frac{\cos(\phi_t)}{\cos(\theta_t)}\omega_t^Z \tag{2.20c}$$
$$\dot{\omega}_t^X = w_t^{\omega X} \tag{2.20d}$$
$$\dot{\omega}_t^Y = w_t^{\omega Y} \tag{2.20e}$$
$$\dot{\omega}_t^Z = w_t^{\omega Z} \tag{2.20f}$$

and $\omega_t^{\{X,Y,Z\}}$ are the angular velocities and $w_t^{\{X,Y,Z\}}$ are the unknown input terms, here modelled as angular accelerations. In contrast to the linear motion dynamics (2.19), this one is nonlinear. Although some nonlinear dynamics is possible to discretise exactly, it is not the case here and some approximate method, like Euler sampling, must be used. It can immediately be seen that there exists a singularity in the dynamics for $\theta_t = \pm\pi/2$. In this case the roll and the yaw angles are undefined. Fortunately, there is another representation of the rotations in $\mathbb{R}^3$, with the help of the quaternions, that does not suffer from these limitations (Kuipers, 1999; Shuster, 1993).

Quaternions are defined in four dimensional space as $q = [q_0\ q_1\ q_2\ q_3]^T$, $q_i \in \mathbb{R}$. To represent the rotation in $\mathbb{R}^3$, $q$ is constrained to the unit sphere, i.e., $q^T q = 1$. The transformation from Euler angles to quaternions is defined as

$$q_0 = \cos(\phi/2)\cos(\theta/2)\cos(\psi/2) + \sin(\phi/2)\sin(\theta/2)\sin(\psi/2) \tag{2.21a}$$
$$q_1 = \sin(\phi/2)\cos(\theta/2)\cos(\psi/2) - \cos(\phi/2)\sin(\theta/2)\sin(\psi/2) \tag{2.21b}$$
$$q_2 = \cos(\phi/2)\sin(\theta/2)\cos(\psi/2) + \sin(\phi/2)\cos(\theta/2)\sin(\psi/2) \tag{2.21c}$$
$$q_3 = \cos(\phi/2)\cos(\theta/2)\sin(\psi/2) - \sin(\phi/2)\sin(\theta/2)\cos(\psi/2) \tag{2.21d}$$

and from quaternions to Euler angles as

$$\phi = \text{atan2}(2(q_0q_1 + q_2q_3), 1 - 2(q_1^2 + q_2^2)) \tag{2.22a}$$
$$\theta = \text{arcsin}(2(q_0q_2 - q_1q_3)) \tag{2.22b}$$
$$\psi = \text{atan2}(2(q_0q_3 + q_1q_2), 1 - 2(q_2^2 + q_3^2)) \tag{2.22c}$$

where atan2( · , · ) is $\arctan$ function defined for all four quadrants. The dynamics of the quaternions can be expressed as

$$\dot{q}_t = \frac{1}{2}\underbrace{\begin{bmatrix} 0 & -\omega_t^X & -\omega_t^Y & -\omega_t^Z \\ \omega_t^X & 0 & \omega_t^Z & -\omega_t^Y \\ \omega_t^Y & -\omega_t^Z & 0 & \omega_t^X \\ \omega_t^Z & \omega_t^Y & -\omega_t^X & 0 \end{bmatrix}}_{S(\omega_t)} q_t = \frac{1}{2}\underbrace{\begin{bmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{bmatrix}}_{\tilde{S}(q_t)} \omega_t \tag{2.23}$$

We see that the dynamics is still non-linear, but the non-linearities are much simpler than for the Euler angles (it is actually bilinear if angular rates are considered as input). A nice property of the quaternion dynamics is that, if we assume zero order hold, i.e., constant

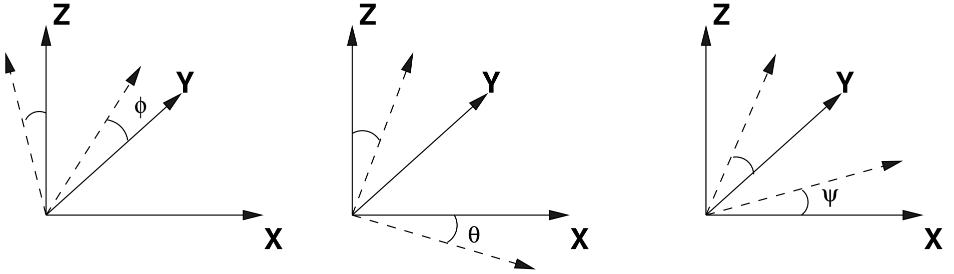


***Figure 2.1:*** *Illustration of Euler angles, from left to right: roll angle ($\phi$, rotation around $x$-axis), pitch angle ($\theta$, rotation around $y$-axis) and yaw angle ($\psi$, rotation around $z$-axis). Solid line coordinate system represents local navigation frame and dashed coordinate system is the body frame.*

angular rate between sampling times, the discrete time model is

$$q_{t+1} = \exp\left(\frac{T_s}{2} S(\omega_t)\right) q_t \tag{2.24}$$

which also preserves the norm of the quaternion. Note that the second form of the dynamics, the one with $\tilde{S}(q)$, cannot be discretised in this way and some other strategy must be applied, e.g., Euler sampling.

Given the Euler angles or quaternions it is possible to calculate any vector coordinates in different systems. The simplest way to perform that is via a rotation matrix, $\mathsf{R}$. As a function of Euler angles, the rotation matrix is defined as three consecutive rotations around each of the three axis, as shown in Figure 2.1. Since the order of rotations is important, different rotation matrices can be obtained for different orders given the same Euler angles. The standard order in aeronautics is $z-y-x$, i.e., rotation is first performed around $z$-axis, then around $y$-axis and finally around $x$-axis. The total rotation can then be composed of three consecutive rotations as (with shorthand notation $s_\cdot = \sin(\,\cdot\,)$ and $c_\cdot = \cos(\,\cdot\,)$)

$$\begin{aligned}\mathsf{R}(\phi,\theta,\psi) = &\begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\phi & s_\phi \\ 0 & -s_\phi & c_\phi \end{bmatrix} \begin{bmatrix} c_\theta & 0 & -s_\theta \\ 0 & 1 & 0 \\ s_\theta & 0 & c_\theta \end{bmatrix} \begin{bmatrix} c_\psi & s_\psi & 0 \\ -s_\psi & c_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \\ &\begin{bmatrix} c_\theta c_\psi & c_\theta s_\psi & -s_\theta \\ s_\phi s_\theta c_\psi - c_\phi s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & s_\phi c_\theta \\ c_\phi s_\theta c_\psi + s_\phi s_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi & c_\phi c_\theta \end{bmatrix}\end{aligned} \tag{2.25a}$$

and as a function of quaternions, the corresponding matrix is

$$\mathsf{R}(q) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(-q_0q_2 + q_1q_3) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(-q_0q_1 + q_2q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{2.26}$$

These rotation matrices express how to rotate a vector given in navigation frame to a vector given in body frame. To emphasise this relationship, if necessary, superscript will be used to denote which frame vector is expressed in. For example the gravity vector, $g$, always has coordinates $[0\ 0\ \ -9.81]$ in the local navigation frame, which we can write as $g^n$. If we want to find its representation in a body frame, $g^b$, we simply multiply the rotation matrix, $\mathsf{R}^{bn}$, with $g^n$

$$g^b = \mathsf{R}^{bn}\, g^n \tag{2.27}$$

Superscript $bn$ denotes the rotation from navigation to body frame. To obtain the inverse rotation, body to navigation frame, $nb$, an orthonormality property of the rotation matrix is used to obtain

$$\mathsf{R}^{nb} = (\mathsf{R}^{bn})^{-1} = (\mathsf{R}^{bn})^T \tag{2.28}$$

i.e., the transpose of the matrix denotes the inverse rotation. In this thesis only $\mathsf{R}$ will denote $\mathsf{R}^{bn}$ and $\mathsf{R}^T$ will denote $\mathsf{R}^{nb}$ unless the frame superscript is absolutely needed.

The next thing that is needed is a measurement or likelihood function. Its task is to describe how the information from the used sensors, i.e., the measurements, is related to the platform's states and the environment that we want to map. It usually has a form as in (2.1b), (2.2b) or (2.3b), but generally any function that relates sensor measurements, platform's states and the surrounding environment can be used in the ML and MAP optimisation criteria. Basically, any function of the form

$$F(x_{0:N}, \mathsf{M}, y_{1:N}) \tag{2.29}$$

can be used. Here, the map, $\mathsf{M}$, denotes the representation of the environment that should be mapped. Also, this function will in most cases correspond to negative log-likelihood, implying that it should attain minimum value for the correct states and map parameters. This representation can be quite different depending on the used sensors, for example a SAR will produce an image which will be considered as a map of the environment. In the next chapter, Chapter 3, where the used sensors are described in detail, concrete examples of these functions will be given. Note also that some of the sensors will not measure the environment, and in that case, the function will simply not depend on the map.

# 3

# Navigation System and Sensors

In this chapter an overview of the typical UAV navigation system and the sensors used in the estimation methods introduced in Chapter 2 is done. The utilised sensors and their properties are described as well as how they are incorporated into the sensor fusion framework.

## 3.1 UAV Navigation System

The definition of the navigation system in a UAV is the system which measures and delivers all interesting navigational quantities to the users of these data. Navigational quantities of interest are usually

- 3D position in some global coordinate system
- 3D velocity
- 3D acceleration
- attitude (pitch, roll and heading)
- 3D angular rate

which can either be directly measured or estimated in some way. In a basic configuration the UAV navigation system can look like the one in Figure 3.1. Basic sensors used here are Inertial Measurement Unit (IMU), Global Positioning System (GPS) and Air Data Computer (ADC). IMU senses specific forces and angular moments that platform exhibits and measures these giving accelerations and angular rates (or velocities) of the platform. In principle, the position and velocity could be calculated directly by integration of the

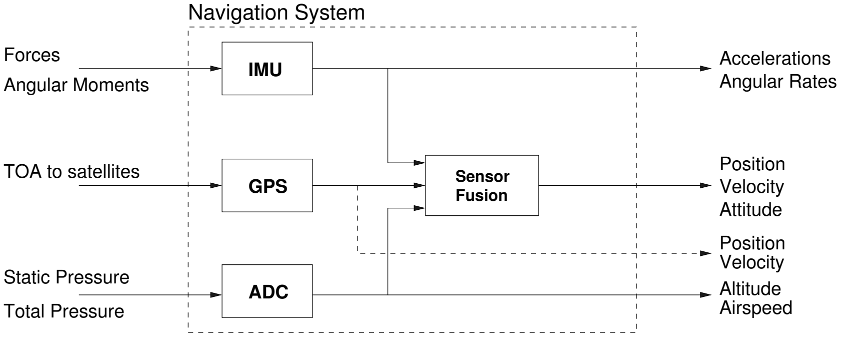


***Figure 3.1:*** *Overview of the typical UAV navigation system. IMU is Inertial Measurement Unit, GPS is Global Positioning System and ADC is Air Data Computer. Dashed output from the GPS is used to indicate that raw GPS position could be used, but it is usually the fused one that is used in practice.*

acceleration based on the solution of (2.17)

$$p_t = p_0 + \int_0^t v_s \, \mathrm{d}s \tag{3.1a}$$

$$v_t = v_0 + \int_0^t a_s \, \mathrm{d}s \tag{3.1b}$$

where $p_t$ is the position, $v_t$ is the velocity and $a_t$ is the acceleration of the platform expressed in some local navigation frame. Since the accelerations are measured relative the so called body frame which rotates together with the platform, a rotation with the attitude must be performed, $a_t = \mathsf{R}^{nb} a_t^m$, where superscript $m$ denotes the measurements and $\mathsf{R}$ is the rotation matrix between body and navigation frame. The rotation is directly dependent on the attitude of the platform which can also be integrated from the angular velocities in a similar way. This approach can be applied when the sensor accuracy in terms of bias and noise is high, otherwise the velocity and position will drift over time rendering the system unusable. The solution is to use other sensors which will estimate the bias and give more stable estimate. More detailed description of IMU is presented in Section 3.2.

Global Positioning System (GPS) is a version of the Global Navigation Satellite System (GNSS). Another example is Russian GLONASS. Here the GPS receiver onboard the platform, receives Time-of-Arrival signals from the satellites orbiting the Earth. Since the satellite positions are known quite precisely, it is possible to, with trilateration, estimate the position (and velocity) of the platform with high accuracy.

Air Data Computer (ADC) measures static and total pressure of the air around the platform. Static pressure is directly proportional to the altitude of the platform, while total pressure is the sum of the static and dynamic pressures. Dynamic pressure is a consequence of the platform's movement through the air. In this way, the platform's airspeed, the speed in relation to the surrounding air, can be calculated. ADC is not utilised directly in this work since other sensors can provide same information, but can easily be incorporated into the presented framework.

The used sensors are able to calculate almost all navigational parameters by themselves, it is basically only the attitude that needs to be estimated in some way from the raw sensor data. However, in order to get good dynamical behavior, e.g., remove noise or the effects of the different sampling times, the sensor fusion methods described in Section 2.1 must be used.

As mentioned above, different sensors, when fused, give the desired performance of the navigation system. In the configuration above, the performance is heavily dependent on the GPS, since it gives the stable, drift-free position and velocity information. This is however not desirable, since GPS is easily jammed or even spoofed. Jamming is a situation when signal is deliberately disturbed leading to the loss of GPS positioning. Spoofing is, on the other hand, a case when a GPS signal is maliciously altered locally such that it gives false and misleading positioning. This is, in a way, a worse case since false position can lead to severe consequences, like UAV crash in urban area. To mitigate these risks, alternative means of navigation must be devised. One way of doing this is introducing other sensors which, in combination with sensor fusion, could give an estimate of the navigational parameters which is drift-free and has good accuracy. Here, two imaging sensors, Synthetic Aperture Radar (SAR) and regular optical camera, will be used. Figure 3.2 shows the addition of these sensors to the navigation system. The reason why these particular sensors are chosen as a complement is that most UAVs have a camera, either as a part of the standard equipment or as a mission specific sensor. SAR is a more mission specific sensor, but it has certain advantages compared to an optical sensor since the wavelengths a radar operates on make it weather-insensitive. This property is important in many situations where high availability and all-weather capability is expected. In Section 3.3 and Section 3.4 the sensor models and functional principles for SAR and optical cameras will be described, and how they are used in the sensor fusion framework.

Another quality that above-mentioned imaging sensors have, is that they give some kind of information about the environment they are observing. In the case of the sensors that will be used here, the information is an image of the surroundings. As already described, this information will allow us to build a map of the environment, at the same time as we estimate the navigation states.

## 3.2 Inertial Measurement Unit

In this section, inertial sensors, which is a collective name for accelerometers and gyroscopes, will be described. These are usually combined in a single Inertial Measurement Unit (IMU) and will measure three dimensional acceleration, $[a^X \; a^Y \; a^Z]^T$, and angular velocity, $[\omega^X \; \omega^Y \; \omega^Z]^T$. This setup is usually called 6 Degrees-Of-Freedom (6-DOF) inertial sensor. The accelerometers measure the specific force in an inertial frame. For example on the surface the earth, at rest the accelerometer will measure earth's gravity force and force exhibited by earth's rotation, so called coriolis effect. The basic functional principle of the accelerometer is a damped mass on a spring. When a force is applied to the

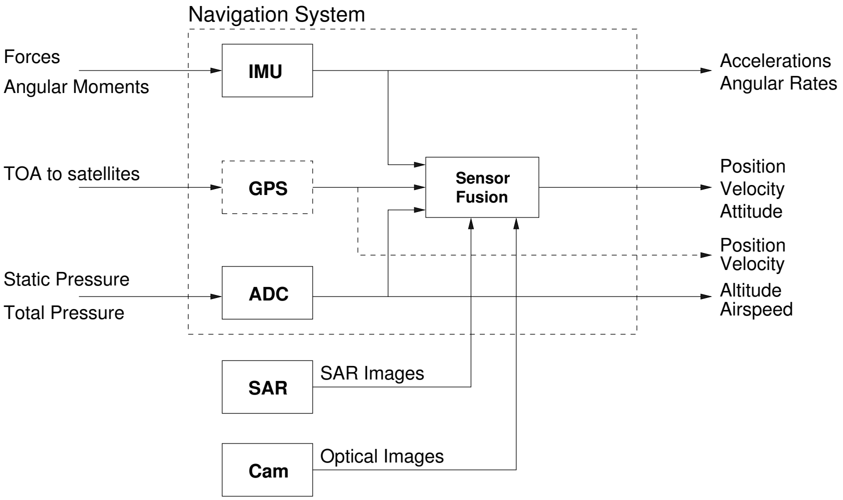

Navigation System
Forces
Angular Moments
IMU
Accelerations
Angular Rates
TOA to satellites
GPS
Sensor Fusion
Position
Velocity
Attitude
Position
Velocity
Static Pressure
Total Pressure
ADC
Altitude
Airspeed
SAR
SAR Images
Cam
Optical Images

***Figure 3.2:*** *UAV navigation system with addition of alternative sensors, SAR and optical camera. Both are imaging sensors, but images have different properties making them complementary. GPS block is dashed out here to emphasise that it can be left out.*

accelerometer, the force component parallel to the longitudinal axis will move the mass elongating the spring. The elongation is simply proportional to the applied force (or its longitudinal component) as

$$F = ma^X = kX \tag{3.2}$$

where $m$ is the mass, $k$ is the spring coefficient and $X$ is the elongation. The technology for building accelerometers based on this spring mass principle has been developed during the recent years resulting in cheaper components. One such technology is Micro Electro-Mechanical Sensors (MEMS) where piezo-electrical effect is used. In this case, the silicon based rod is bent when affected by the force creating current which will be proportional to the applied force. The big advantage of this construction is that it can be made very small and light. This is a very attractive property of MEMS components, since both space and weight are limited in a typical UAV. However, the downside is that they are very sensitive for temperature variations giving large bias and noise terms.

The gyroscopes, or gyros, measure the rotational velocities around the fixed axis in the inertial frame. The basic technology how this is performed can be based on a principle of gyroscopic precession (Meriam and Kraige, 1998) used in mechanical gyros, e.g., the spinning or MEMS gyros, or light interference effect used in fiber optical and ring laser gyros. The precession is a principle where a rotating mass spinning around an axis will start to rotate around another axis when an external torque is applied around the third axis. This is illustrated in Figure 3.3 where the disc is spinning with angular velocity $\omega$ around $x$-axis and the torque $M$ is applied around $y$-axis. This will cause the disc to precess

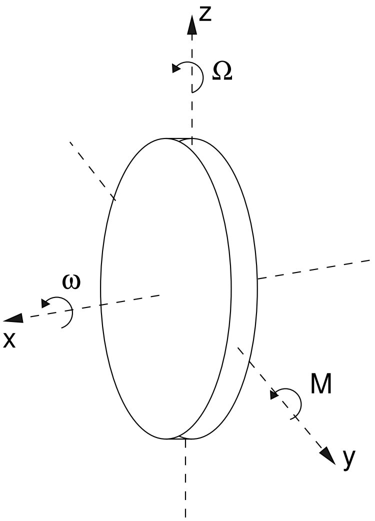


***Figure 3.3:*** *Illustration of the gyroscopic precession principle. A cylindrically shaped mass spinning with angular velocity $\omega$ around $x$-axis will precess with angular velocity $\Omega$ around $z$-axis when a constant torque $M$ is applied around $y$-axis.*

around $z$-axis with angular velocity $\Omega$ and the relationship between these is

$$M = I\Omega\omega \tag{3.3}$$

where $I$ is the disc's moment of inertia. Spinning gyros use real spinning discs to measure the angular velocity while MEMS gyros use vibrating elements instead of rotating, but the underlying principle is the same. Fiber optical and ring laser gyros are using completely different principle based on the Sagnac effect. Here two beams of light are sent through a long optical fiber coil in opposite directions. When the coil rotates the phase shift in these two beams will cause interference pattern that can be observed and measured. The size of the interference is directly proportional to the angular velocity. In order to make errors small, optical gyros must be made large and heavy making them impractical for small UAV applications. On the other hand the accuracy and stability of the optical gyros is the best one with small bias and noise errors. MEMS gyros have the worst performance and purely mechanical ones are in the middle range.

For practical reasons, IMU is rigidly mounted, or strapped, in the platform's frame (body frame), i.e., it has a strap-down mounting. In this way the need for moving and potentially heavy parts is removed. This means that the acceleration is measured in the body frame. In order to use acceleration for navigation accelerations must be expressed in the local navigation frame, (this was used in (3.1)). The implication of this is that acceleration must be transformed to the local navigation frame in order to use it in the navigation application as already mentioned in Section 3.1. The measurement equation (2.2b) for the accelerometer is then

$$y_t^a = \mathsf{R}_t(a_t - g) + b_t^a + e_t^a \tag{3.4}$$

where $\mathsf{R}_t$, as already described, is a rotation matrix that expresses rotation between navigation and body frame, and $g = [0\,0\ -9.81]^T$ is the gravitation acceleration expressed in the local navigation frame. The bias term $b_t^a$ can be included in the model if the estimate of it is not compensated for or if the estimate of it is sought. $e_t^a$ is the measurement noise term, usually assumed to be white and Gaussian with $e_t \sim \mathcal{N}(0, R_t^a)$.

The measurement equation for the gyros is even simpler, it is

$$y_t^\omega = \omega_t + b_t^\omega + e_t^\omega \tag{3.5}$$

where, as in acceleration case, $b_t^\omega$ and $e_t^\omega$ denote bias and noise terms respectively, with $e_t^\omega \sim \mathcal{N}(0, R_t^\omega)$.

Notice also that the output from the IMU can either be used as a measurement $y_t$ or as an input $u_t$ in the dynamics. Since both approaches have their advantages and disadvantages, which one is actually used will depend on the specific application.

## 3.3 Synthetic Aperture Radar

In this section the basic principles of the Synthetic Aperture Radar (SAR) are explained. It starts with the Real Aperture Radar (RAR), continues with the methods to create high resolution images and mentions some of the effects associated with SAR.

### 3.3.1 Real and Synthetic Aperture Radar

In the simplest setup a radar image can be created with the moving platform carrying the side-looking radar and flying above the scene to be imaged. By sending and receiving radar pulses along the trajectory, a range-azimuth image is created, see Figure 3.4. Since the energy in each radar pulse is spread out over the scene, every pulse will produce a one-dimensional image according to a simple principle; each echoed pulse is received and gated in range bins according to the time, $t$, it takes to receive the pulse, using the relation $R = tc/2$, where $c$ denotes the speed of light. This means that each range bin will contain the total energy reflected from the scene on that specific range. When all of these simple one-dimensional images are stacked next to each other, a full RAR image is created. RAR images will have the imaged scene smeared across the azimuth, giving very poor effective resolution which can be seen in Figure 3.5, where a simulated example of the RAR image of two point targets is depicted. The resolution in the azimuth direction for the RAR images is governed by the basic laws of the electromagnetism which state that the radar lobe width is dependent of the antenna size and the wavelength of the carrier according to $\lambda/d$, where $d$ is antenna size and $\lambda$ is the wavelength. This will give the resolution $\lambda R/d$ where $R$ is the range to the imaged scene. We see that by decreasing the wavelength or increasing the antenna we can make the lobe narrower and increase the resolution. The wavelength is usually fixed to a specific value and there is not much freedom of decreasing it, and besides there are other unwanted effects if low wavelengths are used like cloud occlusions. There are also limits regarding how large antenna a flying platform can carry which limits that parameter as well. However, by using the movement of the platform a long antenna can be synthesised and the resolution of the images is drastically increased. This principle leads to Synthetic Aperture Radar (Cutrona et al.,

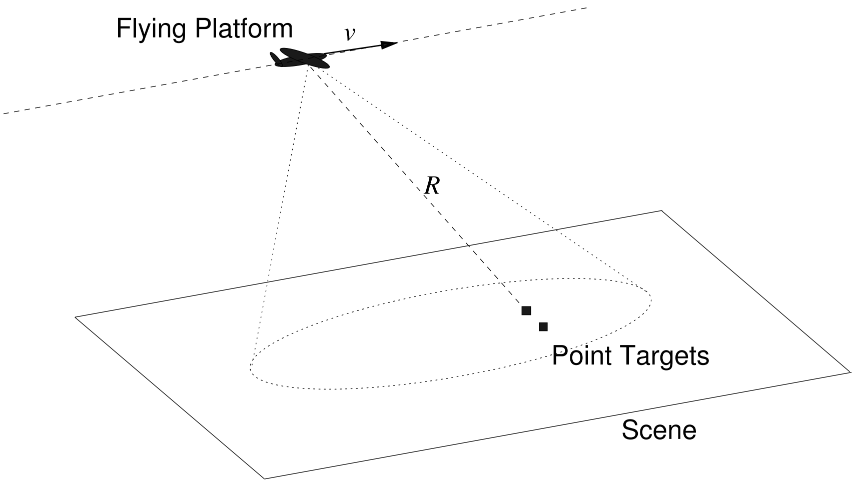

Flying Platform
v
R
Point Targets
Scene

***Figure 3.4:*** *Side-looking radar geometry. The flying platform is moving along the straight trajectory parallel to the scene with speed $v$ while it transmits and receives radar pulses. $R$ is the range from the platform to the target.*

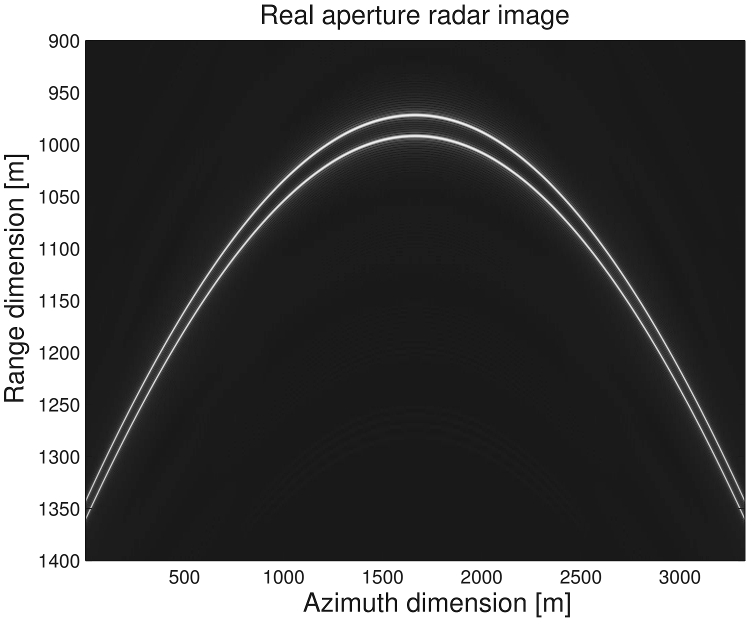


***Figure 3.5:*** *Real aperture radar image of the two point targets in Figure 3.4.*

1961).

As mentioned above the movement of the platform can be utilised to improve the resolution in the azimuth direction. During the motion of the platform the scene, which can be assumed to consist of point scatterers, will travel through the radar lobe and the slant range to it will vary, see Figure 3.4. This slant range variation can be compensated for each radar echo and all of the echos can be integrated in order to produce an image which can be expressed as

$$I^{\mathrm{SAR}} = \int\limits_{D_t} I_t^{\mathrm{RAR}}(R) g(R,t)\,\mathrm{d}t \tag{3.6}$$

where $I^{\mathrm{SAR}}$ is the SAR image, $I_t^{\mathrm{RAR}}(R)$ is the RAR image (or raw radar data), $g(R,t)$ includes possible demodulation to baseband and lobe weight terms, $R$ is the range dimension, $t$ is the azimuth dimension of the raw data (or time dimension), and $D_t$ is the azimuth (or time) domain of the RAR image. Image creation can be performed in the image (or time) domain or in the frequency domain. Some of the most important frequency domain methods are the Range-Doppler (Walker, 1980), the Fourier-Hankel (Fawcett, 1985; Hellsten and Andersson, 1987; Andersson, 1988) and the $\omega$-K migration methods (Rocca, 1987; Cafforio et al., 1991; Milman, 1993). The frequency domain methods are generally fast, but has a downside that they assume straight trajectories in order to work properly. However, in reality the trajectory will never be straight, especially if the flying platform is a small UAV. This will cause image distortions and auto-focusing is more complicated. This opens up for the use of time domain methods, which are slower, but can handle any trajectory shape. This of course is an important benefit, particularly if SAR images are to be used for trajectory estimation, and time domain methods will be considered from now on.

One of the most known time domain methods is back-projection (Natterer, 1986). In the back-projection procedure each saved radar echo, which is one dimensional, is back-projected onto a two dimensional area. In this way a poor quality image of the scene is obtained. Now we can sum up all these back-projected images in order to obtain the full SAR image. This is equivalent to the integration operation in (3.6), except that integration becomes summation due to the discrete data. This process can be described for each pixel $(i,j)$ in SAR image as in

$$I_{ij}^{\mathrm{SAR}} = \sum_{t=1}^{N} z_t(R_t^{ij}) \tag{3.7a}$$

$$R_t^{ij} = \|p_t - s^{ij}\|_2 \tag{3.7b}$$

where $p_t$ is the position of the platform and $s^{ij}$ is the position in the scene which corresponds to the pixel $(i,j)$. $z_t(R)$ is the compensated raw data in sampled form. Figure 3.6 describes this procedure in a schematic way and Figure 3.7 is the resulting image produced with the back-projection method if the same raw data used for RAR from Figure 3.5 is used.

The main downside of this kind of procedure is that the number of operations needed to synthesise an image is proportional to $\mathcal{O}(KMN)$ for a $K \times M$ image created from

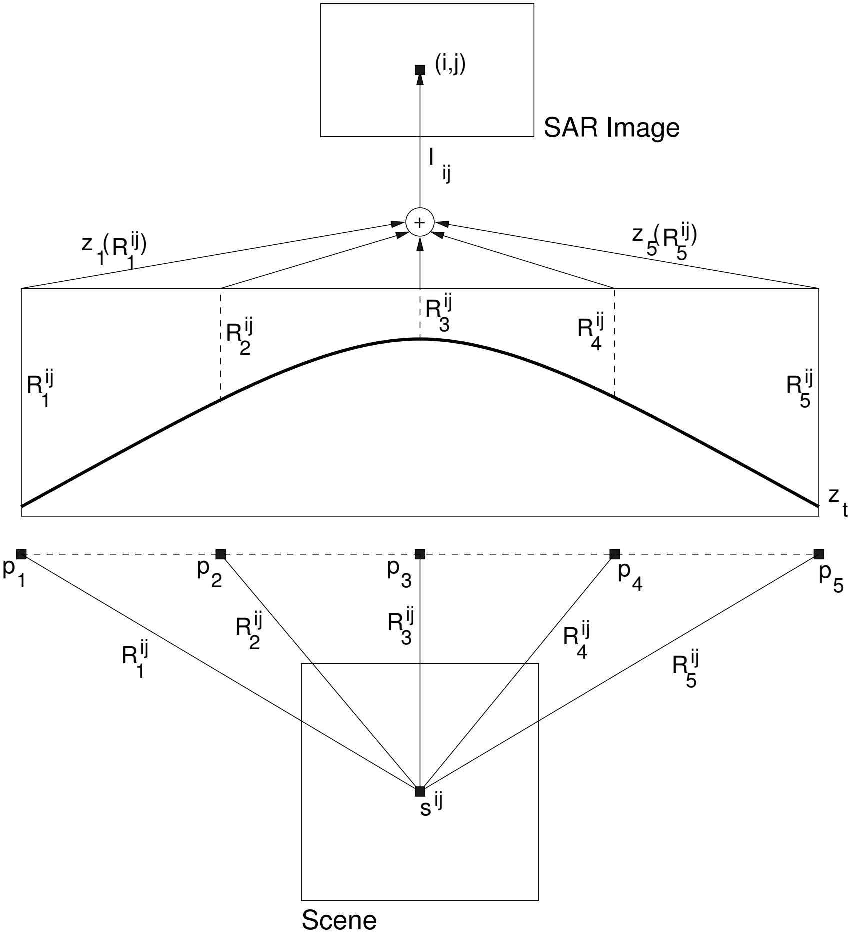

(i,j)
SAR Image
$I_{ij}$
+
$z_1(R_1^{ij})$
$z_5(R_5^{ij})$
$R_1^{ij}$
$R_2^{ij}$
$R_3^{ij}$
$R_4^{ij}$
$R_5^{ij}$
$z_t$
$p_1$
$p_2$
$p_3$
$p_4$
$p_5$
$R_1^{ij}$
$R_2^{ij}$
$R_3^{ij}$
$R_4^{ij}$
$R_5^{ij}$
$s^{ij}$
Scene

***Figure 3.6:*** *Back-projection operation schematically described. The scene consists of only one point target in this illustration. The figure is not to scale.*

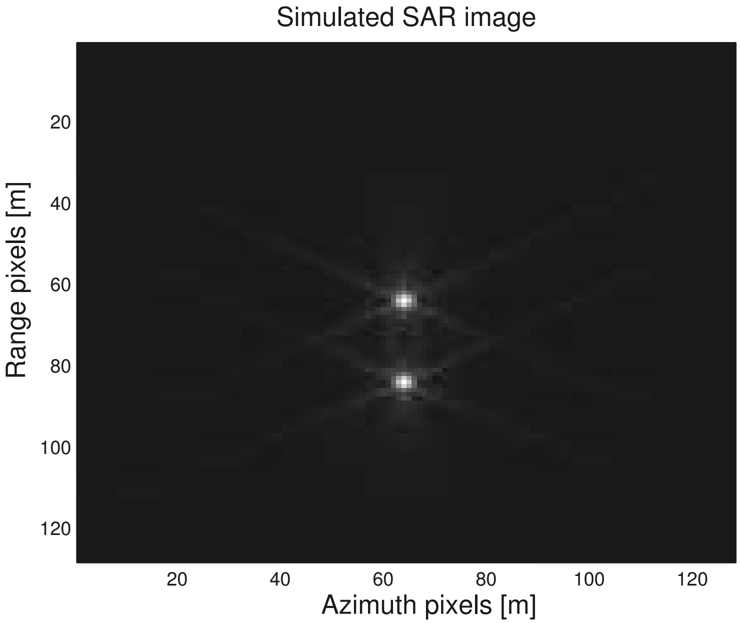
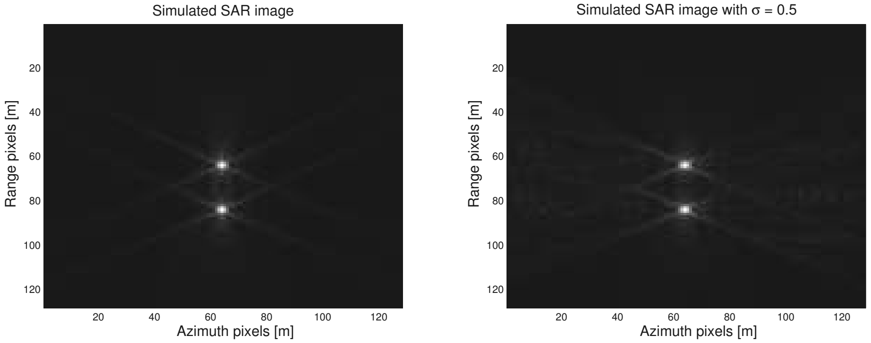


***Figure 3.7:*** *Synthetic aperture radar image of the two point targets from Figure 3.4.*

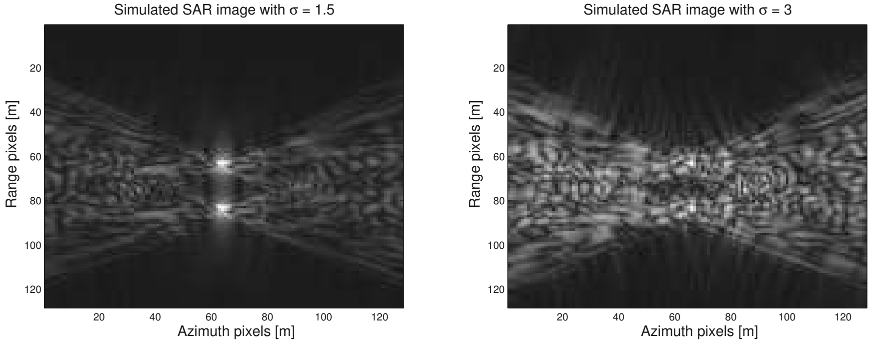
$N$ radar echos. This can be a large number for large images and long aperture times. However during recent years a modification to the original back-projection, called fast factorised back-projection, has been developed. This method can actually create the SAR image in $\mathcal{O}(KM \log N)$ number of operations (Ulander et al., 2003). This implies considerable time saving and together with the development of the computers, this allows to consider real time SAR imaging.

It is now clear that in order to perform the back-projection (or factorised fast back-projection) operation the trajectory of the platform must be known or otherwise the resulting image will be distorted. The image distortion can manifest itself in many ways, from pure translation through geometric distortion to defocusing. Since translation and geometric distortions are hard to measure if the true scene is unknown, these will not be considered here. Defocusing, which is a very common distortion, is measurable, at least seen as an image property. The main source of defocusing of the SAR images is the error in the trajectory estimate used for creation of the images. In the frequency domain methods, a straight trajectory is used, and if the real trajectory deviates from this assumption, it will cause image defocusing. In the time domain methods, despite the fact that the general trajectory form is used, deviations from the real trajectory will cause the back-projected sub images to be shifted. The summation operation of the sub images will then cause defocusing. To illustrate this, the simple two point image from Figure 3.7 can be used. If Gaussian white noise with different variances is added to the cross-track position of the platform, images as in Figure 3.8 are obtained. In Figure 3.8a, the image is created with the same trajectory as data were acquired, which results in a perfectly fo-



***(a)*** *Focused SAR image of two point targets.*

***(b)*** *Unfocused SAR image of two point targets with* $\sigma = 0.5\ \mathrm{m}$.



***(c)*** *Unfocused SAR image of two point targets with* $\sigma = 1.5\ \mathrm{m}$.

***(d)*** *Unfocused SAR image of two point targets with* $\sigma = 3\ \mathrm{m}$.

***Figure 3.8:*** *Example SAR images with different perturbed trajectories.*

cused image. In the other three images the cross-track position noise was $\mathcal{N}(0, \sigma^2)$ where $\sigma = \{0.5, 1.5, 3\}$ $[\mathrm{m}]$ and the images are created under the assumption that the trajectory was linear. This results in defocused images, and the degree of defocusing depends on the noise variance. Much effort has been spent to correct for this, see for example Oliver and Quegan (2004); Wahl et al. (1994); Xi et al. (1999); Morrison and Munson (2002); Xing et al. (2009). Traditionally, these methods are open-loop type, meaning that the image is created with assumptions of linear flight trajectory and focusing is done afterwards in an open-loop way discarding possible flight trajectory information. This is a consequence of the off-line image generating process where the trajectory is no longer interesting. In the setup where SAR images are generated on-line, an idea, as already mentioned, is to use information from the image defocusing and navigation system together. In the next section, different focus measures, that can be used in this procedure are covered.

### 3.3.2 Measuring Focus

#### Sum-Modified-Laplacian

The Sum-Modified-Laplacian focus measure is defined as

$$\begin{aligned}\nabla^2_{\mathrm{ML}} I_{ij} = &|2I_{ij} - I_{(i-1)j} - I_{(i+1)j}| + \\ &|2I_{ij} - I_{i(j-1)} - I_{i(j+1)}| \end{aligned} \tag{3.8a}$$

$$\mathrm{SML}(I) = \sum_{i=2}^{K-1} \sum_{j=2}^{M-1} \nabla^2_{\mathrm{ML}} I_{ij} \cdot \mathbf{I}_{[\nabla^2_{\mathrm{ML}} I_{ij} \geq T]}(\nabla^2_{\mathrm{ML}} I_{ij}) \tag{3.8b}$$

where $I_{ij}$ is the $K \times M$ image grey-scale intensity for pixel coordinate $(i, j)$, $\mathbf{I}$ is the indicator function, and $T$ is the threshold value. This measure has its maximum for the most focused images. Since (2.29) is a minimisation criterion, SML can be inverted or negated to fit into this criterion.

#### Tenengrad

The Tenengrad focus measure is defined as

$$\mathrm{TG}(I) = \sum_{i=2}^{K-1} \sum_{j=2}^{M-1} S^2_{ij} \cdot \mathbf{I}_{[S_{ij} > T]}(S_{ij}) \tag{3.9}$$

where $S_{ij}$ is the Sobel gradient in pixel coordinate $(i, j)$,

$$S_{ij} = \sqrt{(S^r_{ij})^2 + (S^c_{ij})^2} \tag{3.10}$$

where $S^r$ and $S^c$ are row and column Sobel gradients respectively, $T$ is threshold value and $I$ is the $K \times M$ image. The Sobel gradients are obtained by convolving the image with the row and column Sobel kernels $D^r$ and $D^c$,

$$S^r = D^r * I \tag{3.11a}$$

$$S^c = D^c * I \tag{3.11b}$$

$$D^r = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{3.11c}$$

$$D^c = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3.11d}$$

Just as SML, TG has it maximum for the most focused images. This problem is circumvented in the same way as for the SML.

#### Discrete Cosine Transform

Further, the Discrete Cosine Transform (DCT) can be used to measure image focus, see Kristan et al. (2006). The main idea is that focused images have higher frequency components than unfocused images. The focus measure based on DCT is defined as

$$\mathrm{DCT}(I) = 1 - \frac{\sum_{\omega=1}^{T}\sum_{\nu=1}^{T}|D(\omega,\nu)|^2}{(\sum_{\omega=1}^{T}\sum_{\nu=1}^{T}|D(\omega,\nu)|)^2} \tag{3.12}$$

where

$$D(\omega,\nu) = \alpha_\omega\alpha_\nu \sum_{i=0}^{K-1}\sum_{j=0}^{M-1} I_{ij}\cos\left(\frac{\pi(2i+1)\omega}{2K}\right)\cos\left(\frac{\pi(2j+1)\nu}{2M}\right) \tag{3.13a}$$

$$\alpha_\omega = \begin{cases} \frac{1}{\sqrt{K}}, & \omega = 0 \\ \sqrt{\frac{2}{K}}, & 1 \le \omega \le K-1 \end{cases} \tag{3.13b}$$

$$\alpha_\nu = \begin{cases} \frac{1}{\sqrt{M}}, & \nu = 0 \\ \sqrt{\frac{2}{M}}, & 1 \le \nu \le M-1 \end{cases} \tag{3.13c}$$

is the Discrete Cosine Transform of the image $I$ and $T$ is threshold. Even DCT behaves as TG and SML, i.e., it attains the maximum value for the sharpest images.

### Entropy

Another measure of the image focus is image entropy calculated as

$$E_1(I) = -\sum_{k=1}^{256} p_k \log_2(p_k) \tag{3.14}$$

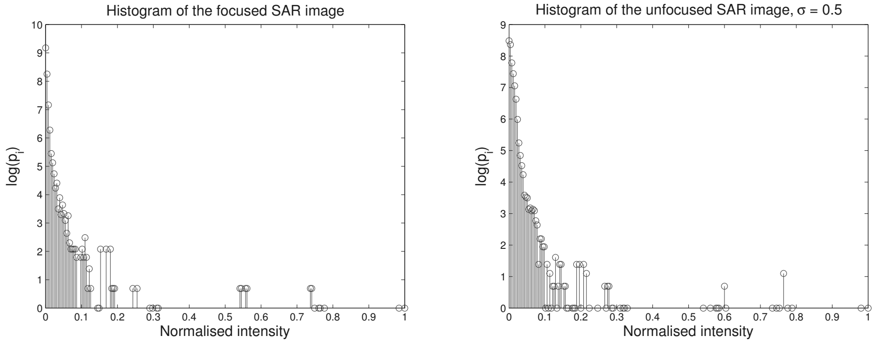
where $p_i$ is an approximated grey level distribution of the $K \times M$ grey-scale image. A grey-scale SAR image is taken as the amplitude image $|I_{ij}|$, where $I_{ij}$ is here considered as the complex-valued SAR image. $p_i$ can be obtained from the image histogram, calculated as

$$p_k = \frac{\{\#\text{ of pixel values } |I_{ij}|\} \in [k-1,k]}{KM} \tag{3.15a}$$

$$k \in [1,256] \tag{3.15b}$$

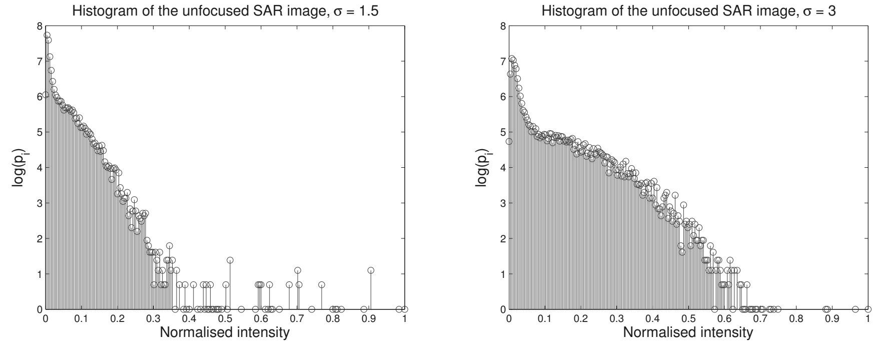
The more focused the image is, the higher the entropy is (Ferzli and Karam, 2005), exactly as for the previous mentioned measures. Note however that entropy measure is primarily used for optical images, whose defocus (or rather unsharpness) have different nature from the defocus of SAR images. In Section 3.3.4 it will be shown that the entropy for a SAR image is lower the more focused the image is. Example histograms and entropy values for the images in Figure 3.8 are given in Figure 3.9.

An alternative definition of entropy (and more frequently used in the SAR context) is



***(a)** Histogram of the focused image.* $E_1(I) = 2.66$.

***(b)** Histogram of the unfocused image with* $\sigma = 0.5\,\mathrm{m}$. $E_1(I) = 3.32$.



***(c)** Histogram of the unfocused image with* $\sigma = 1.5\,\mathrm{m}$. $E_1(I) = 5.72$.

***(d)** Histogram of the unfocused image with* $\sigma = 3\,\mathrm{m}$. $E_1(I) = 7.47$.

***Figure 3.9:** Histograms for the images in Figure 3.8 and corresponding entropy 1 values. Note the log-scale on* $y$*-axis.*

(Yegulalp, 1999; Xi et al., 1999; Morrison and Munson, 2002),

$$E_2(I) = -\sum_{i=1}^{K}\sum_{j=1}^{M} p_{ij}\ln(p_{ij}) \tag{3.16a}$$

$$p_{ij} = \frac{|I_{ij}|^2}{\sum_{i=1}^{K}\sum_{j=1}^{M}|I_{ij}|^2} \tag{3.16b}$$

This entropy, on the other hand, will have its minimum for both optical and SAR images, as will be shown in the evaluations in Sections 3.3.3 and 3.3.4.

## 3.3.3 Evaluation on Test Optical Image

For the purpose of evaluation of the different focus measures, the image in Figure 3.10 is used. The evaluation will be performed for different thresholds where it is applicable and for different degrees of defocus (or unsharpness).

### Threshold Dependence

In order to evaluate the performance of the different measures, the ratio between focused and unfocused images, $F(I_{\text{focused}})/F(I_{\text{unfocused}})$, is plotted for the test image as a function of threshold $T$ ($F$ is one of TG, SML or DCT). Unfocusing is performed with the Gaussian low-pass filter kernel with size $5 \times 5$ pixels and with standard deviation $\sigma = 1.5$. The results are plotted in Figure 3.11. Since entropy does not depend on the threshold, only one value is obtained for this case, $E_1(I_{\text{foc}})/E_1(I_{\text{unfoc}}) = 6.99/6.80 = 1.03$ and $E_2(I_{\text{foc}})/E_2(I_{\text{unfoc}}) = 12.40/12.42 = 0.99$.

### Blur Kernel Dependence

In Figure 3.12 the ratio $F(I_{\text{focused}})/F(I_{\text{unfocused}})$ is plotted for different standard deviations of the $5 \times 5$ Gaussian low-pass kernel, in the range $\sigma \in [0,\ 4]$. Higher variance will yield a less focused image. For the measures that depend on the threshold, the threshold values are chosen based on the results from Section 3.3.3, i.e., the threshold values that give high ratio between focused and unfocused images. For this particular evaluation they are $T_{\text{TG}} = 350$, $T_{\text{SML}} = 270$ and $T_{\text{DCT}} = 2$.

## 3.3.4 Evaluation on Test SAR Images

For SAR images, basically the same evaluation as above will be performed, but in this case defocusing will be caused by the different trajectory errors, which is the main reason for SAR image defocusing.

### Threshold Dependence

To get a feeling for how different focus measures perform on SAR images as a function of the threshold, the same evaluation is done as in Section 3.3.3. The example SAR image used in the evaluation is in Figure 3.13. Defocusing is here obtained with adding range direction noise with standard deviation $\sigma = 1.5$. Results are plotted in Figure 3.14. Exactly as above, since entropy does not depend on the threshold, there is only one value for each $\sigma$. In particular, $\sigma = 1.5$ gives $E_1(I_{\text{foc}})/E_1(I_{\text{unfoc}}) = 1.91/5.12 = 0.37$ and $E_2(\tilde{I}_{\text{foc}})/E_2(\tilde{I}_{\text{unfoc}}) = 4.28/8.10 = 0.53$.

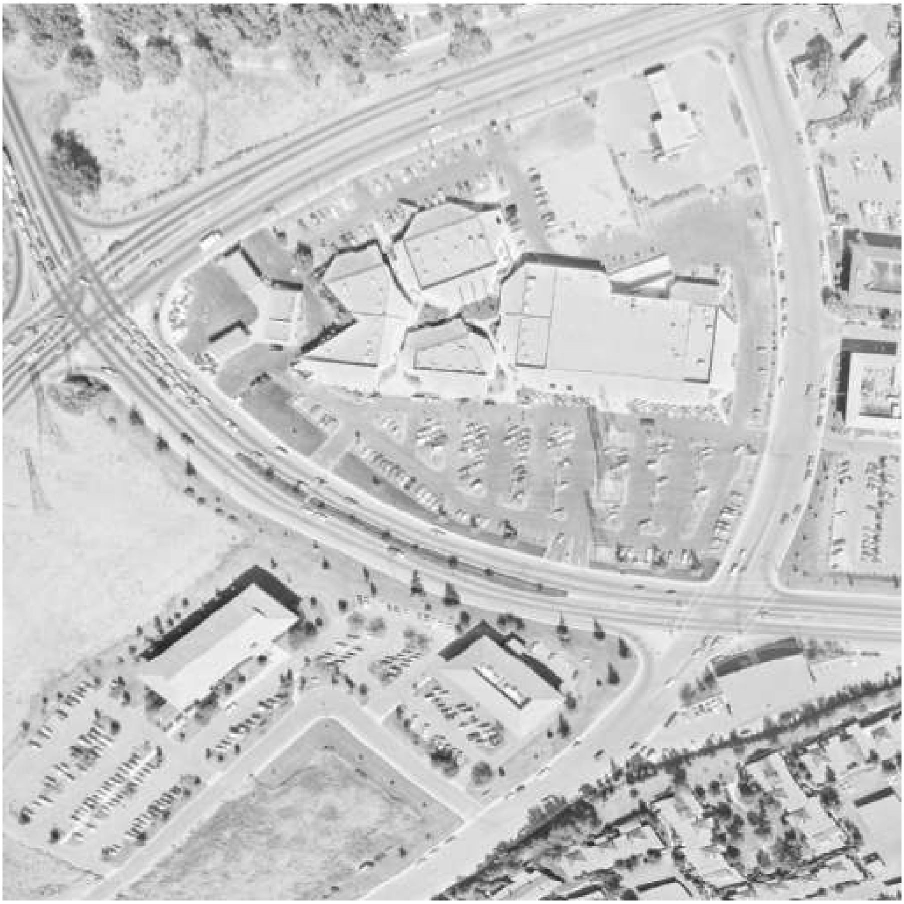
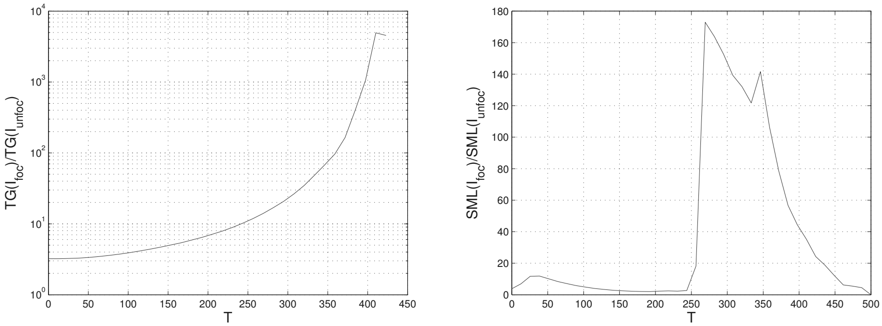
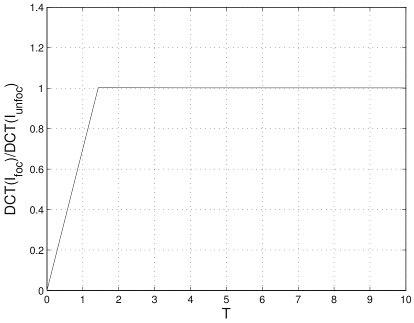


*Figure 3.10:* Test image used for evaluation of the focus measures. Image is courtesy of University of Southern California, Los Angeles, California, USA.



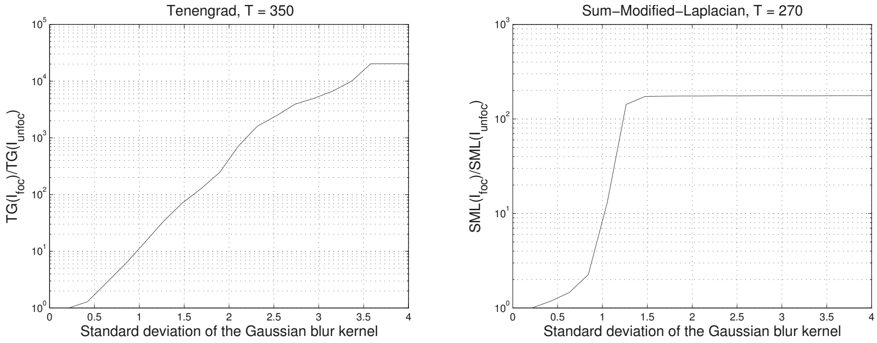
***(a)*** *$TG(I_{\mathrm{foc}})/TG(I_{\mathrm{unfoc}})$, log-scale on the $y$-axis.*

***(b)*** *$SML(I_{\mathrm{foc}})/SML(I_{\mathrm{unfoc}})$*



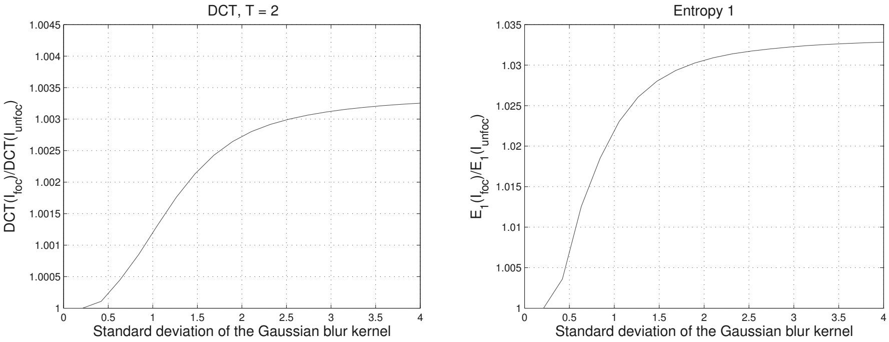
***(c)*** *$DCT(I_{\mathrm{foc}})/DCT(I_{\mathrm{unfoc}})$*

***Figure 3.11:*** *Ratio of the different focus measures for focused and unfocused optical images as a function of the threshold.*



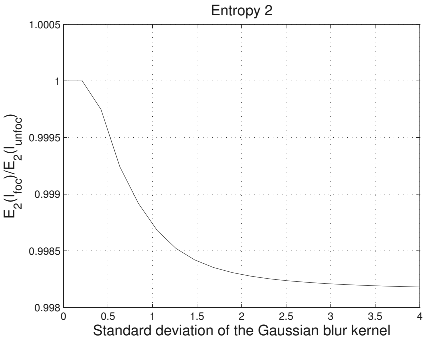
***(a)*** *$TG(I_{\text{foc}})/TG(I_{\text{unfoc}})$, log-scale on the $y$-axis.*

***(b)*** *$SML(I_{\text{foc}})/SML(I_{\text{unfoc}})$, log-scale on the $y$-axis.*



***(c)*** *$DCT(I_{\text{foc}})/DCT(I_{\text{unfoc}})$.*

***(d)*** $E_1(I_{\text{foc}})/E_1(I_{\text{unfoc}})$



***(e)*** $E_2(I_{\text{foc}})/E_2(I_{\text{unfoc}})$

***Figure 3.12:*** *Ratio of the different focus measures for focused and unfocused optical images as a function of the standard deviation of the Gaussian blur kernel.*

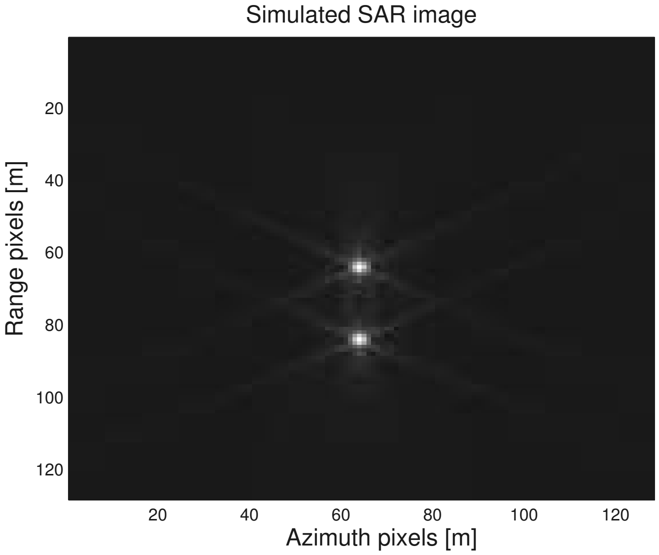


***Figure 3.13:*** *Example SAR image used for evaluation of the focus measures.*

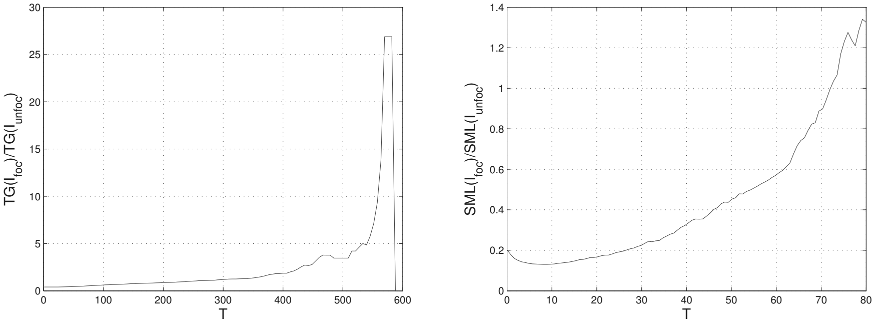
### Position Blur Dependence

Another test of the focus measures that is performed is how they depend on the variance of the noise in the range direction for a given threshold, similar to the evaluation in Section 3.3.3. The simulation is performed with the same noise realisation, but different variances, i.e., position in range direction is

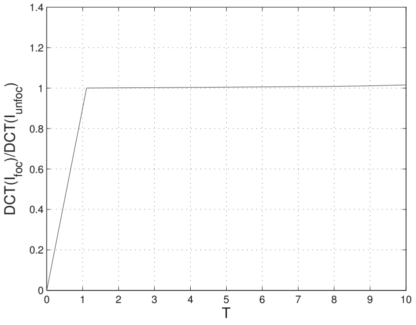
$$Y_t = \sqrt{Q} w_t \tag{3.17}$$

where $w_t$ is $\mathcal{N}(0,1)$ and $Q \in [0,\ 9]$ $[\mathrm{m}^2]$ (standard deviation is between 0 and 3 $\mathrm{m}$). The result is depicted in Figure 3.15. From these plots it can be noticed that DCT and entropy 2 measures behave as expected, i.e., the ratio between focused and unfocused image is monotonically increasing as a function of the standard deviation for DCT and decreasing for entropy 2 measure. For TG, SML and entropy 1, however, there is a difference between optical and SAR images. For optical images, all three measures will have their maxima for focused images, while for SAR images a minimum value is obtained instead. This behaviour can be explained both by the look of the test image, two bright points and lot of dark area, and the nature of the SAR images, i.e., SAR images usually look like they are negative optical images and contain more dark areas than optical images.

In order to obtain an image which is more informative, a scene in Figure 3.16 is created. A smaller image is created to minimise the dark area. This image should be more representative for the SAR images. The simulation above is modified and the maximum value for the standard deviation of the noise is set to 1.5 instead of 3. This is done because the smaller image might cause rand effects which will negatively influence the focus measures. The results are depicted in Figure 3.17. It can be noticed in these plots that TG and SML behave as expected now, but SML has a highly non-convex form. This indicates

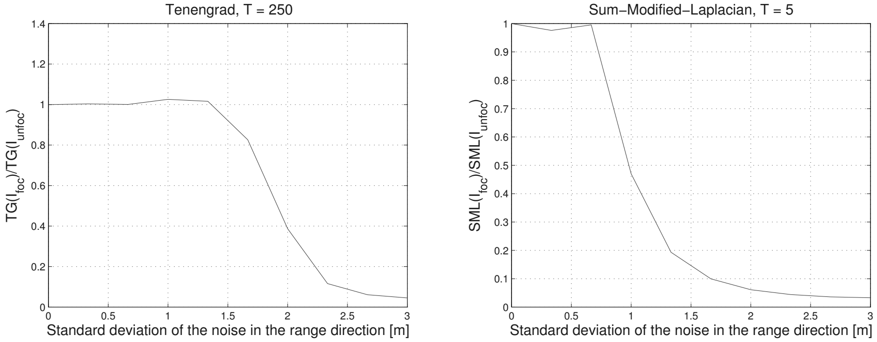


***(a)*** $TG(I_{\mathrm{foc}})/TG(I_{\mathrm{unfoc}})$ ***(b)*** $SML(I_{\mathrm{foc}})/SML(I_{\mathrm{unfoc}})$

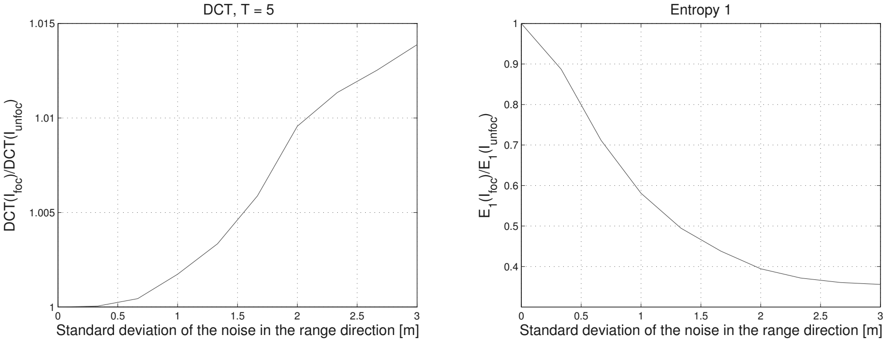


***(c)*** $DCT(I_{\mathrm{foc}})/DCT(I_{\mathrm{unfoc}})$

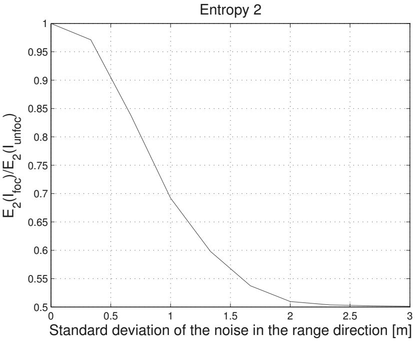
***Figure 3.14:*** *Ratio of the different focus measures for focused and unfocused SAR images as a function of the threshold.*



***(a)*** $TG(I_{\text{foc}})/TG(I_{\text{unfoc}})$ ***(b)*** $SML(I_{\text{foc}})/SML(I_{\text{unfoc}})$



***(c)*** $DCT(I_{\text{foc}})/DCT(I_{\text{unfoc}})$ ***(d)*** $E_1(I_{\text{foc}})/E_1(I_{\text{unfoc}})$



***(e)*** $E_2(I_{\text{foc}})/E_2(I_{\text{unfoc}})$

***Figure 3.15:*** *Ratio of the different focus measures for focused and unfocused SAR images as a function of the standard deviation of the noise in the range direction.*

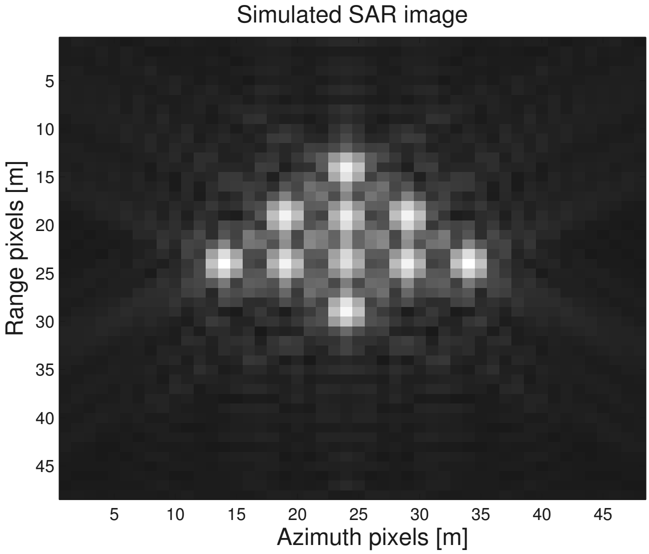


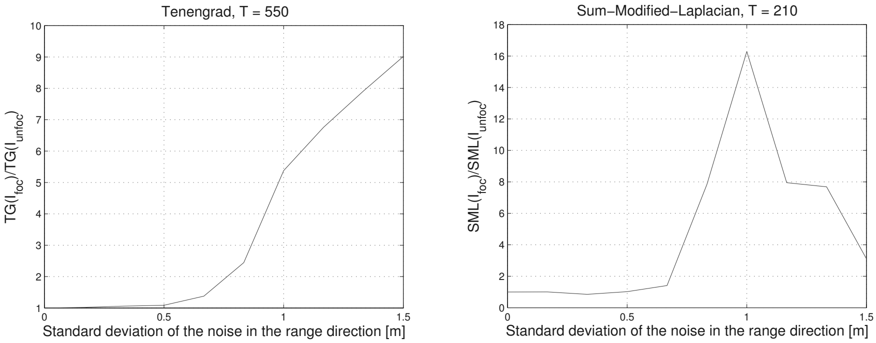
***Figure 3.16:*** *Example SAR image with more informative scene.*

that SML might not be suitable measure for the SAR images. Entropy 2 and DCT still behave as expected, and entropy 1, just as in the previous case, attains minimum for the most focused SAR image.

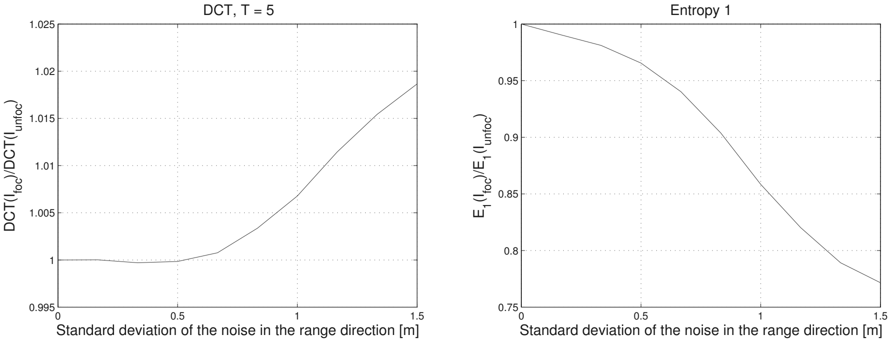
### Evaluation on Perturbed Trajectory

In order to evaluate the focus measures behaviour for different trajectories on the simulated SAR images, simulations with different trajectory errors are performed. The trajectory is simulated with the model (2.19) and trajectory errors consist of different incorrect initial conditions on velocity in azimuth direction, $v_0^X$, and acceleration in range direction, $a_0^Y$. The nominal initial values are chosen as $v_0^X = 100$ m/s and $a_0^Y = 0$ m/s$^2$ and acceleration error is varied between $-0.045$ and 0.045 m/s$^2$ and velocity error is varied between 99.962 and 100.038 m/s. Those values are chosen as $3\sigma$-values of the state errors from the assumed typical navigation system performance in an UAV, see Paper A. The noise, $w_t^a$, is set to zero in these simulations, i.e., the trajectory is completely deterministic. This has been done in order to be able to illustrate focus measures in a two-dimensional plot. On the other side, all focus measures, as functions of the state noise are convex and impose no problems in the minimisation step. In Figure 3.18 trajectory examples with some different acceleration and velocity errors according to Table 3.1 are shown.

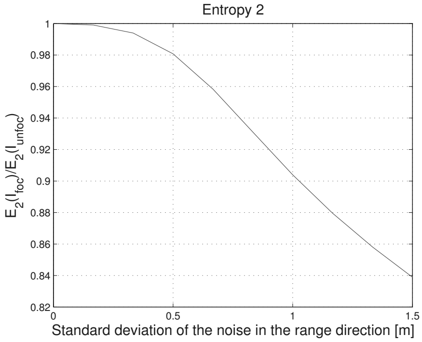
All five focus measures, TG with threshold value $T = 550$, SML with $T = 170$, DCT with $T = 5$ and entropy measures 1 and 2, are compared on the image in Figure 3.16. The thresholds are chosen according to the results from the simulations in Section 3.3.4. Results are depicted in Figure 3.19 where $1\sigma$-, $2\sigma$- and $3\sigma$-standard deviations of the initial states $v_0^X$ and $a_0^Y$ are also drawn. The contours in the plots are the level curves of the focus measures as a function of the error in the initial states. The level curves for the



**(a)** $TG(I_{\text{foc}})/TG(I_{\text{unfoc}})$ **(b)** $SML(I_{\text{foc}})/SML(I_{\text{unfoc}})$



**(c)** $DCT(I_{\text{foc}})/DCT(I_{\text{unfoc}})$ **(d)** $E_1(I_{\text{foc}})/E_1(I_{\text{unfoc}})$



**(e)** $E_2(I_{\text{foc}})/E_2(I_{\text{unfoc}})$

***Figure 3.17:*** *Ratio of the different focus measures for focused and unfocused SAR images with more informative scene as a function of the standard deviation of the noise in the range direction.*

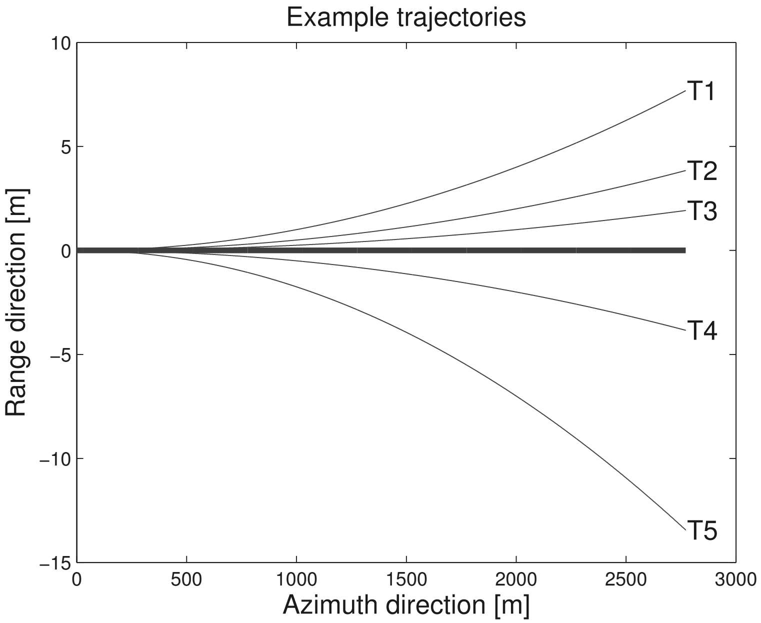


***Figure 3.18:*** *Example trajectories from Table 3.1. Thick line: no errors in initial values.*

| **Trajectory** | **Error in** $a_0^Y$ $[\mathrm{m/s^2}]$ | **Error in** $v_0^X$ $[\mathrm{m/s}]$ |
|---|---|---|
| T1 | $0.02$ | $-0.005$ |
| T2 | $0.01$ | $-0.01$ |
| T3 | $0.005$ | $0.005$ |
| T4 | $-0.01$ | $0.02$ |
| T5 | $-0.035$ | $0.005$ |

***Table 3.1:*** *Errors in initial acceleration and velocity for example trajectories shown in Figure 3.18.*

measures with the maximum for the correct values of the initial states, like TG or SML, are inverted.

Further, plots where only one parameter, $v_0^X$ or $a_0^Y$, is varied are presented in Figure 3.20 (all measures are normalised between 0 and 1). From all these figures it looks like that all measures except entropy 2 have several local minima and are highly non-convex. We also see that TG and SML do not have minimum value for the correct velocity value and that DCT has its minimum in the wrong value of the acceleration. The entropy measures perform fairly well, and entropy 1 has much sharper global minimum than entropy 2. Based on this, the entropy 1 and the entropy 2 measures look as the most attractive measures since they attain minimum value for the correct values of the states, they do not have any threshold to tune and entropy 2 is also smooth and convex in the interesting region of the state values.

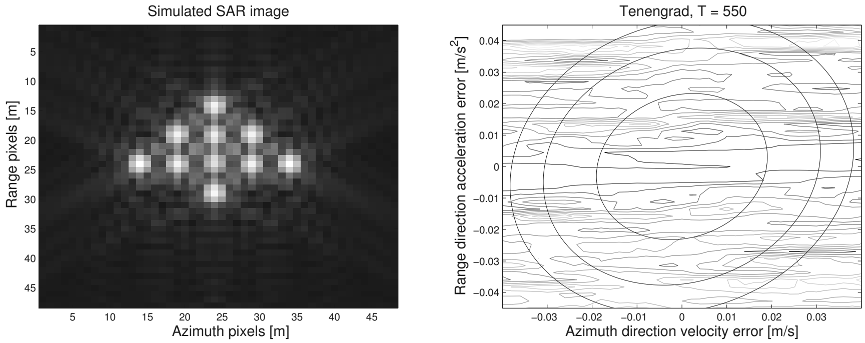
The focus measures' performance is also tested on a more unstructured scene illustrated in Figure 3.21a. This scene is created by randomly placing 150 point targets and assigning them a random reflectivity. The focus measures for this scene are shown in Figure 3.21. Here it can be seen that the measures look even worse and not even the entropy 1 measure has its global minimum for the correct values of the states. Entropy 2 is however still convex and smooth in the vicinity of the correct values of the states, and is still the most promising alternative to use as a focus measure in the minimisation criterion.

The focus measures above are evaluated on images that have no noise, i.e., the images are perfect. In reality, that is not the case, and images contain some noise caused by the noise in the radar measurements. Therefore, the same two scenes are used again, but white Gaussian noise with variance $\sigma^2 = 1.5$ is added to the radar echos. The images obtained with this setup are depicted in Figure 3.22a and Figure 3.23a. The focus measures for these two images are shown in Figure 3.22 and Figure 3.23. It can be seen in these plots that entropy 2 measure is still smooth and convex in the vicinity of the true values of the navigation states. The only thing that happens with entropy 2 is that the value of the function is different, but the principal form is the same as for the noise free case.

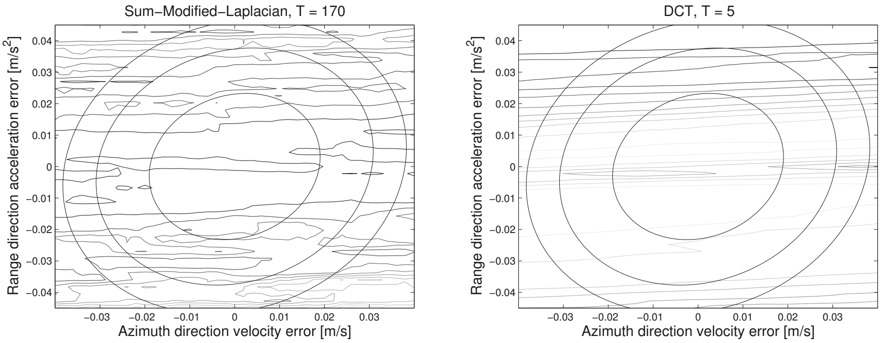
Another thing that can be noticed from these plots is that the entropy 1 measure does not have a pronounced global minimum in the case of structured scene, as it has for the noise free case. It looks similar to the case with unstructured scene. This is not surprising since the unstructured scene will behave like the image noise for the focus measure. It also looks like TG measure behaves better for the noisy case, at least around the true values of the navigation states. The explanation might be that TG measure has a threshold that can filter out the noise.

The conclusion from the evaluation above is that entropy 2 measure works quite fine for both cases, with and without noise in the radar measurements, and it is the smoothest measure of all tested ones. Entropy 1 could be used as well, at least for the scenes with structure, while TG seems to work for the case with noise in images. The drawback of the TG measure is the threshold that must be tuned to the different imaged scenes.

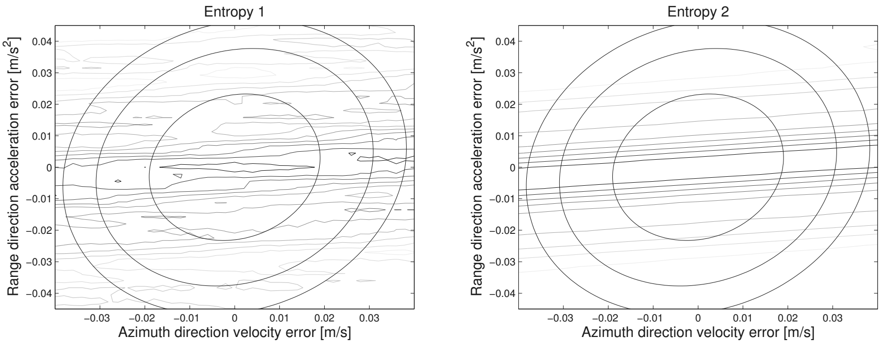
Another interesting thing that can be seen in this evaluation is that there is an ambiguity in the focus measure between initial $X$-direction velocity and $Y$-direction acceleration. It can be seen that almost all measures have a pronounced diagonal form. Very similar



*(**a**) SAR image of the structured scene.* *(**b**) TG focus measure with $T = 550$.*

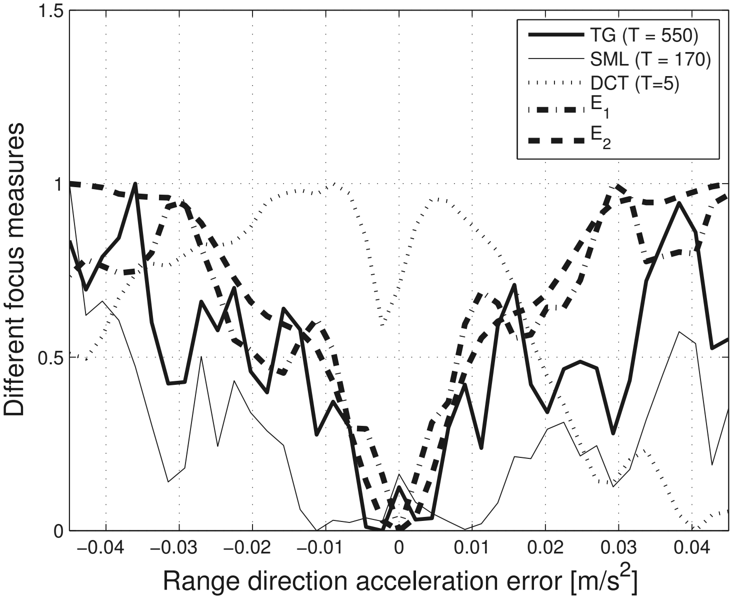


*(**c**) SML focus measure with $T = 170$.* *(**d**) DCT focus measure with $T = 5$.*

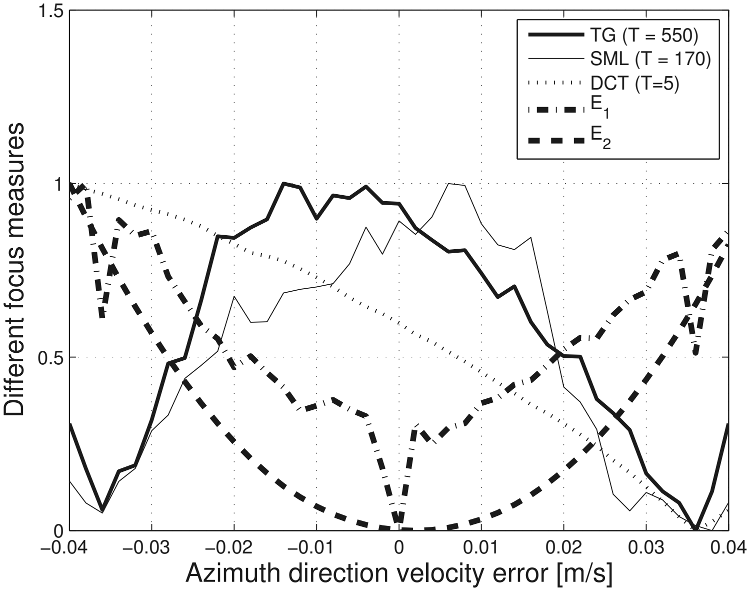


*(**e**) Entropy 1 focus measure.* *(**f**) Entropy 2 focus measure.*

***Figure 3.19:*** *Focus measures for the image with more informative structured scene. Standard deviations of the states are also drawn.*

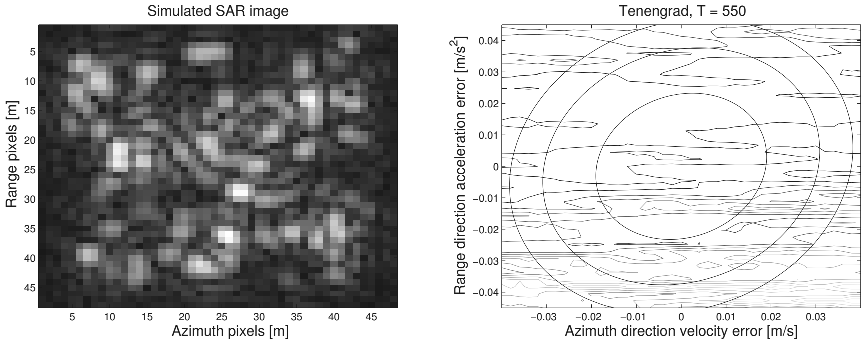


*(a) Focus measures with no velocity error.*

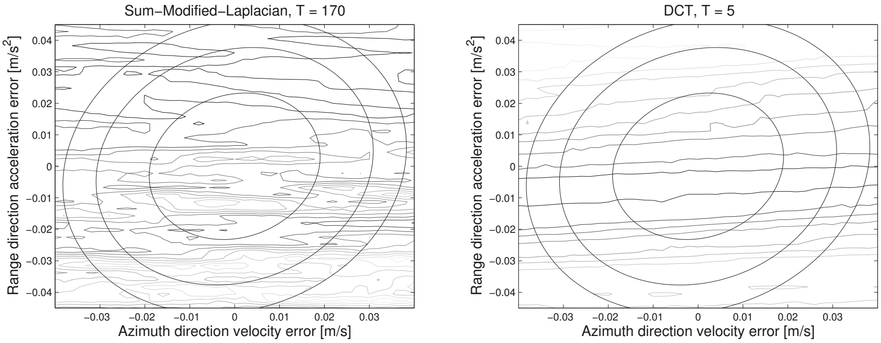


*(b) Focus measures with no acceleration error.*

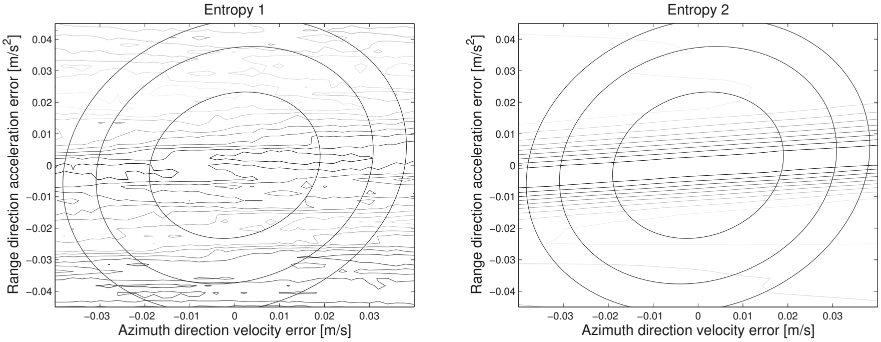
***Figure 3.20:*** *Focus measures for the structured scene with only error in* $a_0^Y$ *and* $v_0^X$*.*



***(a)** SAR image of the unstructured scene.* ***(b)** TG focus measure with $T = 550$.*

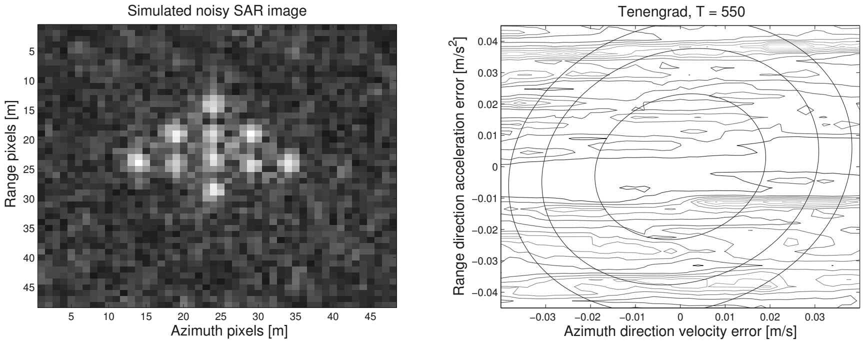


***(c)** SML focus measure with $T = 170$.* ***(d)** DCT focus measure with $T = 5$.*

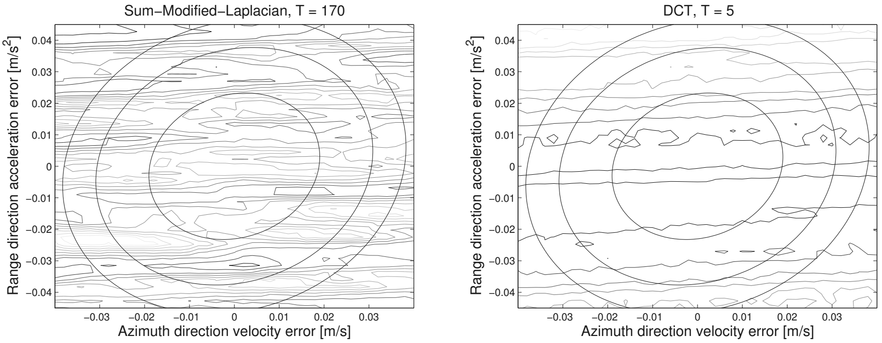


***(e)** Entropy 1 focus measure.* ***(f)** Entropy 2 focus measure.*

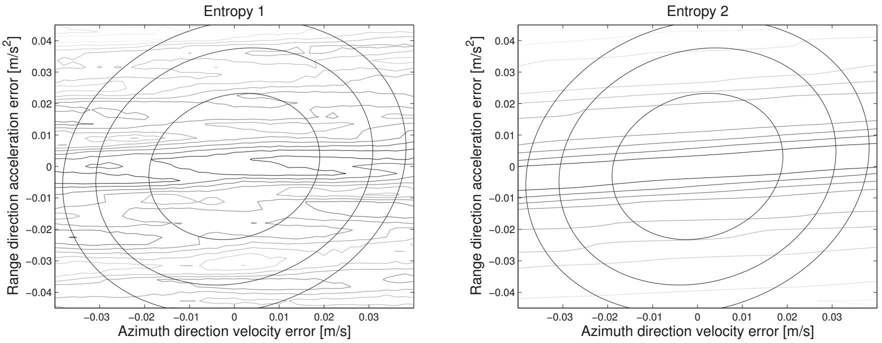
***Figure 3.21:** Focus measures for the image with more informative unstructured scene.*



***(a)*** *Noisy SAR image of the structured scene.* ***(b)*** *TG focus measure with $T = 550$.*

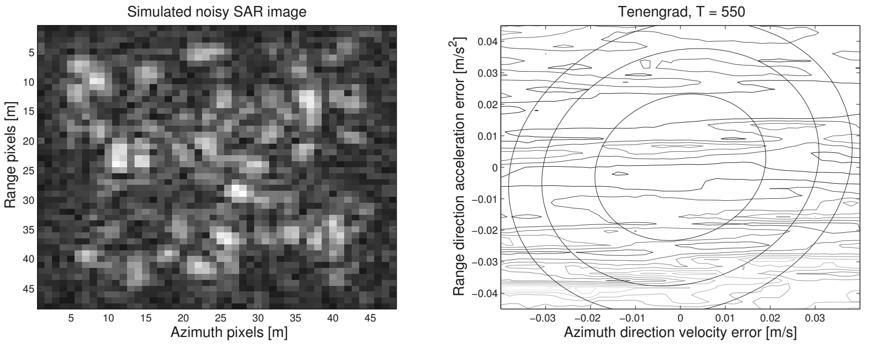


***(c)*** *SML focus measure with $T = 170$.* ***(d)*** *DCT focus measure with $T = 5$.*



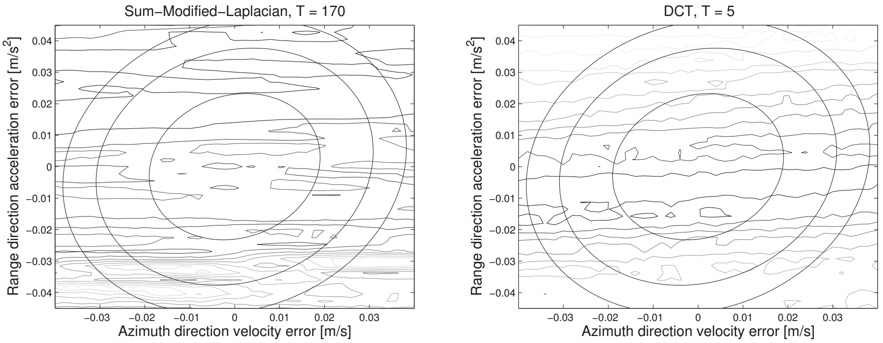
***(e)*** *Entropy 1 focus measure.* ***(f)*** *Entropy 2 focus measure.*

***Figure 3.22:*** *Focus measures for the noisy image with more informative structured scene. White Gaussian noise with variance $\sigma^2 = 1.5$ is added on the radar echoes.*



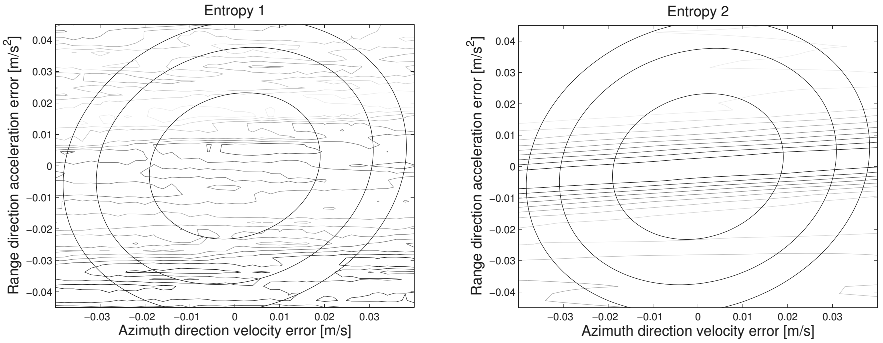
***(a)*** *Noisy SAR image of the unstructured scene.*

***(b)*** *TG focus measure with $T = 550$.*



***(c)*** *SML focus measure with $T = 170$.*

***(d)*** *DCT focus measure with $T = 5$.*



***(e)*** *Entropy 1 focus measure.*

***(f)*** *Entropy 2 focus measure.*

***Figure 3.23:*** *Focus measures for the noisy image with more informative unstructured scene. White Gaussian noise with variance $\sigma^2 = 1.5$ is added on the radar echoes.*

focus value is obtained if acceleration and velocity errors are both negative as well as if they are both positive. This is most clearly visible in the Entropy 2 measure. In practice this means that many different trajectories give equally focused image. This behavior of the focus measure puts limitations on estimation performance with focus measure only. In this case, the sensor fusion framework, where several different sensors are used, allows us to estimate the platform's states better. In this case inclusion of the IMU will improve the estimation performance.

## 3.4 Optical Camera

In the previous two sections IMU and SAR principles have been described and measurement functions for these sensors were defined. The last sensor that is used here is a regular optical camera. In the following section a function of an optical camera will be described. The basic projective geometry model of the pinhole camera will be explained, as well as how this model is used to formulate a measurement function used in the sensor fusion framework.

### 3.4.1 Projective Geometry Model

Generally speaking, an optical camera is a device that creates images of the surrounding environment, much like SAR does. The main difference is that the camera is a passive sensor, i.e., it does not transmit any signal or energy in order to create an image[1]. Instead it receives a visible light and detects the total light energy with an array of photo-sensitive cells. Each cell will be irradiated by the portion of the light from the environment, represented as rays of light, creating the actual image value at this cell. These values are then saved in memory and this is what we call an image. The number of these cells, or pixels, is deciding the image resolution, usually expressed as the amount of pixels in horisontal and vertical directions. For example, so called Video Graphics Array (VGA) format, although originally used to refer to display hardware invented by IBM, is today also used for specific resolution, namely $640 \times 480$ pixels, (IBM, 1992). All modern cameras use optics, consisting of one or several lenses to refract the light, i.e., change its propagation path, in order to place the image plane on the right place inside the camera, i.e., on the sensor array. A quite simple model of the lens, thin lens, can be used to explain this mechanism. As depicted in Figure 3.24, rays passing through the point $\mathsf{P}$, and which are parallel with the optical axis, will deflect its path through the lens and pass through the focal point placed on the focal length, $f_0$, from the lens. In addition, rays passing through the center of the lens, so called Optical Center, $\mathsf{OC}$, will pass unchanged. These two properties of the thin lens will define where image of $\mathsf{P}$, $\mathsf{p}$, will be placed, namely on the intersection of these rays. By changing the focal length we can make the image larger or smaller, which is known as zooming. Note also that the actual image is placed upside-down. With help from the principle of similar triangles we obtain the fundamental equation of the thin lens

$$\frac{1}{f_0} = \frac{1}{Z} + \frac{1}{z} \tag{3.18}$$

---

[1] This is true if we disregard from the flash that many cameras frequently use. However the flash is used to enhance the image quality and is not essential for the image creation

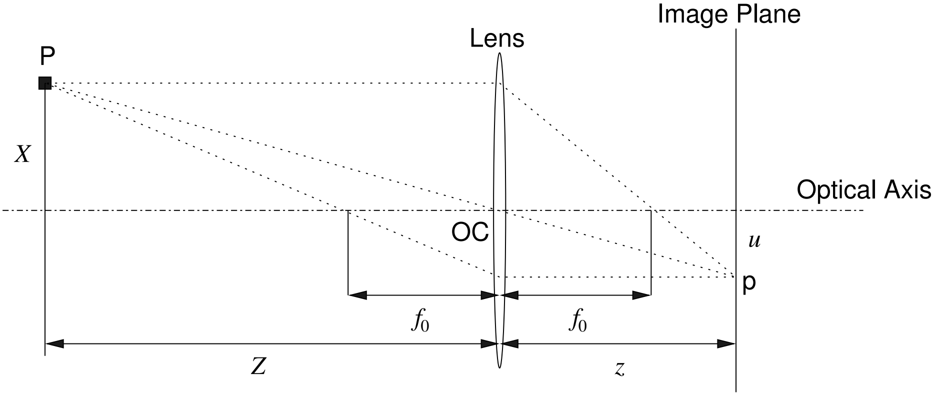


***Figure 3.24:*** *Basic thin lens geometry. Point* P *on a distance $Z$ from the lens is imaged as a point* p *on the image plane placed on a distance $z$. $f_0$ is defined as the focal length and* OC *is the Optical Center of the camera.*

With some algebraic transformations of this equation we can obtain the $u$-coordinate of p as

$$u = -\frac{Xf_0}{Z - f_0} = -\frac{(z - f_0)X}{f_0} \tag{3.19}$$

where the minus sign denotes the upside-down image position.

Although this equation describes a thin lens principle quite accurately, another, and simpler, model will be used to describe the projective nature of the camera. This model is called *Pinhole Camera Model* and is obtained by letting the size of the lens go to zero, (Ma et al., 2004). In that case, only the rays that are going through the optical center will contribute to the image creation. This also implies that $f_0$ will also go to zero and by rewriting (3.19) as

$$-\frac{Z - f_0}{X} = \frac{f_0}{u} \tag{3.20a}$$

$$-\frac{z - f_0}{u} = \frac{f_0}{X} \tag{3.20b}$$

and letting $f_0$ go to zero we obtain

$$-\frac{Z}{X} = -\frac{z}{u} = 0 \Rightarrow u = z\frac{X}{Z} \tag{3.21}$$

This is a perfect pinhole camera projective model. In Figure 3.25 this setup is depicted. Here the image plane is actually moved in front of the optical center in order to simplify expressions (minus sign is omitted in that way) and also, somewhat ambiguously, the distance between the optical center and the image plane is called focal length in the pinhole camera case and will be denoted $f$ instead of $z$. In a two-dimensional image plane, the

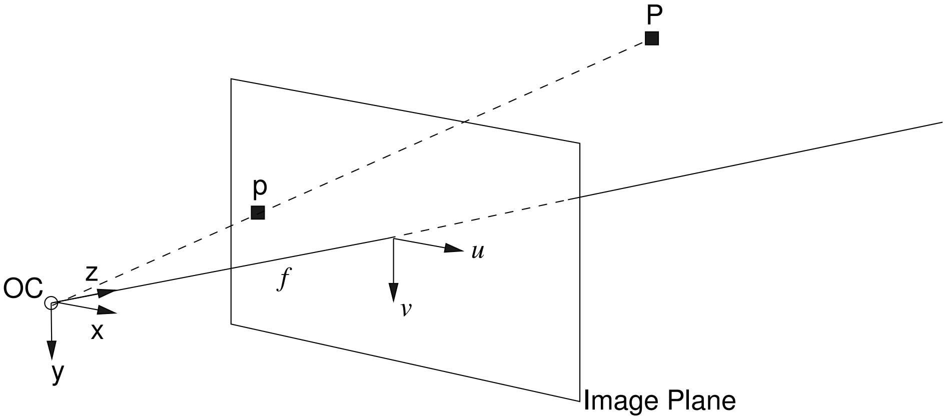


***Figure 3.25:*** *Illustration of the projective geometry of the pinhole camera model.* OC *is the Optical Center of the camera,* p *is the image of the point in 3D,* P*, i.e., its projection in the image plane.* $f$ *is the focal length of the camera, which is a distance between optical center and image plane.* $u$ *and* $v$ *are the normalised pixel coordinates.*

ideal pinhole camera projective model is then

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{3.22}$$

where $[X\ Y\ Z]^T$ are the 3D coordinates of the imaged point P expressed in the camera frame xyz, see Figure 3.25, and $[u\ v]^T$ are the coordinates of the image point p. Often, the focal length of the camera is known, which means that we can compensate for it and obtain the normalised pinhole camera model, where the focal length is equal to one. This model will be defined via an operator $P : \mathbb{R}^3 \to \mathbb{R}^2$ as

$$\begin{bmatrix} u \\ v \end{bmatrix} = P([X^c\, Y^c\, Z^c]^T) = \frac{1}{Z^c} \begin{bmatrix} X^c \\ Y^c \end{bmatrix} \tag{3.23}$$

where we have explicitly written imaged point's coordinates in camera frame, $c$, and here $[u\ v]^T$ is denoting the normalised coordinates of the image point. The main property of the projective model is its inherent depth ambiguity, i.e., without prior knowledge of the scene, it is impossible to resolve if the observation comes from the large scene that is far away, or from a small scene that is close to the camera. This is similar to the focus measure ambiguity mentioned for the SAR images. This behavior can easily be seen in the structure of (3.23). If $X$ and $Y$ are scaled by a constant, making the scene larger (smaller), then it can be compensated by scaling $Z$ with the same constant making the distance to the scene bigger (smaller). This corresponds to moving along the line OC-P in Figure 3.25. If the scene is unknown and we have several camera observations from different viewpoints, this ambiguity can be resolved, and this is known as triangulation, see for example Hartley and Sturm (1997). However, if both camera positions and the scene are unknown, the

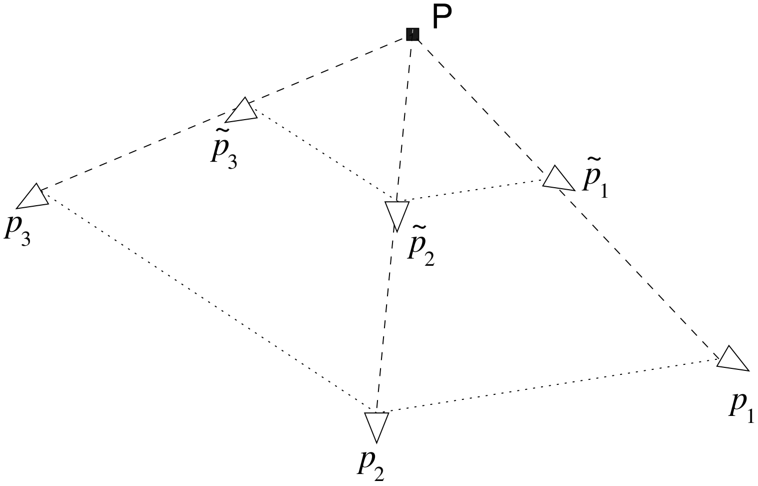


***Figure 3.26:*** *Depth ambiguity of the moving camera. A camera moving fast in positions* $\{p_1,\ p_2,\ p_3\}$ *perceives the point* P *in exactly the same way as a slowly moving camera in positions* $\{\tilde{p}_1,\ \tilde{p}_2,\ \tilde{p}_3\}$.

problem is much more difficult. If camera is moving, the depth ambiguity manifests itself as a possible speed scaling, which in turn causes position scaling. This can be seen in Figure 3.26 where a camera located in positions $\{p_1,\ p_2,\ p_3\}$ gives the same images of the point P as a camera positioned in $\{\tilde{p}_1,\ \tilde{p}_2,\ \tilde{p}_3\}$ but these two cameras are moving with different speeds. In order to resolve this ambiguity, we need additional measurements of the motion. As mentioned in Section 3.3.4, exactly as for the SAR, an IMU can provide that kind of measurements and the sensor fusion framework gives the means necessary to solve the problem.

The normalised coordinates above assume that the actual image has infinite resolution, and they are expressed in metric units, e.g., meters. In practice, the actual digital image consists of discrete image elements, pixels, with coordinates $[p_x\ p_y]^T$ in the image plane. In order to use the projective model we need to transform the pixel coordinates to the normalised coordinates. This will be done by an affine model, involving scaling and shifting. Scaling is done to transform from pixel size to metric size and shifting is done because the normalised coordinates have their origin in the middle of the image plane, i.e., the point where optical axis and image plane intersect, while the pixel coordinates have their origin in the upper-left corner of the image. If these, so called intrinsic camera parameters, are given, the total transformation can be written as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{f}\left(\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}\begin{bmatrix} p_x \\ p_y \end{bmatrix} - \begin{bmatrix} s_x o_x \\ s_y o_y \end{bmatrix}\right) \tag{3.24}$$

where $s_x$ and $s_y$ are the scaling parameters (in $\mathrm{m}$/pixel) and $o_x$ and $o_y$ are the coordinates of the middle of the image plane (in pixels). The intrinsic camera parameters can be estimated, for a particular camera, with readily available software, see for example

Bouguet (2010). This model assumes that there are no nonlinear distortions caused by the lens system for example, but even that kind of calibration is possible to perform giving a slightly more complicated model, see e.g., Ma et al. (2004). In the model (3.23) the 3D coordinates are referenced to the camera frame, but usually both platform's states and imaged environment points are referenced to the local navigation frame. For that reason we need to transform these coordinates to the camera coordinates. Just as in the IMU case, this transformation is performed as

$$m^c = \mathsf{R}^{cn}(m^n - p^n) \tag{3.25}$$

where $m^c = [X^c \; Y^c \; Z^c]^T$ are the environment point's, called landmark, coordinates in the camera frame, $c$, $\mathsf{R}^{cn}$ is the rotation matrix between navigation and camera frame, $m^n$ is the coordinates of the landmark expressed in the navigation frame and, as before, $p^n$ is the position of the platform in the navigation frame. The resulting measurement equation, for one landmark, can finally be written as

$$y_t^m = \begin{bmatrix} u_t \\ v_t \end{bmatrix} = P(\mathsf{R}_t(m - p_t)) + e_t^m = h^m(x_t, m) + e_t^m \tag{3.26}$$

where the frame notation has been omitted, and we added the time dependence and noise which is, as before, assumed to be white and Gaussian, $e_t^m \sim \mathcal{N}(0, R_t^m)$. The Equation (3.26) can be generalised for several landmarks by stacking all the measurements of the corresponding $M$ landmarks as in

$$\underbrace{\begin{bmatrix} u_t^1 \\ v_t^1 \\ \vdots \\ u_t^{\mathsf{N}_{y_t}} \\ v_t^{\mathsf{N}_{y_t}} \end{bmatrix}}_{y_t^m} = \underbrace{\begin{bmatrix} P(\mathsf{R}_t(m^{c_t^1} - p_t)) \\ \vdots \\ P(\mathsf{R}_t(m^{c_t^{\mathsf{N}_{y_t}}} - p_t)) \end{bmatrix}}_{h_t(x_t, \mathsf{M}_t)} + e_t^m \tag{3.27}$$

where $\mathsf{N}_{y_t}$ denotes the number of measurements at time $t$ and the correspondence variables, $c_t^i$, encodes the measurement-landmark assignment, $y_t^i \leftrightarrow m^j$, which gives a subset of the landmarks at time $t$, $\mathsf{M}_t = \{m^j\},\ j \in \{1, \ldots, M | \, c_t^i = j\}$. This relation can be further generalised to collect the measurements for all the times, $1 : N$, by stacking (3.27) to obtain

$$\underbrace{\begin{bmatrix} y_1^m \\ \vdots \\ y_N^m \end{bmatrix}}_{y_{1:N}} = \underbrace{\begin{bmatrix} h_1(x_1, \mathsf{M}_1) \\ \vdots \\ h_N(x_N, \mathsf{M}_N) \end{bmatrix}}_{h_{1:N}(x_{0:N}, \mathsf{M})} + \underbrace{\begin{bmatrix} e_1^m \\ \vdots \\ e_N^m \end{bmatrix}}_{e_{1:N}} \tag{3.28}$$

In this way, a measurement function, or likelihood function, on the form (2.29) is obtained, where all the $M$ landmarks $m^1, \ldots, m^M$ are collected into the parameter $\mathsf{M}$. This parameter serves as a map of the environment that is observed by the camera. This representation is conceptually different from the map obtained with SAR in a sense that the camera based map is parametric, with point landmarks, while the SAR based map is simply an image obtained with nonparametric methods. In any case, since both kind of maps can be related to the platform's states, we can use similar estimation methods for both the states and the map.

### 3.4.2 Image Feature Extraction

In the previous section it was explained how to obtain the normalised image coordinates from the pixel coordinates and how to relate the normalised coordinates to the platform's states. In this section the techniques how to obtain interesting points in the images, represented by pixel coordinates and called image features, will be introduced.

As already mentioned, the observation model for the camera is parametric, meaning that the environment is encoded with discrete points in three dimensions, landmarks. The landmarks are, except in very special occasions, not given in advance, implying that they must be found and initialised during the estimation, i.e., we must use the images to extract the landmarks. In practice, this means that some of the points in images will be remembered in order to track them across several images and in that way establish the measurements of the environment needed for the estimation. The ability to track the landmarks in several images puts some requirements on which points in the images to choose. So the task for the feature extractor is to find image points that are relatively easy to find in as many images as possible. One way of accomplishing that is to find corners in the image. A reason the corner points are a good choice is that they have a distinct gradient change in all directions around a corner point. A simple edge would have change in gradient in only one direction (across edge) and a flat region has no change in gradient, as expected. By image gradient, we consider e.g., Sobel gradient calculated as in (3.11). A simple way to measure this response is to build an approximation to Hessian matrix in each pixel as in

$$H_{ij} = \begin{bmatrix} \sum_w (S^r)^2 & \sum_w S^r S^c \\ \sum_w S^r S^c & \sum_w (S^c)^2 \end{bmatrix} \tag{3.29}$$

where the sum is taken over a window, $w$, centered on pixel $(i, j)$. The size of the window is a parameter that is free to choose and it determines the size of the surrounding where a feature is to be determined. Usually this window is combined with some low-pass kernel, e.g., a Gaussian kernel. By taking a simple transformation of the matrix above

$$C_{ij} = \det(H_{ij}) - \kappa \operatorname{Tr}(H_{ij})^2 \tag{3.30}$$

where $\det(\,\cdot\,)$ denotes the determinant and $\operatorname{Tr}(\,\cdot\,)$ the trace of the matrix, we obtain a function called Harris strength. The scalar $\kappa$ is seen as a tuning parameter which governs the sensitivity between corner and edges. The Harris strength will have a high value for the pixels where the corner is located and we can detect corners simply by seeking the local maxima of $C$. This is exactly the principle behind a very popular Harris corner detector (Harris and Stephens, 1988) and many implementations are available. As an illustration of the Harris corner detector, consider the image in Figure 3.27. A resulting Harris strength and 30 corner features are shown in Figure 3.28.

Although simple to understand and implement, the Harris corner detector provides us only with the corner positions. In order to find the same corner in another image, we need some kind of corner identity as well. Since the Harris detector does not provide that, a common

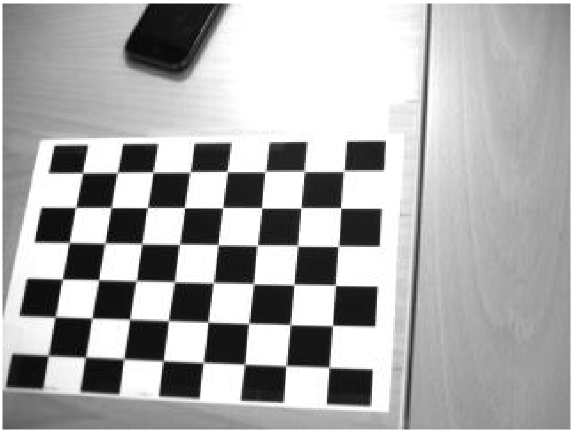
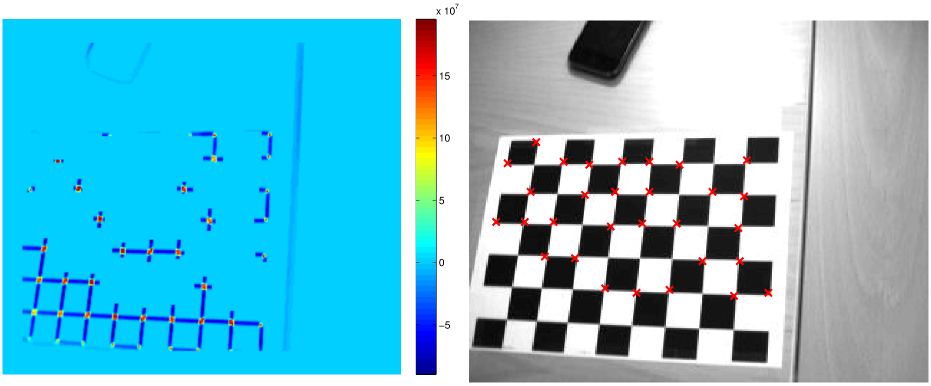


***Figure 3.27:*** *Example image for the illustration of the Harris corner detector.*



***(a)*** *Harris strength for the image in Figure 3.27.*

***(b)*** *30 Harris corners for the image in Figure 3.27.*

***Figure 3.28:*** *Harris strength and features for the example image.*

approach is to use a small image patch around the corner as a feature identity. This allows for feature matching by a correlation method, for example. This method works well for pure translational movement and without changing the scale of the image. To overcome that limitation, other feature detectors, that have scale and rotation invariance, have been proposed, and the most popular ones are Scale Invariant Feature Transform (SIFT) (Lowe, 1999) and Speeded Up Robust Features (SURF) (Bay et al., 2006). These feature detectors fall into the class of blob-detectors and use Difference of Gaussians (DoG) or determinant of Hessian (DoH) approach to detect key-points at different image scales. The key-points are chosen, just as in Harris detector case, as local maxima of the DoG or DoH operators. In addition to the invariance to rotation and scaling, these detectors also provide a feature descriptor, i.e., a feature identity that is needed. For SIFT it is a 128 dimensional vector which is composed of accumulated image gradients from small image regions in the direct surrounding of the key-point. In this way, a compact descriptor representation is obtained and it allows us to use it for feature matching across the images.

Given the feature detector of our choice, together with the measurement model (3.28), we can now use the camera as a sensor in order to estimate the navigation and environment parameters.

# 4

# Concluding Remarks

This chapter concludes the whole thesis by summarising the main conclusions from the publications included in Part II, in a “bird-eye view” fashion, and giving some suggestions about future work. For more detailed treatment of the conclusions and the future work the reader is referred to each paper.

## 4.1 Conclusions

In this work, a sensor fusion and optimisation based estimation approach to navigation and mapping, also called SLAM, is presented. The approach is mainly aimed for applications to unmanned aerial vehicles and, as such, it is using sensors that are available on these platforms. Furthermore, the chosen sensors are independent on the external infrastructure, like satellites in case of GNSS. The benefit this sensor setup gives is the higher availability of the UAV in the cases where the external infrastructure cannot be assumed available. In particular, the sensors that are used here are the inertial and imaging sensors measuring the UAV’s ego-motion and the surrounding environment, respectively. Inertial sensors considered here are the accelerometers and gyros, while imaging sensors are the optical camera and SAR. In the SAR case, images can also be used to obtain the global UAV position by matching these to a map with known global coordinates.

The proposed methods are implemented and demonstrated on both simulated and real data. The results obtained for both considered imaging sensors are showing that they, in combination with IMU, can be used for navigational purposes as well as for mapping of the surroundings. Since the nature of the errors in the imaging sensors are different from the inertial errors, i.e., velocity and position ambiguity versus drift, they serve as a good complement to each other producing metrically correct solutions, i.e., resolving the ambiguities, and reducing the drift caused by integrating the IMU errors.

The structure of the SLAM problem can be exploited, mostly for the inertial/visual and partly for inertial/SAR case, and together with the optimisation formulation used to efficiently obtain the smoothed estimate of the interesting parameters. In the inertial/visual case the problem is formulated as nonlinear least squares for which efficient algorithms can be used. Furthermore, for the inertial/visual case, certain conditional structure can also be used to obtain a good initial solution of the parameters. This is highly desirable since bad initialisation of the nonlinear optimisation problem may lead to convergence to local minima.

For the inertial/SAR case, both sequential and batch formulations of the navigation and SAR image focusing are proposed. For the batch formulation an efficient calculation method of the optimisation-criterion gradient is devised and a quasi-Newton method can be applied to solve the optimisation problem in fewer iterations than only using pure gradient search. In the sequential solution an Extended Kalman Filter is used with a system model where range gradient is estimated from the SAR raw data and used as a measurement function. The optimisation formulation of the joint navigation and image focusing is further successfully extended by using maps to find global position and flight direction of the UAV.

The SAR technique has also been applied to cellular phone time-of-arrival data to estimate the non-line-of-sight signal propagation in urban environment. This can also be viewed as mapping of the dominant reflector environment. In particular, a multistatic extension of the basic SAR back-projection method has been applied to real cellular phone data and preliminary results show good potential.

## 4.2 Future Work

A quite broad area that involves navigation and mapping has been treated in this thesis. In most cases the optimisation framework has been used to estimate the parameters, which usually represent navigational states and some parametrisation of the environment map. Currently, a naïve implementation of these problems is done where standard MATLAB generic optimisation functions are used. By exploiting the structure of the problems, tailored solvers could be implemented which potentially can make execution more efficient from both the execution time and memory usage.

Another interesting option that can be explored is the Moving Horizon Estimation (MHE) where a smoothed estimate within a fixed-length window is obtained and the optimisation methods proposed here can easily be adapted for MHE. This would also give a decreased execution time, since the window size can be kept constant which means the the problem size is also approximately constant.

# Bibliography

L. E. Andersson. On the Determination of a Function from Spherical Averages. *SIAM Journal on Mathematical Analysis*, 19(1):214–232, 1988. doi: 10.1137/0519016. URL `http://link.aip.org/link/?SJM/19/214/1`.

T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics & Automation Magazine*, 13(3):108–117, Septemeber 2006.

H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Proceedings of European Conference on Computer Vision ECCV2006*, pages 404–417, May 2006.

T. Bayes. An Essay towards solving a Problem in the Doctrine of Chances. *The Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

J. Y. Bouguet. Camera Calibration Toolbox for Matlab. `www.vision.caltech.edu/bouguetj/calib_doc/`, 2010.

C. Cafforio, C. Prati, and F. Rocca. SAR data focusing using seismic migration techniques. *IEEE Transactions on Aerospace and Electronic Systems*, 27(2):194 –207, March 1991. ISSN 0018-9251. doi: 10.1109/7.78293.

L. J. Cutrona, W. E. Vivian, E. N. Leith, and G. O. Hall. A high-resolution radar combat-surveillance system. *IRE Transactions on Military Electronics*, MIL-5(2):127–131, April 1961. ISSN 0096-2511. doi: 10.1109/IRET-MIL.1961.5008330.

H. Durrant-Whyte and T. Bailey. Simultaneous Localization and Mapping: Part I. *IEEE Robotics & Automation Magazine*, 13(12):99–110, June 2006.

J. Farrell and M. Barth. *The global positioning system and inertial navigation*. McGraw-Hill Professional, 1999.

J. A. Fawcett. Inversion of N-Dimensional Spherical Averages. *SIAM Journal on Applied Mathematics*, 45(2):336–341, 1985. ISSN 00361399. URL `http://www.jstor.org/stable/2101820`.

R. Ferzli and L. J. Karam. No-reference objective wavelet based noise immune image sharpness metric. In *Proceedings of International Conference on Image Processing, ICIP 2005*, volume 1, pages I–405–8, Septemeber 2005.

R. A. Fisher. On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41:155–160, 1912.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings of Radar and Signal Processing*, 140(2):107 –113, April 1993. ISSN 0956-375X.

*Global Positioning System Standard Positioning Service Performance Standard.* GPS, US Government, 2008.

F. Gustafsson. *Statistical Sensor Fusion.* Studentliteratur, Lund, 2010.

C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of Fourth Alvey Vision Conference*, pages 147–151, 1988.

R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68 (2):146 – 157, 1997. ISSN 1077-3142.

H. Hellsten and L. E. Andersson. An inverse method for the processing of synthetic aperture radar data. *Inverse Problems*, 3(1):111, 1987. URL `http://stacks.iop.org/0266-5611/3/i=1/a=013`.

*IBM VGA Technical Reference Manual.* IBM, 1992.

T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation.* Prentice-Hall, Upper Saddle River, New Jersey, 2000.

R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

M. Kristan, J. Pers, Perse M., and S. Kovacic. A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform. *Pattern Recognition Letters*, 27(13): 1431 – 1439, 2006. ISSN 0167-8655. doi: DOI:10.1016/j.patrec.2006.01.016.

J. B. Kuipers. *Quaternions and Rotation Sequences.* Princeton University Press, 1999.

R. Larsson, Z. Sjanic, M. Enqvist, and L. Ljung. Direct Prediction-error Identification of Unstable Nonlinear Systems Applied to Flight Test Data. In *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

D. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the Seventh International Conference on Computer Vision (ICCV99)*, pages 1150–1157, Corfu, Greece, September 1999.

Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry. *An Invitation to 3-D Vision.* Springer Science+Business Media, LLC, 2004.

J. L. Meriam and L. G. Kraige. *Engineering Mechanics, Dynamics, 4th ed.* John Wiley & Sons, 1998. ISBN 0-471-24167-9.

A. S. Milman. SAR Imaging by Omega-K Migration. *International Journal of Remote Sensing*, 14(10):1965–1979, 1993.

R. L. Jr. Morrison and D. C. Jr. Munson. An experimental study of a new entropy-based SAR autofocus technique. In *Proceedings of International Conference on Image Processing, ICIP 2002*, volume 2, pages II–441–4, September 2002. doi: 10.1109/ICIP.2002.1039982.

F. Natterer. *The Mathematics of Computerised Tomography*. New York: Wiley, 1986.

C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. The SciTech Radar and Defense Series. SciTech, 2004. ISBN 1-891121-31-6.

F. Rocca. Synthetic Aperture Radar: a New Application for Wave Equation Techniques. *Stanford Exploration Project SEP-56*, pages 167–189, 1987. URL `http://sepwww.stanford.edu/oldreports/sep56/56_13.pdf`.

M. D. Shuster. A survey of attitude representations. *The Journal of Astronautical Sciences*, 41(4):439–517, October 1993.

Z. Sjanic and F. Gustafsson. Simultaneous Navigation and SAR Auto-focusing. In *Proceedings of 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.

Z. Sjanic and F. Gustafsson. Navigation and SAR Auto-focusing Based on the Phase Gradient Approach. In *Proceedings of 14th International Conference on Information Fusion*, Chicago, USA, July 2011.

Z. Sjanic and F. Gustafsson. Fusion of Information from SAR and Optical Map Images for Aided Navigation. In *Proceedings of 15th International Conference on Information Fusion*, Singapore, July 2012.

Z. Sjanic and F. Gustafsson. Simultaneous Navigation and Synthetic Aperture Radar Focusing. *Provisionally accepted to IEEE Transactions on Aerospace and Electronic Systems*, August 2013a.

Z. Sjanic and F. Gustafsson. Navigation and SAR focusing with Map Aiding. *Submitted to IEEE Transactions on Aerospace and Electronic Systems*, June 2013b.

Z. Sjanic, M. A. Skoglund, T. B. Schön, and F. Gustafsson. A Nonlinear Least-Squares Approach to the SLAM Problem. In *Proceedings of 18th IFAC World Congress*, Milano, Italy, August/Septemeber 2011.

Z. Sjanic, F. Gunnarsson, C. Fritsche, and F. Gustafsson. Cellular Network Non-Line-of-Sight Reflector Localisation Based on Synthetic Aperture Radar Methods. *Submitted to IEEE Transactions on Antennas and Propagation*, September 2013a.

Z. Sjanic, M. A. Skoglund, and F. Gustafsson. Expectation-Maximisation Maximum Likelihood Estimation for Inertial/Visual SLAM. *Submitted to IEEE Transactions on Robotics*, September 2013b.

M. A. Skoglund, Z. Sjanic, and F. Gustafsson. Initialisation and Estimation Methods for Batch Optimisation of Inertial/Visual SLAM. *Submitted to IEEE Transactions on Robotics*, September 2013.

L. M. H. Ulander, H. Hellsten, and G. Stenstrom. Synthetic-aperture radar processing using fast factorized back-projection. *IEEE Transactions on Aerospace and Electronic Systems*, 39(3):760–776, July 2003. ISSN 0018-9251. doi: 10.1109/TAES.2003.1238734.

D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jr. Jakowatz. Phase gradient autofocus - a robust tool for high resolution SAR phase correction. *IEEE Transactions on Aerospace and Electronic Systems*, 30(3):827–835, July 1994. ISSN 0018-9251. doi: 10.1109/7.303752.

J. L. Walker. Range-Doppler Imaging of Rotating Objects. *IEEE Transactions on Aerospace and Electronic Systems*, AES-16(1):23–52, 1980. ISSN 0018-9251. doi: 10.1109/TAES.1980.308875.

L. Xi, L. Guosui, and J. Ni. Autofocusing of ISAR images based on entropy minimization. *IEEE Transactions on Aerospace and Electronic Systems*, 35(4):1240–1252, October 1999. ISSN 0018-9251. doi: 10.1109/7.805442.

M. Xing, R. Jiang, X.and Wu, F. Zhou, and Z. Bao. Motion Compensation for UAV SAR Based on Raw Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2870–2883, August 2009. ISSN 0196-2892. doi: 10.1109/TGRS.2009.2015657.

A. F. Yegulalp. Minimum entropy SAR autofocus. In *7th Adaptive Sensor Array Processing Workshop*, March 1999.